Principles of Evolutionary Medicine

Second edition

Peter Gluckman
Centre for Human Evolution, Adaptation and Disease, Liggins Institute, The University of Auckland, New Zealand

Alan Beedle
Centre for Human Evolution, Adaptation and Disease, Liggins Institute, The University of Auckland, New Zealand

Tatjana Buklijas
Centre for Human Evolution, Adaptation and Disease, Liggins Institute, The University of Auckland, New Zealand

Felicia Low
Centre for Human Evolution, Adaptation and Disease, Liggins Institute, The University of Auckland, New Zealand

Mark Hanson
Institute of Developmental Sciences and NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton, UK

OXFORD UNIVERSITY PRESS
Preface to the Second Edition

Since the first edition of *Principles of Evolutionary Medicine* was published in 2009, the field has grown considerably. An international society and new journals have been established, there has been a US taskforce on the teaching of evolutionary medicine, and many more medical, anthropology, and public health programs or courses on evolutionary medicine have been introduced. Research in the field has expanded greatly. This growing activity reflects a growing understanding of the value of applying an evolutionary approach to understanding human ecology, health, and disease. This edition reflects both this growth in the field and the continuing need for a summary of evolutionary principles appropriate to the distinct needs of health practitioners, educators, social scientists, and others interested in the human condition. With this in mind, this edition includes considerable revisions to each chapter.

The general structure of the book remains unchanged, with the first six chapters providing a summary of the evolutionary theory relevant to understanding human health and disease. The second part of the book describes how evolutionary principles can be used to understand behavior, metabolism, immunity, and reproduction, and, in an entirely new chapter, cancer. These two sections are bridged by a new chapter that details pathways by which evolutionary processes affect disease risk and symptoms, and how hypotheses in evolutionary medicine can be tested. The final two chapters are considerably expanded; they illustrate the application of evolutionary biology to medicine and public health, and consider the ethical and societal issues arising from an evolutionary perspective. A number of new clinical examples and historical illustrations are included.

With this expansion we have added two new authors who have been active in the field for many years: Dr. Tatjana Buklijas, who has a particular focus on historical and clinical perspectives, and Dr. Felicia Low, whose research has focused on molecular and developmental evolutionary biology. We thank the many readers who sent suggestions after the first edition, and the many colleagues who have helped develop the field and whose research is reflected in many places throughout this volume.

Peter Gluckman

Alan Beedle

Tatjana Buklijas

Felicia Low

Mark Hanson

August 2015
Preface to the First Edition

Evolutionary science can be viewed as the fundamental “organizing principle” of all biology. The biological and biomedical sciences can only be fully integrated with the aid of an evolutionary toolkit. Together with discovery of the nature of particulate inheritance (the gene) and the sciences of molecular and developmental biology, evolutionary biology provides the basis of our understanding both of the function of an organism and of its relationship with its physical, social, and biotic environment. Thus an effective comprehension of human biology, health, and disease requires knowledge of evolutionary principles and an appreciation of how they have shaped biological and biomedical processes at both an individual and a population level. Although this imperative is well appreciated in the other biological sciences, medicine has been slow to recognize evolutionary biology as a fundamental and underpinning science. However, advances in areas such as bacterial evolution, genomics, and epigenomics mean that evolutionary thinking has much to add to modern medicine.

There has been a growing list of books and edited volumes devoted to this topic following the publication in 1994 of the groundbreaking book by Randolph Nesse and George Williams entitled Why We Get Sick: The New Science of Darwinian Medicine. But it seemed to the authors of the present book that there was no integrated or comprehensive textbook which set out the basic principles of evolutionary biology for the medical reader and which focused on how medicine and public health can utilize these principles. Most members of the health professions have had little or no formal exposure to such principles—this volume aims to remedy that deficiency. We hope that it provides just such a toolkit which doctors, medical students, trainees and practitioners in allied health professions, biomedical scientists, and anthropologists interested in human health need in order to utilize evolutionary biology to gain a more complete understanding of the processes that shape the human condition.

Evolutionary biology emerged in the nineteenth century, although its acceptance and integration into modern biology was not without controversy and took many decades. The concept that species were not immutable emerged in Europe as a significant school of thought at the end of the eighteenth century, but it was not until Charles Darwin and Alfred Russell Wallace independently described natural selection in 1858, and thus provided an explanation of how species might change over time and how new species might evolve, that the field of evolutionary science was firmly established. But this science could not progress much further until the nature of inheritance was understood. Gregor Mendel had recognized the particulate nature of inheritance in the late nineteenth century, but the significance of his discovery was largely overlooked. It was not until his findings were rediscovered, the role of the nucleus in inheritance was determined, and chromosomal DNA was identified as the chemical basis of the gene, that there could be a fuller understanding of how evolutionary processes operate. It was the elucidation of the structure of DNA in 1953 which finally allowed evolutionary processes to be understood at a biochemical level. The subsequent explosion of knowledge in molecular and developmental biology has been informed by, and has also informed, progress in evolutionary science.

There are many dimensions to evolutionary biology and it has a number of sub-disciplines. It has to address many questions: how species have formed; how lineages respond to and adapt to
their environment and thus evolve to appear to be “designed” to match their environment; how environmental influences induce the development of a range of phenotypes from a single genotype; why different species even within the same taxa have very different physical, reproductive, and social characteristics; why species have particular life histories; and many more such issues. The answers to these questions inform our understanding of the origin of the particular characteristics of a species and the range of phenotypic variation seen between individual members of that species—in particular in their anatomy and their physiology, the characteristics of their life course, and the manner in which they respond to environmental challenge and opportunity. Thus evolutionary biology is very much concerned with the basis and the significance of individual variation.

Humans, like all other living organisms, have individual characteristics, including the fact that we do not all suffer from the same diseases. This variation is defined in part by our evolutionary history and, conversely, without variation there could be no evolution. Indeed, understanding the significance of individual variation was one of Darwin’s great insights. Thus to understand human biology and medicine fully we must have an understanding of evolutionary principles and how they apply to our species.

Much of medicine is focused on understanding disease causation, for this informs how to prevent disease and how to intervene when it does occur. Medical thinking has a tendency to dichotomize into normal or healthy and abnormal or pathological. But, as this book will make clear, such assessments are contextual—what is a successful adaptation in one context, and so normal under those conditions, may be highly abnormal in another.

The definition of what is health and what is disease when viewed through an evolutionary lens can therefore lead to helpful new perspectives on a potential patient. Variation is a fundamental attribute of biology, and it determines individual risk in response to an environmental challenge—whether this is a parasite like malaria, an environmental toxin like nicotine, or a lifestyle of excessive calorie intake, to name just a few examples. This textbook explains how individual risk is determined by our evolutionary history and how that history has given us a capacity to cope with many challenges but has also placed constraints on that capacity. The consequences of encountering challenges which exceed our adaptive capacity become manifest as disease.

Most medical training focuses on understanding the immediate pathways leading to disease—these are the so-called proximate causes. From this perspective, hypertension arises because of changes in peripheral vascular resistance or because of changes in the renin–angiotensin system secondary to renal disease; sickle cell anemia arises because of a mutation in the hemoglobin gene; appendicitis arises because of inflammation in a gastrointestinal diverticulum; and cerebral palsy can arise because of asphyxia at birth during an obstructed labor. It is this understanding of proximate cause that gives rise to most medical therapies: serotonin reuptake inhibitors to treat depression, antibiotics to treat bacterial pneumonia, angioplasty to improve blood flow through occluded coronary arteries, or cesarean section. But as we shall see, in each case there is a broader dimension.

The proximate explanations reveal how certain symptoms appear and provide the basis of a logical approach to intervention, but there is a further level of enquiry which is valuable. This concerns questions about why certain symptoms appear, why some individuals are at greater risk, why we cannot accommodate easily to certain situations healthily, why we have an appendix which gets inflamed, or why the day you are born was the most dangerous day of your life so far. This evolutionary level of interrogation seeks to understand the ultimate causes of health and disease. Through it, we discover that we get appendicitis because our evolutionary ancestors were leaf-eaters and had a large cecum to help digest cellulose-based foods; we now no longer need this large gastrointestinal organ, but the appendix remains as an evolutionary relic which can become inflamed. This ultimate perspective provides health professionals with better insights into their patient and must improve their management of the case. In many instances an evolutionary perspective leads to a better understanding of which approaches to prevention are more promising and why certain therapies are more likely to work.
Humans now live in very complex environments which are very different from those in which most of our ancestors lived and evolved. The consequent mismatches can challenge our health. We can never escape our biology or our biological past. Evolutionary processes operate to promote passage of genetic information from one generation to the next and evolutionary success is about successful passage of genes within a lineage to future generations. Thus the processes of evolution are focused on what drives reproductive success within a lineage, a concept termed “fitness.” But fitness does not depend necessarily on longevity or health. It involves “trade-offs” which ensure reproductive success even if they incur other costs such as a shorter life. Evolutionary biology is a science that considers how an organism trades off one component of its biology against others to optimize its fitness. Because many modern humans live long lives and medicine is increasingly focused on promoting the quality of life, health professionals cannot ignore the constraints imposed by such evolutionary considerations.

Evolutionary medicine, therefore, is a growing and central discipline that applies evolutionary knowledge to the understanding of human biology, both normal and abnormal. It is an essential science, necessary for a holistic perception of how health and disease emerge. It has application in both individual healthcare and in public health. It adds much to understanding other basic disciplines of medicine, including physiology, anatomy, biochemistry, pathology, molecular biology, population health, and behavioral sciences. Indeed, a complete understanding of these more immediate disciplines is not possible without an understanding of evolutionary biology.

Evolutionary biology is a vibrant, if broad, domain of biomedical science. Some aspects of evolutionary knowledge are not essential or central to understanding the core principles of evolutionary medicine. For example, the subject of macroevolution—the process underpinning speciation and biodiversity—is not central to a medical perspective. Much of evolutionary biology involves quantitative approaches, for example for defining aspects of selection or genetic drift; again these are not essential for the medical reader. Many of the details of the dynamics of selective processes are technical and are not required in applying evolutionary principles in human medicine. We have therefore omitted them from this book, which is intended for the clinician, whether in training or in practice. Most textbooks in evolutionary biology focus on other species and only minimally refer to humans. In contrast, unless there is an essential comparative point to be made, we have tried to use only examples from human biology to illustrate key evolutionary principles.

The book is presented in three parts. In the first we detail the basic principles of evolutionary biology: what biological evolution is, how it operates through the processes of selection, how evolution is reflected in our genome, the relationship between genotype and phenotype, how developmental and evolutionary processes interact, what determines the characteristics of the human life history, and how the evolution of our species has led to features which now become manifest in the doctor’s office or on the hospital ward. An important evolved characteristic of our species is that we live in groups and our social environment and our capacity to develop and apply technology are essential components of our evolution. So it is not possible to discuss biological evolution without consideration of our cultural evolution, and this we do in the first part of the book.

In the second part of the book we describe how these principles can be applied to an understanding of human disease, using four illustrative axes: human reproduction; nutrition and metabolism; biological defense systems; and human behavior. We have intentionally restricted the discussion in this way so that these systems can be elucidated in sufficient detail to highlight how evolutionary approaches to the human condition can be applied in practice. This is not intended to be a comprehensive medical textbook—there are plenty of those—but is intended to give the reader a new understanding which can be applied generally in clinical medicine and which informs other domains of medical science.

In the third part of the book, we synthesize these various strands to provide a systematic evolutionary framework for understanding human health and disease. We propose that each person
preface to the first edition

presenting to a physician has three relevant histories: the history of the complaint itself; the developmental history of that individual; and his or her evolutionary history. All three histories are essential for a comprehensive understanding of the way an individual has responded to his or her environment. We detail the pathways by which individual risk can be influenced by evolutionary processes, pathways which should always be part of a health professional’s reflection on the situation of the patient before him or her.

Evolutionary biology has an intimate relationship with the ecological sciences, and humans must also be understood in their ecological context. Consideration of how our lives progress in any environment, including our social environment, is greatly enhanced by understanding evolutionary biology. In turn, such understanding can contribute greatly to the development of effective public health strategies.

Evolutionary biology as a science has always had, and continues to have, an awkward and complex relationship with broader intellectual and philosophical concerns. For example, it is seen by some to be in conflict with their specific belief systems. Darwin’s propositions when first put forward were clearly at odds with the prevailing concepts of natural theology and of an active creation, the dominant institutional explanations of the natural world in early nineteenth-century Britain. Yet today the majority of scientists find no need to see evolutionary biology as in conflict with their understandings and thus to this volume. One person in particular, Dr. Chris Kuzawa (Northwestern University, Chicago), was critical to the book. Chris made major contributions to Chapters 5 [Life histories] and 8 [Nutrition] in the early stages. Unfortunately other commitments prevented him from playing the larger role which we, and he, had planned. We must also acknowledge colleagues who have read chapters, offered critiques and suggestions, and elucidated matters beyond our expertise. They include Professor Sir Patrick Bateson FRS (Cambridge), Professor Peter Ellison (Harvard), Professor Eva Jablonka (Tel Aviv), Professor Hamish Spencer (Dunedin), Professor Randolph Nesse (Michigan), Professor Paul Rainey (Massey), Professor Russell Gray (Auckland), Professor Wayne Cutfield (Auckland), Professor Murray Mitchell (Auckland), and Professor Des Gorman (Auckland). We thank Dr. Cinda Cupido (Auckland), Dr. Tatjana Buklijas (Auckland), Dr. Felicia Low (Auckland), and Ms. Amanda Calhoun (Yale) for assistance with research.

Peter Gluckman
Alan Beedle
Mark Hanson
October 2008
# Contents

## PART 1  FUNDAMENTALS OF EVOLUTIONARY BIOLOGY

1  **Introduction**  
   1.1  What is disease?  
   1.2  What evolution is: fundamental principles  
   1.3  Time  
   1.4  Constraints  
   1.5  We are not alone  
   1.6  Culture and gene–culture coevolution  
   1.7  How evolutionary arguments fit alongside other biological perspectives  
   1.8  Evolution and medicine  
   Key points  

2  **Evolutionary theory**  
   2.1  Introduction  
   2.2  What does evolutionary theory explain?  
   2.3  How does evolution work?  
   2.4  Areas of debate and the limitations of the adaptationist argument  
   2.5  Conclusion  
   Key points  

3  **The molecular basis of variation and inheritance**  
   3.1  Introduction  
   3.2  Genes and disease  
   3.3  The molecular basis of human genetic variation  
   3.4  Factors affecting genetic variation  
   3.5  Single-gene or Mendelian disorders  
   3.6  No single genes for common diseases  
   3.7  Epigenetic mechanisms as a cause of variation  
   3.8  Non-genetic inheritance  
   3.9  Conclusion  
   Key points  


4 Evolution and development

4.1 Introduction
4.2 Development: pre-ordained or plastic?
4.3 Is development important to evolution?
4.4 Developmental plasticity
4.5 Responses to environmental cues during development
4.6 Epigenetic processes and development
4.7 Learning and instinct
4.8 The evolution of novelty
4.9 Conclusion

Key points

5 The human life history

5.1 Introduction
5.2 General overview of life-history theory
5.3 Body size and shape
5.4 Growth in humans
5.5 Evolutionary analysis of the distinct features of human growth
5.6 Conclusion

Key points

6 Human evolution and the origins of human diversity

6.1 Introduction
6.2 The hominoid clade
6.3 Hominin evolution
6.4 Genomic changes that make us human
6.5 Human adaptation to local selection pressures
6.6 Are humans still evolving?
6.7 Social and medical implications of human diversity
6.8 Conclusion

Key points

PART 2 EVOLUTION IN HEALTH AND DISEASE

7 An evolutionary framework for understanding human health and disease

7.1 Introduction
7.2 Fundamental principles of evolutionary medicine
7.3 Why has evolution left our bodies vulnerable to disease?
7.4 An evolutionary classification of ultimate mechanisms affecting disease risk
7.5 Testing hypotheses in evolutionary medicine

Key points
# Contents

## 8 Reproduction

- 8.1 Introduction 177
- 8.2 Sexual reproduction 178
- 8.3 Why did sex evolve? 178
- 8.4 Sex determination 181
- 8.5 Reproductive strategies 182
- 8.6 Mate choice 183
- 8.7 Sexual differences in the human 186
- 8.8 Sex differences in morbidity and mortality 188
- 8.9 Human reproductive life cycle 190
- 8.10 Conclusion 203

Key points 204

## 9 Nutritional and metabolic adaptation

- 9.1 Introduction 205
- 9.2 Strategies for energy storage 206
- 9.3 Human diet: an evolutionary history 210
- 9.4 How can change in the environment increase disease risk? 218
- 9.5 Conclusion 235

Key points 236

## 10 Coevolution, infection, and immunity

- 10.1 Introduction 237
- 10.2 Coevolution 237
- 10.3 Humans and their associated species 238
- 10.4 The challenge of infectious disease 240
- 10.5 Pathogen emergence 241
- 10.6 Pathogen virulence and transmission 242
- 10.7 Host defenses 244
- 10.8 Public health measures 252
- 10.9 Vaccination 253
- 10.10 Antibiotics 255
- 10.11 Conclusion 258

Key points 258

## 11 Psychology and behavior

- 11.1 Introduction 261
- 11.2 Biological determinants of culture and behavior 261
- 11.3 Evolution of the human brain and behavior 263
- 11.4 Evolution of social behavior 264
- 11.5 Evolutionary perspectives on psychology 278
- 11.6 Evolutionary psychiatry 279
- 11.7 Conclusion 284

Key points 286
## 12 Cancer

12.1 Introduction 287
12.2 Epidemiology of cancer 288
12.3 Ecology of cancer 289
12.4 The biology of cancer 290
12.5 Cancer in the light of evolutionary mechanisms 293
12.6 Implications of an evolutionary approach for the prevention and treatment of cancer 299
12.7 Conclusion 302
Key points 302

## 13 Evolutionary principles applied to medical practice and public health

13.1 Introduction: Understanding health and disease from an evolutionary perspective 303
13.2 Testing evolutionary hypotheses in medicine 305
13.3 Clinical examples 305
13.4 An evolutionarily mismatched or novel environment 305
13.5 Life-history-associated factors 311
13.6 Excessive and uncontrolled defense mechanisms 313
13.7 Consequences of coevolution with microbes 315
13.8 Results of evolutionary constraints 317
13.9 An apparently harmful allele is maintained by balancing selection 319
13.10 The consequences of sexual selection 321
13.11 The outcomes of cladal and demographic history 322
13.12 Value and limits of an evolutionary medicine perspective 324
Key points 326

## 14 Evolution, medicine, and society

14.1 Introduction 327
14.2 Origins of Darwin’s theory 327
14.3 From Darwin to “Social Darwinism” 331
14.4 Eugenics 332
14.5 The “Modern Synthesis,” human evolution, and medicine 335
14.6 Evolution, society, and religion 337
14.7 Evolutionary thought and the human condition 339
Key points 339

References 341
Index 365
Life expectancy in high-income societies has risen dramatically over the past 250 years, and more recently in many low- and middle-income countries, largely as a result of major technological changes, including improvements in public health and in the understanding of the biology of disease, and through concurrent societal changes which have brought a greater emphasis on the value of life. For example, the average life expectancy at birth in pre-revolutionary France was about 30 years—not so different from that of prehistoric (and in particular Paleolithic; Table 1.1) humans—yet now it is over 80 years.

Advances in nutrition, infection control and treatment, trauma care, and maternity and neonatal services have addressed many of the extrinsic causes of death that led to the short lifespan of our Paleolithic forebears. These causes were major contributors to what in our not-so-distant ancestors we would now call “premature” death (Figure 1.1). But as we live for longer, diseases which were previously unimportant become more so; many, such as cardiovascular disease, appear in middle age—by which time the majority of our forebears were already dead. Other morbidities, such as mental disorders, have become more dominant as a result of the pressures of living in the much more complex societies that urbanization and changes in modes of communication have brought.

In addition, our ambitions as organisms have changed: advances in medical care, improved access to knowledge, and individual empowerment have brought a changed individual and societal focus on longevity and quality of life. Modern medicine is increasingly faced with patients’ expectations of being able to live well into their ninth decade and to expect highly interventional medicine, if needed, to maintain quality of life throughout. Increasingly we face disabilities that do not arise from extrinsic causes but instead result from intrinsic ageing of the body’s cells, as reflected in degenerative diseases. Thus, while medicine is now dominated by a population-wide expectation of good health into old age this is, to a large extent, in conflict with the evolutionary processes that have molded our species. This textbook is about the principles underlying those processes and how many disease states can be understood in terms of this conflict. While health and longevity are the primary concern of our patients, neither of these, with caveats to be discussed later, are the primary drivers of evolutionary processes.

Briefly stated, evolution of a species (macroevolution) and evolutionary change within a species (microevolution) operate to produce an organism that is matched or adapted to its environment. That match is not primarily defined by a particularly long or comfortable life for an individual, but rather by the successful passage of that individual’s genes to successive generations. Indeed it is a truism that all organisms on the planet today are here because their ancestors reproduced successfully. Lineages that did not do so are now extinct. This is the core concept of fitness, which is fundamental to evolutionary biology.

Evolution is the process whereby a population changes over time to optimize the fitness of its individual members within a particular environment. Thus Homo sapiens largely evolved by adaptations that maximized its fitness in the environments of eastern Africa, the region in which our species first emerged. Biological fitness for a human was—and
still is—achieved by a strategy of supporting a small number of progeny to grow successfully to adulthood, reproduce, and live long enough to support their own offspring in becoming reproductively competent. Evolutionary pressures on our lineage operated to ensure this; there were few, if any, drivers of health beyond the reproductive period in the life course, or generally of longevity beyond the period necessary to support offspring into adulthood.

The term *environment* occurs frequently throughout this book, and a definition here is thus appropriate. Evolutionary and developmental biologists use the term in a wider sense than in common popular usage, where it relates to issues such as global warming or threatened biodiversity. We define the environment of an organism as the sum of all the external conditions and stimuli that it experiences, including climate, nutrient supply, social structure resulting from cooperation with or competition from other members of its own species, symbiotic relationships with other species, particularly micro-biota, and threats from other species in the form of predation, parasitism, or infection. If the environment changes for the worse, the lineage may be able to cope, but at a cost; or it may go extinct; or individuals may adapt to their new environment or, if mobile, migrate to a more suitable environment. Some species can ensure relative constancy of their environment by constructing it themselves, a process called *niche construction*—examples include the temperature-controlling mounds of termites or the dams constructed by beavers. The human species is a niche constructor *par excellence* through its use of technologies ranging from fire and clothing to urban design and, as described in numerous places in this book, this capability has both positive and negative consequences for our health.

<table>
<thead>
<tr>
<th>Human timescale</th>
<th>Geological epoch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
<td>kya</td>
</tr>
<tr>
<td>Paleolithic</td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>2500–100/200</td>
</tr>
<tr>
<td>Middle</td>
<td>300–30</td>
</tr>
<tr>
<td>Upper</td>
<td>50–10</td>
</tr>
<tr>
<td>Mesolithic</td>
<td>10–6</td>
</tr>
<tr>
<td>Neolithic</td>
<td>10–4</td>
</tr>
<tr>
<td>Bronze Age</td>
<td>5.3–2.4</td>
</tr>
<tr>
<td>Iron Age</td>
<td>3.3–1.6</td>
</tr>
</tbody>
</table>

*Table 1.1* Approximate human and geological timescales expressed in thousands of years before present (kya). Definitions for human timescales can vary depending on region.

*Figure 1.1* Causes of death in the United States have changed in the twentieth century. Data plotted from Hoyert (2012).
1.1 What is Disease?

Whereas modern medicine focuses on the concept of health, evolutionary biology focuses on the determinants of optimal fitness. Importantly, fitness can only be defined in relationship to a particular environment: for example, polar bears have evolved with a physiology suited to the Arctic but not appropriate for a temperate world. This conceptual difference between health and fitness is critical for understanding the human condition and in defining normality and the disease process.

Consider a young man who presents with abdominal pain, bloating, and diarrhoea. He is a recent immigrant from Southeast Asia with no history of these symptoms. He reports that yesterday he shared lunch with work colleagues during which he consumed a couple of glasses of milk and had a plate of ice cream. This was unusual for him, but his colleagues, who are of European ethnicity, were unaffected. Why is this young man made ill by ingesting a normal foodstuff?

Cows’ milk, like the milk of most mammals, is rich in the disaccharide lactose. The sugar transporters in the human gastrointestinal tract cannot move intact lactose across the gut wall, but babies can digest lactose because of the presence of the enzyme lactase, which breaks down lactose into easily absorbable glucose and galactose. In most humans, lactase expression in the intestine disappears after weaning, but human populations with a history of pastoralism—mostly people of northern European or East African origin—have a high prevalence of mutations in the promoter region of the lactase gene, causing the enzyme to be expressed within the intestinal tract throughout life. This enables them to consume milk throughout their lives.

But this young man of Asian origin does not carry the persistence mutation and therefore does not express lactase in his duodenum. The discomfort that he experiences after drinking milk is caused by the osmotic load of the unabsorbed lactose, and by the gas produced by fermentation of the sugar by intestinal bacteria. Regardless of the molecular details, his symptoms arise from a mismatch between his genetic origin—from a population where, historically, consumption of milk after weaning was unknown and lactase persistence is rare—and his current environment where milk is easily available and widely consumed.

His friends could drink milk freely because their forebears inherited the lactase persistence allele which was selected for in populations that herded cows. Because the ability to absorb milk had a nutritional advantage, and allowed individuals to grow, survive, and reproduce more successfully, the mutant form had become the most common allele in their forebears. This example is central to the purpose of this book, because Western medical textbooks often define the inability to absorb lactose as a metabolic disorder—adult hypolactasia—but from an evolutionary point of view this man’s inability to digest lactose is normal and is shared with 70% of the world’s population. It has only become manifest in an environment distinct from that to which he is adapted.

Parenthetically, very rarely there are mutations in the expressed sequence of the lactase gene which means that even the infant cannot digest its mother’s milk. Such congenital hypolactasia is a severe disorder that requires urgent attention and lactose-free nutritional support. Congenital hypolactasia would have been fatal in the past when milk, human or otherwise, was the sole source of infant nutrition, and thus remains a rare autosomally recessive condition.

There are several lessons to be learned here. The first is that our understanding of an individual’s health status may depend on our knowledge of their evolutionary origin and how that interacts with the place where they now live. This concept of an organism matched or mismatched with its environment is fundamental to both evolutionary biology and evolutionary medicine, where mismatch—that is, failure to adapt because of temporal or structural constraints—may lead to pathology.

The second lesson from this case is broader: it is not always easy to define disease. Disease may be caused by an external agent such as trauma or infection, but it can also arise from a mismatch between the physiology of an individual and the environment in which they live. The physiology of an individual is influenced by their evolutionary history. Thus can we really say that the majority of humans have a disease because they do not carry a single nucleotide polymorphism which emerged in a subpopulation
and that causes lactase to persist into adulthood—a deficiency that is of no consequence for their health or fitness in the context of an environment free from cows’ milk? Rather, should we label the species-atypical state of lactase persistence in people of northern European origin as the abnormal condition, and then reflect on the context-sensitive dichotomy between abnormality and ill-health?

The third lesson relates to the impact of novel environments on human health, and how the capacity to cope with a range of environmental exposures is in turn determined in large part by our evolutionary history. Exposure to milk after infancy represents a novel environment for people of Asian lineage and thus they have not evolved the capacity to cope with it (Box 1.1).

**Box 1.1 The Genetics of Drinking Milk**

For most adults of northern European or East African origin, milk is a normal part of the diet. But for the majority of the world’s population, drinking milk would be the precursor to some unpleasant gastrointestinal symptoms. The reason for this is that most adult humans cannot digest the disaccharide lactose, which is a major constituent of milk. The World Health Organization classifies such “lactose intolerance” as a metabolic disorder, although in fact this trait represents the normal and ancestral human condition.

All young mammals rely on their mother’s milk, which is high in the disaccharide lactose, for early nutrition. They can tolerate the lactose because their small intestine contains the enzyme lactase, which breaks down lactose into the two easily absorbed sugars glucose and galactose. But after weaning, production of the enzyme is largely switched off: most animals will never encounter large amounts of lactose again, so why waste resources synthesizing the enzyme? In most human populations, lactase is lost between the ages of 2 and 5 years. Without lactase in the intestine, lactose in the diet cannot be absorbed and causes gastrointestinal upset, both because of its osmotic activity, causing water to be drawn into the gut, and because intestinal bacteria ferment the sugar, leading to gas production, bloating, and cramps.

Yet only a minority of the world’s population continues to express lactase and is able to drink fresh milk in adulthood—that is, they show “lactase persistence.” This ability is genetically transmitted as a dominant trait, and in general lactase persistence in a population correlates with that population’s history of domestication of cattle. Although milk can be used by people who lack the enzyme—the lactose content of milk can be decreased by processes that encourage the growth of lactose-hungry micro-organisms such as by allowing it to sour or by making cheese—individuals who can digest fresh milk benefit from the additional energy obtained from the lactose. Thus, in the absence of major costs of maintaining expression of lactase it is easy to see how strong selection pressure for lactase persistence would have favored the retention of this capacity into adulthood in cultures that domesticated cattle (Figure 1.2).

The genetic basis for lactase persistence in populations of northern European origin has been traced to a single nucleotide polymorphism, C/T (–13910), in a regulatory element upstream of the lactase gene, with the T allele completely associating with lactase persistence (Figure 1.2). Estimates from modern populations for the age of this allele center on about 10,500 years before present, roughly in line with estimates of the introduction of domestic cattle to Europe some 8000–9000 years before present. Furthermore, there is evidence from Neolithic and Mesolithic remains that the allele was not widespread in Europe before that time, suggesting that selection for the allele occurred after the introduction of dairying rather than supporting the competing hypothesis that dairying was only adopted in populations that already carried high levels of the allele.

Europeans are by no means the world’s only cattle herders, and there are pastoralist populations in East Africa who also show lactase persistence (Tishkoff et al. 2007). Those populations do not carry the C/T (–13910) allele, but instead carry several other genetic changes in the same regulatory region upstream of the lactase gene, of which G/C (–14010) is the most common and the most tightly linked to lactase persistence. That allele appears to have spread to its present high frequency in Kenyan and Tanzanian populations within the past 3000–7000 years, one of the strongest examples of recent positive selection in the human genome. The later appearance of the G/C (–14010) allele in East Africa compared with the C/T (–13910) allele in Europeans is consistent with archeological evidence dating the spread of cattle domestication into those areas. Yet another lactase-persistence allele appears to have arisen in Arab populations, possibly associated with consumption of camels’ milk (Enattah et al. 2008). These independent origins of lactase persistence are examples of evolutionary convergence, wherein natural selection arrives at a similar functional trait in different populations owing to similar selection pressures.
European Americans 81–98%
Hispanic Americans 48%
African-Americans 23–30%
Native Americans 0–5%
Asian-Americans 0–5%

17%

European Australians 90–95%
Native Australians 15%

1%

Figure 1.2 Worldwide distribution of lactase persistence.
Physiological systems can generally maintain homeostasis in the presence of a degree of environmental variation, but there are limits to that capacity. When those limits are exceeded, disease can occur. The range of environments to which a lineage has been exposed in its evolutionary history will influence that range of adaptability, and human health can be compromised by living in marginal environments beyond our homeostatic capacity. For example, humans cannot live in an iodine-deficient region because of the importance of iodine for metabolism and brain growth. The Sherpa population of the Himalayan foothills lives in a permanent state of iodine insufficiency, and this is reflected in both developmental disruption (impaired cognition because of poor brain growth) and compensatory plasticity in the form of goiter, in which the thyroid gland is enlarged in an attempt to maintain adequate production of thyroid hormone.

The fourth lesson is that the environments inhabited by humans are not constant, and much of this environmental change comes from the activities of humans themselves. The young man’s discomfort after drinking milk can be linked to two environmental changes caused by human activity. The first is the historical domestication of cattle and the concomitant selective pressure for lactase persistence, together resulting in widespread use of cows’ milk as a component of the adult diet of certain populations. The second is the recent social and technological changes that have encouraged large-scale migration and mixing of individuals from different cultures and evolutionary backgrounds.

Another change in the social environment, discussed in more detail in Chapter 11, is the increasing size of social networks. In Paleolithic times it is thought that an individual would only have interacted with about 150 other humans over his or her lifetime, as our ancestors lived in small, isolated social groups. The invention and adoption of agriculture (see Sections 6.3.10 and 9.3.2) committed humans to living in an increasingly complex and dense social network. This became progressively more complex following the Industrial Revolution, and has of course been magnified still further by modern technologies. The Nobel prize-winning economist Robert Fogel has proposed that the improvements in health and knowledge over the past few centuries are mutually reinforcing, a process he calls techno-physiological evolution, leading to an accelerating rate of technical change and population growth (Figure 1.3). Nevertheless, as we will discuss in Chapter 11, there is growing evidence that the mismatch between the social environment in which we evolved and that in which most humans now live has consequences for mental health. Throughout this book we will see many examples of rapid environmental change leading to pathological consequences because of constraints on our speed or capacity to adapt.

1.2 What Evolution Is: Fundamental Principles

Evolutionary biology is fundamentally concerned with the various processes that have determined the “design” of the human body at all levels, from how we interact as whole organisms with other members of our species to every component and level of our internal biological organization. Design is a frequently used term in the evolutionary literature. It is a metaphor, used as shorthand to describe the various processes by which a species evolves, such that its characteristics—anatomical, physiological, biochemical, maturational, and behavioral—fit the environment in which the population lives. It does not in any way imply the existence of a designer; it just happens that it is much easier to describe processes using this metaphor, but we must always remember that it is a metaphor and no more.

Each evolved characteristic of a person is often described as an adaptation, although strictly adaptation as used in evolutionary biology refers to those evolved elements (or traits) that can be shown to have promoted fitness. A further commonly used metaphor is strategy, which allows us to describe the functional significance of these adaptations. We will expand on these concepts in Chapter 2.

In evolutionary thought it is important to avoid the trap of describing an evolutionary process, or evolution itself, as having a purpose or direction: to do so is a form of teleology. One of the dangers of the design and strategy metaphors, or of speaking about higher or lower species, is that they can encourage
1.2 WHAT EVOLUTION IS: FUNDAMENTAL PRINCIPLES

such thinking. There is a major difference in the thought processes underlying the statement that “limbs evolved for walking,” which is teleological, and the evolutionary statement that “there was progressive selection over time on the traits associated with the ancestral fin, and the adaptive advantage associated with effective terrestrial movement led to cumulative selective change resulting in the formation of the limb.” But we can see how clumsy the second statement is, so it is easy to be sloppy and say that a process or a structure “evolved for ....” It is an almost unavoidable temptation, but when we use such language we must also remember that it does not imply purpose.

The beginning of modern evolutionary theory in the late eighteenth century was based in the growing acceptance of two fundamental concepts. The first was gradualism, the idea that the geological features of the planet are the result of slow processes operating over “deep time.” The second concept was that biological species are not immutable but that, with time, new species could emerge, evolve into other species, or become extinct, and that in biology, as well as geology, deep time provides a setting for such gradual change. Although macroevolution is a large component of evolutionary biology, it is not a major focus of this book beyond a brief consideration of the evolutionary history of the hominin clade since the last common ancestor that humans and their direct ancestors shared with chimpanzees (Box 1.2; Chapter 6). But Charles Darwin recognized that species could change their characteristics over time, and we call this within-species change microevolution. Importantly, macroevolution and

Figure 1.3 Anthropogenic changes in the human environment. From Fogel (2004), with permission.
microevolution are a continuum, not distinct processes. Classically, evolution is therefore defined as “change with time within a population of organisms over generations,” and the insights of Charles Darwin and his correspondent and competitor Alfred Russell Wallace were to see that this change is a result of selection acting on heritable variation in the traits within a population.

Most evolutionary biology has focused on genetic inheritance. Indeed, it was the discovery of the gene as a mechanism for inheritance which eventually moved selection to the center of biological thinking. However, as we will discuss in Chapters 2 and 3, there are other modes of inheritance, and increasingly these have been incorporated into evolutionary science.

1.2.1 Selection

Selection describes the processes by which one individual is more likely to reproduce successfully than another within a population because of the possession of a particular advantageous trait. It is itself a word which has metaphorical danger. If there is a genetically inherited component to that particular trait, then that particular trait will become concentrated in the next generation of that population. Darwin’s seminal contribution (Darwin 1859) was to recognize the importance of selection. He introduced his most famous book, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (shortened in later editions to *The Origin of Species*), with a discussion of artificial selection, which is the form of selection known to plant or animal breeders who select for a particularly desirable trait by breeding from individuals with a variation in that trait in the desired direction. Artificial selection, as with the other forms of selection, can only operate to the extent that it is heritable, and it was Darwin’s observations of animal breeding which gave him insights into this process, even though at the time there was no widely accepted theory of how inheritance might operate (Darwin did not know about the work of Gregor Mendel; see Section 2.3.4).

In *The Origin of Species* Darwin makes one of the greatest intellectual leaps in science to recognize that selective processes also operate in nature. He
recognized that natural variation in a trait might make an individual organism more or less likely to survive and reproduce successfully in a given environment, and that therefore as that trait became concentrated in the population over time the lineage would evolve to be well adapted to a particular environment. He called this process natural selection by analogy with artificial selection. But whereas artificial selection has a direction determined by the selective agent (the breeder), any direction in natural selection is solely determined by the selecting environment.

Darwin recognized that selection is about reproductive success. Natural selection is one mechanism for achieving this, but in his second major book, The Descent of Man, and Selection in Relation to Sex (Darwin 1871), Darwin described another mechanism, whereby reproductive success is not related to characteristics affecting survival but rather to sexual dominance and choice. In a species where there is competition (usually among males) for mating opportunity, individuals may fight for sexual dominance and thus there is sexual selection for body size, weapons, and strength. For example, the male elk is much larger than the female and has developed large antlers for battling other males. In other animals, particularly birds, mating choice by the female may be dependent on the physical appearance of the male and this can lead to the evolution of sexually dimorphic extravagances such as the peacock’s tail. We will further discuss the different forms of natural selection, sexual selection in relationship to human biology, and the different levels at which selection can act in Section 2.3.2.

1.2.2 Not Everything has Evolved by Selection

As we discuss in Section 2.4.3, not every change in a lineage’s properties is due to selection. If a small part of the population is permanently isolated from the rest of a population, and this new isolate does not possess the full range of genetic variation, this means that: (1) the range of genetically determined traits that it demonstrates will differ, and (2) the path of subsequent selection for that isolate will differ from that in the parent population. The mechanism leading to the phenotypic consequences of such limited sampling from a parent population is called genetic drift. Such drift explains a number of differences in disease susceptibility and patterns in different populations that evolved through a population bottleneck (see Section 3.4.1).

In the main, mutations occur randomly, and many mutations have no biological effect either because they do not change the amino acid coded for or because they do not affect a regulatory or expressed component of biological significance in the environment in which the mutation occurs. Such mutations are known as neutral mutations. However, if the environment changes the consequences may then become apparent—we will explore this phenomenon in relation to the evolutionary explanation for scurvy in Box 6.3.

1.2.3 Variation and Inheritance

The key features of the preceding discussion are the themes of variation and inheritance. Recognizing their importance was another of Darwin’s seminal contributions. Up to Darwin’s time, biologists (and doctors) had largely been concerned with classifying all living organisms; they therefore concentrated on similarities—as defining the average for a species—rather than variations (Box 1.3). Darwin shifted the focus to individual variation within a species. The most remarkable aspect of Darwin’s contributions and those of the other early evolutionists, including Alfred Russell Wallace, is that they made their observations in the absence of any understanding of how variation and inheritance might operate.

It was the discovery of particulate inheritance—the idea that characteristics could be inherited as discrete units—that led to the development of genetics and the study of mutation. The field was slow to develop because the principles of inheritance elucidated by Gregor Mendel were ignored for 20 years until their rediscovery at the beginning of the twentieth century. There were bitter disagreements between selectionist biologists and genetic biologists until the “Modern Synthesis” (see Sections 2.3.4 and 14.5) brought these two fields together in an integrated concept.

The discovery of DNA and the power of molecular biology added impetus to this integration. In particular these discoveries provided a mechanism
1.2.4 Development and the Life Course

One area of biology was more difficult to include in the synthesis between molecular and selectionist thinking, namely the impact of developmental processes operating from conception to maturity. Development is not simply a matter of a fertilized ovum growing and dividing according to a pre-programmed mechanism. There are complex pathways of differentiation from a single cell into an adult human and there are distinct components to the life course: from pre-implantation embryo, to implanted embryo, to fetus, to neonate, to dependent infant, to juvenile, adolescent, and adult. In recent years, major progress has been made in integrating our understanding of developmental plasticity—a set of processes which have themselves evolved—with the rest of evolutionary thought, and this remains perhaps the most contentious and complex component of evolutionary theory, with ongoing robust debate (Laland et al. 2014).

Organisms have different biological strategies at different times in their lives. Some organisms have very distinct forms and can have very complex life courses: for example, the human parasites causing malaria (a protozoan) and schistosomiasis (a worm) have very different body forms, or morphs, according to whether they are living in humans or in their invertebrate vectors (mosquitoes and water snails, respectively). Humans too have distinct phases for heritable variation, but molecular architecture also provides powerful evidence for how evolution has progressed. When used in relation to speciation, for example, the evidence from DNA goes far beyond the fossil records with which Darwin and his colleagues had to content themselves. In Chapters 3 and 6 we review how we can use our understanding of molecular biology both to explain and demonstrate evolution and to explore evolutionary history.

Subsequent evolutionary research has essentially built on these fundamental principles to describe the origin of variation: the extent to which it is driven by mutation, the link between genotype and phenotype, the role of chance, the speed of evolution, the level at which selection operates, and the modes of inheritance.

Box 1.3 Seeking Similarities or Seeking Differences

Humans have long classified living things on the basis of their appearance, behavior, or habitat, but our present understanding of biological species owes much to the work of two scientists whose lives were separated by just 100 years, the eighteenth-century Swedish naturalist Carl Linnaeus (1707–78) and the nineteenth-century English naturalist Charles Darwin (1809–82).

In 1758, Linnaeus published his *Systema Naturae*, a book describing all the then known species of plants and animals and also containing a scheme for classifying them: the Linnaean system of taxonomy with its hierarchy of species, genera, families, orders, classes, phyla, and kingdoms. Linnaeus introduced the binomial system of species names—for example *H. sapiens*—and proposed that species are the fundamental unit of classification. Although he recognized that species are usually classified by morphology—members of a species tend to resemble each other more than they resemble members of other species—he appreciated that the underlying definition of a species is the ability of individuals within a species to interbreed and produce viable offspring. Linnaeus’s world view was that species were fixed and immutable, and that each species had been created by a higher power to fill a place in the sequential ladder of life.

Any biography of Charles Darwin will recount the story of how his desire to be viewed as a credible biologist was dominated by a 7-year study of the taxonomy of barnacles. That endeavor taught Darwin that within a given species there is a wide spectrum of variation, and that it is often difficult to decide where variation within a species shades into a distinct species. The appreciation of such variation helped to reinforce his developing view that species were not immutable and that selective retention of beneficial variation might be a mechanism for one species to change into another.

The Linnean Society of London, the world’s oldest learned society devoted solely to natural history, takes its name from Linnaeus and maintains his botanical, zoological, and library collections. The papers by Darwin and Wallace that first proposed the theory of evolution by means of natural selection were read at a meeting of the Linnean Society on July 1, 1858.
in their life courses. For example, the nutritional processes of the pre-implantation embryo, fetus, pre-weaning infant, and post-weaning child are all very different. The human fetus is nourished across the placenta so both placental physiology and the mother’s adaptations to pregnancy are attuned to maximize the nutrition of the fetus. The fetal gastrointestinal tract is quiescent until birth, at which time it is activated and uniquely adapted for the digestion of human milk, for example of lactose using lactase. After weaning, human gastrointestinal physiology changes yet again, one feature being that the bulk of the world’s population loses lactase activity.

One component of evolutionary biology is to consider how different phases in the life course evolved and how these different phases are inter-related. An organism’s life-course strategy is ultimately an issue of the optimal use of energy supplies to maximize reproductive success. Selection therefore operates on components of this balance, such as investment in growth, patterns of development, approaches to reproduction, social structure, and patterns of ageing, all aimed at maximizing fitness within the environment of the population. Trade-offs must be made, given the limiting availability of energy and the constraints of time—the risk of death from predation or disease before reproduction. Body size itself is constrained by mechanical and thermoregulatory issues and nutrient availability. The element of evolutionary biology related to consideration of these general strategies and trade-offs is known as life-history theory (Stearns 1992). Humans have very distinct characteristics to their life-history traits, and these and their health implications are discussed in Chapter 5.

Just as selection is the process of the interaction between inherited determinants of phenotype and the environment, development is not a solely intrinsic process in which a single ovum grows into a mature organism. Rather, the developing organism is subject to external influences which affect its later phenotype (Chapter 4). Vulnerability to such influences is enhanced during particular critical windows of development occurring at different stages according to the nature of the environmental stimulus. At the most pathological level, exposure of the developing embryo to certain drugs can interfere with cell replication and interaction. For example, the anticancer drug methotrexate, which crosses the placenta and interferes with folic acid metabolism, can cause spontaneous abortion (at high doses) and fetal defects (at moderate to low doses), particularly when given around the time of conception or in the first trimester of pregnancy. This is a pathological example, but environmental variation within the normal range can also affect fetal development. For example, exposure to severe famine during pregnancy can lead to offspring with a low birthweight, and as we will see in Chapter 9 this has lifetime health consequences, including a predisposition to a greater risk of non-communicable disease (Roseboom et al. 2006).

### 1.3 Time

Organisms face challenges over a number of timescales. These include immediate physical challenges, such as intraspecific competition for energy and reproductive opportunity, attack by predators (themselves engaged in energy harvesting), and changes in the environment. The latter may be short term (e.g., daily temperature fluctuations), medium term (e.g., seasonality of food availability), or long term (e.g., climate change causing the disappearance of preferred food sources or habitat).

Organisms have evolved a hierarchy of responses to these challenges. Some classical homeostatic responses are very rapid and highly reversible over seconds to hours, for instance those mediated by the central nervous and endocrine systems. Some involve structural change or long-term readjustment of set points for homeostatic feedback systems (called rheostasis or homeorhesis), and these operate over hours. Many longer-term but within-lifetime effects are initiated in early life through the processes of developmental plasticity. Yet other responses are beyond the timescale of individual lifetimes and involve natural selection, resulting in genetic change over a few or many generations (Figure 1.5).

An example of a short-term (homeostatic) response is sweating or shivering in response to changes in temperature. Increased muscle size, perfusion with blood, and changes in myofiber metabolism after physical training provide an
example of a medium-term and reversible plastic response affecting tissue function or organization. Stunting represents an example of developmental plasticity. Children who are chronically undernourished, either in utero or in infancy, may irreversibly reduce their degree of somatic growth. Reduced body mass during childhood and adulthood reduces nutritional needs, thus allowing the individual to better cope with an environment that is likely to be nutritionally limited over their lifetime. Here we have the evolved processes of adaptive developmental plasticity acting chronically within a lifetime to promote survival. While the propensity to be stunted in response to developmental undernutrition is an inherited capacity, stunting itself is generally not. Nevertheless, stunted mothers may give birth to children who are smaller because of the phenomenon of maternal constraint (see Sections 8.9.5 and 9.4.3.2), and due to the similarity of the environment to which mothers and children are exposed one often sees intergenerational stunting.

In this book we will often consider the evolution of phenotypic traits that are inherited across generations—this has been the mainstay of much evolutionary biology. Humans from lineages that remained in tropical Africa from our initial appearance as a species have generally different body shapes from those from lineages that have lived in higher-latitude environments for many millennia. Here selective pressures are likely to have favored different body shapes to aid thermoregulation in the very different climates these lineages have faced.

As another example of temporally different classes of response, consider the hierarchy of responses affecting that most extensive and sensitive of human organs—the skin. We respond to a pinprick with immediate withdrawal, on a timescale of less than a second, mediated by cutaneous nociceptors. On a slightly longer timescale of minutes to hours, another insult—ultraviolet radiation—causes sunburn characterized by erythema, pain, and edema. That also invokes a behavioral avoidance response: we cover our skin. The sensitivity of skin to ultraviolet radiation is determined by its melanin content, and for many people with light skin the melanin content can be increased by a medium-term adaptive response that increases output from melanocytes, so a suntan develops over days to weeks. And long-term genetic adaptations of human populations to the amounts of sunlight at different latitudes have created a variety of human skin colors caused by variation in the amount and type of melanin in the skin. It is likely that the selective pressure for lighter skin color in populations living at higher latitudes arose from the need to preserve vitamin D synthesis in the skin in regions with lower availability of sunlight. However, this can result in vitamin D deficiency in darker-skinned migrant populations, especially when traditional concealing clothing is worn by women (see Section 13.4.1).

It is also important to appreciate how responses on the shorter timescales can be modified by an individual’s previous developmental experience or evolutionary history. For example, susceptibility to sunburn is modified by past exposure (suntan) and by population origin (skin color). Susceptibility to the effects of milk consumption is modified by developmental stage (lactase is lost during development)
and by ancestral nutritional history. These examples underline the importance of a complete knowledge of an individual’s history for understanding their current health status.

1.4 Constraints

Evolution is not without limits. Such constraints may arise from the nature of physical processes. For example, insects are—perhaps fortunately for human survival on the planet—limited in size by gravity, the mechanics of their exoskeleton, and the physics of oxygen diffusion along their tracheal network. Or constraints can arise because evolution generally works in an incremental way: for example, the human eye is poorly “designed,” but gradual evolution towards the structure of the superficially similar but optically superior octopus eye would require non-functioning intermediate forms, so we have retained the organizational structure of our eye in a form modified from functioning eyes in our precursor lineage (see Section 2.3.1.3). We must remember that most of the discussion so far has been about qualitative aspects of the environment, but constraints will also act when there is a mismatch between the rate of environmental change and that at which natural selection can operate, or simply from an extreme degree of environmental change.

The complexity of the link between the genotype and phenotype places all sorts of constraints on how selection can operate. As will be reviewed in Chapter 3, the concept of a single gene relating to a single trait is in general misleading, as coding genes can produce multiple products due to alternative splicing. The growing understanding of regulatory networks within the genome highlights the importance of the phenomena of epistasis (multiple genes acting together to produce a trait; Section 3.6) and pleiotropy (one gene has effects on multiple traits; Section 5.2.4).

Single-gene effects may lead to mutations and disease, but selection operates on the integrated phenotype and this creates limitations on the rate and nature of change between generations. Consider childbirth in humans (see Section 8.9.5). The shape of the human pelvis is quite distinct from that of other apes. This is because we are bipedal and the mechanics of efficient walking and running have led to selection of relatively rotated hips and a narrower pelvis. Yet there has to be a trade-off, because humans are also distinguished by a large brain. Thus if birth occurred at the same stage of neuronal maturation as in other apes, with well-developed motor function at birth, the fetal head would be too large to pass through the pelvic outlet. Hence, while other apes are precocial (having relatively mature offspring at birth capable of a level of independent activity), humans have developed secondary altricial characteristics and are extremely dependent on the mother for a long period after birth. Here is an example where a constraint was imposed on how human evolution could proceed by the development of bipedalism early in the evolution of the hominin clade (see Section 6.3.2).

There are also constraints imposed by biochemical and physiological processes, which have in turn been imposed by the past evolution of the lineage. For example, human physiology has evolved such that we cannot live and reproduce successfully above about 5000 m, and cannot survive at all without supplementary oxygen above 9000 m. In contrast, the bar-headed goose migrates regularly over the Himalayas at altitudes of up to 10,000 m.

Constraints and exaptations (features that perform a function but were not initially selected for their current use; for example, the three small bones of the middle ear that originally evolved in jaw-boned fish as part of their jaw hinge mechanism), and “historical contingencies” (chance events that have determined, for example, why terrestrial life is based on L- rather than D-amino acids and why the human and cephalopod eyes look superficially similar but have fundamentally different architectures), have also played a role in determining our form and function.

1.5 We Are Not Alone

Humans do not live in isolation from other species. Here it is useful to introduce the concept of coevolution, a process where two species exert reciprocal selective pressures which affect the evolution of both. The lactase story shows how some humans have evolved to live alongside cows, and after 10,000 years of artificial selection for milk or
meat yield, modern breeds of *Bos taurus* are very different from the wild cattle first domesticated by Neolithic humans. Another example of coevolution is a predator–prey relationship in which the prey evolves some defensive or escape mechanism and the predator in turn evolves a means of overcoming the prey’s response. In predator–prey interactions, coevolution can lead to an evolutionary arms race in which the target species must continually change to maintain its fitness relative to the predator species it is coevolving with. This is the so-called Red Queen effect, named after the character in Lewis Carroll’s book *Through the Looking Glass* who complains of always having to run to stay in the same place.

The biggest group of organisms affecting our biology, for good or bad, is our associated microorganisms. Infectious disease was a major cause of mortality until the introduction of mass vaccination programs and antibiotics very recently in human evolutionary history, and the threat is still with us: the last few decades have seen the emergence of new viral diseases, such as acquired immune deficiency syndrome (AIDS) and severe acute respiratory syndrome (SARS), as well as the resurgence of bacterial diseases once thought to be conquered, such as tuberculosis, now in a lethal multidrug-resistant form. Humans, in common with other vertebrates, have evolved an adaptive immune system to resist infection with micro-organisms, and our antimicrobial phenotype has recently been extended by the use of biotechnology to develop antibiotics and antimicrobial and antiviral drugs. But our pathogens are also engaged in this arms race, and can use the advantages of vast population sizes and simple genomes to rapidly evolve mechanisms for evading immune surveillance and neutralizing our antimicrobial drugs. As described in Chapter 10, the eventual outcome of that competition is often uncertain.

Of course, not all micro-organisms are pathogenic. Some, such as yeasts and lactobacilli, have achieved a certain utility in food and beverage preparation, and others are used as chemical reactors in industrial processes after fierce bouts of artificial selection to increase their yield. Still other micro-organisms are engaged in a mutual relationship with humans not dissimilar to that between humans and their domesticated animals. The many hundred species of bacteria that comprise the human gut flora perform critical functions, including digestion of some food components, production of vitamins, protection against pathogenic micro-organisms, and fine-tuning of the immune response, in return for life in a stable and protected environment with a regular supply of nutrients (see Section 10.3). The consequences for our health of perturbations in the composition or activity of the gut flora are being increasingly recognized, and are leading to new insights and approaches to therapy (e.g., fecal transplants).

### 1.6 Culture and Gene–Culture Coevolution

Culture can be defined as the sum of socially transmitted information obtained through the processes of teaching, imitation, and other forms of social learning (see Box 2.6). This includes not only skills involving physical constructs such as tools and technologies, but also art forms, beliefs, and the social mores of a species. Culture itself evolves, but often horizontally within a generation, rather than solely vertically as in biological evolution. Humans have particular and unique capacities to develop material, behavioral, and social cultures, because of their evolved cognitive, language, and manual capacities (Chapter 2). But just as culture evolves, so too can it influence biological evolution, and vice versa—this interaction is known as *gene–culture coevolution* (Richerson and Boyd 2005). The example of the development of lactase persistence is a classic example. Humans, through their cognitive and social capacities, developed the cultural practice of herding and milking cows. This in turn led to the biological evolution of a lineage with persistent lactase expression, which in turn favored more dairy husbandry, and so on.

### 1.7 How Evolutionary Arguments Fit Alongside Other Biological Perspectives

When considering biological phenomena it is important to appreciate that evolutionary questions cannot and should not be considered separately from other perspectives. A valuable approach is
Box 1.4 Applying Tinbergen’s Four Questions

Tinbergen’s questions can be explicitly applied in a systematic manner to many biological phenomena. Consider how and why we sweat when we are frightened. We know that sweating is regulated by innervation of the sweat glands, and the mechanistic or proximate cause must be stimulation of the sympathetic nervous system (Question 1).

We also know that infants do not have a mature thermoregulatory system, and indeed the sweat glands are not active until innervation is completed some months after birth. But in terms of the reflex relationship between fear and sweating, it must also rely on maturation of the capacity of the infant to perceive a threat (Question 2).

The function of this behavior is likely to have had adaptive value as part of a flight response: it allows the organism to lose heat as it prepares to take flight (Question 3).

The evolution of the fear and flight response is key to any species that faces the risk of predators. All mammals have such a flight response, which is generally associated with activation of the sympathetic nervous system. A successful flight response will require the capacity to lose heat. Homeothermic species must be able to thermoregulate, and most mammals lose heat either by sweating and/or panting. Humans, with little hair covering their bodies, primarily sweat to lose heat. Thus, sweating may have evolved as part of the thermoregulatory system but has been integrated into the fear response (Question 4).

1. What is the mechanism underlying the phenomenon of interest?
2. How does the phenomenon develop during the lifetime of the individual? That is, what is its ontogeny?
3. What is the function of the phenomenon? How does it serve the organism’s interests?
4. How did the phenomenon evolve? What is its evolutionary history, are there analogous phenomena in other species, and what is their evolutionary relationship to the human? What is the evidence for a selected origin?

These questions are often implicit in the discussion throughout this book. Another way of considering this range of perspectives is to consider causation at two levels. At one level there are the molecular, anatomical, physiological, and pathophysiological mechanisms that lead to any biological phenomenon. Thus, insulin resistance leads to type 2 diabetes mellitus, infection with the human immunodeficiency virus (HIV) may progress to AIDS, a vertebral disk prolapse leads to back pain and sciatic nerve injury, and inflammation of the appendix leads to appendicitis. These are called proximate causes. But there is another level of explanation: why is it that some people are prone to develop obesity and insulin resistance, why is it that humans are susceptible to HIV infection, why is it that so many humans get back pain, and why do we have the appendix, a useless organ that gets infected? These questions are about ultimate causes: the answers lie in the domain of evolutionary biology and are the topic of this book.

1.8 Evolution and Medicine

Modern biology rests on two related concepts. The first is the description of the gene and its structure; the second is the description of the evolution of the whole organism. Despite their close interrelationship, it is only recently that genetics and evolutionary biology have been seen to be fully compatible. This recognition, which occurred in the middle of the twentieth century, is so important that it is termed the Modern Synthesis. Although this synthetic view is well understood in the biological sciences, its relevance to human biology and medicine is still emerging.

Medical science has become dominated by relatively reductionist approaches. That is, it has
tended to regard individual levels of organization (the gene, the cell, the tissue), or individual organ systems, or the different disciplines (physiology, biochemistry, anatomy) as quite distinct. This approach, while necessary at one level, is not good for medicine or for the patient. Just as an integrated holistic approach to medical care and public health is optimal, so a multilevel approach aids our understanding of the etiology and mechanisms of disease.

An evolutionary perspective changes the way doctors think about health and disease (Gluckman et al. 2011a). It not only helps to identify research questions but also allows engagement with individual patients in ways that promote understanding of their current health status and also contribute to the design of appropriate interventions in public health. Because so much of public health depends on an ecological perspective—that is, an understanding of the social and broader environments—evolutionary biology also has particular application in public health, and many examples will be given in other chapters of this book.

Human biologists and anthropologists apply evolutionary principles as a fundamental component in their research and teaching, but their focus is, by definition, on the normal human condition. Medicine has been slow to recognize the importance of such thinking to concepts of normality versus abnormality, the limits of adaptation, and the consequences of maladaptation in the etiology of disease.

In medicine we are concerned primarily with the current state of the individual, but to understand that state we must understand that person’s biology and the context in which she or he lives. Traditionally, taking a medical history has focused on the story of the patient and their environment from the time of their first symptom. But, as this textbook will emphasize, that history must also extend to understanding the development of the individual from their conception, and beyond to transgenerational influences mediated through genetic and cultural inheritance. An evolutionary perspective ensures this integration: in this book we will see how an evolutionary perspective assists the doctor or health professional in developing such an approach and in placing individual components of human biology in perspective.

The immediate determinants of an individual’s biological (health) status—that is the world they live in (the environment), how they live in it (their behavior), and how they function (their physiology)—are all proximate causes. Modern medicine has to a large extent been concerned with such proximate causes and with defining whether a given phenotype, or indeed a patient’s response, is “normal” or not. This textbook does not focus primarily on acute pathophysiology or the details of environmental stimuli such as stress or infection, but rather on explaining how an individual’s response to them evolved and how this influences the potential for health or disease, and thus decisions about its management. In other words, we are concerned primarily with ultimate causes: how evolution has led to the persistence of a particular trait or set of traits. This leads to further questions about whether the adaptation is helpful or not under present circumstances, and whether the limits of adaptability have been exceeded.

Key Points

- Evolution acts primarily by selection on heritable variation among a population of organisms.
- Selection maximizes fitness (lifetime reproductive success) and not necessarily health or longevity.
- Selection is not the only evolutionary process.
- Health and disease are influenced by an individual’s evolutionary and developmental background in the context of his or her current environment.
- Consideration of ultimate causation provides additional insights to understandings of proximate mechanisms of health and disease.
- Concepts of health and disease are altered by taking an evolutionary perspective.
**CHAPTER 2**

**Evolutionary Theory**

**2.1 Introduction**

Evolutionary biology in essence provides explanations for two sets of phenomena: firstly how a plethora of species emerged from a single common ancestral species in a series of descendant and radiating lineages (macroevolution), and secondly how organisms come to be well matched to face the threats and opportunities in the environments they inhabit (microevolution). No area of science has produced such a voluminous popular literature, or one as polarized. Perhaps the best comparison would be with the debates surrounding Copernicus and Galileo in the sixteenth and seventeenth centuries. Both these scholars challenged and ultimately overturned a belief-based system which placed the Earth at the center of a religiously defined cosmos.

In a similar fashion, evolutionary biology has gradually removed the need, once central to the Western religious tradition, to believe that a deity created humans as a special form of organism or controlled the design and macroevolution of the millions of living species, either extinct or extant, that have inhabited the Earth over the past 4 billion years. In contrast to the belief that a deity or supernatural force intervened in the making of the natural world, known as creationism, evolutionary biology requires no agents external to the interaction between the environment and the organism to explain biological complexity or the match between function and biological form. This, however, does not mean that an acceptance of evolution and religious belief are incompatible, and indeed most major religions accept evolutionary theory within their theological framework. (For more on evolutionary biology and religion see Chapter 14.)

But it is not just its relationship with religious traditions that makes evolutionary biology a particularly difficult area of academic research: it is demanding also because of the range of disciplines it must embrace, from the most molecular to the mathematical, as well as comparative and historical approaches. Box 2.1 defines several terms that are commonly used in evolutionary biology, while Box 2.2 uses the story of “Darwin’s finches” to illustrate the complex and gradual nature of research in evolutionary biology. The dimension of time creates particular difficulties, given that evolution primarily deals with change over multiple generations which cannot be directly observed except in rapidly reproducing organisms. Thus with respect to evolution in organisms more complex than bacteria, empirical data have generally been considered difficult to obtain because of the generation times required.

In the case of human evolution, research approaches are largely restricted to the interpretation of retrospective analyses. Later in the book we will discuss approaches to the testing of hypotheses within evolutionary medicine (see Section 7.5). By the very nature of this approach there are often limited data sets and the interpretation may have a speculative element. However, more recently, evidence of rapid or contemporary evolution has been obtained in a number of vertebrate species, and studies of molecular evolution point to recent and potentially ongoing evolutionary change in humans (see Section 6.6).

A potential danger in evolutionary theory is to consider only selection-driven explanations for a trait when other possibilities such as genetic drift and neutral change also exist and need to be taken into account. This matter is considered further in
**Box 2.1 Definitions**

**Adaptation:** a trait that is increasing evolutionary fitness.

**Adaptive radiation:** rapid appearance of multiple related species from a common precursor, each taking a different ecological niche. It is thought to occur when a species enters a new ecosystem.

**Character displacement:** a process by which traits, in particular those related to resource exploitation such as jaws and beaks, diverge over evolution to reduce interspecies competition (see example in Box 2.2).

**Eukaryote:** an organism consisting of one or more cells that contain a membrane-bound nucleus and other organelles.

**Fitness:** the potential for an individual to survive and reproduce successfully; the measure of individual fitness is its reproductive success.

**Gene:** the structural and functional unit of heredity. Genes may be defined as specific sequences of nucleotides that are used as templates for the transcription of RNA. Functionally they include protein-coding genes, non-coding regulatory regions (“promoters”), and possibly, and more controversially, other sequences that modulate transcription, either by enhancing or repressing it.

**Prokaryote:** a single-celled organism in which subcellular membrane-bound organelles, including the nucleus and mitochondria, are absent.

**Reproductive isolation:** the occurrence of different species living in the same area but which are prevented from interbreeding. Reasons may include: incompatibility between mating times, environments (locations) and behaviors (rituals); incompatibility between sexual organs; inviability or sterility of offspring.

**Speciation:** the emergence of a new species.

---

**Box 2.2 Darwin’s Finches: What Galápagos Birds Teach us About the Origins, History, and Future of Evolutionary Thought**

In 1835, HMS *Beagle*, a survey ship with the young naturalist Charles Darwin on board, anchored in the Galápagos, an archipelago some 600 miles off the coast of Ecuador. These 16 principal, and many smaller, volcanic islands have never been connected to the mainland so their flora and fauna consist of self-introduced species. Isolated in the novel and distinct environments of their respective islands, animals and plants adapted to local pressures over many generations. The vice-governor of the Galápagos mentioned to Darwin that each island has its own turtle species; Darwin was also struck by the diversity of mockingbirds. But it was the existence of 14 species of finches on 14 islands that inspired Darwin to propose natural selection as the key mechanism of evolution.

Or so the legend goes. As a matter of fact, Darwin only realized the importance of what he saw in the Galápagos upon his return to Britain; the finches showed such a striking variation in body size and the size and shape of their beaks that he initially believed them to belong to different families (Davis 2013). Furthermore, at the time Darwin was still a proponent of intelligent design, a framework proposing that the Earth was populated from several divine “centres of creation.” Because Darwin did not expect the Galápagos to be one of these “centres,” he never recorded the island of origin for his specimens of finches. This detail illustrates the importance of a prior hypothesis for the result of empirical research. It was only after the ornithologist John Gould, who examined the entire *Beagle* collection, alerted Darwin to his findings that the finches gained a place in his writings—though only in his account of the *Beagle* voyage (Darwin 1839) and not in the *Origin of Species*.

In the century that followed, as scientific expeditions travelled to the Galápagos to study its fauna, finches became a Darwinian icon: the most famous case study demonstrating key concepts such as reproductive isolation, adaptive radiation, character displacement, and speciation (Sulloway 1982; Box 2.1). Most recently, the genome of five of the finch species became the subject of an epigenetic study. The study indicated that epimutations were more common than genetic change (here copy-number variations, understood as more useful measures than point mutations), that their number correlated better with phylogenetic distance, and that specific epimutations were associated with evolutionarily important genes (Skinner et al. 2014).
Section 2.4.3. However, despite these difficulties, the body of empirical observation from diverse species gives a very solid basis for developing a framework for applying evolutionary biology to medicine. Indeed progress in molecular evolution, whereby interpretation of genomic structure can be used to reconstruct lineages, has overcome a number of these practical problems.

The argument, common among creationists, that much is uncertain in evolutionary biology reflects a misunderstanding of how science works. Evolutionary science is no different from any other mainstream field of biology or medical research: the core elements are well established, while debates and competing viewpoints drive the field to address new hypotheses with new research methods. As any science progresses, there are still frontiers where more research is needed and where the relative merits of one view versus another remain to be resolved.

An analogy would be a disease such as type 2 diabetes mellitus. No one would debate the role of insulin resistance in its etiology, but there remains considerable uncertainty as to the ultimate underlying pathogenic mechanisms. For example, what are the relative roles of genetic and/or epigenetic factors and of lifestyle? Is the primary defect in the pathway of insulin’s action or in the function of pancreatic beta cells, and what is the role of vascular endothelial dysfunction or inflammatory cytokines? Similarly, there will be different views among health professionals about the optimal treatment approaches for certain individuals. But all this intense debate does not prevent a consensus about the basic pathophysiology.

In this sense, then, uncertainty and debate are features of all healthy sciences, and evolutionary biology and medicine are no exception. In the case of evolutionary biology, the scientific debates are complicated by the fact that they not only have intellectual implications but also societal and religious consequences (Chapter 14). Again, an analogy might be between orthodox views of the role of insulin and dietary control in diabetes therapy compared with the views of those who reject orthodox medical approaches in favor of faith healing.

### 2.2 What Does Evolutionary Theory Explain?

Everyone is familiar with the wide variety of life forms on our planet—microbial, plant, and animal. Nearly 2 million living species have been named so far, and it is estimated that there are perhaps 10–50 million extant species. These numbers are dwarfed by estimates of the number of species that have become extinct: probably only 1 in 1000 of species that have ever lived are still alive. Of these, a scant 250,000 or so have been preserved as fossils.

The diversity between species can appear great, as between a whale and a beetle for example, or small, as between two species of bacteria. Evolutionary theory provides an explanation for how this huge diversity of present and past life forms arose from ancestral species. Evolutionary theory is concerned with the diversification of life on Earth but does not attempt to provide an explanation for how life originally arose, which remains an important but separate question (Box 2.3, Figure 2.1).

Evolutionary theory makes much use of the concept of adaptations, meaning features that make an organism better able to survive and reproduce in the environment in which it lives. Such adaptations may sometimes give the impression that an organism has been “designed” for its environment. Yet as we will see, evolutionary theory is able to explain adaptation to an environment by invoking a rather simple process of natural selection—differential survival, heritable traits, and subsequent reproduction—acting on phenotypic variation within a breeding population.

It is critical to recognize that the meaning of the word adaptation in an evolutionary sense refers to its effect on fitness, although there are other medical and physiological uses of these terms. Clarity is essential in the use of these terms. Evolutionary fitness is defined as the potential for an individual to survive and reproduce successfully: thus the fitness of an individual is measured by its lifetime reproductive success, the number of offspring it produces that are themselves able to reproduce. This evolutionary definition of adaptation is distinct from that often used in physiology and medicine to refer to homeostatic responses. It is also a more nuanced
Box 2.3 The Origins of Life

The origins of life have been the subject of much cosmological, biological, and philosophical speculation, and nothing can be definitive. It has been suggested that the first replicators must have been inorganic crystals, a role later taken over by the nucleic acids. The source of energy driving replication may have been the sun but it may also have been the heat from the Earth’s core: the latter is increasingly favored given the increasing knowledge of thermophilic bacteria which live in extreme heat and make use of hydrogen sulfide. The origins of DNA, RNA, and high-energy phosphate bonds and amino acids are even more speculative.

The famous experiments of Urey and Miller trying to replicate the primeval environment resulted in the formation of several amino acids from a primitive carboniferous atmosphere containing methane and ammonia. Others have suggested that carbonaceous meteorites were the source of the earliest organic molecules, including amino acids, but that just shifts the cosmic location of some initial synthesis.

A generally held, but not proven, hypothesis is that RNA replication evolved before DNA replication, and that RNA replication may have been autocatalytic. Indeed, it has been reported that, under the right circumstances, RNA can be...
formed from precursor molecules *in vitro* provided a catalytic enzyme is present, but then the question becomes what is the source of the initial catalyst. The details of this “RNA world” are highly speculative. The single-helix structure of RNA, in contrast to the double-helix structure of DNA, means that the replication of RNA is much more error-prone and mutation rates are much higher. Thus, only simple viruses with small genomes are able to use RNA as their replicator. But RNA can take on tertiary structures that may have allowed it to be catalytic before that role was adopted by proteinaceous enzymes. From this RNA world, possessing autocatalysis and replication and the capacity to code primitive proteins, may have evolved the capacity to synthesize DNA.

DNA sequence data have allowed estimation of the points of divergence between the lineages of the organisms, including plants, fungi, animals, and prokaryotes, extant today. The earliest fossil evidence of simple prokaryotic cells is dated to about 3.5 billion years ago, the age of the earliest stromatolites, which are still formed by some cyanobacteria. The organization of the prokaryotic kingdoms is complex and uncertain, probably reflecting horizontal transfer of genes between emergent species. Two major groupings are recognized: the Archaea, which includes many anaerobic organisms living in extreme environments, and the Bacteria, which includes the bacteria with which we are traditionally familiar. These major classes of bacteria diverged perhaps 3.8 billion years ago from the common precursor to all forms of life, known as the Last Universal Common Ancestor. The origin of the Earth itself dates to about 4.5 billion years ago.

The first eukaryote evolved between 2.7 and 1.5 billion years ago. The origin of the nucleus in a eukaryotic cell (a cell with a nucleus) is uncertain, but sequence data suggest that it is most likely to have resulted from the engulfing of an archaeal micro-organism, which ultimately formed the nucleus, by a bacterium. Hence the human genome is more closely related to the Archaea than to the Bacteria.

Mitochondria evolved following the engulfment of another form of bacterium, probably related to the rickettsia, a process known as *endosymbiosis*. Chloroplasts in plant cells also arose by endosymbiosis of a cyanobacterium.

The first vertebrates evolved some 530 million years ago (Mya), and it was probably at least 900 Mya that the first multicellular organisms appeared. Dinosaurs (and their descendants, birds) along with other classes of reptiles diverged from the lineage which was to lead to mammals some 310 Mya, and the large dinosaurs emerged about 250 Mya. Monotremes shared a common ancestor with birds and reptiles 180 Mya, and placental mammals including marsupials shared a common ancestor as recently as about 140 Mya. Some 75 Mya all species of extant primate shared a common ancestor. The last common ancestor (LCA) of *H. sapiens* and the chimpanzee and bonobo (the last of the apes to share a common ancestor with humans) existed about 7–10 Mya (Chapter 6). Figure 2.1 illustrates these evolutionary relationships.

concept than the use of adaptation in biology and in popular science to refer to the degree of match between a phenotype and its environment. Such well-matched phenotypes may well have evolved through adaptations as evolutionarily defined, but only if the mechanism involved a direct or indirect fitness advantage for a trait with heritable characteristics. We use the term adaptation in this book in the strict evolutionary sense.

If a particular genotype in a population increases in frequency in the next generation relative to the average frequency of all genotypes in that population then that genotype has greater fitness. For example, if under a particular set of selective conditions the majority of individuals in a population (for simplicity we will use an example of an asexually reproducing organism) produce two viable offspring, whereas individuals carrying a particular mutation produce three viable offspring, then the *relative fitness* of the favored individuals is 1.5. Accordingly, the proportion of that genotype in the population will, all other things being equal, increase by 50% in the next generation. It is important to remember that fitness is a composite measure of survival to reproductive age (as an individual that dies before reproducing will have zero fitness) and reproduction rate (the number of offspring produced by an individual that has reached reproductive age).

Fitness is measured in a particular environment, and a genotype with greater fitness in one environment may be inferior to another genotype when placed in a different environment. For example, the short stature characteristic of African and Southeast Asian rainforest hunter-gatherers (“pygmy phenotype”) may represent a successful adaptation for living in forest environments (Perry et al. 2014), but in the Western world there is considerable evidence
that being very short generates social and economic disadvantage. By way of another example, fair skin coloring appears to have evolved as an adaptive advantage for humans living in high latitudes to allow vitamin D synthesis, but as any sunburnt northern European knows it is somewhat disadvantageous when such individuals translocate to the tropics.

The calculation above is a simple example of the direct fitness of an individual organism (e.g., a bacterium) that reproduces by asexual reproduction, but in the case of sexual reproduction the genetic structure of each parent and the effect on their fitness and that of their offspring must be considered: this brings into play considerations of mating success, sexual selection, and recombination, and these will be considered in more detail later. Beyond the direct fitness of individual parents, we must remember that an individual shares components of their genotype with their close relatives, and any action by an individual which improves the reproductive success (i.e., direct fitness) of a close relative will consequently improve the indirect fitness of that individual by ensuring that at least some of those shared alleles (one of the different forms of a gene that can occur at a specific locus on a chromosome) will be transmitted to the next generation. The sum of direct and indirect fitness is called inclusive fitness, and evolutionary behavioral sciences rely heavily on this concept to explain many social behaviors (see Chapter 11).

2.3 How Does Evolution Work?

Charles Darwin described evolution as “descent with modification.” Here, descent refers to ancestry through a lineage (potentially the common ancestry of all species) and modification refers to the various changes that occurred in the descendants of these ancestors as they adapted to new and changing environments. In more modern terminology, we now accept the insights of Darwin and Wallace that evolution occurs because of the following mechanisms and processes (Figure 2.2):

- Individual members of a species vary in ways that affect their survival and reproductive success (i.e., their fitness) (variation). In Darwin and Wallace’s framing this was initially seen primarily in terms of excess progeny competing for resources and

![Figure 2.2 Evolution operates through processes of variation followed by selection causing differential survival or reproductive ability.](image-url)
surviving predation. Later Darwin was to recognize that differential reproductive success could occur in other ways (see Section 2.3.2.3).

- Some subset of this variation is heritable (inheritance).
- Differential survival and reproduction among these variant forms leads to over-representation of successful forms in the next generation (selection).
- Selection over generations leads to change in the composition of the population (evolution).

It is important to understand that evolution operates on populations of organisms over generations. Individual organisms do not evolve during their lifetime: they may be able to change some aspects of their phenotype in response to the environment (a process known as plasticity, which is discussed in Chapter 4), but those changes are not heritable and are not passed on to their offspring. (There may be some exceptions to this rule for epigenetic marks crossing generations following the induction of these marks during developmental exposures; see Section 3.8.3.)

Although natural selection operates at the individual level by favoring particular characteristics (traits) in the phenotype which contribute to improved survival or reproductive success, for a population to evolve those phenotypic traits must be associated with heritable factors in the genotype to ensure that the selected trait is passed on to the next generation. In other words, selection must result in changes in allele frequencies in the gene(s) that determine those traits, and many evolutionary biologists have defined evolution in terms of changes in allele frequency in a population. However with rising interest in the potential for epigenetic inheritance even in mammals (Laland et al. 2014), and because of the complex roles of development and environmental interactions in evolutionary processes (see Chapter 4), we prefer Darwin’s original definition.

In the following subsections we look at the processes of variation, selection, and inheritance in more detail.

### 2.3.1 Variation

Before Darwin, biologists were enthusiastic classifiers; indeed, Darwin himself spent many years classifying barnacles, perhaps to demonstrate to fellow biologists that he had empirical skills. The system of classification invented by Carl Linnaeus in the eighteenth century relied on shared physical characteristics to place organisms into orders, families, genera, and species, with a nominated specimen (the type specimen) considered as the standard for identification of other members of the species. The focus was therefore on establishing similarities between individuals of a species. But one of Darwin’s critical contributions was to emphasize the importance of variation among individuals of that species as the raw material for selective mortality and reproductive success, allowing populations gradually to change their proportion of different variants over time. If there is no variation between individuals in a population, there are no discriminatory traits for selection to act on and, therefore, no capacity for evolution to take place.

We can easily appreciate how widely organisms from the same species can vary: humans, that along with the domestic dog represent one of the more variable of mammalian species, vary markedly both in physical appearance—height, build, skin, hair, and eye color—and in less obvious characteristics such as blood group, tissue type, lactase and amylase expression, and disease susceptibility. There is no physical type specimen for *Homo sapiens* and recent genomic data have underlined that there is no type specimen for the human genome either. Examination of other species by morphological and molecular methods also reveals wide variation among individuals.

What is the source of this phenotypic variation among individuals in a population? The principal sources of heritable (genetic) variation—genetic novelty—are *mutation* and (in sexually reproducing organisms) *recombination*. Additionally, as we discuss in more detail in Chapter 4, environmental factors during the development of an individual can modify the phenotype by the process called *developmental* (or phenotypic) *plasticity;* the ability of developmental processes to generate phenotypic novelty for selection to act on has become an important topic in evolutionary developmental biology in recent years.

#### 2.3.1.1 Mutation

A mutation is a change in the base pairs in the DNA sequence, and therefore by definition it is heritable
when it occurs in the germline. So a somatic mutation which occurred in a population of dividing cells in the body, for example in the intestinal mucosa, might have a dramatic effect on those cells (it could produce a cancer) but clearly cannot be passed to the next generation. Here we are concerned only with heritable changes. The scale of such possible changes ranges from a single base substitution, giving rise to a single nucleotide polymorphism or SNP, through the duplication of blocks of tens to thousands of base pairs, to deletions, duplication leading to copy number variation, or inversion of whole genes or gene-sized blocks of non-coding DNA. Larger changes compatible with life, such as duplication or deletion of whole chromosomes (aneuploidy), are not infrequent, particularly in the sex chromosomes, but cannot be passed on to the next generation. The molecular basis of mutation is discussed in Chapter 3.

Because a new mutation typically affects only one copy of the DNA, individuals carrying a new mutation, or their offspring, will be heterozygous for that mutation. The consequences of the mutation will depend on a number of factors, such as whether it occurs in a coding or non-coding region of the genome and whether its phenotypic effect—if it has one, for as explained in Chapter 3 many mutations are phenotypically silent—is dominant or recessive. The existence of silent and recessive mutations suggests that hidden variation is common in the genome; indeed, most single-base substitutions are thought to be selectively “neutral” or only mildly deleterious and do not significantly affect the fitness of the carrier. This means that they will be passed on to the next generation.

2.3.1.2 Recombination

Sexually reproducing organisms inherit their parents’ genes, not their genotypes (nor their phenotypes). This is because of the genetic variation introduced by two processes during sexual reproduction: first, meiosis during gametogenesis involves recombination, exchange of genetic material between pairs of homologous chromosomes before the cell divides, and secondly the fusion of the two gametes, each with a parental haploid genome.

The process of recombination can generate high levels of variation in offspring because it generates new combinations of alleles on a chromosome (see Section 3.3.5). This is particularly likely to have an effect on the phenotype if the genes affected have additive or subtractive effects on some trait. For example, consider four genes all determining a trait such as susceptibility to a particular disease, with alleles A, B, C, and D each independently increasing susceptibility by the same amount and alleles a, b, c, and d independently decreasing susceptibility by that amount. The parental genotypes might be AbCd and aBcD, each with moderate susceptibility, but recombination during gametogenesis might produce gametes with genotypes ABCD and abcd, with the potential to confer high and low disease susceptibility, respectively, on offspring receiving those genotypes.

Because some sites in chromosomes are more likely than others to be preferred sites for recombination (“hotspots”), some regions of the genome are more likely to stay linked to each other through recombination. Thus selection on one gene can also effectively select on neighboring genes if they remain linked. This phenomenon of hitchhiking is important in studies of molecular evolution in so-called “selective sweeps” seeking sections of DNA which have been positively selected; linkage disequilibrium refers to the bioinformatic estimate of how closely genes stay linked through recombination. In such screens, the molecular approach identifies regions of DNA containing potentially more than one gene, one of which at least has been subject to selection. This has been a powerful tool for identifying evidence of recent human evolution (see Section 3.4.2).

2.3.1.3 Constraints on Variation

Constraints on variation may explain why organisms often show examples of poor “design.” They appear not to be as ideally suited to their environment as they could be. Why should variation be constrained? There are a number of reasons, which, as we will see in later chapters, explain some aspects of human disease.

First, at least in multicellular organisms, accumulation of genetic variation and its filtering by
selection are generally slow processes, so that the response to a new selective pressure such as environmental change will also be slow. This lag time can lead to a mismatch between an organism and its environment if environmental change occurs on a faster timescale than evolutionary processes.

A second constraint occurs because evolution by variation and selection demands continuous “tinkering,” and does not permit partial dismantling of an existing variant of modest fitness in order to construct a better solution, as that would have the effect of reducing fitness in intermediate generations (see Figure 2.4 in Box 2.4). An example of this effect is the human eye. The eye has an adequate, but flawed, design which routes nerves and blood vessels over the top of the retinal photoreceptors (Figure 2.3). However, it is improbable that it could evolve towards the structure of the octopus eye (where the retina faces the “right” way) without passing through a non-functional and non-adaptive form.

We have discussed the highly variable nature of individual phenotypes and genotypes, but a third constraint derives from the limited pool of available genetic variation in any given generation. A population may fail to meet an environmental challenge simply because it lacks the appropriate range of genetic variants for selection to work on; such constrained variation is more likely in populations which have gone through a bottleneck because of the small range of alleles in the surviving population and the increased likelihood that genetic drift will have further reduced diversity (see Section 3.4.1).

A fourth constraint on the generation of new phenotypic variation is the existence of a variety of processes such as canalization, by which distinct genotypes tend to produce the same phenotype (this is sometimes termed robustness; Bateson and Gluckman 2011). A canalized trait is one that has self-correcting mechanisms to ensure that a relatively narrowly defined phenotypic outcome

---

**Figure 2.3** Building plans of the vertebrate eye (left) and the cephalopod eye (right). In vertebrates, in which photoreceptor cells differentiate from the central nervous system, nerve fibers pass over the top of the photoreceptors. The consequence is that there is a blind spot where nerve fibers pass through retina, forming the optic nerve, and the arrangement of layers with the superficial blood vessels creates a risk of retinal detachment. In cephalopod and arthropod eyes photoreceptors differentiate from the epidermis and nerve fibers lie behind the retina, so the blind spot is absent.
The concept of a “fitness” or “adaptive” landscape was first proposed by the pioneer of the Modern Synthesis (Section 14.5), Sewall Wright, to visualize the relationship between genotypes and fitness. A “fitness landscape” is a three-dimensional, “mountainous” landscape in which genotypes are organized on the horizontal ($x$–$y$) plane while fitness is plotted on the vertical ($z$) axis, appearing as mountain peaks. The proximity between peaks denotes similarity between underlying genotypes. Adaptation is imagined as climbing to the higher position on this map.

Figure 2.4 represents an adaptive landscape of eye evolution. Height indicates optical quality and the horizontal plane evolutionary distance. Each individual peak represents evolutionary “improvement” on a particular basic plan for a light-sensing organ. An eye can evolve “uphill”, as its fitness increases continuously, but generally cannot evolve from one peak to another, as to do so would require traversing a valley of lower fitness as one underlying structure is dismantled and replaced by another. For a long time the “fitness landscape” was a theoretical concept, but recently empirical (genomic) and theoretical developments have enabled construction of real fitness landscapes, in turn encouraging studies of evolutionary predictability (de Visser and Krug 2014).

**Box 2.4 Fitness Landscape**

Figure 2.4. An adaptive landscape of eye evolution. Adapted from Dawkins (1996). Figure originally drawn by Professor Michael Land, University of Sussex, with permission.

develops. Canalized traits are buffered against perturbations from the environment or from mutations, and this means that some novel mutations which arise remain silent. Such silent mutations will not be exposed to selection, and populations will tend to accumulate hidden genetic variation in canalized traits. However, under some circumstances this variation can be exposed by stressors. Robustness and
Fundamental to evolutionary biology is the concept of differential fitness, namely that some individuals in a population have phenotypic characteristics that make survival and successful reproduction more likely than it is for others. To the extent that this variation in success between individuals can be traced back to inherited (genetic) variants, this leads over time to a change in the genetic characteristics of the population’s lineage. Selection refers to the processes which determine this differential fitness.

2.3.2 Artificial Selection

In The Origin of Species Darwin intentionally started with a description of livestock and pigeon breeding. He did so to illustrate what is termed artificial selection. Here the breeder identifies a characteristic that is desirable (or undesirable) and chooses to actively mate particular individuals to try and amplify (positive selection) or remove (negative selection) that trait from the lineage. Such artificial selection is active because there is an active agent—the breeder—selecting a characteristic that it is hoped to change in the population for a specific purpose.

Obviously the strategy can only be successful to the extent that the characteristic is genetically determined and not constrained by other factors. For example, it is not possible to select for five-legged horses because the basic vertebrate body plan is bilaterally symmetrical, and therefore a variant with genetic determinants for five-legged mammals does not exist. Similarly, there is a limit on the body size which can be selected because of other anatomical or physiological constraints such as bone strength or cardiovascular function. Despite billions of dollars being spent in the thoroughbred horse industry, the lack of strong genetic determinants of speed in the horse means that bloodlines have relatively modest impact on the chances of winning the Derby. On the other hand, where there are strong genetic determinants rapid selection is possible. For example dog breeds have been developed by breeders over a very short time, producing breeds as distinct as the Chihuahua and the Great Dane, even though these and all other domestic dog breeds have been derived from wolves first domesticated about 15,000 years ago.

2.3.2.2 Natural Selection

The bulk of Darwin’s argument, also put forward in parallel by Alfred Russell Wallace, was that there is a natural variation in phenotype and within an environment some variants would be more likely than others to survive and to reproduce successfully; their young in turn would be more likely to survive to reproduce. This process did not require any external active agent, because under most conditions each species produces more offspring than the environmental resources can sustain. The presence of variation in phenotype means that some would be more able to cope with the environment and others less so. Thus ecological circumstances act as a selective agent and would lead to population change if there were heritable components associated with this differential fitness. Darwin termed this natural selection, while the philosopher and socio-political theorist Herbert Spencer later called this process “survival of the fittest” (Section 14.3), a term that resulted in unfortunate misapplications. The basic principles developed by Darwin are remarkable in that particulate inheritance was not understood until the early twentieth century and there were a long series of intellectual debates before it was recognized that genetics and evolutionary biology were fully compatible.

Darwin originally saw natural selection as a very slow process, and indeed it often is, but there are situations where natural selection can act within a few generations. Speed of selection will clearly depend on the trait, its genetic determinants, and the strength of selection pressure created by the environment. In terms of human biology, modern molecular genetics has started to identify traits that may have been subject to relatively recent selection; for example, the ability to digest lactose as an adult (Box 1.1), latitude-based changes in skin color (Sections 6.7 and 13.4.1), and resistance to locally common infections (Section 7.4.6).
Natural selection is the central process by which organisms come to be adapted to their environment. At its heart is the principle that it leads to a change in allelic frequencies across generations within a lineage, and in mathematical terms this can be used to quantify and define it. It is beyond the scope of this book to describe in detail the mathematical theories of measuring fitness and selection—we present the fundamentals in Box 2.8—but there is a large theoretical and empirical literature which does so. It is important to recognize that while much experimental work focuses on a single gene and its multiple alleles, most genes may have multiple effects (pleiotropy) or may modify the actions of other genes (epistasis). We explore these concepts in the setting of genes causing disease in Sections 5.2.4 and 3.6, respectively.

Natural selection works by repeatedly favoring the persistence of alleles that have positive fitness effects for the organism within the environment it inhabits. Hence, similar organisms living in different environments will evolve in different ways if the allelically determined variation has differential effects on fitness in these various environments. Thus fitness cannot be defined with respect to a single allelic variant; it must be interpreted in relation to conditions which include the physical and ecological environment and the influence of the rest of the genome. So, for example, an allele which favors tall stature, which might otherwise confer greater fitness by giving greater strength, may not do so if the individual is in a nutritionally limiting environment; indeed, it has been observed that during a famine the largest individuals may have a lower survival rate (Bateson 2001).

Fundamental to selection is the presence of genetic variation within a population. Different alleles can be maintained in a population if they have differential advantage in relation to the environment: this is termed balancing selection. Section 3.5.2 describes how balancing selection can maintain apparently harmful alleles within a population.

Evolutionary biologists use a number of terms to categorize forms of natural selection which describe either the different outcomes of natural selection or the different processes that are associated with differential fitness. Some of these are detailed in Box 2.5.

2.3.2.3 Sexual Selection

Although Charles Darwin is best known for *The Origin of Species*, in a second great book, *The Descent of Man, and Selection in Relation to Sex*, he also recognized that differential fitness could be also created by active intraspecies competition affecting which members of the species could mate. This type of within-species competition can operate in one of two ways. The first is through intraspecific competition between members of the same sex. If males compete with each other for the right to mate, then the nature of the competition will affect the male phenotype favored and developed across generations. For example, in deer stags use their antlers in the rutting season, fighting to gain mating rights over females in a herd. There are genetic determinants of antler size, and over time there will be selection pressure to grow larger antlers, an ornament that has no other role in helping the stag to adapt to its local environment or ecology. At the same time, however, it is important to note that there are constraints on antler size which limit the extent to which natural selection can shape their form. It has even been suggested that the antlers of the Irish elk evolved to be so large that they created a disadvantage and contributed to the species’ extinction 11,000 years ago.

A second type of within-species competition operates when members of one sex choose their mate according to a set of specific characteristics that are perceived as attractive. The most well-known example is the tail of the peacock. Males parade and fan their tails and the female (the peahen) is thought to choose according to the extravagance of the display. Thus a peacock’s tail has evolved to be long and extravagant. But this example leads to the question of why females might be attracted to such a seemingly burden-some trait in the first place. Is it simply because the longer tail is more attractive or does the tail act as a marker of some more relevant characteristic? For instance, a longer tail creates a handicap and an adult male with a long tail signals his overall health and strength, akin to a human running a marathon and winning despite carrying barbells. This latter hypothesized mechanism is known as the handicap hypothesis. There are advocates of
both explanations and it seems likely that both operate in different situations.

Sexual selection involving male versus male competition will often lead to differences in body size (as larger males are generally more successful in such competition), and in general there tends to be greater sexual dimorphism in body size in species with social systems where dominant males compete for mating rights over multiple females. Thus in species of primate where there are harem-like societies, such as the gorilla, body size determines which male has mating opportunities, and thus the male is much larger than the female. The presence of some dimorphism in human adult body size has been used to infer that early H. sapiens had a social system involving dominant males. This is possible, but there may be other reasons for the difference.

Sexual selection based on mate choice need not involve physical characteristics. For example, in bower birds the males compete for females by showing off collections of objects which they have found and displayed in their bowers, while in song birds the male competes by a vocal display.

Because it involves intense selection in each generation, sexual selection can lead to relatively rapid evolution of traits which improve reproductive attractiveness or competitive abilities. Thus explanations based on sexual selection are often considered where there is evidence of relatively fast evolution of a trait.

Sexual selection can lead to traits that do not obviously improve the match with, or adaptiveness to, the environment itself: they are instead adaptations to the social pressures faced by members of that
species, whether from competition with members of the same sex or choice on the part of members of the opposite sex. It is probable that a number of human physical characteristics developed through sexual selection, in particular distribution of body hair and other secondary sexual characteristics (see Section 8.7).

2.3.2.4 Levels of Selection

While it would seem self-evident that selection acts at the level of the individual, since it is the phenotype of the individual which determines whether that particular individual reproduces within an environment, a range of other possibilities need to be considered (Okasha 2006). In theory selection can act at any level of biological organization if there is differential survival associated with that level. So selection can act at the level of a cell, as in the case of somatic mutations leading to cells with differential survival, such as cancer cells. Selection can also act at the level of a group or population, and it is the question of the relative importance and significance of that higher level of selection that has been the subject of much controversy.

The primary challenge to the concept of the individual as the unit of selection came originally from consideration of populations of animals living in groups and of how cooperative and apparently altruistic behaviors arose. That natural selection does not always operate at the level of individual reproductive fitness is particularly apparent for some eusocial (colony-living) animals where the majority of colony members are in fact sterile: insects such as honey bees are the extreme examples where reproduction is limited to the queen bee and a small number of fertile males. The implication—and the issue that has generated the most debate—is that groups of individuals may act as a higher-level unit upon which selection acts, a process that has been termed group selection. Superficially, it would appear to be a simple matter: if a behavior advantageous to a subpopulation (a deme is a definable subpopulation that might be a unit of selection) is exhibited by that group, but not in the members of competing subpopulations, then that behavior will be more likely to be favored by natural selection and thus transmitted to subsequent generations. (Here we are focusing on the direct or indirect genetic determinants of behavior, rather than simply cultural transmission.) As one contentious hypothetical example, it has been suggested that if religious belief had genetic determinants and if religious belief made societies more healthy, or at least more stable, and also encouraged behaviors that favor reproduction, such as forbidding contraception, then groups with religious belief—and the individuals and genes that comprise them—would be positively selected.

Debate on group selection has most visibly focused on two problems: how altruistic behavior between individuals could have evolved, and the problem of the “freeloader.” Over the past 30 years there have been strong advocates for the idea that altruistic behavior must have developed by group selection. However, there are compelling arguments against such a view. Essentially the counter-argument runs as follows: imagine a population which is mainly composed of altruists who share resources or protect each other. But within the population there are some selfish non-altruistic freeloading individuals (cheaters) who are happy to take all the resources (food, shelter, mating rights, etc.) rather than share them; they may prefer to hide to save their skins rather than warn the population of predators. The altruistic members of the population are more likely to be harmed or have their resources curtailed to the benefit of the cheaters. Over time the altruistic group loses out, they are less fit and so the cheaters come to dominate. It is this type of argument that has been used to argue that group selection, if operative at all, is a relatively weak force.

If group selection is weak, how then can we explain the evolution of altruistic behavior? The concept of kin selection was first developed by the quantitative population geneticists Ronald Fisher and J. B. S. Haldane, and later formalized by William Hamilton (Hamilton 1964a, b). These and other workers pointed out that genes are shared between relatives. In a sexually reproducing species such as the human, fitness is defined by our capacity to pass our genes on to our descendants. But our brothers and sisters share half our genes and thus are also capable of passing these on to their descendants; our cousins share a quarter of our genes, and so forth. Thus we can improve our fitness, since it is defined by the successful passage of our genes to
subsequent reproducing generations, by encouraging our close relatives to reproduce. Inclusive fitness is the term used to describe this concept; that is, the impact of a given allele on both the carrier’s own fitness and that of neighbors or relatives carrying the same allele. The assumption is then that individuals act not only to maximize their individual fitness but their inclusive fitness: within families this is the process of kin selection. As Haldane is famously said to have put it: “I would be prepared to lay down my life for two brothers or eight cousins!”

Kin selection could explain the evolution of altruism because it develops in social animals such as insects in which there is a high degree of relatedness between individuals. The haplodiploid nature of bee genetics, whereby females are diploid and males are haploid, means that female workers share up to three-quarters of their genes with the queen’s other female offspring, more than the 50% they could share with their own offspring were they able to mate. This could explain why it is in their evolutionary interest to forego reproducing themselves but to support the survival of the queen’s offspring. Similarly in humans or in other mammals living in small groups, siblings have a high genetic relatedness with each other’s offspring, and an individual’s genes can be transmitted to future generations not only by promoting the reproductive fitness of his or her own offspring, but also by promoting the fitness of nieces and nephews (see Section 11.4.1). We consider the additive (and complicating) role of uncertain paternity in primate social structures in Section 8.5.

Nevertheless, the evolution of altruism does not necessarily require a group-selective explanation: reciprocal cooperation may lead to behavior which can be interpreted as altruism. There is advantage in cooperating in situations where A is more likely to help B, if B had helped A in the past. As there may be situations where help is essential for A to survive and reproduce, then alleles controlling cooperative behavior will be more likely to survive and to increase in frequency. For example this is seen in the sharing of blood meals by vampire bats with those members of their colony who failed to get a meal in the usual manner. Even so, vampire bats are sensitive to which animals have and have not shared meals in the past, weeding out any would-be cheaters. This is not group selection but selection for a form of delayed self-interest that benefits an individual’s fitness, since the cost to a bat of sharing food is offset by its expectation that it will benefit from reciprocal behavior in the future. There is an extensive literature in game theory to identify how competition and poor cooperation can be dealt with through the evolution of policing, reciprocation, and punishment, leading to evolutionarily stable strategies, and this will be discussed in Chapter 11 with respect to its implications for understanding human behavior.

Following the work of the famed evolutionary biologist George C. Williams, who parenthetically is regarded as one of the fathers of evolutionary medicine, the problem of the freeloader was considered to be so large that group selection was considered to be practically impossible. However, in recent decades a more nuanced view has become more favored, and while it is accepted that in most situations individual-level selection is the dominant influence allowance is made for some influence of group selection. These can coexist and this coexistence is known as multi-level selection.

2.3.2.5 Genes as Units of Selection

Implicit in the discussion above is the notion that selection acts on the phenotype and eventually benefits the organism’s lineage by matching its phenotype to the environment. For instance, in a species subject to the pressure of predation, selection might result in adaptation at the level of the individual, such as a change in muscle biochemistry allowing faster escape, or perhaps at the level of the group, such as a change in behavior whereby some individuals reduce their apparent individual fitness by acting as sentinels to warn the group of the presence of a predator.

A different perspective was provided by Richard Dawkins with his important concept of the selfish gene. While not denying that selection acts on the phenotype, Dawkins argued that the ultimate beneficiary of adaptation is the gene itself, in that genes are the agents which over the long term survive, or fail to survive, because of the benefits they confer, or the limitations they impose, on the phenotypes that they generate (Dawkins 2006). In this view, genes
are considered as the permanent “replicators” and phenotypes as the temporary, generation-based “vehicles,” formed by the products of coalitions of genes, which they inhabit. Even the genotype can be considered as temporary, since in sexually reproducing species it is destroyed in each generation by meiosis and recombination. The selfish-gene metaphor therefore characterizes an adaptation as something which increases the ability of a gene to promote its own survival without necessarily promoting the survival of the organism, group, or even species.

Most commonly, of course, the interests of a gene and its “vehicle” (the organism) do coincide, but they may not. We have already mentioned one example, where an individual organism limits its foraging time and increases its predation risk, thus reducing its own fitness, by acting as a sentinel. This apparently unselfish behavior can be explained from the viewpoint of the selfish gene by considering a closely related group who share many alleles. By protecting other group members from predation, the sentinel is helping to ensure survival of its genes—even though they happen to be residing in other members of the group. As recognized by William Hamilton (see Section 2.3.2.4, and the discussion of Hamilton’s rule in Section 11.4.1), the degree of relatedness should determine the extent of such altruistic effort. Hamilton further suggested that alleles that were somehow able to “signal” their presence in an individual would be at an advantage in triggering altruistic behavior toward that individual from other individuals carrying the same allele.

More extreme examples of genes acting “selfishly” can be found in the various phenomena collectively known as intragenomic conflict, in which genetic elements proliferate at the expense of other genes in the same genome. These have been largely identified and studied in bacteria and insects. Such elements may do so by ensuring that they are over-represented in the gametes, or by promiscuously copying themselves to increase their abundance in the genome. Meiotic drive or segregation distortion occurs when an allele is able to “cheat” on the normal process of Mendelian segregation (which ensures that each of the two alleles in a diploid genome has a 50% chance of appearing in a gamete), increasing its chances of being passed on to the next generation. A good example is the multilocus gene complex SD in Drosophila, which includes the driver Sd and its target Rsp (Larracuente and Presgraves 2012). Segregation distortion of sex-linked genes can bias the sex ratio of the offspring. Transposable elements (see Section 3.3.4) are the most abundant class of selfish genetic elements that promote their own transmission by promiscuous replication.

2.3.2.6 Extended Phenotype

We introduced the concept of natural selection by considering a simple scenario, namely one in which an individual’s fitness is solely determined by his or her phenotype. We have used phenotype as a description of the totality of an organism’s behavioral and biological traits. However, the impact of the genotype can extend beyond the physical limits of the organism, and these can also have an impact on the direction and pace of evolution. This concept of the extended phenotype was a second important contribution of Richard Dawkins (Dawkins 1982).

The body, the phenotype, is something genes build for themselves to make their way through the world. The extended phenotype is everything that the genes can influence to gain advantage. It includes the effect of the social behavior of one individual on other individuals in the population, the manipulative effects of parasites and pathogens on the behavior or life cycle of their hosts, the creation of artifacts—spiders’ webs and birds’ nests—and, most spectacularly, the engineering of the environment by building structures such as termite mounds or beaver dams through niche construction. Niche construction is an important concept because in this situation the organism changes the environment in such a way that it in turn can have important influences on selection pressures (Odling-Smee et al. 2003). For example, housing and clothing reduce the influence of the thermal environment on survival. The ideas of extended phenotype and niche construction should not be taken to extremes: the acid test is the presence of a direct correlation between variation in the structure or behavior and variation in the transmission of the genes responsible for that structure or behavior.
Humans create their own environments. Our capacities to make clothing, to use fire, and to build shelter are examples of ways in which we use our technological capacity to allow us to survive and flourish in a variety of environments. Many other species create a built environment in which they breed and generally live: a termites’ nest is brilliantly designed by the colony to maintain a constant internal temperature and a beavers’ nest provides both protection and a constant environment in which to breed. In most species such niche construction is stereotypic and is likely to be genetically determined, possibly having arisen through genetic assimilation (Chapter 4). Humans could be considered as niche constructors but they do not produce a stereotypic niche; rather, they use their technological and innovative capacity to find ways to modify their environments, expanding the environmental range in which they can live.

2.3.3 Culture as an Evolutionary Force

In the previous section we discussed the impact of the genotype beyond the physical limits of the organism through the concepts of extended phenotype and niche construction. We are clearly social animals where social behavior is both a major component of our phenotype and a major component of our selective environment. Also, human ability to produce material culture—from stone tools to airliners—and our ability to modify our environment to suit our physiology has allowed the human species to spread across the planet.

Indeed, culture and human biology are deeply intertwined. First, culture creates selective pressures and thus may change the direction of evolution, which in turn may change culture. This phenomenon is known as gene–culture coevolution. Consider the example with which we introduced this book, lactose intolerance. Only those with the genetic variant allowing for lactase persistence could nutritionally thrive on a diet including cows’ milk. When certain populations domesticated cattle, those individuals thrived and the population underwent selection for the persistence of the allele. In turn, as the population became more able to utilize dairy products, farming of cattle would be promoted. There may have been other consequences of agricultural settlement. The presence of the heterozygous state for sickle cell anemia confers protection against malaria, even though homozygosity is highly disadvantageous (see Section 7.4.6). Changes in farming practice in West Africa led to more stagnant pools of water, and thus a higher incidence of malaria, and this may have in turn favored an increase in the gene frequency of an allele conferring relative protection against malaria. So we can see how interactions between inherited culture and genetic variation can shape human evolution.

Second, some argue that cultural change should be studied in the same way as biological evolution. Richard Dawkins introduced the concept of the cultural equivalent of a gene, termed a meme. As initially proposed, memes were true units of cultural inheritance, capable of self replication, with variation due to errors in duplication, and which due to the nature of a society would have differential “reproductive” success (in other words a kind of fitness for memes, how well they persisted and spread within a population). Examples of memes include “tunes, catch-phrases, clothes fashions, ways of making pots or building arches.” But this concept is now seen by most evolutionary psychologists as nothing more than a useful metaphor.

Box 2.6 explains that our current understanding of culture focuses on the mode of transmission rather than the content, seeing culture as any information that can be transmitted through social learning. In contradistinction to genetic inheritance, cultural inheritance need not involve vertical transmission between generations. Horizontal transmission is the norm; for example, when young people adopt a particular form of dress we can see that it is rapidly transmitted through the peer group.

But culture itself undergoes evolution. Every aspect of human culture from belief, to art, language, music, and technology shows a process of change which is termed cultural evolution (Box 2.7). There is variation in a culturally determined characteristic, and there will be selection by the society as to which variants are preferred and thus which spread successfully. However, the fidelity of replication need not be sustained, unlike most genetic replication for which there are repair mechanisms which generally maintain fidelity. Thus change can be rapid: witness how quickly the change in the
The meaning of the term “culture” has been redefined and contested since the early days of anthropology in the nineteenth century, from “the complex which includes knowledge, belief, art, law, morals, custom, and any other capabilities and habits acquired by man as a member of society” (Tylor 1871) to “the total shared, learned behavior of a society or a subgroup” (Mead and Metraux 1953), and then “an historically transmitted pattern of meanings embodied in symbols” (Geertz 1973).

One influential modern definition—coming, importantly, from evolutionary science rather than the background of purely cultural anthropology—places the stress on the transmission rather than the content of culture, by defining it as “information that people acquire from others by teaching, imitation, and other forms of social learning” (Boyd and Richerson 2005). This definition allows animals to have culture too: New Caledonian crows, famous for their toolmaking, acquire those skills not just by trial and error but also by observing their parents (Holzhaider et al. 2010). The ability of blue tits to open the foil tops of traditional British milk bottles spread rapidly when these tops were introduced. It is unlikely that each bird initiated the skill de novo; rather, many would have gained the skill from observation of others. There has been much study of cultural transmission in non-human primates, such as the classic example of potato washing in macaques in Japan. Scientists were in the habit of placing potatoes on a sandy beach to attract the animals out of the forest. Once one female macaque had learned to remove sand from her potatoes by washing them in the sea, other monkeys and eventually all the monkeys in the troop adopted the same behavior. Also, chimpanzees in different parts of Africa use sticks to extract termites, but they do so in different ways, reflecting different cultural traditions within different troops.

Age when women choose to have their first child has become distributed across Western society (see Figure 6.5).

The very nature of culture means that its specific manifestations are not genetically determined, although the capacity to exhibit culture must have a genetic basis. For example, as discussed in Section 7.4.7, the “acting out” behavior of young males, whether on the sports field or in driving fast cars, has its origin in the evolved determinants of male behavior related to displaying evidence of relative fitness, the result of sexual selection. Culture itself unconsciously affects fitness: it generates behaviors which promote potential reproductive success for the individual through creating rules for membership of a group and promoting reciprocal

### Box 2.6 What is Culture?

The meaning of the term “culture” has been redefined and contested since the early days of anthropology in the nineteenth century, from “the complex which includes knowledge, belief, art, law, morals, custom, and any other capabilities and habits acquired by man as a member of society” (Tylor 1871) to “the total shared, learned behavior of a society or a subgroup” (Mead and Metraux 1953), and then “an historically transmitted pattern of meanings embodied in symbols” (Geertz 1973).

One influential modern definition—coming, importantly, from evolutionary science rather than the background of purely cultural anthropology—places the stress on the transmission rather than the content of culture, by defining it as “information that people acquire from others by teaching, imitation, and other forms of social learning” (Boyd and Richerson 2005). This definition allows animals to have culture too: New Caledonian crows, famous for their toolmaking, acquire those skills not just by trial and error but also by observing their parents (Holzhaider et al. 2010). The ability of blue tits to open the foil tops of traditional British milk bottles spread rapidly when these tops were introduced. It is unlikely that each bird initiated the skill de novo; rather, many would have gained the skill from observation of others. There has been much study of cultural transmission in non-human primates, such as the classic example of potato washing in macaques in Japan. Scientists were in the habit of placing potatoes on a sandy beach to attract the animals out of the forest. Once one female macaque had learned to remove sand from her potatoes by washing them in the sea, other monkeys and eventually all the monkeys in the troop adopted the same behavior. Also, chimpanzees in different parts of Africa use sticks to extract termites, but they do so in different ways, reflecting different cultural traditions within different troops.

### Box 2.7 Studying Cultural Evolution

Culture is predicted to undergo evolution in much the same way as genetic evolution, where cultural features that are tested against the environment evolve at a different rate from those not tested against the environment. However, using a scientific framework to support this has been difficult since humans tend to be unpredictable in their beliefs and behaviors.

A study on Polynesian canoe designs provides empirical evidence for this argument (Rogers and Ehrlich 2008). Using a large historical data set characterizing detailed features of canoes, researchers examined a total of 134 functional and symbolic features. A functional trait would, for example, be whether canoes were made up from a single trunk or several, or whether they were sewn together using coconut fibers.

These features would have been important in the seaworthiness of the canoes and thus the survival of their builders. Symbolic traits examined included paintings and shell decorations, which have no impact on a canoe’s sturdiness. Statistical analysis showed that functional traits evolved more slowly than symbolic traits, indicating that features that improved survival, migration, and reproduction tended to be preserved. This is a more scientific manifestation of the adage “if it ain’t broke, don’t fix it.” Thus, in a way, natural selection was acting to prevent the persistence of inferior designs. More importantly, it demonstrated that evolution of culture can be measured, and that this process can indeed be considered akin to the way natural selection operates in genetic evolution.
altruistic behavior (examples being religion and food sharing) or which allow the potential reproductive value of the individual to be demonstrated (e.g., art and fashion).

2.3.4 Inheritance

Inheritance is an essential building element of evolutionary theory. We know today that genes are discrete segments of the DNA chain packaged in chromosomes, that the specific order of base pairs within the DNA carries the genetic information, and that the double-stranded nature of DNA ensures accurate transmission of that information during reproduction. But at the time when Darwin was developing his ideas, the nature of inheritance was not understood. Darwin himself proposed a rather bizarre model of inheritance called pangenesis, which involved accumulation in the gonads of so-called gemmules shed from body tissues. Nevertheless, he realized that any mechanism of inheritance that involved mixing or averaging of the characteristics of the parents (blending inheritance) presented a problem for his theory of descent with modification, since the advantage of any new variation would be diluted out within a few generations.

It is part of the tradition of evolutionary biology that although it was during Darwin’s lifetime that Mendel made his fundamental discovery that the individual units of genetic information remained intact and unmixed during transmission from one generation to the next (particulate inheritance), his findings were unknown to Darwin and only came to be recognized after the theory of inheritance came together with cell biology in the late nineteenth century. They formed the basis of modern genetics (a term introduced by William Bateson in 1906), along with the rapid recognition of the role of the chromosome in inheritance, and in 1909 the unit of particulate inheritance was given the name “gene” by Wilhelm Johannsen. Johannsen also recognized that there was not a direct link between a gene and a trait and he coined the terms phenotype and genotype.

These accumulated concepts led to the rapid application of quantitative techniques to the study of inheritance. After initial debate over whether continuous variation of a trait (as is usually observed in biology; for example, height or weight in humans) could be generated through particulate inheritance of the type described by Mendel (whose peas were dichotomously yellow or green, round or wrinkled, etc.), the matter was firmly resolved in 1918 by the great mathematical biologist Ronald Fisher, who demonstrated that Mendelian inheritance acting across many genes, each making a small but additive contribution to the trait, could nonetheless in combination account for continuous variation. During this period it became clear that genetic theory and evolutionary theory were not distinct but were in fact rather tightly linked. Indeed, it was genetic biology that provided a mechanism by which the effects of selection operated. Complex models were developed and evolutionary biology became formalized in mathematical terms (which are beyond the scope of this book) (Box 2.8). Since this Modern Synthesis of evolution and genetics, a common definition of evolution is “change in the allele

---

**Box 2.8 Equations of Evolution**

Evolutionary biology has a strong quantitative heritage. A key element in the Modern Synthesis was the contribution of biometricians showing that Mendelian genetics and natural selection were compatible concepts. Quantitative population genetics became a major component of evolutionary research. For a diploid species it is important to distinguish between the *allele frequency* (the proportion of that allele across all gene copies in the population) and the *genotype frequency* (the proportion of a population with a specific allele pair at a particular locus). It is their interdependencies and how these are defined under different conditions that are at the heart of quantitative evolutionary biology. Evolutionary change can be defined as a change in the genotype and allele frequencies from generation to generation. A change in such frequencies is taken, all else being equal, as evidence for selection. These formal models generally require a number of assumptions about the population under study which are often not met, particularly if the population is small.

The **Hardy–Weinberg equilibrium** is a simple statement of how allele frequencies in a population have stable
proportions in a population at equilibrium (that is, not undergoing active positive or negative selection). It is based on several assumptions, including that all individuals within the population can mate at random (that is there is no assortative mating or male dominance), and that the population is in equilibrium. In the simplest case of two alleles, $A$ and $a$, there are three possible genotypes: $AA$, $Aa$, and $aa$. If the allele frequency of $A$ is $p$ and that of $a$ is $q$, then the genotypes are distributed in each new generation as $p^2$, $2pq$, and $q^2$. Importantly, for a population in equilibrium both the genotype and allelic frequencies remain in the same ratios over successive generations. The mathematics is obviously more complex when there are multiple alleles and multiple loci involved, but the principles remain. The mathematics underlying the relationships between two loci in the Hardy–Weinberg model allow estimates of linkage disequilibrium; that is, the deviation of the two-locus allele frequencies from the product of the respective single-locus allele frequencies (it is thus a measure of recombination).

It is therefore possible to examine the allelic and genotype frequencies in a population of interest to see if they are compatible with the Hardy–Weinberg formula. If not, then one of the assumptions has been broken: mating may not be random, the population is too small and genetic drift may be occurring, there may be migration of others into the population, or there is a selective process conferring unequal survival/reproduction rates on some of the genotypes. This latter possibility means that if other assumptions have been dealt with, say by having a very large population to study, evidence for selection can be inferred from examining allelic and genotypic frequencies.

Conversely, where the population is in equilibrium but there is a clear disadvantage to possessing the homozygous state of one allele, then it may imply that the heterozygotic state of the persisting disadvantageous allele confers a selective advantage. This is balancing selection; the classical example of this is the heterozygotic advantage in resistance to malaria of carrying one allele coding for the hemoglobinopathy underlying sickle cell anemia.

Allelic frequencies often differ between subpopulations of a species, reflecting founder effects, drift, non-random mating, or selection acting on the population. Examples of each are found within the chapters of this book. There are many mathematical techniques used in evolutionary biology to define and dissect out these different mechanisms.

The study of variation is central to evolutionary biology. It is of historical interest that analysis of variance techniques were first developed and used by Fisher to define and mathematically separate genetic and environmental determinants of phenotype and their interactions in artificial selection (animal breeding) experiments, and to seek evidence in wild populations for natural selection. Variations may be induced by genomic, developmental (epigenetic), or environmental factors, and it is clear that it is difficult if not impossible to fully separate these effects as they are highly interdependent; indeed, much of modern evolutionary developmental biology is based on the inability to do so. However, there are obvious practical advantages in trying to understand the relative importance of these different factors in generating variation in a trait.

A major driver in early quantitative genetics was animal and plant breeding. A key question was to what extent is the variance in phenotype based on genetics. The heritability (generally indicated as $H^2$) of a trait was defined from these simple estimates of variance as the ratio of the variance due to additive effects in the genotype (i.e., ignoring epistatic effects) to the total phenotypic variance of the population. The most common way of estimating $H^2$ is to look at the regression between parental and offspring measures of a trait. If there is a high correlation between the measures in the two generations, it would imply a strong genetic basis for that trait, and if there is no correlation then there is no genetic contribution to the variation and, from the breeder’s perspective, no value in a selection program.

There are important limitations on the estimates of heritability, and most of the literature assumes two sources of variation—genetic and environmental—and considers the latter largely as a noise factor. The recognition that not all that is inherited across generations need be genetic but might be epigenetic has confounded matters; even more confounding has been the growing recognition of the role of developmental plasticity and the developmental environment itself (see Chapter 4). By definition, $H^2$ is determined by examining a population in a particular context, and under different environments the observed heritability may be very different. The heritability of adult height in a population living in circumstances of variable food insecurity and frequent infection during childhood will be very different from that of another population living under optimal conditions in childhood. Obviously a breeder with a plant or animal population will seek an optimal environment. Such estimates are not possible in humans.

But constraints may also play a role. A plant might be selected for optimal growth, and there may be a high variance in the founder population which has a high genetic component. But after generations of selection all the plants may have reached very similar heights because of some other constraints: the genetic variance on height is effectively reduced to zero! This example, while using artificial selection, shows that evolutionary processes themselves can affect estimates of heritability.
The effects of selection can be measured mathematically using these approaches. The classic breeder's equation is \( R = h^2 S \), where \( R \) is the response to selection and \( S \) is the selection differential (the difference between the mean trait value for the whole population and the mean trait value for the individuals who contribute to the next generation). For example, if the heritability of height in a particular species is 50%, with the other 50% of variation caused by environmental effects \( (h^2 = 0.5) \), then if a breeding program attempts to increase the height of a population of mean height 100 cm by breeding only from individuals 120 cm tall \( (S = 20 \text{ cm}) \), then \( R = 10 \text{ cm} \) and the mean height of the next generation will be 110 cm.

The selection coefficient, \( s \), measures the strength of selection acting on one genotype versus another genotype. If the relative fitness (see Section 2.2) of a wild-type genotype is set at 1, and a mutation in the genotype acts to reduce fitness, then the relative fitness \( W \) of the new genotype will be less than 1, and \( s = 1 - W \). The value of \( s \) indicates the strength of selection, ranging from 0 (no selection) to 1 (complete selection, that is the new genotype is eliminated from, or becomes fixed in, the next generation). Thus, the value of \( s \) provides an indication of how quickly an allele will be eliminated from (or spread through) the population. For example, an advantageous new mutation with \( s = 0.1 \) might take about 100 generations to reach 80% allele frequency in a population, but an order of magnitude longer if \( s = 0.01 \). A selection coefficient of 0.1 is considered high. There are difficulties in estimating selection coefficients in human populations because of the various non-genomic factors that can affect fitness. Nevertheless, one of the best-established is the value of \( s = 0.05–0.1 \) for the mutation underlying lactase persistence (see Box 1.1), which spread rapidly through pastoral populations in northern Europe and Africa.

The classic tool in humans for trying to separate genetic from environmental factors has been twin studies, where monozygotic twins with the same genotype and assumed to have the same prenatal environment are evaluated in comparison with dizygotic twins who do not have the same genotype. Studies have then tried to look at the role of the environment by looking at monozygotic twins reared apart. However, monozygotic twins need not be genetically identical, and indeed there is growing evidence that they are not (see Box 4.4). Indeed, their fetal environments are not identical: the nature of placentation is such that each twin does not receive identical nutritional support in utero. On the other hand, it can be argued that each is in a similar but atypical developmental environment; that is, there may be influences from early life of developing in a polytocous rather than monotonous environment. For these reasons, twin studies, while useful, are not absolute in their interpretation.

frequencies within a population over time.” As we will see in Chapter 3, molecular genetic studies of allele frequencies in humans provide evidence for the selective pressures that have shaped populations in the past and continue to do so today.

### 2.4 Areas of Debate and the Limitations of the Adaptationist Argument

Although the concept that organisms evolve through time is now accepted by biologists as fact, there is still debate on the specifics of how evolution operates, as well as some misconceptions about its nature.

#### 2.4.1 Does Evolution Have a Direction?

The popular but erroneous view of evolution is that it has a direction, proceeding from “lower” to “higher” species and that evolution implies “progress.” It is very easy to fall into the trap of teleology: believing that if some feature of an organism performs a function then it has been chosen to achieve some goal or purpose. The extensive use of metaphors in evolutionary biology can unfortunately make this more likely. And it is all too easy to suppose that every trait is an adaptation, the result of positive selection. Let us look at some of these misconceptions.

First, there is no reason to suppose that evolution has a direction, and the notion of progress, with the implication that “higher” organisms are in some way superior, leads us back towards the anthropocentric view that humans are at the pinnacle of evolution. The trend towards increasing complexity in living organisms (from bacteria to mammals) is not in itself an argument for progress, because bacteria may represent the minimum level of complexity for an independent organism and therefore selective
pressure can only result in increased complexity. Sometimes evolution leads to greater simplicity, as in the loss of eyes in the case of cave-dwelling fish.

Second, perhaps part of the problem lies in the metaphorical term natural selection, which carries with it the idea that some external process is responsible for “selecting” the fittest individuals. Yet natural selection is merely the result of differential reproductive success among individual organisms, a numerical construct that requires no external selector. Another related issue is the notion that evolution “improves” the species. Perhaps it does if improvement is defined narrowly as increasing the match between the species and its current environment so that reproductive success is maximized in that environment. But it must be remembered that such a match only exists as long as the environment does not change. In reality, all species are evolving in parallel, and many adaptations do not improve the fit but merely maintain it against a shifting backdrop.

More importantly, that so-called improvement is defined only in terms of reproductive success, and other dimensions such as the longevity or health of an individual organism are ignored by natural selection in so far as they do not affect fitness. This tension between reproductive fitness and health is a key concept in evolutionary medicine and is the focus of Chapter 7. Importantly, evolution does not act for “the good of the species”—an argument which was frequently used in the debates over levels of selection. While group selection may possibly be to the advantage of a particular group, that is not the same as a (flawed) teleological argument regarding the “good of the species,” and in any event, as we have discussed, in practical terms most selection occurs at the level of the individual.

2.4.2 Selection is Not Random

One of the commonest erroneous statements about evolutionary processes is that they are entirely random. Whereas the preceding discussion has emphasized that evolutionary processes do not have a direction, this does not mean that there are no constraints on what is possible. This is particularly evident in concepts such as stabilizing selection, where extremes in traits in either direction are selected against, and indeed we see that in many phenotypic traits. While most mutational and recombination events are random, selection is not. Selection can only act where a fitness advantage exists. Thus over time there is an improved match between the phenotype and the selective environment. Further advantageous mutations can only build on what has already evolved. Thus, for example, the fins of the earliest vertebrates became the four legs of terrestrial mammals and all vertebrates have the same basic body design of a dorsal neural tube and a ventral gut.

Virtually all traits involve the cumulative effects of multiple mutations and selection effects and thus there is a contingency to the process. Convergent evolution is the process by which similar phenotypic traits, serving the same function, evolve independently as a response to similar environments. Frequently quoted examples are the wings of birds, bats, and insects, and the similar eyes of vertebrates and cephalopods; these organs perform the same functions but in each group these structures are profoundly different from each other and arise via different developmental pathways. Biological observation of convergent evolution is often quoted as consistent with evolution by descent but inconsistent with any suggestion of “design,” because why would any rational designer use different templates to achieve the same functional outcome?

There are many examples of convergent evolution in human populations. Lighter skin color, probably an adaptation to reduced sunlight in human populations who migrated away from equatorial regions, has arisen by different biochemical mechanisms in Europeans and East Asians (see Section 6.5.2). The persistence of lactase in adulthood, an adaptation to pastoralism allowing adults to consume milk well beyond the age of weaning, is found in those European, East African, and Middle Eastern populations with a history of domesticating milk-producing ruminants (cattle, goats or sheep, and camels), and each population has a different causative allele with a different ancestry (see Box 1.1). Physiological and genetic studies of Andean and Tibetan populations show that the two groups have adopted different solutions to the problem of delivery of oxygen to the tissues at altitudes above 4000 m. In Andeans, the solution favors higher blood hemoglobin levels
and consequently higher arterial oxygen content, whereas the Tibetan population is characterized by high blood flow in association with high endothelial synthesis of the vasodilator nitric oxide, particularly in the pulmonary vasculature, as well as by higher capillary density in muscle (see Section 13.11.2).

2.4.3 Is Selection the Only Mechanism of Evolution?

In *The Origin of Species* Charles Darwin wrote: “Variations neither useful nor injurious would not be affected by natural selection, and would be left a fluctuating element . . .” That remark foreshadows the proposal by Kimura of the *neutral theory*, which challenged the role of natural selection as the major force in evolution. Kimura argued that random mutations and drift account for most of the variation between individuals at the level of DNA and amino acid sequences. Many mutations in the DNA sequence will be neutral (meaning without effect on fitness, or “neither useful nor injurious”) because they do not occur in a functional (i.e. protein-coding or regulatory) region, because of the degeneracy of the genetic code (most of the 20 amino acids are coded by more than one of the 64 possible codons), or because the amino acid specified by the mutated codon is functionally identical to the amino acid specified by the original codon (thus the term “synonymous” mutation), leaving the function of the protein unchanged. Kimura showed that, in the absence of selection, the frequency of a neutral allele in a population will drift stochastically through generations (because of random mortality among individuals) until it reaches 0% (i.e., it is lost from the population) or 100% (i.e., it is fixed in the population). The smaller the population, the fewer the generations needed for elimination or fixation. If two subpopulations arise from division of the original population, perhaps by migration of a few individuals, then because of the random nature of drift it is entirely possible that in time an allele could become lost from one population and fixed in the other. The two populations would therefore come to have different genotypes without the involvement of any form of selection.

Three important practical consequences arise from the neutral theory. First, if molecular variation (mutation) occurs at a constant rate and is not subject to selection, then the amount of sequence variation between two species will act as a *molecular clock* of the time since their divergence: this clock is used extensively in modern phylogenetic analysis. Second, in the example of formation of a subpopulation by just a few individuals (a *population bottleneck*), random drift will be a particularly important determinant of the subpopulation’s subsequent genotype, leading to rapid deviation from the genotype of the main population (the *founder effect*). Bottlenecks and founder effects are discussed in more detail in Section 3.4.1. Third, a mutation that might be neutral in many environments may become advantageous or disadvantageous in another.

*Neutral mutations* can, in the absence of positive selection for the trait, explain why a feature present in a past member of a lineage may become atrophic; for example, the coccyx is the atavistic remnant of a tail. Was it lost by positive selection because it impeded some aspect of bipedal locomotion or was it simply lost by neutral mutations because a tail had no advantage once our ancestor species became terrestrial? More clearly, loss of the capacity to synthesize vitamin C was covert and of no significance until humans started to sail the seas without access to fresh vegetables and fruits: in the nutritional environments of our ancestor species, vitamin C was ubiquitous, so there was no ongoing selection to ensure a capacity to synthesize it, and so neutral mutations to its critical synthetic enzyme could accumulate until its endogenous synthesis was no longer possible (see Box 6.3).

*Genetic drift*, or random drift, occurs when a subpopulation becomes isolated from its parent population and within that smaller population the range of allelic variations is less than in the parent population. Thus some potential genotypes are left in the parent population but not in the subpopulation (see Section 3.4.1). Such genetic drift is likely to have been an important part of human genetic differences across populations as they spread from Africa 60,000 years ago, and where very small numbers formed founder populations via bottlenecks the impact of drift may have been very large. Thus drift explains the low genetic diversity among Native American populations, exemplified by the low frequency of blood type A and the virtual absence of blood type B.
Gene flow refers to the effect where many sub-populations may be not entirely isolated and sexual reproduction between members of two groups that are normally isolated allows alleles in one population that may not be present in the second population to enter the second population. As human populations have become more mobile and less separated, gene flow is playing a major role in the changing genetic structure of populations.

2.4.4 Is Every Feature of an Organism an Adaptation?

In this chapter we have described how evolution operates by natural selection on variant phenotypes, resulting in adaptations that increase relative reproductive success. Yet the limitations of retrospective analysis mean that there must be debate about the extent to which any particular phenotypic feature is the result of natural selection, and about the extent to which other processes, such as drift, constraint, and phenotypic plasticity (see Chapter 4), can account for the variety of traits seen in living organisms.

The debate on adaptationism—the attempt to account for any trait by postulating its adaptive advantage—originated in criticism of the tendency of sociobiology to seek adaptive explanations for virtually every aspect of human behavior (see Box 11.1). The controversy reached its peak in a famous paper by Gould and Lewontin (1979) that used an architectural analogy (the decoration of the archways of St Mark’s Cathedral in Venice) to argue that features which currently perform a useful function (as surfaces for painting, in the example) are not necessarily specifically designed for that purpose, but may arise as consequences of another function (the support of the ceiling by arches, in the example). The paper took a further biological analogy from Voltaire’s Dr. Pangloss (satirized for his view that all was for the best in the best of all possible worlds), pointing out that while the human nose might function to support spectacles it had of course evolved to optimize breathing and thermoregulation.

Gould and Lewontin adopted the architectural term “spandrel,” the approximately triangular area between two adjacent arches and the ceiling they support, to refer to traits that have arisen as by-products of other adaptations. A biological example might be that blood did not evolve to be red just because we see red as a sign of danger, as in hemorrhage: its redness arises from the spectroscopic properties of the oxygen-carrying protein hemoglobin and our behavior, “seeing red,” is a spandrel.

Gould also introduced the related term exaptation, referring to a trait that currently performs a particular function but originally arose as an adaptation for another function. One example is the feathers of birds, which probably evolved as insulation in reptilian ancestors but were later exapted for flight. Extended gene families, in which multiple genes with different functions evolved from the same ancestral gene after gene duplication, can be considered a form of exaptation at the molecular level: one example is the steroid hormone receptors (Bridgham et al. 2006). The key message from this debate is that the current form and function of a trait must be considered in the context of the complete evolutionary history of the organism that carries it, rather than just on the present apparently adaptive significance of the trait.

It is easy to speculate on the origin of a trait, but adaptive claims need to be well supported by evidence and alternatives need to be examined, and this is not always straightforward. The fundamental criterion for an adaptation is that it must be shown to increase the fitness of an organism which carries it. In practice, formal measurement of reproductive success is difficult in complex and slowly reproducing organisms, and especially so in humans, and so other criteria must be used. Furthermore we must consider not just the possibility of direct effects on fitness, but also indirect effects.

The criteria that are used to infer an adaptive process include the use of proxy measures of fitness, such as growth and survival, or what might be called the design criterion: the biological plausibility of the putative adaptation occurring in response to the selective pressures arising during the phylogeny of the species. This of course is tricky: trying to imagine what a trait would be like if it had been actively selected for easily leads to teleological arguments. Such speculations without evidence are termed “just-so stories,” after the children’s tales by the English writer
Rudyard Kipling which purported in a fanciful way to explain why something is the way that it is, for example How the Leopard Got His Spots. Key therefore to evolutionary biology in general, and evolutionary medicine in particular, is to consider how hypotheses can be validated—this is challenging when the bulk of evolutionary biology is based on historical analysis. However, a range of approaches can be used, and these are discussed in Section 7.5.

The power of molecular biology allows an analytical approach to understanding the relationship between organisms, both within and between species, and thus to infer the likely evolutionary processes involved. But with respect to much of this book it needs to be kept in mind that while a given adaptive scenario may be likely, if not compelling, the evidence is nearly always inferential. Thus adaptive arguments in human biology must always be considered, to some extent, to be hypothetical. For example it is generally considered that bipedalism is the earliest defined characteristic of the hominin clade and yet, as will be discussed in Chapter 6, adaptive explanations for its origin remain speculative and many other, equally plausible, speculative arguments have been put forward. So while there is a consensus that in some way bipedalism provided an advantage in the shift from an arboreal to a terrestrial existence, in what precise way must remain speculative.

Thus, whenever in this book an adaptive argument for the origin of a trait is proposed, the caveat exists that while the particular argument represents the collective view of evolutionary biologists, it cannot necessarily be proved or disproved. We must always remember the potential for a plausibly adaptive trait to be a spandrel, to be an exaptation, or to have arisen through random drift and neutral mutation rather than through selection.

2.4.5 How do Species Evolve?

This book is mainly concerned with microevolution: the genotypic and phenotypic changes that have occurred since the emergence of H. sapiens, and their consequences for our health and disease. But the adaptive environment of hominids has included, and still includes, many other species of various different kinds including, for example, the gut microbiota. Indeed in this case there is an intimate co-dependency of evolutionary processes for both the host and the microbiota, as the phenotype of one species influences the selective environment for the other—we term such processes coevolutionary.

The processes that have determined how species have diverged since the beginning of life on the planet are termed macroevolution. Central to macroevolution is speciation, the process that causes a new species to emerge and that thereby leads to a branching point in phylogeny. Further divergence and branching of the resulting two lineages eventually leads to the taxonomic diversity we can observe today; that is, the huge variety of animal, plant, and microbial species.

Before discussing speciation, we need to have a clear idea of what constitutes a species. In our daily lives, this appears self-evident: humans, cats, spiders, and pine trees are separate species. But what about different sorts of cats? Are alley cats and pedigree Siamese different species? Given such substantial variation between individuals within a species, where do we draw the line between one species and a closely related species? In fact, biologists have several different definitions of a species, often depending on the disciplinary background of the proposer, whether evolutionary biologist, taxonomist, molecular geneticist, or botanist.

We will use the biological species definition proposed by Ernst Mayr in the middle of the twentieth century—that a species is a group of actually or potentially interbreeding populations which are reproductively isolated from other such groups—while accepting the existence of other definitions based more on phylogeny, and also the difficulty with the Mayr definition when applied to species which do not engage in sexual reproduction, such as most bacteria, many plants, and a few vertebrates. This difficulty in definition highlights the point that macroevolutionary and microevolutionary processes are part of a continuum underpinned by the same general principles.

Therefore reproductive isolation is in practice the key test to distinguish two species, and this accords with our day-to-day experience: in general, different species do not mate with each other, and if they
do so then the hybrids are usually sterile, as are, for example, the mules that are the progeny of mating between horses and donkeys. Note also that the definition does not include any notion of morphological similarity—different populations, even if individuals vary within a population and populations vary between each other, are considered as a single species as long as fertile interbreeding between the individuals of those populations can take place. Alley and Siamese cats can interbreed—usually to the chagrin of the owner of the Siamese—and so are the same species. This point is particularly important when considering the wide diversity of human populations: despite our differences, there is only one extant species of *Homo*.

How does speciation occur? Our definition of species implies that a reproductive barrier must arise between two populations for two different species to emerge. Such a barrier can be most easily imagined as geographical: some form of physical separation, either large scale because of topographical or climatic change or small scale because of local variation in the habitat of a slowly dispersing species. This is called allopatric (“different homeland”) speciation. Imagine a single population that becomes separated, perhaps as rising sea levels create an island. Over time, the two populations will diverge, either because of natural selection causing different adaptations to the different environments or, particularly if one of the populations initially consisted of a few founder individuals, because of genetic drift. Eventually the diversity will reach a point where the two populations cannot interbreed even if the physical barrier separating them is removed, and a new species will have formed (Box 2.2 and Figure 2.5).

There is broad agreement among evolutionary biologists that allopatric speciation is responsible for much of the current biodiversity. More controversial is the role of sympatric (“same homeland”) speciation, in which a population diverges into two reproductively isolated populations without the existence of any physical barrier preventing mating between the two groups. The driver for sympatric speciation might be, for example, adaptation to a different choice of food resources, creating “specialists” with high fitness on one or the other resource, whereas “generalists” who could use either resource but with lower average fitness would be out-competed as long as both resources continued to be plentiful.

Nevertheless, since gene flow resulting from mating among the diverging populations would tend to act against speciation by reducing the necessary divergence, the existence of sympatric speciation has been questioned by many evolutionary biologists. The counter-argument is that such mating would be reduced if the emerging “specialist” populations tended to cluster on or around their preferred resource, thus effectively reducing sympatric speciation to a special case of allopatric speciation.

![Figure 2.5 Model of allopatric speciation. The ancestral species is divided by geographical isolation, leading to reproductive isolation and speciation. Even if mixing subsequently occurs by removal of the geographical barrier, interbreeding does not occur.](image-url)
Whatever the mechanism, it is clear that speciation is a gradual process. Estimates of the rate of speciation distinguish between the time for speciation (that is, the time taken for reproductive isolation to emerge) and the biological speciation interval (that is, the average time between speciation events). Estimates for these values in natural populations range from tens of thousands of years to several million years for the time for speciation, depending on the type of organism studied; estimates of the time between speciation events produce values of a few million years. Yet, slow as these processes might sound, they are more than adequate to explain the biodiversity that has arisen in the past 3.8 billion years (Figure 2.1).

Speciation is the fundamental event responsible for biodiversity, because reproductive isolation prevents gene flow between species, preserving the genotypic, phenotypic, and morphological diversity that is produced by selection as those species evolve in a range of environments. The rates of speciation quoted above emphasize the slow, gradual, and continuous nature of evolution: this is termed **gradualism**.

Yet there have been suggestions that evolution operates less smoothly, that species remain unchanged over long periods of time but that sudden large changes then occur which give rise to new species. This hypothetical process is referred to as **saltationism** (from the Latin ‘saltus,’ meaning a jump) or macromutationism, and is based on the frequent lack of intermediate forms between species, both in the fossil record and among extant species. Saltationism supposes that a macromutation—a single mutation with a large effect, most likely disturbing development of the early embryo—could produce phenotypic changes large enough to create a new species. Although most such mutations would be expected to be highly deleterious and the resulting phenotypes to be non-viable, occasionally a new phenotype might not only be viable but would actually represent an “improvement” over existing phenotypes, with a set of features that qualify it as a new species. In the memorable words of Richard Goldschmidt, an early proponent of saltationism in the 1930s, the new phenotype would be a “hopeful monster.”

The discovery in the 1990s of **Hox** genes, a cluster of genes which code for transcription factors and appear to regulate development of the fundamental body plan, caused some biologists to suggest that they are candidate gene targets for saltational evolution. Indeed, mutation of **Hox** genes in invertebrates can result in large changes in phenotype such as altered numbers of body segments or additional pairs of wings, and changes in the arrangement or copy number of particular **Hox** genes appear to be associated with major morphological changes during evolution. Nevertheless, evidence for macromutation as a frequent mechanism for speciation remains scant, and most evolutionary biologists are now skeptical of the role of hopeful monsters in evolution.

A concept often confused with saltationism is Gould and Eldredge’s theory of **punctuated equilibrium**, which also challenges neo-Darwinian gradualism. In contrast to saltationism, **punctuated equilibrium** accepts the steady accumulation of small mutational changes as the fundamental mechanism of evolution, while proposing that morphological change corresponding to such molecular variation is generally suppressed, and that species change little over geological time. Then, **punctuated equilibrium** proposes that speciation occurs over relatively short periods (“short” in this context meaning thousands of years) in which hidden genetic variation is expressed, often in a small subpopulation which is under some form of environmental stress, and that major trends in adaptation depend on the formation and extinction of similar but non-identical species during such speciation events. Punctuated equilibrium remains controversial, although perhaps less so than in the years immediately after it was proposed in 1972, as it becomes appreciated that its fundamental mechanisms of variation and selection do not conflict with Darwinian theory.

**Introgression** refers to evidence of horizontal gene transfer between species that can occur either because the hybrid is sexually fertile or in asexual species because there can be transfer of genetic material between species by a variety of mechanisms, such as horizontal gene transfer (Crisp et al. 2015). The presence of introgression highlights the complicated nature of the definition of species. The most common situation of introgression in sexually reproducing species is likely in the early stages of speciation when species are not completely isolated.
Recent molecular studies have shown evidence of introgression between *H. sapiens* and other *Homo* species (Chapter 6).

### 2.4.6 How Fast is Evolution?

We are accustomed to thinking of evolution as a leisurely process: the “slow and gradual accumulation of numerous, slight, yet profitable, variations” described in *The Origin of Species*. Indeed, as already discussed, speciation in multicellular organisms is usually a process that is too slow to study in the laboratory, the field, or indeed the lifetime of a researcher. Nevertheless, trends towards formation of separate species can be observed in rapidly reproducing organisms subjected to artificial manipulation. For example, formation of subpopulations characterized by distinct food preferences and reduced interbreeding ability has been observed in insect species placed on new plant hosts.

Microevolutionary adaptation to a new environment is a much faster, and easily demonstrable, process. This is sometimes termed *contemporary evolution*. Rapid adaptation implies strong selective pressure, or in other words a marked change in an organism’s environment to which it must adapt. Chapter 10 describes the evolutionary strategies adopted by viral and bacterial pathogens to evade host defenses, such as the rapid appearance of subtypes of HIV or *Salmonella* within the course of a single infection, and the evolution of antibiotic-resistant strains in bacteria. But rapid adaptation can also be seen in some vertebrates. Small male fish (guppies) that are brightly colored in a predator-free environment evolve over a few generations towards a duller coloration when predators are artificially introduced into their tank. Conversely, the dull-colored male fish from a high-predation environment evolve rapidly as a result of sexual selection (female fish prefer flashier mates, it would seem) to produce more brightly colored guppies when predators are removed.

Adaptation to new environments can also be detected in humans, although with our longer generation time of 20 or so years and low fecundity, the time for a new variant to reach detectable levels in the population is much longer. In Chapters 3 and 6 we discuss how humans have adapted over the past 50,000 years to the new environments caused by our dispersal over the planet, and present the population and molecular genetic data underlying some of these changes.

### 2.4.7 How do we Explain Traits That Appear to Reduce Fitness?

Species may have characteristics that seem disadvantageous even within the environment in which they evolved. The menopause is essentially unique to humans: women cease to reproduce, often decades before the end of their lifespan. Superficially this could be seen to reduce fitness, but because humans are slow-maturing animals and childbirth was in the past a hazardous process, a mother’s direct fitness might have been enhanced if she ensured a situation where she lived long enough to nurture her youngest child to independence. Mathematical models suggest that this might be a partial evolutionary explanation of the origin of menopause. But there is an alternative, not mutually exclusive explanation, because by stopping reproduction a female is able to support her own daughters as mothers and thus might increase her inclusive fitness. Modeling suggests that the combination of both these advantages provides a fitness advantage (Shanley et al. 2007). Other evolutionary explanations for the menopause are also plausible (see Sections 5.5.3 and 8.9.7). Similarly, earlier in this chapter we discussed sexual selection as a mechanism which can lead to extreme traits and thus to “handicaps” such as a peacock’s tail.

Elsewhere in this book we will discuss the extent to which the human phenotype has been affected by sexual selection (Sections 7.4.7 and 8.7). The mere presence of sexual dimorphism in body size or our secondary sexual characteristics and the peculiarities of the distribution of human body hair appear to be based on sexual selection. More controversially, we will discuss the role of sexual selection in generating human behavior, a matter that has given rise to much debate.

### 2.5 Conclusion

This brief overview of the major components of evolutionary biology has focused on those elements
2.5 Conclusion

Evolution cannot be seen as a random process. Natural and sexual selection operate on the variable phenotypes of individuals within a population. If there is differential passage of genetic variants to the next generation, through either differential survival or reproductive success, the lineage evolves. By its very nature, most evolutionary change is subtle over a single generation and the nature of the change is constrained by the genotype (a mutation can only act on the pre-existing genetic structure), the constraints of gene-repair mechanisms which reduce the probability of a persistent mutation, and the various integrative and physical constraints (e.g., a large body cannot be supported on miniscule legs). Most gross mutations are likely to be harmful and thus advantageous variants in gene structure will likely be subtle. The emergence of complexity over evolutionary time is discussed in Box 2.9.

Because selection can only act on what is there, and is dependent on the survival differential between organisms within the same environment,

Box 2.9 The Evolution of Complexity

There is no progression in evolutionary processes as natural selection has no inherent direction; but yet over evolutionary time organisms that seem more complex have emerged. How can we reconcile these two views?

The first problem is how to define complexity: more genes, the development of the nucleus, multicellularity, more complex regulatory systems? Another way to view it is to consider, as John Maynard Smith and Eors Szathmáry did in their book Major Transitions in Evolution (Maynard Smith and Szathmáry 1995), whether there have been some key “watershed events” in evolution. They propose these to include:

1. Replicating molecules become populations of molecules in compartments.
2. Entities capable of independent replication become chromosomes.
3. Development of RNA as a replicating molecule, of DNA as the gene, and of proteins as enzymes.
4. The formation of nucleated cells with mitochondria (eukaryotes) from prokaryotes.
5. Development of sexual reproduction: from asexual clones to sexual populations.
6. The development of multicellularity and tissue differentiation, evolution from protists to animals, plants, and fungi.
7. Evolution of non-reproductive castes: from solitary individuals to colonies.

Common features for each of the “major transitions” include the loss of independence, where entities once capable of independent replication can, after the transition, only replicate as part of a larger unit; division of labor/task specialization; and changes in information storage and transmission, which applies both to the genetic code and to language. In each case once a “breakthrough” has been made the range of evolutionary possibilities changes. Importantly, microevolution and macroevolution are not distinct forms of evolution; rather macroevolution is nothing more than the accumulation of microevolutionary change until effective reproductive isolation is present between populations. There is no reason to believe these transitions represent anything different.

So how does the appearance of complexity emerge? The simplest explanation is a passive one, as suggested by Stephen Jay Gould, who likened it to the path of a drunken man walking down a pavement. He has a wall on one side which stops him wandering to one side but he can weave back and forth over the pavement and road on the other. The evolutionary space is always open for more complexity to arise.

But with drift and selection some evolved features may also get lost; for example, some parasites dispense with some synthetic capacities of their own to rely on the host just as primates have lost the capacity to make vitamin C through neutral drift.

The vast majority of species have always been prokaryotic organisms, and this remains the case. Complexity theory shows that noisy self-replication coupled with finite resources and a capacity to gain information about the environment (i.e., a simulation of natural selection) leads to greater genomic complexity. Computer models suggest that self-organization will emerge out of complex systems, and that this evolution of modularity in turn allows more complexity to emerge through duplication. This duplication in turn allows greater complexity to emerge from a relatively small genome (see Chapter 3).
organisms from a lineage gradually become better matched or adapted to their environment. This assumes that the environment does not change. If the environment does change, then a different phenotype underpinned by genetic elements may be more advantageous and be selected. Thus humans living in different regions of the world may have differing characteristics, selected for in the diverse environments to which they migrated within Africa over the past 150,000 years and elsewhere over the past 60,000 years. In Chapter 6 we will discuss human diversity and the selection pressures that have operated to cause it.

Before discussing human evolution in some detail, the next two chapters will first build a more complete model of how the phenotype emerges. First we discuss the molecular basis of evolution, that is how the genome changes in relation to evolutionary pressures, and then the other set of major processes which drive the emergence of the phenotype, namely developmental plasticity and particularly epigenetic processes.

As in any field of science, concepts are continually being refined and debated and new experiments devised to test them. Our ability to study evolution at the molecular level in addition to at the level of the whole organism has greatly accelerated our understanding. Equally, it is important to recognize that not every characteristic of an organism necessarily evolved for adaptive advantage. It is easy to fall into the trap of developing an adaptive argument for every characteristic: strictly, a trait should not be termed adaptive unless it is shown to have conferred some fitness advantage. Often such proof is difficult to achieve in practical terms and assumptions must be made, but they must be made with caution.

As this chapter has attempted to demonstrate, by its very nature some aspects of evolutionary theory may not be amenable to empirical enquiry. This situation is not unique—it is common to most sciences. Whereas the areas of debate are often seized upon by those whose belief systems are threatened by the science of evolution, we hope that the reader can already see that, from what we do know, knowledge of evolutionary principles is enormously helpful for understanding human biology in health and disease.

### Key Points

- Evolutionary science explains how the huge diversity of present and past life forms arose.
- Heritable variation between particular characteristics (traits) of individuals causes differential reproductive success (fitness), leading to the accumulation of beneficial variations (adaptations) in subsequent generations.
- Changes in an individual’s genotype (caused by mutations or recombination) are the basis of heritable variation. For selection to act, those changes must cause differences in the phenotype.
- Selection acts on phenotypic characteristics influencing survival and reproduction (natural selection) or the ability to obtain a mate (sexual selection).
- It is a fundamental principle of evolutionary medicine that selection acts to optimize reproductive success, not necessarily the health or longevity of an individual.
- Random genetic drift can influence the evolution of a species, particularly in the presence of founder effects and population bottlenecks.
- While evolution does not have a purpose or a direction there are constraints on evolutionary possibilities, including those imposed by limits on variation and by the evolutionary history of a lineage.
- Not all the characteristics of an organism need have an adaptive explanation.
- Many adaptive arguments, no matter how plausible, must remain hypothetical rather than proven.
- Evolutionary thinking should avoid the trap of teleology.
CHAPTER 3

The Molecular Basis of Variation and Inheritance

3.1 Introduction

Each individual carries a unique genome consisting of the DNA in his or her chromosomes. Humans are about 99.5–99.9% similar to each other genetically, but that 0.5–0.1% difference, corresponding on average to about one nucleotide every few hundred base pairs of DNA, underlies the phenotypic diversity of humans. This ranges from superficial differences such as body shape, skin color, and hair color, to hidden ones such as the ability to digest various sugars, and—significantly—susceptibility or resistance to disease.

In this chapter, we briefly consider the molecular basis of human variation and describe how such variation may be studied. We then review how genomic information can be used to identify the selective pressures that humans have experienced during their expansion across the planet. Next we consider monogenic inherited diseases, with particular reference to why adaptive evolution has failed to eliminate such diseases from the population. We then focus on common multifactorial diseases, highlighting the difficulties in identifying their genetic determinants and discussing whether past selective pressures may have led to an accumulation of alleles that are deleterious in a modern environment. Finally we consider other methods of transmission of biological information across generations.

3.2 Genes and Disease

Some inherited diseases are caused by defects in single genes, resulting in absent or aberrant expression of the protein coded for by the gene, or failure of the regulatory function which that gene performs. Although individually most such diseases are rare, mutations in over 2000 genes are now known to cause disease in humans. An obvious question is why the disease-causing alleles of those genes have not been eliminated by selection in the past, given that their effects on survival and/or fitness are often severe. The reasons include the following: (1) even if the gene defect has highly deleterious effects on fitness, the prevalence of the disease may be maintained by new mutations that replace the alleles eliminated by selection (e.g., hemophilia); (2) the effects of the deleterious allele may not become apparent until after reproductive age, so that the parent may pass on the allele to a child before selection has had a chance to operate (e.g., Huntington’s disease); or (3) the deleterious effects of the disease-promoting allele are confined to or expressed most strongly in homozygotes, but heterozygotes for the allele have some selective advantage over non-carriers and this causes the frequency of the allele to be maintained in the population (so-called balancing selection; e.g., the protective effect of heterozygosity for the sickle cell allele against malaria).

The situation is different for the common diseases, such as type 2 diabetes mellitus, cardiovascular disease, rheumatoid arthritis, many forms of cancer, and psychiatric disease. Here, no single gene is responsible for the disease phenotype. Rather, the cause is multifactorial and generally involves the interaction of multiple susceptibility alleles, each adding only a small relative risk, with environmental and developmental factors also making important contributions. The identification of genes predisposing to common diseases is a difficult task.
and has only begun to yield results in the past few years with the advent of rapid methods for identifying and analyzing sequence variation in large cohorts of cases and controls (Wellcome Trust Case Control Consortium 2007). Another complication with these common diseases is disentangling the influence of cultural or behavioral transmission; for example, the apparent genetic transmission of obesity in a family may arise from transmission of unhealthy eating patterns, rather than susceptibility alleles, from parents to children.

This model of multiple alleles, each with a small effect, conferring a predisposition to disease which is modulated by environmental factors also applies to many common non-disease traits. So, despite the continuing flow of media reports, there are no single genes “for” tallness, intelligence, athletic ability, or creativity to be discovered, and the ability to manipulate such traits in individual human offspring remains the stuff of science fiction.

3.3 The Molecular Basis of Human Genetic Variation

3.3.1 What is a Gene?

The human diploid genome contains about 6.3 billion base pairs in men and 6.4 billion base pairs in women (the difference being due to the different sizes of the Y and X chromosomes). Less than 2% of this actually represents protein-coding sequences, containing the surprisingly small number of 20,000–22,000 genes. The remainder consists of sequences coding for functional RNA, regulatory elements, and a large proportion (80–90% of total bases) of non-coding (sometimes wrongly called “junk”) DNA.

Until recently the most common definition of a gene was a segment of nucleic acid that, together with its regulatory elements, coded for a protein. Such coding genes are composed of exons—the segments coding for amino acids—separated by intervening segments of DNA called introns. After transcription, the segments of the resulting RNA corresponding to the introns are removed by splicing to leave a messenger RNA that binds to ribosomes and codes for the protein. As shown in Figure 3.1, a typical coding gene consists of a promoter, which is the control region responsible for binding various transcription factors that switch gene expression on or off, followed by the thousand or so base pairs of the coding sequence arranged in 10–15 exons, all extending across several thousand base pairs of DNA (the average protein molecule in a cell is about 300 amino acids long, requiring 900 base pairs of coding DNA). The number and size of introns is highly variable even among genes coding for similar-sized proteins. Genes often have associated enhancer regions that also bind transcription factors to modulate expression; enhancer regions may be intragenic (located in introns) or remote (thousands or hundreds of thousands of base pairs) from the coding region.

There have been numerous definitions of a gene since the term was first coined by Johanssen, largely dependent on the biological discipline of the definer. For example, the distinguished evolutionary biologist (and one of the founders of evolutionary medicine) George C. Williams defined a gene as “any portion of chromosomal material that potentially lasts for enough generations to serve as a unit of natural selection” (Williams 1966). The geneticists who wrote the Guidelines for Human Gene Nomenclature defined a gene as “a DNA segment that contributes to phenotype/function” (Wain et al. 2002). A biological philosopher might define a gene as “a segment of DNA that carries biological information.” Certainly the recognition that segments of DNA could code for RNAs that are not transcribed into proteins but into regulatory RNAs of different types required a broader definition of a gene (see Box 3.1). The concept that there are both coding (for proteins) and non-coding genes, together with the vast array of functions of different classes of RNAs, has wide implications, not only for the definition of a gene and understanding gene function but also for the mapping of disease-causing mutations.

Additionally, mitochondria contain a separate but small genome of about 17,000 base pairs, with only a tiny amount of non-coding DNA. This genome codes for 13 proteins, all of which are involved in mitochondrial electron transport, and for the transfer RNAs necessary to translate the slightly different genetic code used by mitochondrial DNA. The mitochondrial genome in all eukaryotes has evolved separately from the nuclear genome. One of the
selective pressures on the mitochondrial genome has been for small size, and therefore efficient use of the translation machinery it encodes, leading to different codon usage. The unique properties of mitochondrial DNA—which is only inherited from the mother—make it useful in tracing maternal ancestry (see Chapter 6).

### 3.3.2 Mutations

Mutation is the ultimate cause of diversity in the genome. The term mutation covers a broad variety of events that vary in scale, cause, and phenotypic effect. Many, if not most, mutations are neutral in that they have no readily discernible effect on the phenotype at the time they occur. Some are deleterious to a greater or lesser extent and we would expect them to be eliminated by natural selection, whereas others are beneficial in the sense of increasing an organism’s fitness in its present environment, and we would expect such mutations to spread in the population; the reasons why such expectations might be wrong are discussed in Section 3.4.

Mutations are only relevant to evolutionary change when they occur in the germline, that is, the cell lineage giving rise to the gametes. Mutations in the remainder of the body (the soma) can cause disease such as cancer (Chapter 12) in that individual, but are not transmitted to the offspring.

Two broad classes of sequence variation can be distinguished among the 0.1–0.5% sequence diversity between individuals: single nucleotide polymorphisms (SNPs), which account for up to 0.1% of the differences, and structural variations, which account for up to 0.4% of the differences. The examples in Figure 3.2 show how sequence variation can lead to biologically relevant variation in phenotype.

### 3.3.3 SNPs

A SNP is a single-base substitution of one nucleotide for another at a particular site in DNA. The human
Box 3.1 Trash or Treasure: the ENCODE Project

A particularly intriguing result from the massive efforts of the Human Genome Project, which was first initiated in 1990 to sequence the 3 billion base pairs of DNA in the human genome, was that we possess around only 21,000 genes, with more than 98% of our DNA not coding for proteins. The dominance of the protein-centric view within biology led to such non-coding DNA being viewed as devoid of purpose and termed “junk” DNA. Nevertheless, it became increasingly evident that identifying the sequences and locations of genes could give few insights into how gene expression is regulated and how that could underpin human traits and diseases. The majority of disease-associated single nucleotide polymorphisms (SNPs) identified in multiple genome-wide association studies were located within non-coding loci rather than in specific genes, suggesting that clues to the genetic basis of disease may lie within these regions of the genome.

This provided impetus for the launch in 2003 of the ambitious Encyclopedia of DNA Elements (ENCODE) project that aimed to catalog the entire array of functional elements of the human genome. Using next-generation sequencing technologies, hundreds of researchers analysed 147 cell types to systematically generate 1640 genome-wide data sets for regions of transcription, DNA methylation, histone modification, long-range chromatin interactions, and transcription factor-binding patterns and activity. This culminated in 2012 in the simultaneous publication of a compilation of articles and reviews in several prominent journals (The ENCODE Project Consortium 2012). The analyses generated a staggering amount of data, considerably expanding our understanding of the regulation of gene expression. For example, it was shown that the non-protein-coding intergenic spaces were in fact occupied by enhancers, promoters, and regions encoding putative regulatory RNAs. Ultra-deep sequencing of RNA transcripts suggested that about three-quarters of the genome is amenable to transcription, with considerable overlap in regions being transcribed, raising questions about the current definition of a “gene.” Interactions between transcription factors were revealed to be combinatorial in nature, demonstrating context and region specificity. And, mapping of the spatial relationships between transcription start sites and distal elements showed that such long-range interactions could not be readily predicted by linear proximity. Since many of the newly identified potential regulatory sites had previously been associated in genome-wide association studies with pathologies such as Crohn’s disease and other inflammatory conditions, the ENCODE project was hailed for providing much needed data for understanding the role of genetic variation in these regions in human disease.

Perhaps the most notable—and widely publicized—of ENCODE’s findings was that specific biological function could, in fact, be assigned to more than 80% of the human genome. This provocative claim gained much traction within both scientific circles and the broader popular media, with some commentators touting the ENCODE results as sounding the “death knell” for the notion of “junk” DNA. However, many have questioned the validity of ENCODE’s definition of “functional,” which relied broadly on a capacity to encode a defined product such as a non-coding RNA or protein, or to demonstrate a biochemical property such as protein binding or methylation. This, it was argued, reflected the lack of consideration or appreciation of evolutionary concepts; instead, a “selected effect” concept of function, whereby a sequence and its function have been preserved by selection, was more befitting in this context. It has also been pointed out that the term “junk” DNA in fact referred to sequences for which gain or loss has little impact on fitness, and therefore that ENCODE’s definition of biological function by no means negates the concept of junk DNA (Doolittle 2013; Graur et al. 2013).

Despite these controversies, the ENCODE data have provided an invaluable head start to elucidating the complex network of controls over how genetic information in every cell is expressed, improving our fundamental understanding of human biology and the genetic basis of disease. Ongoing work is expanding the scope of the catalog with data for additional functional elements and cell types, and knowledge gleaned from this project could find applications in pharmacology and stem cell therapy.

Genome contains one SNP every few hundred nucleotides, with current estimates suggesting about 15 million common SNPs in the human population—“common” being defined as an allele frequency of over 0.5%. SNPs occur in both protein-coding and non-coding sequences; however, because SNPs in coding sequences are more likely to have phenotypic effects that will be exposed to selection, less variation is found in these sequences. SNPs that do occur in coding regions are divided into synonymous and non-synonymous polymorphisms. Synonymous SNPs do not affect the amino acid sequence of the
3.3 The Molecular Basis of Human Genetic Variation

The protein product because of the degeneracy of the genetic code, whereas non-synonymous SNPs affect the integrity of the protein product, either because one amino acid is replaced by another or because a stop codon is introduced which causes premature truncation of the protein. SNPs in non-coding regions that have regulatory effects can also cause significant phenotypic variation. One example is lactase persistence, which is caused by one of a number of adjacent SNPs in an enhancer region some 14,000 base pairs upstream from the lactase gene itself (see Box 1.1 and Figure 3.2a).

SNPs can arise from two distinct processes: mistakes in copying the DNA sequence during replication, and chemical or physical mutagenesis caused by environmental chemicals or ionizing radiation. Although the DNA polymerase that copies DNA at cell division is extremely accurate, having an error rate of only $10^{-9}$ to $10^{-11}$ per nucleotide, each replication of the diploid genome requires copying of $6 \times 10^9$ nucleotides, suggesting that an error will occur every few replications. Similarly, although complex mechanisms have evolved to detect and repair damage to DNA caused by chemical and physical mutagens, these mechanisms are also not entirely free from error.

Estimates of the mutation rate (the likelihood that a base substitution will occur at a particular site in

Figure 3.2 Genetic variation leads to biologically relevant variation in the human phenotype. (a) The T allele of a single nucleotide polymorphism 13,910 base pairs upstream of the start site of the lactase gene correlates perfectly with lactase persistence (LP). Homozygosity for the C allele predicts lactase non-persistence (LNP). Data are from 331 mostly Finnish individuals. Plotted from data in Enattah et al. (2002). (b) Copy-number variation for the amylase gene AMY1 predicts salivary amylase content (left) and correlates with population history of starch consumption (right). From Perry et al. (2007), with permission.
the germline per generation) can be made using a number of approaches. The most recent estimates come from directly sequencing the whole genomes of parent–offspring trios. Such studies suggest a calculated mean mutation rate of slightly over $1 \times 10^{-8}$ per site per generation (Campbell and Eichler 2013). The difference in the proportion of mutations arising from the male and female gametes is interesting: overall, more than 75% of the new mutations arose from the paternal germline, with the number of such mutations increasing with paternal age. This would be expected from the additional mutational load in the male germline that arises from the lifelong cell division involved in spermatogenesis, compared with the fixed number of cell divisions needed for oogenesis.

Early estimates of the mutation rate were based on the prevalence of disease caused by a particular mutation. The number of new mutations per generation must equal the loss of mutated alleles from the population by death from the disease before reproduction (the mutation–selection balance). This approach was used by J. B. S. Haldane in the 1930s, with X-linked hemophilia B as the phenotype, to calculate a mutation rate for the affected gene of $2 \times 10^{-5}$ per generation. Since we now know that the gene coding for factor IX, which is mutated in hemophilia B, has a coding sequence of around 1400 base pairs, this estimate fits remarkably well with the values obtained from molecular studies.

Another way to estimate the mutation rate is to compare non-functional sequences (to eliminate the effects of selection) in species whose divergence times are known. Using this “molecular clock” approach with human and chimpanzee sequences gives a rate of base substitution of about $2 \times 10^{-8}$ per site per generation. Since we now know that the gene coding for factor IX, which is mutated in hemophilia B, has a coding sequence of around 1400 base pairs, this estimate fits remarkably well with the values obtained from molecular studies.

The best-known example of a disease-causing indel is the three-base-pair deletion removing the phenylalanine residue at position 508 of the cystic fibrosis transmembrane conductance regulator (CFTR) protein, one of the mutations causing cystic fibrosis (see Box 3.5).

**Indels** are short insertions or deletions (most frequently of one to three base pairs) within the DNA sequence. If they occur in a protein-coding region, indels are more likely to be deleterious than SNPs because they have the potential to cause insertion or deletion of one or more amino acids (if the indel is three or a multiple of three bases long), or cause frameshift mutations (if the indel is not a multiple of three bases) that may completely disrupt the production of the protein.

The presence of VNTRs leads to subsidiary bands (“satellites”) on DNA separations in the laboratory. The length of each repeat and the number of repeats vary widely: microsatellites have repeats of one to six base pairs and 10–30 repeats are typical; minisatellites have repeats of up to 100 base pairs and up to 1000 repeats are possible; and satellites have repeat units of several hundred base pairs. The variation of satellite DNA patterns
between individuals forms the basis of DNA fingerprinting used in forensic investigations.

Although most VNTRs are neutral, with little or no phenotypic effect, certain microsatellites are particularly dynamic and can expand rapidly in repeat number over a few generations. For example, rapid expansion of the CAG trinucleotide microsatellite within the coding region of the huntingtin gene causes Huntington’s disease, and the number of repeats predicts disease progression.

_Transposable elements_ are dispersed but repetitive DNA segments that can move within the genome by a “copy and paste” mechanism, which inserts a new copy of the element at a random site. Over 40% of the human genome consists of approximately 1.5 million copies of retrotrotransposons (transposable elements that replicate via an RNA intermediate). Many retrotrotransposons have originated from insertion of retroviruses into the genome (endogenous retroviruses or ERVs). Most retrotrotransposons are of ancient origin and have mutated to such an extent that they have lost the ability to be copied and inserted; however, several retrotrotranspon families are of more recent origin, some being young enough to be human-specific, and are still active in retrotrotransposition. Indeed, some retrotrotransposons have inserted into the human genome recently and this leads to populations that are polymorphic for the presence of an element at a particular chromosomal location, and hence the element can be used as a genetic marker of ancestry.

Retrotransposons can disrupt protein-coding or regulatory regions at the site of insertion, altering gene expression and creating a disease phenotype (Box 3.2). Retrotransposons may also have wider effects on genome structure by promoting recombination and deletion events that have evolutionary consequences. For example, a retrotrotransposition event caused deletion of the active-site exon of the human cytidine monophosphate-N-acetylneuraminic acid hydroxylase gene (CMAH) some 2–3 million years ago (Mya), after the divergence from our last common ancestor with chimpanzees (see also Section 6.4). This causes humans to be deficient in the sialic acid N-glycolylneuraminic acid (Neu5Gc), which is a common constituent of cell-surface glycoproteins in other mammals, with postulated effects on pathogen resistance and brain development (Varki 2009).

A remarkable example of an evolutionary role for an endogenous retrovirus is the syncytin genes which originally coded for viral proteins that promote entry of the virus into the host cell through fusion of viral envelope and host cell membranes. In mammals, syncytin genes have been co-opted to promote fusion of placental trophoblast cells to form a key component of the placenta. This is the syncytiotrophoblast, a multinucleated layer between embryonic and maternal tissue that mediates exchange of material between the mother and the fetus. It is likely that this exaptation of viral genes was a key event in the evolution of placental mammals (Lavialle et al. 2013).

---

**Box 3.2 Alu-mediated Mutations and Human Genetic Diseases**

_Alu_ elements are short sequences of DNA (about 300 base pairs) originally characterized by sensitivity to a particular restriction enzyme. There are over a million _Alu_ elements in the human genome, all derived from a transposon which inserted into a precursor species to primates and then replicated across the genome. Because they are widely distributed across the genome they make good marker sequences for following ancestry.

The distribution of _Alu_ elements throughout the genome can give rise to various human genetic diseases in two ways. The first is through simple insertion of an _Alu_ element with its disruptive consequences to normal coding. Diseases that have been attributed to _Alu_ insertions include breast cancer (involving the _BRCA2_ locus), hemophilia (_F9_), and neurofibromatosis (_NF1_). The second mechanism by which _Alu_ elements can induce the onset of inherited disease is through unequal homologous recombinations that cause insertion or deletion mutations. This has been shown to be responsible for a number of diseases including alpha-thalassemia (affecting alpha-globin) and hypercholesterolemia (affecting the low-density lipoprotein receptor) (Kaer and Speek 2013).
The extent of large-scale structural polymorphism in the human genome has only been appreciated recently. The basis of structural polymorphism is segmental duplication, whereby approximately 5% of the human genome consists of duplicated sequences in large blocks of thousands to hundreds of thousands of base pairs. Because of the similarity of these repeats, they are vulnerable to error (so-called non-allelic homologous recombination) during recombination, promoting deletion, duplication, or fusion of sequence elements that may be large enough to contain one or several genes.

The resulting changes in gene dosage may be clinically significant. For example, copy-number variation in the α7 nicotinic acetylcholine receptor has been implicated in several neuropsychiatric disorders (Gillentine and Schaaf 2015). Multiplication of the gene for the drug-metabolizing enzyme cytochrome P450 (CYP) 2D6, for which up to 13 copies have been reported in some people, leads to the so-called ultrarapid metabolizer phenotype (Ingelman-Sundberg 2005). If people with this phenotype are treated with standard doses of drugs metabolized by CYP2D6, they may experience treatment failure (as reported for some antipsychotics and antidepressants) or toxicity (as reported for the analgesic codeine, which is converted to its active metabolite morphine by this enzyme). Variation of copy number in the gene for amylase, which digests complex carbohydrates, is seen between human populations with nutritional histories of high or low starch consumption (Figure 3.2b).

### 3.3.5 Recombination as a Source of Variation

Individuals inherit their genes, not their genotypes, from their parents. This is because the meiotic cell division that forms the male and female gametes in sexually reproducing organisms involves two processes which shuffle parental alleles into a unique combination in each egg or sperm. The first is recombination, during which homologous chromosomes exchange segments of the DNA sequence, thereby creating new allele combinations on each chromosome (Figure 3.3). The second process is independent assortment of the grandpaternally and grandmaternally derived chromosomes into haploid gametes.

Each chromosome pair undergoes on average one or two recombination events per meiotic cell division. Recombination occurs preferentially (up to a thousand times more frequently) at short sections of DNA called recombination hotspots, compared with the intervening long segments in which there is low recombination activity. There are about 30,000 hotspots in the human genome, but exactly which ones are used in a particular meiotic division appears to be dependent on which variant of a DNA-binding protein called PRDM9 the individual happens to carry (Ségurel et al. 2011). The PRDM9 genotype varies by ancestry, and consequently so does the profile of recombination hotspots used. The PRDM9 genotype is completely different in humans and chimpanzees, and so is their profile of recombination hotspots; the human profile appeared about 1 Mya. PRDM9 is under strong selective pressure; the evolutionary reasons for this are unclear, but may involve the maintenance of a repertoire of potential new hotspots and thereby the potential for recombination and the generation of variation in the form of new allele combinations.

Two regions of the genome undergo little or no recombination, and therefore allelic combinations on these pieces of DNA are transmitted unchanged through generations. The Y chromosome in males has no homologous chromosome with which to recombine (with the exception of small regions showing sequence identity with the X chromosome), and it can be used in phylogenetic studies as a marker of male ancestry, that is, the patriline. Mitochondrial DNA is not involved in meiosis, does not recombine, and is inherited only from the mother via the thousands of clonally reproduced mitochondria present in the oocyte. Female ancestry—the matriline—can therefore be deduced by study of allele combinations present on mitochondrial DNA. Chapter 6 describes how study of these elements of the genome has led to conclusions about human evolution and dispersal.

### 3.3.6 Haplotypes and Linkage

Each member of a pair of chromosomes in a diploid cell contains a particular set of sequence variants
(or alleles) because each chromosome has been inherited from either that individual’s mother or father. These variants are physically linked by their presence on a single inherited piece of DNA, and the set of variants (or the subset of variants, if only a portion of a chromosome is being considered) is referred to as a haplotype (Figure 3.4). The term haplotype is a contraction of haploid genotype; for a diploid cell, the combination of the haplotypes of any pair of chromosomes provides the genotype. For example, if a chromosome has four polymorphic sites, each of which can exist as two variants (P/p, Q/q, R/r, and S/s), then the haplotypes of a chromosome pair in an individual might be PQRS and pqrs; here, one chromosome carries variants P, Q, R, and S and the other carries variants p, q, r, and s. Use of haplotypes for phylogenetic and association studies (see Box 3.6) increases the statistical power and reduces the sample size needed to detect significant associations. Haplotypes are often more useful in this context than individual SNPs, because disease states often involve more than one SNP.

Recombination during meiosis means that these sets of sequence variants are shuffled into different combinations on the chromosomes of the parent’s gametes by crossing over between the chromosomal pairs (Figure 3.3). In other words, haplotypes are broken up by recombination. Returning to our example of the diploid chromosome pair with haplotypes PQRS and pqrs, recombination might produce haploid chromosomes PQRs and pqrS in the gametes which then contribute to the next generation. Recombination can therefore create new
variation in each generation by creating new combinations of alleles.

If recombination is entirely random, then all combinations of alleles P/p, Q/q, R/r, and S/s will occur in the gametes with equal frequency. However, alleles at loci that are physically closer together on the DNA strand are less likely to be separated during recombination than are those that are further apart, as they are more likely to be present on the same chromosome and thus to be co-inherited. Such alleles are said to possess linkage disequilibrium. Linkage disequilibrium is also affected by the presence or absence of recombination hotspots between the two loci, and by the evolutionary time since the alleles at the two loci first evolved: even for closely spaced loci, linkage disequilibrium inevitably decays over time as recombination events separate them.

Nevertheless, when chromosomes or parts of chromosomes from several individuals are sequenced and compared, patterns of allelic variation are less than would be expected by chance. In other words, some degree of linkage disequilibrium is always present, and some allele combinations are frequent and some rare. Recent studies have demonstrated the existence of haplotype blocks, discrete segments of the chromosome each characterized by a few (typically three to five) distinct but common haplotypes, with each chromosome being a mosaic of the possible haplotypes in each block (Figure 3.4). Haplotype blocks are believed to be separated by recombination hotspots. The existence of this modular structure of chromosomes implies that variation within individual modules, that is haplotype blocks, can be identified by the presence of a few haplotype-specific variants (typically SNPs), potentially simplifying studies of population structure and ancestry.

Knowledge of haplotype structure and linkage disequilibrium is essential in mapping the genetic
loci associated with disease causation (see Box 3.6), in establishing patterns of selection, and in studies of population history (Chapter 6).

### 3.3.7 How Different are any Two Individual Genomes?

We have previously discussed how much the genomes of parents and their children differ (Sections 3.3.3 and 3.3.4). The Human Genome Project resulted in the publication in 2004 of the “reference” human genome, which is actually a composite sequence assembled from several anonymous donors (International Human Genome Sequencing Consortium 2004). Since then, the genomes of numerous people of different ethnic origin have been published, allowing estimates of how genomes of unrelated individuals differ from each other (Gonzaga-Jauregui et al. 2012).

It is found that genomes are highly variable, with on average over 4 million differences between each individual genome and the reference sequence, including over 3 million SNPs and more than a thousand significant structural variants, including CNVs. These add up to over 20,000 differences in coding regions, of which about half would lead to protein sequence variants. Altogether, about 15 million SNPs have been identified that differ from the reference genome. Non-synonymous SNPs are not evenly distributed across all protein-coding segments of the genome: some genes are highly conserved whereas others, such as those related to immune function and environmental sensing, are particularly variable.

Surprisingly, more than a hundred of the protein sequence variants in each individual sequenced were such as to cause loss of function of the protein product, with about a quarter of these potentially deleterious changes being homozygous at the particular allele, meaning that the gene would be non-functional. Since the individuals concerned were apparently healthy at the time of sampling, this may reflect redundant gene function, low penetrance (Section 3.5.1), a mild phenotypic effect lying within the range of normal variation, or delayed onset of any resulting disorder (Shen et al. 2013).

### 3.4 Factors Affecting Genetic Variation

In Section 3.3 we discussed the mechanisms underlying the surprisingly large amount of genetic variation among humans. In this section we discuss how evolutionary forces—such as genetic drift and selection—have determined the distribution of genetic variation within and between population groups.

The amount of genetic diversity in a population is determined by population size and history, and by the population’s environment and its ability to adapt to changes in that environment. Knowledge of present diversity is not just of academic interest: it not only tells us about human biogeography in the past but can also provide information about the selective pressures that have shaped human phenotypes to the various environments in which people live. Moreover, it allows us to place past adaptive changes in the context of the environment in which they occurred, and to predict whether such changes may no longer be adaptive in our modern environment.

#### 3.4.1 How Non-adaptive Mechanisms Affect Genetic Variation

Genetic variation in a species can arise not only by adaptation but also by chance (genetic drift) or by demographic events (bottlenecks and founder effects).

The fate of a new mutation (a new allele) in an individual lineage is, to a large extent, determined by genetic drift. The immediate survival of a new germline mutation depends on chance—that is, on whether it happens to be present in a gamete that is successful in forming a member of the next generation. What happens in the next few generations depends largely on two factors: first, whether and how strongly the mutation is favored or disfavored by natural selection, and secondly the effective population size (Box 3.3).

Consider first the case of a new allele with only mildly beneficial or deleterious effects, or neither (a “neutral” allele). Because individuals carrying such alleles have little or no fitness advantage or disadvantage compared with others in the population,
their survival or otherwise will be determined by chance. In a small population, such random drift will cause the new allele either to disappear (0% frequency in the population) or to become fixed (100% frequency in the population). Thus, small populations are more likely to be homozygous than larger populations. In a larger population, random variation is likely to maintain the allele at an intermediate frequency for a longer period, with little change in frequencies between generations, and the population will have a higher level of heterozygosity (Figure 3.5). These predictions are essentially those of the “neutral theory” of evolution (see Section 2.4.3).

Selection and drift have complex interactions. Consider a new allele which confers a marked fitness advantage to its bearer compared with others in the population. Whether selection predominates over drift depends on the size of the fitness difference—formally known as the selection coefficient—relative to the effective population size. If both the selection coefficient and the population size are large, then selection will be stronger than drift.

Modeling shows that even mildly advantageous alleles will increase rapidly (in evolutionary terms) in frequency in an ideal population, that is, one that is infinitely large. For example, an allele that imparts a 1% advantage in fitness will increase to 50% frequency in a population within a few hundred generations. In humans, changes in allele frequency on that scale would take several thousand years, and indeed this is the timescale observed when we look

**Box 3.3 Effective Population Size**

Population size is a major predictor of genetic diversity, measured as the variation in allele frequency caused by genetic drift. But what do we mean by population size? For humans, this question is likely to elicit the answer of about 7 billion, the total number of humans now living. But for the purposes of estimating the extent of genetic drift, the effective population size is very much smaller than the actual (or census) population size. This is because simple models used to calculate drift make a number of assumptions, including no overlap between generations, constant population size over time, random mating among members of the population, and no difference in reproductive success between individuals in the population. Brief reflection shows that these assumptions are not applicable to most organisms, and especially not to humans.

There is much overlap between human generations, with evolutionary consequences. The human population size has expanded at least 1000-fold within recorded history. We do not mate randomly; human expansion out of Africa resulted in a number of partly isolated subpopulations within which gene flow was limited (at least until the establishment of modern patterns of travel and migration). The human preference for assortative mating—the tendency for people to choose their mates according to phenotypic similarity—further tends to reduce diversity. There is variation among individuals in reproductive success: some individuals have many children and some have few. Since that variation is unequally distributed among the sexes, being greater in men than in women (see Sections 8.5 and 8.7), measures of drift will even vary according to whether autosomes or sex chromosomes are analysed.

The effective population size is therefore defined as the size of the idealized population (i.e., a population that conforms with the assumptions made earlier) that shows the same amount of genetic drift as the real-world population. Because of these considerations, the effective population size is not equivalent to the actual number of individuals in a population. Modeling shows that the factor that has most influence on effective population size is often the extent of historical fluctuation in the population, with the historically smallest population sizes having the largest effect.

There have been many estimates of the total effective population size of humans, using various genetic markers, and the results are always in the thousands rather than the billions, showing that the present extent of human diversity reflects expansion from a small population in the past. Although the discussion above uses the total human population as an example, these arguments can be applied to any identifiable subpopulation. Indeed, comparisons of the diversity (effective population size) of worldwide subpopulations can provide valuable evidence about the history and origin of different human groups. For example, estimates of the effective population size involved in peopling the Americas are in the range of 100–200, but this does not mean that only that number of individuals were actually involved in migration across Beringia some 10,000 to 15,000 years ago (see Section 6.5.1).
at human alleles that have undergone recent selection, such as that for lactase persistence. However, since it is rare for a new allele to be highly beneficial, and effective population sizes are generally much smaller than actual population sizes, modeling also shows that random drift is usually stronger than selection and most beneficial alleles will be lost from a population. The converse is true: drift may occasionally cause deleterious alleles to become fixed in the population.

In general, in the absence of selective pressure, allele frequencies within a population will tend to drift towards fixation (homozygosity) because of random drift, and most new mutations will be lost. Although natural selection will tend to increase the frequency of beneficial alleles in the population, and decrease the frequency of deleterious alleles, this directionality is strongly modulated by the random element introduced by genetic drift.

A demographic event that reduces population size will also reduce genetic variation as individuals with particular sets of alleles are lost from the population. The consequently smaller effective population size will also tend to favor drift over selection as the driver of change in allele frequency in the new subpopulation. Such demographic events are conventionally classified as population bottlenecks and founder effects (Figure 3.6 and Box 3.4). A population
bottleneck occurs when a previously larger population is reduced in number by some environmental event, for example disease or famine. Even if the population size recovers, the survivors will have only a small proportion of the genetic diversity that was present in the original population. A variation of a population bottleneck is a founder effect, which occurs when a small part of a larger population becomes genetically isolated, perhaps by migration or by some geographical event such as rising sea level. The isolated subpopulation will contain only a fraction of the diversity present in the main

Figure 3.6 Bottlenecks and founder effects. A reduction in effective population size, whether because of reduction in total population size by, for example, disease (a bottleneck) or because of isolation of a subpopulation (founder effect), will reduce genetic diversity. Subsequent population expansion is derived from only a small sample of the genetic diversity present in the original population.
population, and even in the absence of selection the allele frequencies of the two populations will tend to drift apart.

Analysis of effective population size as revealed by a number of human genomes (Li and Durbin 2011) suggests that humans seem to have gone through a number of population bottlenecks in their evolutionary history: possibly one about 3 Mya, one in African populations from 150 to 50 thousand years ago (kya), and a final one in European and Asian populations starting about 40 kya (presumably the founder effect associated with migration from Africa). Particular human populations will have their own unique histories of bottlenecks and founder effects, as illustrated in Box 3.4.

### 3.4.2 Signatures of Selection

How can we tell when selection rather than drift or neutral change has occurred?

Selection will actively affect allele frequencies by increasing the frequency of advantageous alleles in the population (positive selection) and decreasing the frequency of deleterious alleles (negative selection).

As described in Section 3.4.1, the frequency of an allele is a balance between random drift and selection, with population size playing a major role in determining which factor predominates.

The process whereby a beneficial allele increases in frequency in the population because of selection is referred to as a selective sweep. A further consideration relates to the source of variation that is available for selection to act on. Mutation is a source of variation, but it is necessarily a slow process (Section 3.3.2). If a population is faced with sudden environmental change, does it have to wait for the occurrence of a new mutation that will “fix the problem” and allow it to adapt to the new environment? The alternative is for selection to act on standing variation within the population—all populations contain such pre-existing genetic variation as a result of earlier, mostly neutral, mutations that are on their way towards elimination or fixation as a result of weak selection and/or drift.

Consideration of the dynamics of adaptation from new mutations or from standing variation suggests that the latter will be faster (because the population does not have to “wait” for an advantageous
new mutation to occur) and will have a particular genomic signature. A further feature of adaptation from standing variation is that it allows for the possibility of polygenic adaptation, where selection occurs simultaneously on multiple alleles that each have an epistatic effect on a particular trait, albeit each allele has only a small individual effect.

Tests for selection generally compare observed patterns of diversity with the patterns predicted by neutral evolution, with the assumption that any difference must be a result of selection (Lachance and Tishkoff 2013). Such tests analyse a large amount of DNA sequence data and are therefore dependent on advanced statistical and computational techniques that are part of the discipline of bioinformatics.

One test for selection is to compare the pattern of synonymous and non-synonymous mutations in protein-coding regions within and across species. For humans, such cross-species comparisons can include the chimpanzee as well as archaic humans such as Neanderthals. Since synonymous changes have no effect on the sequence of the protein product, they are assumed to reflect neutral evolution, whereas non-synonymous changes are assumed to reflect selection. A drawback of this method is that it is limited to the small proportion of the genome that codes for proteins, whereas it is becoming increasingly apparent that selection on regulatory and structural regions of the genome also has important evolutionary effects.

A more general method is to study patterns of linkage disequilibrium and haplotype diversity. The principle behind this approach is that an allele strongly favored by selection will rise rapidly in frequency in the population, bringing with it adjacent regions of the chromosome on which it lies (genetic hitch-hiking). Therefore a selective sweep will result in reduced genetic diversity around the target allele. In practice, this reduced diversity is detected as a long haplotype, because if the selective sweep is recent there will have been little time for the surrounding region to be perturbed by mutation or recombination (Figure 3.7).

A further method compares allele frequency distributions across populations, with the expectation that regions under selection in one population

---

**Figure 3.7** The principle underlying linkage disequilibrium. The top row shows an ancestral chromosome in which a mutation has occurred in a particular gene. The chromosome also contains along its length several genetic markers (1–8). The mutation is advantageous and positively selected for, ensuring its presence in subsequent generations. However, recombination events occur between the gene and the markers; the likelihood of a recombination separating the gene and the marker is proportional to their physical separation on the chromosome. Over generations, markers close to the mutation retain linkage disequilibrium with it more strongly than do more distant markers.
will have different allele frequencies from those expected as a result of neutral variation. Alleles in regions not subject to selection will generally be at fixation, whereas regions under selection will contain a higher than expected number of variant alleles.

Bioinformatic methods such as these can be used to infer the age of origin of alleles of interest, to identify targets of selection, and, when applied to the whole genome, to assess past population structure in terms of patterns of divergence and the occurrence of population bottlenecks. When they are used to generate maps of recent positive selection in geographically separated human populations they reveal both shared and region-specific signals. Such analyses have shown that little recent human adaptation appears to have arisen through strong selection after a new mutation, but rather has the characteristics of selection acting on standing variation (Pritchard et al. 2010).

3.5 Single-gene or Mendelian Disorders

3.5.1 From Genotype to Phenotype

So far in this chapter we have discussed the basis of genetic diversity. To appreciate how genetic variation may contribute to physical characteristics or disease we need to know how changes in the nucleotide sequence of DNA are reflected in the phenotype. It is important to note that most genetic variation does not result in any detectable change in the phenotype: as described in Section 3.3.7, there is a considerable difference between apparently healthy humans at the sequence level, and some quite marked changes that are predicted to knock out expression of specific genes do not have obviously deleterious effects. This may be caused by incomplete penetrance of the defects (see later) or by the robustness of developmental processes (Chapter 4).

But there are numerous examples where change in the DNA sequence in a particular gene can be directly linked to a deleterious effect on an individual’s phenotype, most likely because of an alteration in a protein-coding or regulatory sequence causing the amount or activity of a protein product to be altered. Such changes may be dramatic, such as a complete absence of the functional protein (e.g., phenylketonuria), or more subtle, such as production of a protein with a slight change in its function (e.g., thalassemia). In these Mendelian disorders the phenotype is determined by a mutation at an individual locus that is inherited in the simple manner described by Gregor Mendel. Although individual Mendelian disorders are rare, in total they may affect up to 10 in every 1000 live births.

What controls whether a variation in a single gene that affects the integrity of a protein will result in disease? There are three interlinked concepts:

1. Gene dosage. When the affected locus is on an autosome, that is, not a sex chromosome, expression of the protein will depend on whether both alleles (homozygosity) or only one (heterozygosity) carries the causative variation.

2. Dominance or recessivity. An allele at a particular locus is said to be dominant when its phenotypic effect overcomes the effect of the other (recessive) allele. Dominance may occur for several reasons. Consider a mutation in one allele at a locus within a gene coding for an enzyme or structural protein that results in the production of no, or a non-functional, protein. The other allele is unaffected (wild type) and so the genotype is heterozygous at the locus. We might expect functional protein to be made at half of the normal rate, and if this is sufficient for its physiological function then the phenotype will be normal and the wild-type allele is dominant. This situation is referred to as haplosufficiency. For example, in phenylketonuria the mutant allele of phenylalanine hydroxylase is recessive because the amount of enzyme produced by the other functioning allele is sufficient for normal metabolism of phenylalanine. The genotype would need to be homozygous for the mutant allele, meaning that no functional protein would be made, to express its phenotypic effect. Conversely, if the amount of functional protein produced by the single wild-type allele is insufficient for its physiological function, a situation referred to as haploinsufficiency, then the mutant allele will be dominant even when the genotype is heterozygous. For example, the breast cancer susceptibility genes BRCA1 and BRCA2 are inherited in a
dominant manner, implying that a single functioning copy of these genes is not sufficient for their normal tumor suppressor function. We can also imagine a situation where a non-functional protein coded by the mutant allele actually interferes with the function of the normal protein produced by the wild-type allele. In this case too the mutant allele will be dominant in a heterozygote; this is called the dominant negative effect. Some forms of von Willebrand disease, a bleeding disorder, are characterized by a defective protein product that arises due to interference by the mutant allele with the assembly of the multimers of wild-type von Willebrand factor required for blood clotting. A mutated allele may be also dominant because the individual normally only carries a single copy of the relevant gene: this is particularly the case for mutations on the X and Y sex chromosomes in human males that lead to conditions such as hemophilia B and red–green color blindness.

3. **Penetrance** refers to the correspondence between a genotype and the phenotype predicted from that genotype. In the context of a disease-causing mutation, if carriage of that mutation (whether homozygous or dominant) always leads to clinical expression of the disease, then the mutation is said to show complete penetrance. If, on the other hand, only a proportion of individuals carrying the mutation develop the disease then the mutation is said to show incomplete penetrance. If, on the one hand, only a proportion of individuals carrying the mutation develop the disease then the mutation is said to show incomplete, or reduced, penetrance. Incomplete penetrance can result from genetic, epigenetic, or environmental factors, and can vary according to a patient’s age or gender. Genetic factors include modifier genes and copy-number variation; epigenetic factors include skewed X chromosome inactivation and imprinting defects; and environmental influences are likely to be many but certainly include diet, developmental influences such as childhood trauma, and even altitude (Cooper et al. 2013). For example, the penetrance of Huntington’s disease depends on the number of CAG repeats carried by the individual and a person’s age—the disease only emerges in mid-life—whereas the apparent penetrance of phenylketonuria depends on the presence of phenylalanine in the individual’s nutritional environment.

### 3.5.2 Why hasn’t Selection Eliminated Monogenic Disease from the Population?

If it is axiomatic that disease reduces fitness, at least indirectly, as might be presumed to be the case for a dominant allele causing early onset of a severe disease, then why have the alleles that cause monogenic disease not disappeared from the population? There are a number of possible reasons for this.

Firstly, the selection pressure on an allele will depend on whether it is dominant or recessive, and whether it is expressed homozygously or heterozygously. Recessive alleles may be invisible to selection unless they are expressed homozygously, which is one of the reasons why deleterious alleles can persist within a population in carriers who do not exhibit the disease but can pass it on to their offspring if they mate with another individual who also carries the recessive allele.

Secondly, the deleterious allele may be maintained by **recurrent mutation**. That is, even if copies of the deleterious allele are lost from the population because individuals carrying them die before reproducing, the allele is created anew at some finite rate by mutation. If the prevalence of a disease is constant in the population, then this allows the mutation rate to be calculated (Section 3.3.3).

Thirdly, the effects of the deleterious allele may not become apparent until after peak reproductive age, so that selection against the allele is weak or non-existent. This is the case, for example, in Huntington’s disease, where familial transmission accounts for more than 95% of new cases and de novo mutations are rare.

Fourthly, the deleterious allele may be maintained in the population by **balancing selection**, a process that actively maintains more than one variant of an allele in the population. The most usual reason for balancing selection is **heterozygote advantage**, where individuals who are heterozygous at a particular locus are fitter than individuals who are homozygous. The best-known example of heterozygote advantage in human medicine is sickle cell trait in individuals of African ancestry, where heterozygotes for the sickle cell allele (an A→T SNP) at the beta-hemoglobin locus are more resistant to malaria than are homozygotes for the normal allele, while
homozygotes for the sickle cell allele are highly susceptible to sickle cell disease (Section 13.9.1). It is important to note that the deleterious allele is maintained in the population because of the selective pressure maintained by malaria infection; the allele frequency of the sickle cell allele in Americans of African descent is lower than in their ancestral populations in West Africa, perhaps because malaria has been eliminated from North America.

Other examples of Mendelian diseases for which heterozygote advantage has been postulated include cystic fibrosis (Box 3.5) and Tay–Sachs disease. Balancing selection may also contribute to diversity in major histocompatibility complex (MHC) genes.

**Box 3.5 Cystic Fibrosis: Exemplar of a Mendelian Disease**

The fundamental defect in cystic fibrosis is dysregulation of secretion of salt and water from epithelial cells because of the absence of a protein, the cystic fibrosis transmembrane conductance regulator (CFTR), that transports chloride ions across the cell membrane and also influences the regulation of other ion channels. Because of the consequent faulty ion transport, the mucus produced in the bronchial tree is thick and difficult to clear, leading to frequent lung infections, particularly with *Pseudomonas aeruginosa*. Secretion of digestive enzymes from the pancreas is also affected, causing inflammation of that organ as well as malabsorption of nutrients from the intestine. Other symptoms arise from secretory defects in the intestine, gall bladder, skin, and reproductive organs, and fertility is considerably reduced. Affected individuals have a much-reduced life expectancy, and in the past most did not survive beyond childhood, although recent advances in treatment such as lung transplantation, improved antibiotic regimens, and digestive-enzyme replacement therapy mean that many individuals with cystic fibrosis survive into their fourth or fifth decade.

The gene for CFTR is located on chromosome 7q31.2. Over 1000 mutations have been identified in this gene. The most frequent, occurring in 60–80% of cases, is a three-base-pair deletion causing the loss of a phenylalanine residue at position 508 of the protein (Δ-F508); the faulty Δ-F508 protein is destroyed by the cell’s quality-control mechanisms before it reaches the cell surface. Cystic fibrosis is a recessive disorder: homozygotes have a complete lack of functional protein whereas heterozygous carriers have one functional copy of the gene and appear phenotypically normal.

Cystic fibrosis is the most common lethal genetic disorder in European populations, with an average prevalence of 1 in 2500, implying an allele frequency of 0.02. However, the disease is much less common in other populations, with allele frequencies of 0.01 in Africa and 0.002 in East Asia. Within Europe, there is a clinal gradient of allele frequency, decreasing from north to south. The age of the Δ-F508 mutation in European populations has been estimated to be at least 600 generations (about 15,000 years).

The early lethality of homozygosity for cystic fibrosis raises the question of why the allele has not been removed from the population by negative selection. Proposals for a hypothetical heterozygote advantage for cystic fibrosis alleles must provide a plausible molecular basis as well as explain the high prevalence in northern Europe (e.g., Alfonso-Sánchez et al. 2010). Early speculation invoked protection against diarrheal disease. Bacteria that cause secretory diarrhoea, such as *Vibrio cholerae* and some strains of *Escherichia coli*, produce toxins that interact with CFTR on the intestinal epithelium. Furthermore, the bacterium that causes typhoid uses wild-type CFTR to enter gastrointestinal epithelial cells, and cells expressing the Δ-F508 variant contain many fewer bacteria. Nevertheless, the European preponderance of cystic fibrosis poses a problem for this model: diarrheal disease is common worldwide and shows an opposite geographical gradient to that of cystic fibrosis, being more prevalent in tropical regions. Moreover, it is unlikely that cholera was present in Europe before the early part of the nineteenth century, making it highly unlikely that cholera provided selective pressure favoring cystic fibrosis alleles in European populations before that time.

An alternative hypothesis suggests that tuberculosis may be the disease against which cystic fibrosis provides heterozygote advantage (Meindl 1987). The historical and geographical distribution of cystic fibrosis is more consistent with that of tuberculosis than of diarrheal diseases. Several studies have reported a reduced frequency of tuberculosis in cystic fibrosis patients and their parents (who must be heterozygotes). Indeed, reduced arylsulfatase activity in patients with cystic fibrosis may inhibit growth of *Mycobacterium tuberculosis* as it lacks arylsulfatase activity and cannot metabolize sulfate, and this provides a molecular explanation for the phenomenon. Modeling studies have demonstrated that the European tuberculosis epidemic that began in the seventeenth century and continued until the late nineteenth century would have provided sufficient selective pressure to account for the current prevalence of cystic fibrosis (Poolman & Galvani 2007).
(Section 10.7.2), where such diversity maximizes the range of epitopes to which the immune system can respond.

Finally, the changing environment of humans alters the background on which selection operates. We can envisage two opposing factors that will affect how selection operates on single-gene-associated disease. First, improvements in medical care mean that individuals with diseases that in the past severely limited fitness are now able to have children, reducing the selective pressure that tended to eliminate the causative alleles from the population. Conversely, improvements in public health to reduce the burden of infectious disease mean a reduction in the power of the forces that have driven balancing selection, decreasing the heterozygote advantage that has maintained potentially deleterious alleles in the population.

### 3.6 No Single Genes for Common Diseases

In contrast to the numerous but individually rare Mendelian disorders are the common human disorders such as cardiovascular disease and type 2 diabetes. These diseases are generally multifactorial, with susceptibility determined by the combined influence of many genes, all with a small effect, and with a strong influence of developmental and environmental factors determining penetrance and therefore whether and how an individual will be affected.

Although each allele affecting a complex trait will be inherited in a simple Mendelian manner, the various genes involved in a complex trait may not interact additively, meaning that the relationship between genetic variation and phenotypic variation will be non-linear. In other words, the effect of an allele on expression of a disease will be affected by the presence or absence of other alleles as well as by environmental factors, ensuring that disease susceptibility will be inherited in a non-Mendelian way. This interaction between genes is termed epistasis. The complexity induced by such interactions is the reason why the search for susceptibility loci for complex traits has been less successful than for simple Mendelian disorders, where the identification of the single locus for the disease is relatively straightforward.

If certain diseases are common, could the alleles causing them also be common: the “common disease/common variant” hypothesis? Or could they be caused by multiple, but individually rare, alleles at disease loci? This question has implications for the strategies of searching for such alleles (Box 3.6). In fact, both situations seem to occur. For example, some of the heritable variation in levels of plasma low-density lipoprotein cholesterol is caused by multiple rare alleles in the NPC1L1 gene, whose product mediates intestinal cholesterol transport and is the molecular target of the cholesterol-lowering drug ezetimibe (Cohen et al. 2006). Conversely, a common variant in the FTO gene, for which about 20% of people of European origin are homozygous (allele frequency 0.43 by Hardy–Weinberg equilibrium), increases the risk of obesity and therefore of type 2 diabetes (Loos and Yeo 2014).

Another common SNP associated with obesity and insulin resistance has recently been found in the control region of the melanocortin-4 receptor (MC4R) gene, which is involved in regulating appetite and energy balance (Loos et al. 2008). The effect of the MC4R risk allele, the frequency of which is about 0.25 in northern European populations, is additive to that of the FTO risk allele. The MC4R polymorphism, which presumably modulates the expression of the receptor protein in some way, is of particular note because loss-of-function mutations in the coding region for the receptor protein have long been known to be associated with a rare and dominantly inherited syndrome characterized by early onset hyperphagia and morbid obesity, thus revealing overlap in the genetic causes of monogenic and multifactorial metabolic dysfunction. Whether this will be a recurring pattern remains to be elucidated.

But even for FTO, currently the best-characterized gene increasing the risk of obesity and type 2 diabetes, the effects on disease risk are modest and population-specific. For example, in Europeans, homozygosity for the FTO risk allele increases adult body mass by only about 2.5 kg and increases the risk of obesity by 1.5-fold, and FTO status explains only 0.3% of interindividual variation in body mass index (BMI). Yet such allele frequencies and increased risk are not seen in all human populations, with people of recent African origin having
Box 3.6 Searching for the Genetic Causes of Complex Disease: Linkage and Association

There are two broad approaches to identifying genes associated with inherited disease. The first is linkage mapping, which attempts to find correlations between patterns of disease occurrence in families and patterns of transmission of genetic markers, such as known genes with obvious phenotypic effects, or particular DNA sequences. In this approach, the extent of linkage disequilibrium between known marker alleles and an unknown disease-causing allele provides information about the location of the disease allele. If the marker alleles are inherited in the same pattern as disease susceptibility then the disease-causing allele must be physically close to the markers since they are not being separated by recombination. Knowledge of the genes present on the chromosome region identified by the marker furnishes a list of candidate genes whose contribution to disease causation can be examined by molecular biological methods and by assessment of biological plausibility.

Linkage mapping is most effective for highly penetrant monogenic disorders, since there are many sampling and statistical issues in analyzing patterns of multilocus genetic variation. Consequently, many early linkage studies that claimed to identify genes involved in complex multifactorial disorders have not been replicated.

The second approach is association studies, which analyze unrelated people with and without a disease (cases and controls, respectively) for the presence of particular alleles. If an allele is found more often in cases than in controls, then it—or a closely linked allele—is assumed to play a role in the disease. The availability in the last few years of large data sets of human genetic variation and of array technology allowing simultaneous analysis of hundreds of thousands of SNPs, together with assembly of large cohorts (thousands or tens of thousands of individuals) of cases and controls, has paved the way for genome-wide association studies (GWAS) that have identified genetic variation contributing to several multifactorial diseases. Such studies can take the form of broad scans looking for associations with common diseases (e.g., Wellcome Trust Case Control Consortium (2007), which identified disease loci for coronary heart disease, type 1 and type 2 diabetes, rheumatoid arthritis, Crohn’s disease, bipolar disorder, and hypertension), or more targeted studies of specific diseases or even phenotypic traits (e.g., Baron-Cohen et al. 2014).

GWAS have so far identified over 2000 genetic loci associated with over 300 diseases or traits. Nevertheless, it is important to remember that association studies do not directly identify causal alleles. The haplotype blocks identified by the arrays used in GWAS are characterized by a small number of representative (so-called “tag”) SNPs, so the association identified is with a narrow genomic region rather than a particular variant within that region; more detailed sequencing studies are required to demonstrate the specific risk allele.

A further challenge for GWAS in determining causative genetic variation for complex diseases or traits is so-called missing heritability, best illustrated by studies of tallness. Adult height in humans is highly heritable—if your parents are tall then it is likely that you will be too. About 60–80% of variation in height is heritable, and the remainder is environmentally determined by, for example, nutrition. GWAS have identified about 50 loci affecting height, but their individual effect size (how much difference in height can actually be attributed to the locus) is small, and taken together these loci explain only about 5% of interindividual variation in height, leaving 95% unexplained.

Such missing heritability, found in most GWAS, has been referred to, using an astrophysical analogy, as the “dark matter of the genome.” Part of the explanation appears to be that the stringent statistical tests used in GWAS to eliminate chance associations result in the discovery of a few rare loci, each with relatively large (but still small in absolute terms) effects (Golan et al. 2014). If the analysis is extended to include the contribution of multiple common loci, each with small (but cumulatively large in absolute terms) effects, then the missing heritability reappears (e.g., Yang et al. 2010). Additionally, it may be necessary to model the effects of epistatic interactions between the loci discovered (Wei et al. 2014a). The implication is that the genetic architecture underlying many traits and diseases is even more complex than previously envisaged, and that identification of numerous contributing variants of small effect size will require studies that, for statistical reasons, approach practical limitations on sample size.

A further difficulty is that the effect size (penetrance) of any locus may be affected by individual phenotypic robustness, further confounding the analysis of association. Robustness is an organism-level trait that reflects lower responsiveness to environmental or genetic perturbations, and possibly also lower susceptibility to disease. Including assessment of individual robustness into the analysis of association studies may expose some of the missing heritability (Queitsch et al. 2012).

A related complication with association studies is the existence of interactions between genetic variation and environmental factors in determining epigenetic variation (Teh et al. 2014). Given that such epigenetic variation might itself have phenotypic consequences (Section 3.7), simply exploring genetic variation in isolation from considering gene–environment interactions may reduce the power of the analysis.
low frequencies of the FTO risk allele (Loos and Yeo 2014).

Importantly, the penetrance of risk alleles may be modified by environmental factors. For example, physical activity is partially protective against the increased susceptibility to obesity conferred by the FTO risk allele, an effect that appears to be mediated by the effect of FTO mutants on satiety and food intake. The effects of several genes that increase the risk of type 2 diabetes are modified by birthweight (Freathy et al. 2007), which is a proxy for early-life nutrition. In the musculoskeletal system there are interactions between birthweight, vitamin D receptor, or calcium-sensing receptor genotype, and risk factors for osteoporosis in older adults (Figure 3.8).

Again, the question arises: why, even if these alleles are only slightly deleterious, have they not been eliminated by selection during human evolutionary history? The reasons are similar to those discussed for monogenic disorders, but with some nuances. First, slightly deleterious alleles, particularly if they are common in the population, may not affect fitness to a sufficiently great extent that there is strong selection pressure against them, especially if they are only deleterious in combination with other alleles. And, of course, effects which appear after reproductive age, as is generally the case with most chronic diseases, will affect fitness even less. Second, alleles that were once advantageous or neutral in the environments in which humans evolved may now be deleterious in the very different environments in which most of us now live. Thus, it is the interaction with evolutionarily novel circumstances, such as increased lifespan or an energy-rich diet, which causes genetic variation expressed in the phenotype to manifest as ill-health. Applied specifically to energy metabolism, this is the basis of the thrifty genotype hypothesis discussed further in Section 9.4.1.

There is conflicting evidence for and against selection in some of the susceptibility genes for common diseases. For example, the allele of the PPARG gene that increases susceptibility to type 2 diabetes is the ancestral allele—the one found in our nearest relative, the chimpanzee—whereas the protective allele is human-specific (Di Rienzo and Hudson 2005). Similarly, the risk allele of the apolipoprotein E gene (APOE) for cardiovascular

![Figure 3.8](image-url)  
**Figure 3.8** Birthweight modifies the relationship between lumbar spine bone mineral density and vitamin D receptor genotype in elderly men and women. Among individuals in the lowest third of birthweight, bone mineral density was significantly higher among individuals of genotype BB. In contrast, bone mineral density was reduced in individuals of the same genotype who were in the highest third of birthweight. From Dennison et al. (2001), with permission.
disease and Alzheimer’s disease may be the ancestral allele, since it may have a “thrifty” role in lipid metabolism and also protect against childhood diarrhea. Different human populations apparently show climate-related adaptation in genes associated with some of the common metabolic disorders, suggesting that metabolic function may be subject to the same clinal variation as skin color and body shape (see Section 6.7), and that such variation may underpin population differences in disease susceptibility (Hancock et al. 2008). However, attempts to detect selection at a number of specific risk alleles associated with type 2 diabetes in several populations have been unsuccessful, leading to the conclusion that ancestral selection for thrifty alleles does not account for current susceptibility to that disease (Ayub et al. 2014).

Indeed, approaches targeting one or a few genes may be too simplistic. What would we expect to see if selection was acting on a polygenic trait such as susceptibility to type 2 diabetes? As described earlier and in Box 3.6, it is likely that such quantitative traits are controlled by multiple (possibly hundreds) loci of small effect. Such loci have yet to be identified, and assessment of allele frequency changes at each of them will be a challenging task.

### 3.7 Epigenetic Mechanisms as a Cause of Variation

Phenotype is not determined by genotype alone. There are several hundred distinct cell types in the human body, and the differentiation and proliferation of each cell type requires fine-tuned control of gene expression despite all these cells having the same genotype. The epigenetic regulation of gene expression refers to a set of interrelated molecular mechanisms that establish and maintain patterns of gene expression without altering the DNA sequence. These patterns, once established, can be copied through mitotic cell division and therefore affect all daughter cells in a lineage. There is also some evidence that they can persist unchanged through meiosis, creating a non-genomic form of transient inheritance (Section 3.8).

Four primary mechanisms have been implicated in epigenetic regulation: DNA methylation, histone modifications, the activity of non-coding RNAs (ncRNAs), and three-dimensional organization of chromatin and association with the nuclear membrane (Figure 3.9). These mechanisms are interdependent: so, for example, ncRNAs may direct histone modifications, which themselves can induce DNA methylation. The effect of these epigenetic changes is to alter the packing of the chromatin and make it more or less accessible to transcriptional activation.

DNA methylation (Jones 2012) involves the covalent addition of a methyl group moiety to the cytosine nucleotide to give 5-methylcytosine (5mC). 5mC is widely considered to be the fifth base of the genome. In mammals, DNA methylation generally occurs on cytosine bases at which guanine is located at the adjacent 3' position, so-called cytosine-guanine (CpG) dinucleotides. Many of these CpGs are clustered in CpG islands at the transcription start sites of genes. While most CpGs in the genome are methylated, CpGs in CpG islands generally are not; their methylation tends to be linked to long-term repression such as that required for cell-type differentiation, imprinted genes, and for inactivation of the X chromosome.

CpG dinucleotides are depleted throughout the rest of the genome because 5mC tends to undergo spontaneous or enzymatic deamination to thymine, a functional nucleobase, which hinders detection by corrective DNA repair mechanisms. The occurrence of this mutation in somatic cells can promote oncogenesis, and in the germline, pathology.

The palindromic nature of CpG sites within double-stranded DNA enables hemimethylated DNA (i.e., DNA where CpGs are methylated on only one of the complementary strands) to be targeted by methyltransferases that methylate the nascent daughter strand. In this way, extant methylation patterns are maintained during cell division. The effect of DNA methylation on transcriptional activity varies according to context. In general, methylation of DNA promoters blocks the binding of transcription factors and recruits methyl-binding proteins that attract other repressor complexes, leading to downregulated gene expression. In contrast, intragenic methylation is associated with active transcription, although this may vary depending on the region of the gene body at which methylation occurs.
one-quarter of DNA methylation in human embryonic stem cells, with the proportion rising to nearly two-thirds in mouse germinal vesicle oocytes. In the male murine germ cell, non-CpG methylation first accumulates but then declines upon resumption of mitosis in the neonatal period. However, little is currently known about its biological significance (Pinney 2014).

5mC is amenable to further oxidation by the ten–eleven translocation oxygenase family of enzymes into 5-hydroxymethylcytosine (5hmC), which has been detected in mouse embryonic stem cells and in the human brain, where its level is correlated with development of the cerebellum. The functional significance of 5hmC remains to be fully elucidated, although it appears to serve as an intermediate in DNA demethylation—itself a poorly delineated process at present—but may also represent a distinct epigenetic mark. Further iterative oxidation products of 5hmC have also been recently identified in mouse embryonic stem cells.

Post-translational modifications to the amino acid residues on the amino-terminal tails of histones, the proteins around which DNA is wrapped to form nucleosomes, constitute another process by which gene expression is regulated (Bannister and Kouzarides 2011). These reversible, covalent modifications include acetylation, methylation, ubiquitination, and phosphorylation. Some modifications, including acetylation and phosphorylation, attenuate the positive charge of histones and disrupt their interaction with DNA, thus increasing access to transcriptional machinery, although in some cases such as arginine methylation the effect may be residue-specific. Other modifications, such as lysine methylation, mediate the binding of a panoply of chromatin-associated effector molecules that influence transcriptional activity.

There is a high degree of cross-talk between the different modifications; the net effect is to induce reconfiguration of chromatin structure—unwinding results in euchromatin that is permissive to gene expression, while condensation forms heterochromatin that is transcriptionally silent. The preservation of histone epigenetic marks during DNA replication is not well understood, but appears to involve a highly coordinated process of chromatin disassembly and reassembly facilitated by histone
chaperone proteins and chromatin remodeling complexes.

Chromatin itself has a complex three-dimensional structure within the nucleus, creating loops which bring linearly separated chromatin segments together, and also bring DNA sequences on different chromosomes into close association. There appears to be functional significance to this higher-order structure. Of particular interest is the association of long segments of chromatin with the nuclear membrane, which modulates gene expression (Collas et al. 2014).

There is increasing interest in the activity of ncRNA as another level of epigenetic regulation (Morris and Mattick 2014). Numerous ncRNAs are known to be transcribed in tissue- and cell-specific patterns to dynamically regulate cell differentiation and development. In line with this, the ENCODE project (Box 3.1) to map the functional elements of the human genome has revealed that intergenic regions are highly populated with ncRNA transcripts. They are categorized into subclasses including microRNAs, long ncRNAs and piwi-interacting RNAs, depending on characteristics such as length and function. MicroRNAs, typically about 22 nucleotides long, decrease gene expression by binding to cognate mRNAs, preventing their translation or prompting their degradation. On the other hand the long ncRNAs, generally more than 200 base pairs long, regulate transcription, translation, and chromatin remodeling via cis- and trans-acting mechanisms, depending on the proximity of the target genes to their transcription site.

It has been suggested that ncRNA-mediated activity has been afforded a further level of control by enzymatically directed RNA editing. This process most commonly involves adenosine-to-inosine conversions, and may have been particularly important during the evolution of cognitive function. Finally, the recent discovery of mRNA methylation adds yet another layer to the complexity of the regulation of gene expression (Yue et al. 2015).

### 3.8 Non-genetic Inheritance

Is variation in DNA sequence the only way in which information can be transmitted across generations, and if so can features of a phenotype induced in one generation be transmitted to future generations? Such suggestions have long been regarded with suspicion by biologists, and are referred to pejoratively as “Lamarckian” (Jablonka and Lamb 2005; see Chapter 14). Yet there is now strong evidence in model organisms, and more limited data in humans, that environmental influences can affect phenotypes across several generations (Jablonka and Raz 2009). The risk of cardiovascular disease and mortality from diabetes in adults has been associated, in a sex-specific manner, with the level of food supply experienced by a person’s paternal grandparents before puberty. Higher adiposity has been observed in neonate grandchildren of women exposed to famine in the western Netherlands at the end of the Second World War. Greater paternal and grandpaternal age at conception has been shown to make cumulative contributions to telomere length (Eisenberg et al. 2012).

What mechanisms could explain such phenomena? There are several non-mutually exclusive pathways by which induced phenotypes can be inherited across generations (Table 3.1). While the discussion in this section is focused on biological mechanisms, it is also important to consider the

<table>
<thead>
<tr>
<th>Type</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical Mendelian genetic inheritance</td>
<td>Allele frequency changes</td>
</tr>
<tr>
<td>Maternal effects (F_0 \rightarrow F_1)</td>
<td>Maternal phenotype affects offspring phenotype independent of genotype</td>
</tr>
<tr>
<td>Grandmaternal effects (F_0 \rightarrow F_1 \rightarrow F_2)</td>
<td>Effect on developing oocytes of the (F_1) generation</td>
</tr>
<tr>
<td>Adaptive maternal effects</td>
<td>Modulation of developmental plasticity in offspring by cues in the parental environment</td>
</tr>
<tr>
<td>Niche re-creation</td>
<td>The developmental environment that induces the phenotype is re-created in each generation</td>
</tr>
<tr>
<td>Germline-mediated epigenetic inheritance</td>
<td>Persistent epigenetic mark in the germline transmitting through meiosis and zygosis</td>
</tr>
<tr>
<td>Ecological inheritance</td>
<td>Effects of the constructed environment on evolutionary dynamics</td>
</tr>
<tr>
<td>Cultural inheritance</td>
<td>Oral, written, and electronic transmission of knowledge and social attitudes</td>
</tr>
</tbody>
</table>
possible influence of common environmental factors across generations, including both ecological and cultural factors.

### 3.8.1 Parental and Grandparental Effects

Direct parental effects are common in biology, and encompass a broad class of situations where the parental (most commonly maternal) phenotype influences the phenotype of the offspring independent of genetic transmission. In humans, such effects may be a reflection of maternal physiology: for instance, smaller mothers tend to have smaller babies because of the process of maternal constraint (see Sections 8.9.5 and 9.4.3.2), which probably acts through limitation of nutrient supply to match fetal size to maternal size rather than to genetic growth potential. The development of offspring can also be affected by the composition of breast milk and by transmission of the maternal gut microbiota (Section 10.3).

Parental effects can also be associated with epigenetic changes in the offspring. Indeed, an increasing number of factors in the maternal environment have been shown to affect the epigenetic state of the offspring, with likely phenotypic consequences (e.g., Godfrey et al. 2011; Teh et al. 2014). Some of these effects may be adaptive for the offspring (adaptive developmental plasticity is discussed in Chapters 4 and 9).

Maternal effects can transmit over more than one generation because the egg that will give rise to the F2 generation is formed by primordial oogenesis while the mother (F1) is a first-trimester embryo. This can lead to grandmaternal effects, because the developing oocyte is exposed to the maternal grandmother’s environment. For example, there is strong evidence from animal studies that phenotypic characteristics such as blood pressure, heart dimensions, metabolism, and hypothalamic–pituitary–adrenal axis settings in offspring can be altered by endocrine or nutritional manipulations of their grandmothers, without further environmental challenges to the mothers in the intermediate generation (see, e.g., Vickers 2014). Thus, for example, a nutritional challenge to a woman during pregnancy not only affects the developmental plasticity of her fetus, but also potentially produces epigenetic effects in her fetus’s ova that will contribute directly to the epigenome of her grandchildren. This is effectively two generations of apparent epigenetic inheritance, but only through the female in the intermediate generation.

In contrast, production of sperm occurs over the few months before ejaculation, so much more recent aspects of a father’s lifestyle and environment can influence his offspring. Paternal condition at conception appears to have effects on the health and epigenome of the offspring (Lane et al. 2014). Male rats fed a high-fat diet father female offspring that are impaired in insulin secretion and glucose tolerance, and that show epigenetic and gene expression changes in their islets. More complex crossing experiments in mice demonstrate that traits such as resistance to obesity and appetite regulation are passed through the paternal lineage for at least two generations. Obesity or diabetes induced in the father can affect body composition and metabolism of the offspring even when the mother was healthy.

### 3.8.2 Niche Re-creation

Beyond these direct parental effects, epigenetic marks may be re-created in each generation without transmeiotic transmission. Consider that smaller female babies will themselves have a higher risk of growing into smaller adults, as prenatal constraint can lead to a greater risk of stunting. In turn, as shorter adults, their offspring are also at greater risk of being growth retarded. If that growth retardation is associated with epigenetic changes, then similar epigenetic changes would occur across generations even though the epigenetic mark itself did not pass generations—rather, the developmental environment (“niche”) inducing the phenotype and epigenotype of the offspring is re-created in each generation.

A striking example of niche re-creation in animals involving both behavior and epigenetic change is maternal care in rats (Zhang et al. 2013). Rat mothers differ in the amount of attention they pay to their pups, with some mothers licking and grooming their pups often and other mothers doing so only rarely. Pups who receive high levels of attention from their mothers show reduced stress responses and greater confidence (measured by their willingness to explore new situations) in later life. Moreover, because female pups that
receive high levels of maternal attention themselves become high-attention mothers, the trait is transmitted between generations. It is also acquired in early life—pups from low-attention mothers who are fostered into the litters of high-attention mothers take on the behavior of their littermates.

At the molecular level, higher attention correlates with lower DNA methylation at the glucocorticoid receptor promoter in the brain. Downregulated methylation is seen in offspring cross-fostered from low-care mothers to high-care mothers, showing that intergenerational transmission is caused by maternal–infant behavioral factors rather than gametic transfer of the DNA methylation profile, which is created afresh in each generation (Weaver et al. 2004). Experiments in which the methylation change has been reversed pharmacologically also reversed the behavioral patterns. And finally, the observation that corresponding changes in glucocorticoid receptor expression can be detected in humans with a history of childhood maltreatment has clear implications for treating depressive disorders (see Box 11.5).

### 3.8.3 Epigenetic Inheritance

Is there evidence for true transgenerational epigenetic inheritance in mammals, in other words, for transmeiotic inheritance of epigenetic marks in the germline (so-called epimutations)? A major impediment to the acceptance of germline-mediated epigenetic transmission as a biologically plausible mode of inheritance has been how environmentally induced epigenetic marks could persist to be passed on to offspring. DNA methylation marks are clearly maintained through mitosis, but there has been much debate over whether this extends to meiosis. It was once thought that all epigenetic marks are erased during gametogenesis and embryogenesis (the relatively low number of imprinted genes excepted). However, it is now known that epigenetic marks in sperm can persist through the compaction of DNA by protamines, which replace histones during spermatogenesis as the major chromatin-associated protein (Jenkins and Carrell 2012), and recent evidence indicates that some methylation marks can in fact survive erasure in primordial germ cells in humans (Tang et al. 2015).

Demonstration of transgenerational epigenetic inheritance requires, at a minimum, transmission of the induced phenotype in a further generation (F₃) that has been wholly unexposed to the cue. Experimental studies focusing on the paternal line, and on the contribution of sperm, have the advantage of ruling out other indirect pathways of inheritance such as grandmaternal effects and niche re-creation. In recent years, evidence has accumulated for the association of male-line phenotypic inheritance with epigenetic changes in the sperm of both fathers and their offspring. For example, extensive studies using a rat model of maternal exposure to fungicide have shown that F₃ and F₄ offspring exhibit a variety of adult-onset diseases and alterations to gene expression in the brain. Notably, epigenomic changes in germline-derived somatic cells and tissues have been linked to epimutations in F₃ sperm (Skinner 2014).

The mechanisms of such inheritance remain unclear, although both epimutations and subsequent genomic mutations have been implicated (Skinner et al. 2015). Moreover, the small amount of cytoplasm in the sperm head contains microRNAs that have putative roles in early embryonic development. In humans, smoking alters these sperm microRNAs, which might subsequently influence the development of the fetus (Marczylo et al. 2012). MicroRNAs have also been implicated in two mouse models of transgenerational inheritance: one has effects on tail pigmentation and the other on cardiac growth (Rassoulzadegan and Cuzin 2015). These microRNAs may be the mode of transmission, allowing methylation and histone changes to be re-established in the next generation.

Thus, while the precise mechanisms of molecular epigenetic inheritance remain poorly understood, there is proof of concept that phenotypic effects of ancestral exposure to environmental cues can be transgenerationally transmitted via epigenetic modifications of the germ cell. This could offer molecular insights into phenotype-driven evolutionary processes (Section 3.8.5).

### 3.8.4 Ecological and Cultural Inheritance

A further mode of non-genetic inheritance relates to transmission of the “extended phenotype” of
the organism beyond its biological boundaries. Organisms not only adapt to their environment, but also modify the environment around them, a process known as *niche construction* (see Section 2.3.2.6). In this context, *ecological inheritance* refers to the persistence of such modifications in a way that can affect the evolutionary outcomes of future generations.

The other key system of intergenerational influence is *cultural inheritance*, when people learn ways of thinking and acting from other people (Section 2.3.3). Cultural inheritance does not simply involve children learning from their parents; individuals may learn from other members of their parents’ generation or from their peers. The behaviors acquired may have impacts on health that are positive (e.g., attitudes to education or career opportunities) or negative (e.g., copying of damaging patterns of eating, smoking, or drug abuse) (Box 3.7). It is therefore easy to see how, for example, obesity may aggregate in families because of transmission of attitudes to exercise and eating patterns between generations.

### Box 3.7 Myopia

Myopia, or nearsightedness, is a visual condition in which an individual loses the ability to view distant objects clearly without wearing corrective lenses. In myopia, the eyeball is too long and the ciliary muscles in the eye, which are responsible for focusing, are unable to change the shape of the eye sufficiently to allow light rays to fall onto the retina. Initially, myopia was thought to be a genetically determined condition, but it now appears that juvenile-onset myopia is caused by an interaction of genetic, environmental, and cultural conditions.

Although children with at least one myopic parent are more susceptible to myopia than those with non-myopic parents, the mechanism of the heritability of the common form of myopia remains unclear. A single gene mutation has been found in one particularly severe form of myopia—it involves a loss-of-function mutation in the gene for prolyl-3-hydroxylase, which is involved in the biosynthesis of the connective tissue protein collagen. The mutation is inherited in an autosomal recessive manner (Mordechai et al. 2011). However, most myopia does not have a simple single-gene explanation.

The prevalence of myopia varies ethnically and geographically, attesting to strong contributions of environment and culture. There is a higher prevalence of myopia among Caucasians compared with African-Americans, as well as a greater number of myopic individuals among schoolchildren than in non-schooled urban and rural children. In some populations, such as the Inuit and the rural Taiwanese, myopia has become almost universal following the social and environmental changes of the twentieth century (Lin et al. 2004). There is an especially high prevalence and incidence rate of myopia in Singaporean and Chinese children, with over 40% of 7-year-old Singaporean children attending school being reported as myopic.

Although the time that children spend on close-up work and reading does not correlate strongly with onset of myopia, time spent outdoors does correlate. Although it is still unclear what underlies this association—it might be exposure to strong light, or the chance for the eyes to focus on a distant horizon—opportunities for outdoor activity tend to be determined by societal or familial factors, providing evidence for a role of cultural inheritance in determining the prevalence of myopia (Dolgin 2015).

### 3.8.5 Non-genomic Inheritance and Evolution

An outstanding question now beginning to receive attention is the role of these additional inheritance systems in evolution (Danchin et al. 2011). Natural selection operates on inheritable variation in the phenotype, and if epigenetically driven variation in the phenotype can be inherited then we would expect such variants to be substrates for selection. It has already been indicated that aspects of the life experienced by the adult can be passed through the germline to the next generation. This appears to break Weismann’s barrier (see Section 4.2 and Figure 14.1) and was formerly a heretical concept in evolutionary biology. However, it is now clear that that barrier is not absolute and, as already discussed, germ cells can be influenced by the environment, with long-term consequences.

Further challenges to classic dogma that all mutations are random come from observations that epigenetic processes can affect the likelihood of mutation at some genomic sites, either protecting
the site from the risk of mutation or increasing the chance of a mutation that may have phenotypic effects (e.g., Zemojtel et al. 2011; Hernando-Herraez et al. 2015). Such processes could theoretically permit environmental factors to induce a mutation which might be adaptive, allowing evolutionary change to operate much more quickly than purely through random occurrence of helpful mutations, but also allowing it to be directed—contradicting classical evolutionary theory.

Disentangling the relative contributions of genetic and non-genetic inheritance, and considerations of whether such variation can become “fixed” or assimilated in the genome (Chapter 4), are major research priorities.

### 3.9 Conclusion

The phenotypic variation among humans is to a significant extent underpinned by differences in genotype. Variation is introduced into the human genome at a rate of 50–100 new mutations per generation, and such variation provides the substrate for selection, although only a small proportion of mutations result in changes in gene expression.

Although interindividual variation in humans is sufficiently large that there is no “normal” individual, genetic disease arises from variation that is in some way disadvantageous in a particular environment. Diseases with simple genetic determinants are relatively rare. Many monogenic disorders result from \textit{de novo} mutations, but evolutionary considerations such as balancing selection and post-reproductive onset of pathology can explain the apparent puzzle of the persistence of monogenic disorders in the population. In contrast, our most common diseases today, such as type 2 diabetes and cardiovascular disease, have complex genetic, environmental, and (as we will show in Chapter 4) developmental determinants.

The genetic component of increased susceptibility to such disorders—as for most complex phenotypic traits—is likely to involve the combined effects of numerous and relatively common disease-susceptibility alleles, each having a small individual effect. This complexity has broader implications. For instance, it shows why eugenics will not work (Section 14.4): selective breeding is not likely to reduce the prevalence of these diseases. Moreover, it suggests that there will not be many “magic bullet” gene therapies for most diseases. Nevertheless, identification of disease-associated genes for such complex disorders provides clues to the pathophysiology involved, and indicates potential targets for therapeutic intervention.

Studies of human genetic variation have implications beyond medicine. Analysis of patterns of variation in the human genome allows the identification of genes that have undergone selection during human evolution. These studies can provide clues about what make us human—that is, the changes that have occurred since the time of our most recent common ancestor with other primates—and about the selection pressures that have shaped the way humans have adapted to the multiple and changing environments in which we now live.

---

**Key Points**

- The human genome varies by about 0.5% between individuals. This variation arises from SNPs and from insertions, deletions, inversions, and duplications of longer sequences.
- Patterns of sequence variation between populations provide information about human origins and the selective pressures that humans have experienced.
- Only a small proportion of sequence variation causes detectable phenotypic change.
- Although monogenic diseases such as cystic fibrosis are individually rare, several mechanisms account for the fact that selection has not altogether eliminated the deleterious alleles from the population: recurrent mutation, delayed onset after reproductive age, and heterozygote advantage.
| Genetic susceptibility to common diseases such as type 2 diabetes probably arises from the combined effects of numerous alleles contributing to risk, each having small effects, coupled with environmental and developmental factors. |
| Not all inheritance is genetic. Disease risk can be transmitted by cultural or behavioral factors, and there is increasing evidence for transgenerational persistence of epigenetic modifications of DNA and gene expression. |
4.1 Introduction

In 1909, soon after the rediscovery of Mendel’s observations of the particulate nature of inheritance, the Swedish plant physiologist Wilhelm Johanssen coined the terms *gene* to describe the unit of particulate inheritance, *phenotype* to denote an observable characteristic or set of characteristics of a population, and *genotype* as the unobservable biological basis of the phenotype. We now understand genotype to comprise the genetic makeup of an individual cell or organism, and phenotype as its set of observable traits at any level from the whole organism to the molecule. Similarly to Johanssen, we recognize that one genotype can be associated with a range of phenotypes and that multiple genes may contribute directly and indirectly to any trait.

At a cellular level, all cells in the body have the same genotype but each cell type—neuron, hepatocyte, pancreatic beta cell, or neutrophil—has a very different phenotype. Much research in developmental biology has focused on the regulation of cell differentiation by factors in the cellular microenvironment, including the role of other cells, the extracellular matrix, and growth factors, which induce changes in patterns of gene expression that are replicated in subsequent generations of differentiated cells to produce the mature phenotype.

Developmental processes are equally important in understanding the mechanisms by which an organism’s integrated phenotype can vary despite a constant genotype. Identical twins are never absolutely identical (Spector 2012; see Box 4.4). Moreover, as they develop through childhood and adolescence they change further, so that they become increasingly distinguishable from one another, making it clear that the developmental environment plays a large role in determining phenotype.

It has been recognized for over a century that many invertebrates and some fish and amphibians can develop distinct stereotypic phenotypes, or *morphs*, as a result of developmental influences. This phenomenon of distinct morphs emerging from the same genotype is known as *polyphenism*. Here the developing organism responds to an environmental stimulus early in its development and becomes committed to a pathway of development towards one of two, or several, possible morphs. For example, a female bee larva can develop into a queen bee or a worker bee. It develops into a queen if, at the earliest larval stage, it is fed exclusively on royal jelly, and into a worker if it is fed otherwise. Recent studies show that this switch is mediated by *epigenetic* processes affecting the expression of genes critical to determining the phenotype (Foret et al. 2012). Given the role of the worker bees in determining how the larva is fed, it indicates the complex ways in which the environment can influence early development.

More commonly, and inevitably in mammals, developmental influences do not lead to stereotypically distinct morphs; rather the outcomes are a continuous range of phenotypes which can be defined by the magnitude of particular traits. The range of phenotypes induced in response to variation in a component of the developmental environment is called a *reaction norm* or *norm of reaction* (Schlichting and Pigliucci 1998; Box 4.1). The concept of the reaction norm is an important part of the theoretical basis of evolutionary developmental biology, often concatenated to “evo-devo.”

We now recognize that the process of developmental plasticity, the capacity to adjust the phenotype arising from a single genotype by changing the pathway of development in early life, is a key evolutionary
The developmentally induced range of phenotypes that arise from a single genotype can take many forms, including changed physiology, changed behavior, changed anatomy, or changed life-history traits. The specific outcome will depend on the nature of the developmental environment, and the specifics of the inherited genotype and perhaps the epigenotype which create a number of developmental and genetic constraints.

The term norm of reaction or reaction norm is used to describe the range of possible phenotypic outcomes induced by developmental plasticity. A non-plastic reaction norm is one where the phenotypic trait is fixed across the range of developmental environments. Developmental plasticity itself (which is effectively described by the reaction norm) can be the subject of selection: the possession of this developmentally contingent process of plasticity provides adaptive advantage in variable or potentially variable environments. But clearly what has adaptive advantage in one environment may be disadvantageous in another: this is the concept of mismatch described in several chapters of this book.

Because of the important role of quantitative approaches in the development of evolutionary thought, how one might express a reaction norm graphically has become a topic in its own right! The different approaches are beyond the needs of the reader and we will describe the most common form, where the magnitude of the mature trait is shown as a function of the range of environmental conditions during development.

In Chapter 9 we discuss the influence of the fetal environment on later insulin sensitivity. High levels of maternal nutrition are likely to produce offspring who develop obesity and insulin resistance. Maternal undernutrition, acting through different mechanisms, also makes the offspring more likely to develop insulin resistance and visceral obesity. Simple reaction norms might be drawn to demonstrate these relationships (Figure 4.1).

However, under some situations developmental plasticity does not produce a continuous range of phenotypes, but rather induces a discontinuous set of alternative stereotypic forms: these are termed polyphenisms. Environmentally induced polymorphisms are common in insects but are also seen in some vertebrates, for example environmentally induced sex determination in reptiles where the temperature of the sand in which the eggs are laid influences the sex of the offspring which hatch.

In mammals we are generally concerned with developmentally continuous variations in phenotype, but polyphenisms can exist. One example is seen in the Himalayan rabbit, where coloration of the hair in the ears and extremities is dependent on different temperature thresholds during rearing. Similar effects are seen in some Siamese cats, mice, and guinea pigs, and are related to mutations in the genes causing melanization, making them sensitive to temperature.

Stabilizing selection refers to a tendency to select against extreme phenotypes (where extreme is relative to the selective environment), meaning that for a given trait the bulk of variation is clustered around an average phenotype (the middle of the reaction norm) and therefore genetic diversity is reduced. The Russian biologist Ivan Ivanovich Schmalhausen (1884–1963) argued that a consistency in

![Figure 4.1](image-url) Simple reaction norms showing variation in the risk of outcomes in later life with variation in maternal nutrition during pregnancy. Note that because obesity and insulin resistance may arise through different mechanisms, the shapes of the reaction norms may be different.
4.2 Development: Pre-ordained or Plastic?

The question of when a life starts, and how development proceeds, has fascinated thinkers for thousands of years. The Greek philosopher Aristotle believed that a new human being arose from the mixture of male and female elements: the female contributed menstrual blood as the building material from which the male semen, as the active part, generated the new form. This ancient concept was termed *epigenesis*, where the “formed” develops from the “unformed,” and described development as both gradual and plastic. Despite its gross inaccuracies, this idea shares some similarities with modern thinking.

The Aristotelian view remained influential into the early modern era when, with the development of light microscopy, the question of life before birth began to be examined in detail for the first time, although still mostly in animal fetuses. In the seventeenth and eighteenth centuries debates raged about what was seen under these simple microscopes, and its significance. Aristotle had postulated that internal forces within the embryo would drive the emergence of both the form and the soul: a vegetative force in all organisms, a locomotor force in animals as well, and a rational force in humans are unlikely to be sustained over more than one or a few generations.

process (West-Eberhard 2003). This is because it allows organisms to try to maintain or enhance fitness by matching their traits to different environments and adjusting regulatory processes to changing circumstances on a timescale that is intermediate between those of selection and homeostasis (Section 1.3). These processes are essential components of the evolutionary repertoire, providing both a way for organisms to respond to challenges within their genotypic constraints and a mechanism by which evolutionary opportunity and novelty can arise.

This chapter focuses primarily on the key role that developmental plasticity plays in human biology, but to understand this we also need to consider some broader conceptual perspectives. The rapidly developing field of evo-devo provides a new way to understand how environmental influences during development can act within an evolutionary context, a problem which has confounded evolutionary thought since the time of Lamarck. We now recognize that characteristics acquired during development are not heritable, but this does not mean that the environmental influences experienced by one generation do not have an impact on subsequent generations. In fact, this is precisely what selection accomplishes—shaping of the genetic repertoire across generations as organisms adapt to their environment—although as we have seen in Section 3.8 genetic inheritance is not the only form of heritability. Nevertheless such modes of intergenerational influence have less fidelity and are unlikely to be sustained over more than one or a few generations.

In contrast to canalization is the phenomenon of *genetic assimilation*, the concept that environmental change can sometimes expose a phenotype (in other words, a section of the reaction norm) that has not previously been observed. To take the rabbit example, if all Himalayan rabbits lived and were reared in warm temperatures, the capacity to have black ears and extremities would never have been observed even though the genetically informed developmental potential existed. If the environmental change persists, and if there is adaptive value to this exposed component of the reaction norm, then selection will favor that component, thus changing the phenotypic pattern of the population. A further example of genetic assimilation is the famous experiment by Waddington on *Drosophila* subjected to heat shock during development which developed different vein patterns on their wings (see Box 4.3).

There is ongoing debate on whether genetic assimilation is simply selection on covert genetic variation under new environmental conditions or whether *de novo* incorporation of phenotypic variation induced by developmental plasticity can become incorporated into the genotype. For example, it has been suggested that environmentally induced epigenetic change can bias mutations due to the greater propensity of methylated CpGs to mutate (Bateson and Gluckman 2011).
alone. But how could such a complicated structure as a developing fetus, with all its limbs and organs, its beating heart, even its spontaneous movements, arise out of next to nothing? The epigenetic philosophy of development seemed implausible. A contrary position—the *preformationist* view—arose that all the complex structures were already there, it was just that they were very small. In 1694, the “animalculist” Nicolaas von Hartsoeker suggested that he could see, through the “skin” surrounding the sperm head, a tiny man within. This, he claimed, would then grow into the man himself. Others, the “ovists,” claimed to see such pre-formed human beings in the egg. The disagreement between epigeneticists and preformationists ran deep because it was a reflection of a then current conflict between materialism and orthodox Christianity. The epigeneticists seemingly won the argument, and the key question changed. Embryologists no longer aspired to explain the source of embryonic organization; rather their goal now was to establish its laws.

Some of the best-known images of development from the nineteenth century are those of the German scientist, and Europe’s most famous prophet of Darwinism, Ernst Haeckel (1834–1919). Haeckel argued that ontogeny (i.e., development) is the brief and rapid recapitulation of phylogeny (i.e., evolution). This erroneous but highly influential hypothesis implied that humans in their embryonic development repeated the most important steps through which their adult ancestors passed during the evolution of their species. And because all vertebrates have similar ancestry, the early stages of development must be similar, with increasing disparity occurring over time. In 1868, Haeckel first published drawings of embryos in which human and animal forms start the same and then diverge in their development. Haeckel’s opponents accused him of making drawings of embryos from different species more similar than they were, but Haeckel replied that they were meant to be illustrative and not exact. These images, despite their problematic origin, continued to be printed and to court controversy through much of the twentieth century (Hopwood 2015).

By the 1880s, the German cell biologist August Weismann (1834–1914) had made a conceptually critical contribution by recognizing that germ cells were largely protected from the environment and that inheritance operated through this germline (Section 14.4). Thus influences affecting the soma—the body of the organism—would not be passed to the next generation because the germline was considered immune from the environment. This was held to be an absolute statement and it is only in recent years that exceptions to this rule have been identified: these will be discussed later in this chapter.

Developmental biology (then known as embryology) next passed through a stage of brilliant experimental work in which biologists such as Wilhelm Roux, Hans Driesch, and Hans Spemann, although not always in agreement, identified the early processes of cell differentiation and embryonic organization. Studies of development progressed slowly until the advent of molecular biology, when work with *Drosophila*, the roundworm *Caenorhabditis elegans*, and other model species allowed an exploration of the role of specific genes in development (Wolpert et al. 2002). We now know much about the role of early gene expression in regulating polarity and patterning within the developing embryo. Recognition of the molecular basis of segmentation in the organization of organisms and the role of homeotic genes such as the *Hox* genes in developing segments were momentous discoveries (Section 4.8).

### 4.3 Is Development Important to Evolution?

Early embryologists generally studied model organisms in which development appeared to be largely immune to the environment; where there were such environmental effects, they saw them as contributing “noise” and as a source of unnecessary variation. Further, geneticists were focused on concepts of genetic determinism and, when genetic and evolutionary concepts were reconciled in the Modern Synthesis (Section 14.5), the view that development was largely irrelevant to evolution was reinforced. Because the Modern Synthesis was heavily dependent on population genetics and selection was seen to act on the mature, reproducing phenotype, the theorists of that period believed that development could largely be ignored (Hamburger 1980).
During the second half of the twentieth century political and other concerns conspired to reinforce this shift away from seeing development as a critical component of the evolutionary process. In the 1930s and 1940s there had been an enormous amount of pioneering work by Russian scientists, who had pondered the problem of variation; they argued that natural selection must work on the phenotype not the genotype and that as developmental processes were key in the linkage between the two, development itself merited greater attention (Box 4.2). The Russian theorist Ivan Schmalhausen (1884–1963) made substantial conceptual advances in this field, recognizing the importance of developmental influences in giving rise to a range of phenotypes. He focused on the ways in which natural selection acted on the phenotype, considering how it can change the range of phenotypes (or the reaction norm) within a population; that is, change the range of plastic potential. This idea was a forerunner to modern concepts that developmental plasticity itself must be an evolved process.

In Britain, Conrad Waddington (1905–75) had been thinking along similar lines. In 1940, he introduced the term epigenetics (“on top of genetics”) to denote mechanisms and processes by which, during development (epigenesis), genes bring about phenotypic effects (Waddington 1940b). He focused on many of the same issues as Schmalhausen, and in particular how development could produce what was an apparently stereotypic outcome much of the time and yet be influenced by environmental factors (Gilbert 1994). He developed the concept of the epigenetic landscape (Figure 4.2). This metaphor illustrates simply how the program for development, presumed to be inherent in the genotype, could be modified by environmental influences to take the developing organism down different paths: a concept he called canalization (Box 4.1). Inherent in the canalization metaphor remains the concept that aspects of development are protected from the environment, because even though the developing organism can pass down alternative channels the sides of these channels are steep so that it cannot diverge from the path. So developmental processes are kept robust, but could also be plastic when required.

Waddington then made an experimental observation which is regarded as classic because it suggested that environmental influences do not just impact upon development but may also affect the phenotype in subsequent generations. The explanations that flowed from his observation underpinned much of the growth in interest in the interface between evolutionary and developmental biology. When Waddington exposed larvae of the fruit fly *Drosophila* to heat he found that some of the adults that emerged showed differences in their wing-vein pattern. After repeatedly selecting for those flies that showed the heat-induced change in phenotype over several generations, Waddington found that he had a lineage of flies that showed the mutated phenotype spontaneously, i.e., without exposure to the disruptive heat stimulus. Waddington called
this process *genetic assimilation*, but because he conducted these experiments before the availability of modern molecular technologies he was not able to investigate the underlying mechanisms further. Our current understanding of this is discussed in Box 4.3.

### 4.4 Developmental Plasticity

Developmental plasticity refers to a set of processes that act in early development and influence the phenotype for the rest of the organism’s life (Bateson et al. 2004). These may act in some taxa to induce alternative morphs, but in humans we are concerned with a continual range of phenotypic outcomes. It is distinct from other forms of plasticity such as use-hypertrophy and disuse-atrophy in a mature organism (particularly obvious in the musculoskeletal system), compensatory hypertrophy (which can occur in some organs such as the heart), and processes of cellular hyperplasia (which can accompany repair in some tissues).

During development, environmental influences can induce a degree of plasticity in what are otherwise robust developmental processes. These plastic effects can occur at several levels. The first is early in embryogenesis, when environmental influences can affect the allocation and development of embryonic cells. The nutritional, endocrine, and chemical environment within the fallopian tube and the uterus can affect the allocation of early blastocyst cells to the inner cell mass (which will become the embryo itself) versus the trophectoderm (which will form the fetal side of the placenta). This process may be continued into the various stem cell lineages which persist in the body, and thus produce permanent changes in the ability to repair damage, for example vascular endothelial progenitor cells and cardiomyocyte stem cells.

The differentiation of specific cell types from an undifferentiated progenitor demonstrates the power of epigenetic control: a lung cell, a liver cell, and a lymphocyte are very different in structure and function and have very different patterns of gene expression, yet of course they possess the same genotype. The development of specific organs and tissues can be affected by the developmental environment, so aspects of the fetal environment can induce differential growth of organs. For example, growth-retarded offspring born to women who have had placental insufficiency may have fewer nephrons, and this is one reason why they are at a higher risk of developing hypertension later in their lives (Benz and Amann 2010).

Importantly, many plastic changes involve permanent changes in the settings of homeostatic processes upon which short-term survival and performance depend. For example, the thermal environment of the infant is thought to affect for life the number of sweat glands that are innervated and thus able to be used for thermoregulation. In Chapter 5 we discuss how an individual’s
Box 4.3 Genetic Assimilation and Developmental Buffers

Around 1940, Conrad Waddington observed that a particular wing trait in *Drosophila*, called *crossveinless* (because the connecting veins in the wing do not form properly), could be induced by exposure of the larvae to high temperature, and that after several generations of selection the trait was expressed without exposure to high temperature (Waddington 1940a). Waddington called this process genetic assimilation (Waddington 1953). He interpreted the initial expression of the wing deformity as a disturbance of canalization—the heat treatment shifting the trajectory of development to a new course—and its fixation by continued environmental stress over several generations as the result of strong selection for the new phenotype. The observation remained a biological curiosity for 50 years until modern molecular genetics provided a possible explanation for the mechanism.

The heat-shock proteins (Hsps) are a family of proteins that are formed in large amounts in many organisms after exposure to environmental stressors such as high temperature. These proteins act as molecular chaperones, binding to and protecting various critical signaling molecules in the cell; they are also important in the biology of fever (see Section 10.7.1). Even in the absence of stress, Hsps are among the most abundant proteins in cells. Studies in *Drosophila* showed that reduction of the amount of Hsp90 by genetic or pharmacological means caused many different sorts of abnormalities in the flies (Rutherford and Lindquist 1998). Selection for particular phenotypes over several generations caused them to be assimilated and expressed even if levels of Hsp90 were returned to normal: an exact parallel with Waddington’s observations. In this situation, Hsp90 appears to act as a buffer against the expression of genetic variation, presumably by its interaction with signaling molecules, and when the amount of Hsp90 is reduced and buffering capacity is exceeded the variation is expressed. Such buffering is known as genetic accommodation (maintaining a robust phenotype in the presence of genetic variation).

These experiments involved artificial selection, but the process of genetic assimilation itself may play a role in the evolution of novelty. The inducing condition creates an advantageous phenotype through the processes of developmental plasticity. One can envisage that those members of the species that exhibit that phenotype with a lower level of stimulus are favored in selection. This progresses over generations if the stimulus is maintained, until a situation is reached where the phenotype is maintained or fixed without the need for an external stimulus.

An alternative explanation is that the stimulus in early development has induced an epigenetic change or epimutation. This induced the phenotypic change which could be re-induced in subsequent generations by the continued environmental stimulus, or perhaps by epigenetic inheritance. Fixation may then occur, potentially by a random mutation that induces the same phenotype or due to a biased rate of mutation at the site of the epigenetic change (Bateson and Gluckman 2011). These are speculative mechanisms, but methylated CpGs are known mutational hotspots due to the capacity for methylated cytosine to undergo spontaneous or enzymatic conversion to thymine, a functional nucleobase that may escape the DNA repair mechanisms (Pfeifer 2006). In line with this, CpG hypermutability has been shown to explain the decrease in frequency of amino acids coded for by codons with CpGs (Misawa et al. 2008).

These models suggest that phenotypic changes can lead to genotypic change, rather than phenotypic evolution being solely dependent on random mutational change. One example of a human characteristic that may represent the outcome of such assimilation is the thickened heel pad of infants at birth: how is it that evolution selects for thickened heels when walking and the need for thickened heels come some months later? Clearly there would have been advantage in having thickened heels to increase mobility and reduce the risk of foot injury. The classical view would be that a mutation spontaneously occurred that led to heels being thicker, whereas the emergent view would view the relationship between walking and heel pad thickness appearing under natural selection as similar to that between heat stress and vein distribution in the artificially selected fruit fly. This is an exciting, controversial and rapidly emerging area of study.

Reproductive strategy can be determined by the early life environment and in Chapter 9 we discuss how the capacity to regulate metabolism is environmentally determined in early life.

Developmental plasticity has potential adaptive value; perhaps there would be an advantage in being infinitely plastic but it is limited to the early period of life, for several reasons. First, there are inherent constraints; once certain features have developed plasticity may no longer be possible. Once cells are committed to one lineage or another, it is not possible for a developmental influence to subsequently
change that commitment. Second, maintaining the capacity for developmental plasticity may have energetic costs that cannot be afforded.

Thus development is characterized by critical windows. These depend on the nature of the cue and the organ system involved. For example, in rats neurogenesis is largely complete by 3 weeks’ gestation (i.e., about the time of birth), and developmental cues could not be expected to have a major effect on neuronal number after that age. In mammals there is distinct anatomical and functional organization of the hypothalamus between the two genders. In the rat there is a critical period between day 1 and day 5 of neonatal life when exposure of the female brain to testosterone will lead to its being masculinized, and the rat will grow up with some male-like reproductive and neuroendocrine characteristics. The equivalent period in the human appears to be much earlier in fetal life and is not as well defined.

4.5 Responses to Environmental Cues During Development

Developing mammals obtain information about their environments through their mothers in the form of nutrients, hormones, or other substances that cross the placenta or are passed to the infant in the milk. Aspects of the mother’s behavior can also play a role here, whether in terms of the mother’s degree of stress or her nursing and nurturing behavior after birth. The developing fetus responds to this information by making developmental “decisions” about its optimal phenotype (Gluckman and Hanson 2005). This use of the word decision is again metaphorical. It is short-hand for saying that organisms have evolved with the capacity to respond differently to different circumstances with long-term implications for the phenotype. This capacity presumably evolved because it had fitness advantages in that it matched the individual better to its environment. These decisions may not be simple, because the information about the environment can vary in degree and in accuracy. Further, not all environmental stimuli induce a plastic response whereby the developmental program is maintained: some stimuli actually disrupt the normal pattern of development.

4.5.1 Developmental disruption

Not all environmental cues acting in development produce a potentially adaptive or plastic response. Severe environmental influences can disrupt development by interfering with processes of gene expression, cell proliferation, or migration. This is the domain of teratology: the study of developmental disruption which may lead to congenital abnormalities. Anomalies may arise because of genetic mutations (e.g., heart disease, mental retardation, and other features of trisomy 21), or because of environmental influences. The two processes are not really separable because the incidence of mutational disorders such as trisomy 21 increases markedly in women over the age of 40 (particularly when combined with advanced paternal age), suggesting that some change in the environment of the maturing ovarian follicle has induced the mutation.

Purely environmental influences include the teratogenic effects of drugs such as thalidomide (which interferes with limb development, leading to phocomelia) and rubella embryopathy (in which infection of the fetus by one of the few viruses which can cross the placental barrier leads to mental retardation, deafness, congenital heart abnormalities, and interference with long bone growth of the fetus producing severe growth retardation). There remains much controversy about the potential disruptive effects of toxic chemicals such as dyes, heavy metals, and insecticides. For example, there appears to be a higher incidence of hypospadias (incomplete development of the urethra so that the meatus is not at the tip of the penis) in boys who are born to women living close to certain waste-disposal sites in Europe. There is no adaptive basis for any of these outcomes; all involve disruption to the developmental program.

4.5.2 Potentially Adaptive Responses in Development

The capability to alter the trajectory of development, and thus the mature phenotype, in response to environmental influences acting early in development is essentially a universal phenomenon in biology: it is seen in plants as well as animals. Natural selection has presumably favored such developmental
plasticity because it offers the potential for adaptive advantage. It allows the organism the opportunity to mold its later phenotype in response to environmental cues so as to improve its chances of survival and reproduction. Thus the phenotype which develops from a genotype is context-specific; there are many possible phenotypes and their range will change as the environment during development changes.

There are two basic types of non-disruptive developmental responses, although the division is somewhat arbitrary and both may operate together, and they can be viewed as part of the continuum of life-course-related responses acting within the individual (Gluckman et al. 2005c; Chapter 5). Sometimes the developmental environment creates challenges to which the individual must respond immediately so as to increase the chance of survival and thus maintain the opportunity to reproduce—we call this class of responses immediately adaptive responses. At other times the opportunity arises to enhance fitness by matching the phenotype better to the anticipated environment: this second class of responses are made for anticipated need or advantage later in the life course, and are called predictive adaptive responses. They may or may not involve an immediate change in phenotype.

4.5.2.1 Immediately Adaptive Responses: Coping with the Consequences

The developing organism is not passive. As the human fetus develops it gains many homeostatic capacities which operate before birth, particularly in late gestation. For example, if faced with a transient shortage of oxygen due to temporary compression of the umbilical cord, it will reduce limb and breathing movements and redistribute blood flow in an attempt to maintain oxygen delivery to the brain, heart, and placenta, all essential for its immediate survival. But there are more sustained threatening scenarios where the embryo or fetus makes plastic responses over a longer time frame in an attempt to survive the environmental threat (here the intrauterine environment influenced by the mother). These, generally more severe, immediate adaptations allow for fetal survival but may leave the developing individual with a potentially disadvantageous phenotype with which it must cope for the rest of its life. So trade-offs and “decisions” with long-term consequences start well before we are born. Section 13.5.2 discusses several of these trade-offs with respect to pre-term birth.

The most well-understood examples concern reduced nutrition during embryonic or fetal life. This might occur during a period of famine or if the mother has disrupted placental function, for example due to malarial infection (invasion of the maternal placenta by malaria parasites) or pre-eclampsia (see Section 8.9.4). But even in apparently uncomplicated pregnancies in developed societies many women consume an unbalanced diet lacking in key micronutrients, or are too thin or too fat. These situations pose the developing mammalian fetus with a challenge. Should it reduce its growth in response to the poor nutritional supply? If it does, it may be born smaller and be more likely to die in infancy or (in the wild) be more subject to predation, and so have reduced fitness. But the alternative may be to die in utero, so the decision is clear. Humans born smaller have consistently increased neonatal and childhood morbidity and mortality (Figure 4.3), they are more prone to infection, and may have impaired cognitive development and/or stunting in height. A range of compromises must be made, in metabolic homeostasis and growth and in the function of critical organs including the brain.

However, the detrimental effects of reduced growth can be minimized by disproportionate reduction of growth in non-critical (for the fetus) organs such as the kidney, allowing preservation of nutrients for the brain and heart. Being born with fewer nephrons means that those present must filter the blood at a higher rate and this will increase the risk of hypertension later in life, but this will probably be in the post-reproductive phase and hence will not reduce fitness. The fetal challenge may also impair the development of the pancreatic islets and of skeletal muscle. This will limit both insulin secretion and glucose uptake into muscle. Such individuals will be potentially insulin deficient and also insulin resistant; this will reduce their early growth and increase the risk of diabetes later (see Chapter 9).
There has been some debate over who benefits from this adaptive response: is it the mother or her fetus? Could maternal survival be the real driver of this response? This may well be the case in a polytocous, frequently reproducing species which can manipulate its reproduction much more than a slow monotocous reproducer can. Unlike many polytocous species, humans do not completely cease reproduction during conditions of famine or employ embryonic diapause as a way to match the progression of pregnancy and birth to nutritional environments. So even under severe famine the mother’s state does not necessarily outweigh the fitness advantages of attempting reproduction. In humans, lactation is remarkably well sustained during famine, suggesting that the mother is prepared to sacrifice resources to meet the evolutionary drive to reproduce. In these cases the trade-off is between the risk of not reproducing now and not surviving to reproduce, or delaying reproduction in the expectation of better conditions later. These issues are considered further in Chapters 5 and 8.

### 4.5.2.2 Predictive Adaptive Responses

In some situations, embryos and fetuses may make developmental responses that are not for immediate survival but appear to be for longer-term advantage. These predictive responses are hypothesized to be plastic “decisions” made by the embryo/fetus/neonate in response to how it interprets the current environment as a predictor of the future one (Gluckman et al. 2005b; Bateson et al. 2014; Figure 4.4). In essence, the growing organism anticipates or forecasts its future on the basis of nutritional or hormonal signals it receives from its mother, either in utero or during lactation, and adjusts its phenotypic development accordingly.

The fetus must respond to information over a longer timescale than just a single point in time. The mother acts as an integrating transducer of environmental information, and this has the advantage that the fetus will adjust its phenotype for an average environment rather than responding to every minor change in maternal circumstance (e.g., periodic changes in glucocorticoid levels as a result of exercise and minor daily stresses). The origin of this form of potential adaptation lies in trying to select a phenotypic trajectory that will optimize the offspring’s fitness opportunities, and thus it needs to have an integrated interpretation of its environmental forecast. This implies a degree of inertia in the prediction–response relationship (Kuzawa 2005). There are now considerable data to show that maternal body composition at the beginning of pregnancy is a major determinant of pregnancy

![Figure 4.3](image-url)

**Figure 4.3** Neonatal mortality in the USA depends on birthweight. The shapes of the curves for two calendar years nearly 50 years apart are similar, indicating that the relative decline in mortality remained uniform across all birthweights. Plotted from data in Wilcox (2001) and Mathews and MacDorman (2013). Note that the optimal weight for neonatal survival was approximately 4 kg in 2010, whereas the mean birthweight was lower. Other studies support this finding, suggesting that the optimal birthweight, in terms of lowest perinatal mortality, is much higher than the median population birthweight—between the 80th and 84th centiles for the population (Vasak et al. 2015). That the median population birthweight is lower than this is a consequence of maternal constraint, the evolutionary significance of which is described in Section 8.9.5.
4.5 Responses to Environmental Cues During Development

Outcomes (birth size and gestational length) and this may reflect that the embryo can use some correlate of that state as a measure of environmental conditions over a long timescale.

These predictive responses are integrated responses affecting multiple components of the phenotype (Gluckman et al. 2007a). The best-documented cues are nutritional or related to maternal “stress” via the HPA axis and they lead to alterations in endocrine, cardiovascular, metabolic, and reproductive function, and in the development of adipocytes and myocytes. The period of plasticity may vary according to the physiological system involved, but in humans it clearly extends at least from conception until after weaning.

There is not an absolute distinction between predictive and immediately adaptive responses, which probably involve overlapping epigenetic mechanisms. Thus those who are born smaller because of immediate responses may also have altered metabolic performance as a result of predictive responses. Indeed, there is an inverse correlation between birth size and the lifetime risk of type 2 diabetes (Hales et al. 1991). These plastic processes allow for responses within the unexceptional range of developmental exposures and represent an essential component of generating the reaction norm.

Equivalent anticipatory responses are common in other taxa, for example in polyphenic species such as the African locust where phenotype induction by early life environmental signals occurs at a time when there can be no immediate advantage. The decision by the locust larva to develop into the migratory phenotype, with its omnivorous diet, fat-based metabolism, and large wings, is made at a time when it cannot fly. The alternative, solitary, form has a selective diet, a glucose-based metabolism, and small wings and is a form that is advantageous when population density is low and food is plentiful. The choice of morph is made using chemical cues secreted by the mother in the egg cast, which are detected by the larva as it eats through the cast after hatching. Thus developmental plasticity is a tool used across taxa to allow an individual within a species to match its phenotype to the environment it will inhabit (Low et al. 2012).

The hypothesis of predictive adaptive responses has been somewhat controversial. Direct proof in humans is not practical, but very strong indirect evidence of a fitness value of such responses has been visible.
shown by studies of infants who suffer severe malnutrition (Forrester et al. 2012); these studies, which demonstrate how an evolutionary medical hypothesis can be tested, are described in Box 7.2. Later in the book we shall expand further on how predictive responses can play a major role in influencing the age at puberty (Section 5.4.2.2) and in conferring a greater risk of metabolic and cardiovascular disease (Chapter 9).

Environments can shift within a single lifetime, and predictions made in early life may not accurately forecast later experience. However, mathematical modeling shows that the fidelity of prediction need not be high for these to confer a fitness advantage, particularly when environments shift on a time base equivalent to the generation time of the species (Jablonka et al. 1995). While these mechanisms evolved in invertebrates and persist in vertebrates, including humans, the challenge for the fetus is that its ability to read the future environment is confounded by the imperfect transduction of environmental information from mother to fetus—she may consume a diet unrepresentative of the contemporary population, or have hypertension or placental insufficiency. Thus, although mammals may have lower fidelity in their predictions, anticipatory plasticity evolved and has been sustained because it confers sufficient advantage.

### 4.6 Epigenetic Processes and Development

Epigenetic processes (see Section 3.7) are critical to development and are the major molecular mechanism underpinning developmental plasticity.

Following fertilization, widespread removal of epigenetic marks occurs as both maternal and paternal genomes undergo extensive demethylation—the former by passive depletion of 5mC and the latter by a combination of pathways that involve active demethylation, including enzymatic oxidation of 5mC (Seisenberger et al. 2013). This ensures pluripotency of the developing zygote. Some epigenetic marks, mostly those on imprinted loci, survive this process to preserve parent-of-origin genomic identity. Then, de novo methylation occurs in the inner cell mass of the developing embryo. Cells destined for the soma then acquire lineage-specific methylomes, while cells committed to become primordial germ cells undergo a second round of more extensive epigenetic reprogramming. This time marks at imprinted genes in the germline are also erased, and the epigenome is then re-established in a sex-specific manner.

DNA methylation silences the expression of specific genes during the development and differentiation of individual tissues. This is the basis of cell differentiation, and pluripotent stem cells retain the ability to make epigenetic changes at the appropriate time. The difference between two cell types from the same individual lies in the epigenetic profile that determines which genes are functional and under what conditions. For example, the expression of the homeobox (Hox) gene Oct-4, a key regulator of cellular pluripotency in the early embryo, is permanently silenced by hypermethylation of its promoter around embryonic day 6.5 in the mouse, while HoxA5 and HoxB5, which are required for later stages of development, are not methylated and silenced until early postnatal life.

For some genes there also appear to be gradations of promoter demethylation associated with developmental changes in the role of the gene product. The δ-crystallin II and PEPCK gene promoters are methylated in the early embryo, but undergo progressive demethylation during fetal development and are fully demethylated and expressed in the adult. Thus functional changes in different cell lineages are established at different times during development of the embryo. The established pattern of DNA methylation is then copied during mitosis by DNA methyltransferase activity. This provides an “epigenetic memory” of patterns of gene regulation, and hence cell type and function, which once established during development is passed through subsequent cell divisions.

This discussion has focused on cell differentiation, but the same processes can inform the function of regulatory systems within these cells. This immediately suggests a mechanism by which the environment may induce stable changes to cell function which persist into adulthood, and therefore a way in which environmental challenges at different times during development may produce different phenotypic outcomes, and so a differential risk of disease.
Gene-promoter methylation is important for asymmetrical silencing of parentally imprinted genes, for X-chromosome inactivation, and for silencing retrotransposons, some of which are viral DNAs that have invaded the human genome (Section 3.3.4).

Genomic imprinting, in which the expression of an allele depends on its parent of origin, represents a special case of epigenetic regulation of gene expression (genomic imprinting should not be confused with the behavioral imprinting studied by Konrad Lorenz, most famously with geese). Imprinting is most frequently mediated by allele-specific DNA methylation, and incorrect establishment has been linked to several rare imprinting disorders. For example, under normal conditions only the paternal allele of insulin-like growth factor 2 (IGF2) is expressed, but its biallelic expression results in Beckwith–Weidemann syndrome (Brown et al. 1996). The result is a neonate who is abnormally large with hyperinsulinemia. This leads to hypoglycemia and thus a risk of brain damage; these individuals also have an increased risk of developing cancers.

Other imprinting disorders include Prader–Willi and Angelman syndromes, both involving the same locus on chromosome 15q encompassing several imprinted genes (Cassidy et al. 2000). Prader–Willi syndrome is the result of loss of the active paternal allele (the maternal allele normally being silenced) and Angelman syndrome is the loss of the maternal allele (because there is one gene at the locus which is normally silenced on the paternally inherited chromosome). The phenotypes that arise are quite different: Prader–Willi syndrome is characterized by obesity and hyperphagia, mild mental retardation, hypotonia, and hypogonadism, whereas Angelman syndrome shows developmental delay, hand-flapping, seizures, and a happy demeanor.

Mouse embryo studies have shown that assisted reproductive techniques affect parental imprinting status and imprinted gene expression. There is weak evidence in humans that the incidence of imprinting disorders, and particularly Beckwith–Weidemann syndrome, is increased in offspring conceived using assisted reproductive techniques. There is also emerging evidence that the risk of other conditions such as cardiovascular disease and type 2 diabetes is increased in people conceived by in vitro fertilization, and a component of these conditions is thought to involve non-imprinted genes (Chen et al. 2011).

Epigenetic mechanisms allow another level of control on top of the genetic code itself (Box 4.4). Three aspects of the process are of interest in the current context. The first is that epigenetic changes in mammals do not usually involve the coding regions of the genes themselves; rather they affect the gene promoters and enhancer regions well upstream. This means that they do not affect gene transcription until the appropriate transcription factor is present. This allows for the phenotype to be affected in a contingent way, as it will alter the future responses of the individual to an environmental challenge which changes levels of the relevant transcription factors. The second aspect is related: namely, it provides an explanation for how a challenge during development, which did not overtly disrupt that development, can have distant effects at a later stage in the life course. This therefore provides a mechanistic basis for predictive adaptive responses. The third consideration arising from these epigenetic concepts is that they stand distinct from genetic determinism. They offer ways in which interventions could potentially reverse epigenetic changes set up in early life, if they can be identified and rectified within the critical window of plasticity. This will be challenging, but not impossible, as laboratory methods are available for assessing DNA methylation and histone modification, and there is abundant experimental evidence that epigenetic state can be pharmacologically and nutritionally manipulated or reversed (Vickers and Sloboda 2012). Epigenetic changes are a hallmark of cancer and may provide a therapeutic target; some compounds that function as inhibitors of histone deacetylase have already been approved for use in the treatment of some types of cancer, and more have entered clinical trials in oncology (West and Johnstone 2014).

4.7 Learning and Instinct

Learning is a particular and distinct form of plasticity which is intimately connected with the more structural (e.g., synaptogenic) and molecular
Learning is generally transmitted by instruction or observation between individuals, but some behaviors appear to be “instinctive” rather than learned. The transmission of learned behavior was considered independently by several theorists—Lloyd Morgan, Osborne, and later Baldwin—at the end of the nineteenth century and perhaps unjustly has become known as the Baldwin effect (Robinson and Dukas 1999). In essence, the Baldwin effect proposes an evolutionary mechanism for how a learned behavior can become innate or instinctive. Once a particularly advantageous behavior (e.g., predator avoidance or a new way to exploit a food source) has been “discovered,” individuals who have an improved capacity to learn that skill will be selected until the behavior becomes an integral part of the species’ genetic repertoire. This is very similar to the mechanisms and phenomena that Waddington had termed genetic assimilation (Box 4.3). In this way, behavior can shape the course of evolution of a species.

The concept of the Baldwin effect is still controversial: there is debate about the extent to which a learned behavior can eventually be assimilated into the genome and about whether selection can act on specific components of behavior or just on general learning abilities. In addition it has been argued that...
the change from learning to genetically determined behavior (instinct) will not necessarily give greater fitness except in very stable environments where change is extremely slow.

4.8 The Evolution of Novelty

A major question in macroevolution is how phenotypic novelty is achieved. This question is largely beyond the scope of this book, and we will only summarize the issues here because they highlight the importance of adopting a developmental perspective.

The evolution of a phenotypically novel feature, such as the development of limbs in the shift from an aquatic to a terrestrial vertebrate, is dependent on selection acting on phenotypic variation which is or becomes underpinned by genomic variation. As discussed in Chapters 2 and 3, that variation in turn is largely a result of mutation and recombination. But any selected variant must be viable, and this generates bias in what can be selected. As an adult phenotype is dependent on the successful and viable development of the organism, studies of how macroevolution occurs have increasingly focused on embryological development and the science of developmental biology.

Random mutation has generally been seen as the powerhouse for the generation of novelty, but there may also be other factors at play which facilitate the appearance of new forms. Indeed, most random mutations will be either covert and neutral (unless the environment changes to favor the changed phenotype) or, if they involve a coding region, they may well be lethal. Mutations affecting transcriptional regulation are less likely to be lethal as they will affect the place, magnitude, or timing of the expression of a gene and are more likely to result in viable modification of structure or function. The possible role of cis-regulatory (non-protein coding) mutations in morphological evolution is explored in Box 4.5.

The early embryo differentiates into a number of compartments because of the expression of diffusible factors called morphogens which give it polarity, laterality, and its dorsal–ventral dimensions. In turn, this allows the embryo to start to differentiate. A key feature of embryonic development is the distinct pattern of “segmentation” both dorsoventrally and rostro-caudally, and this gives each phylum its distinct organizational characteristics. During development, clusters of genes are activated in strict sequence, determining the overall body pattern in terms of numbers of segments and the orientation of each particular segment; then subclusters of genes specify localized structures such as the particular arrangement of organs and tissues within each segment (Wolpert et al. 2002).

Box 4.5 The Origin of Variation

Until around 40 years ago, the conventional explanation for the source of genetic novelties in evolutionary change was mutations in genes coding for the proteins of the altered structure. Then a number of researchers in the emerging field of evolutionary developmental biology began to present a new source of evolutionary mutations: regulatory genes. They maintained that regulatory DNA, known as cis elements, had a profound effect on changes in morphology and body plan, and that mutations in this DNA could cause large changes in morphology. For example, in the freshwater and marine varieties of the stickleback fish, the marine species maintain their pelvic skeleton, which serves as a sort of armor, whereas some freshwater species have evolved the total or partial loss of their skeletal armor. The relevant Pitx gene was found to be expressed in the pelvic region of the marine species but not the freshwater species, although the sequence of the gene itself remained the same in both groups. This too seems to be evidence that mutations in regulatory cis-element DNA, not coding genes that would actually alter the Pitx gene sequence, were responsible for these new adaptations.

The role of cis elements is a subject of debate, with some investigators maintaining that cis-regulatory mutations are significant but do not represent the foundation of morphological evolution. They point to the many more well-documented examples of novelties arising from mutations in coding genes, and point out that few of the examples of cis-regulatory mutations that cause morphological change involve clearly adaptive traits.
Two classes of genes are critical to specifying this body plan: the segmentation genes, which determine how body segments are divided, and the Hox genes, which determine how each individual segment develops. The homeobox is a DNA-binding element in a protein which attaches to promoter regions of other genes and activates them. Hox genes therefore code for protein transcription factors which, often working in combination, switch on cascades of gene expression that specify the proteins, and hence the structures, that will be produced in the various segments of the developing embryo. Early work on Hox genes was performed in Drosophila, which is easy to study because of its rapid life cycle and well-defined adult segments, but very similar Hox genes are also present in the genomes of mammals. The Hox genes in flies are collinear: arranged on a single chromosome in the order of the body parts which they specify. Although the genes have been duplicated in mammals to form clusters on four chromosomes, their basic arrangement within each cluster is the same (Figure 4.6).

Segmentation or compartmentalization allows a limited number of genes to serve different functions by being linked together in modules which may evolve functionally for different purposes. Hence fins, wings, and limbs all involve homologous genes in comparable segments evolving differently in different organisms. At one level these modules make the phylum fundamentally robust; at another level mutational variation within a module can allow phenotypic change that is more likely to be viable. Mutations of the pattern-forming genes can cause large—but often viable—morphological changes. For example, much early work on Drosophila pattern-forming genes was carried out on a mutant with an extra pair of wings in place of stubby structures called balancing organs (halteres). The collinearity of the relevant genes meant that when the gene which would normally have specified the balancing organ was mutated, the next gene in the sequence came into action and specified a pair of wings instead.

There is a growing body of knowledge that such processes are fundamental components of macroevolution. But equally they can explain certain birth defects in humans. For example the homolog of Pax6 is involved in formation of the eyes in invertebrates and Pax6 mutations in humans can cause aniridia, which manifests as alterations in the structure and function of the eye. Mutation of the human homolog of the Drosophila segmentation gene patched causes the genetic disorder basal cell naevus syndrome, which is characterized by multiple and proliferating skin cancers. Early nutritional deficiencies in the embryo may affect the expression or action of pattern-forming genes, possibly explaining, for example, the gross neural tube defects and anencephaly caused by folate deficiency before neural tube closure is complete. The relationship between multicellularity and organizational complexity is discussed in Box 4.6.

4.9 Conclusion

Development and its contribution to adaptive processes, and the ways in which the environment affects it, were more or less written out of scientific thought during the twentieth century. But we now realize that even if we could succeed in identifying
all the genes which influence a complex trait, we would still not be able to describe or predict the expression of that trait in any given individual. This is just as true of disease. It is the triad of genes, development, and environment that is responsible for the adult phenotype of an individual organism, and therefore susceptibility to disease. The consequences of thinking about the phenotype in this context, rather than focusing on the environment or the genotype alone, should be apparent.
Somatically acquired characteristics cannot be inherited, but environmental influences in one generation can influence subsequent generations through the processes of developmental plasticity and direct or indirect epigenetic inheritance. Modern ideas about non-genomic inheritance, whether mediated by cultural influences, epigenetic changes, or effects on the developmental environment, do not support Lamarck’s concept itself, but they nonetheless endorse the idea that phenotype is influenced to a significant degree by these processes, not just by Mendelian genetic inheritance.

Molecular developmental biology has given us many clues about the processes which mediate such effects, especially epigenetic processes. Measurement of epigenetic marks such as DNA methylation in early life now offers the possibility of early diagnosis of disease risk resulting from inappropriate developmental influences, and also of novel interventions and therapies. This area of medicine is the subject of much current research, and clinical application will be realized in the near future.

The challenge for biomedical science is to marry the concepts of evolutionary genetics, developmental biology, and environmental science into a unified vision of what makes us what we are, and how to address the challenges which we face as a species.

**Key Points**

- The triad of genes, development, and environment is responsible for the adult phenotype of an individual organism.
- Developmental plasticity is the capacity to adjust the phenotype arising from a single genotype by changing pathways of development in early life. It allows organisms to maintain or enhance fitness by matching themselves better to different environments, and to adjust to changing circumstances on a timescale intermediate between that of selection and homeostasis.
- The modularity of structure and function allows for duplication, and this can provide an important source of evolutionary novelty.
- Epigenetic molecular processes are central to the mechanisms of developmental plasticity.
- Developmental plasticity, which has an adaptive origin, must be distinguished from developmental disruption, which does not. However, some developmentally plastic responses, while being adaptive in origin, can lead to maladaptive outcomes and may result in impaired health later in life.
5.1 Introduction

Every species has one or more stereotypical patterns that describe its life cycle, and the term life history is used to describe this constellation of key characteristics, including the patterns of growth, development, maturation, reproduction, and mortality, which define progress through life (Stearns 1992). Some elements of the life history are common to both males and females, others are sex-specific. Depending on the species, some stages and transitions within the life history can be plastic. For example, the age of puberty in humans is highly plastic reflecting experiences early in life, and there is growing evidence that such plasticity has evolved for potential adaptive advantage.

There is an enormous diversity in life-history strategies across the animal kingdom, but there are also unifying patterns and similarities. For example, from bacteria to the giant sequoia there is a tight relationship between lifespan and body size. Mice are not only much smaller than dogs, humans, or whales, but they also have shorter lifespans. Naturally, their offspring must be born much smaller and, perhaps partly as a result, they also have much higher extrinsic mortality from dangers which large size protects against, such as dying at the hands—or talons—of a predator. Given the high likelihood that a newborn mouse will end up as a meal for a predator, the mouse has evolved a polytocous trait, such that rather than giving birth to a single offspring it has large litters at frequent intervals in the expectation that a few will survive to adulthood to reproduce. Contrast this with the life strategy of an elephant or a human and we begin to see how the characteristics of a species should be envisioned as encompassing much more than just its adult form.

Instead, an organism should be viewed as having a strategy of development, growth, reproduction, and longevity which has been adaptively shaped by the environments in which it evolved, and thus by natural selection.

Such a comparative approach is a powerful tool, revealing regularities linking traits across the animal kingdom. Defining these relationships is valuable for two reasons. First, it allows the identification of general patterns hinting at deeper constraints or underlying processes. Such regularities might, for instance, reflect the limitations of the mechanical properties of bone, the finite energy available for distribution and use within a body of a given size (see later), or something as simple as the need for more time for larger-bodied species to reach that larger size. Second, it allows us to identify deviations from the normal trend, which highlight those instances when selection has pushed a species into an unusual state for that trait. This is illustrated by the example of brain size (see Box 5.5).

Life-history theory is an important sub-field of evolutionary biology, and its application provides a valuable means for understanding important aspects of biology and behavior. Understanding how humans fit within the general mammalian pattern as well as those instances in which we deviate from it has been essential to understanding the evolution of the human life course.

Like other selected traits, life-history traits are subject to variation. But within this potential variation there are interactions or trade-offs which constrain possible life-course strategies. For example, across species (and often within a species) there is commonly an association between an earlier age at sexual maturation and a reduced adult body
size. Accelerated maturation—such as the processes of metamorphosis or puberty—tends to be observed when there is a high threat of predation or death. This relationship results because the potential advantage of a larger body size in later life is "traded off" against the greater cumulative risk of death from disease or predation during a longer period of maturation. Understanding such trade-offs between life-history traits is essential for interpreting many biological strategies. We will return to these themes of shared pattern and uniqueness throughout this chapter.

But, as has been suggested, life histories are not immune from environmental influences. This is most obvious in polyphenic species in which environmental signals early in development can cue very different patterns of phenotypic development. One well-studied example is the female honey bee larva, which will develop into a worker or a queen depending on the nutrition provided by worker bees at critical larval stages, and this is effectively determined by the colony’s needs at the time. The phenotypic effects of the differential feeding of bee larvae are mediated by epigenetic mechanisms (Foret et al. 2012).

More subtle effects of environmental signals on life-history traits are universal—for example effects on fecundity, rate of maturation, and timing of metamorphosis are common. There is growing evidence that environmental factors also influence human life-history traits, such as the dramatic reduction—but not complete obviation—of human fertility during famine (Peng 1987), and there is also evidence that the age of puberty and lifespan are affected by conditions in early life.

Human life history is typified by an individual being born to a singleton pregnancy after a gestation of approximately 280 days. This is followed by a long phase of post-natal nutritional dependence on the mother and a prolonged childhood with sexual maturity delayed until more than a decade after birth. Females generally enter puberty before males and there is only modest sexual dimorphism in adult body size, which has been suggested to reflect the evolution of hominin mating systems (Section 8.6). Each human female has only a few children, and there is high parental investment in each child such that there is a comparatively high probability that children will live to be able to reproduce; there are estimates that perhaps over 50% of infants born survived to adulthood in prehistoric societies (Hassan and Sengel 1973). Family structures generally involve some form of pair bond between a male and female(s). But whereas males are capable of reproduction into old age, females experience menopause and terminate reproduction before the end of their intrinsic lifespan. The slow intrinsic rate of senescence means that lifespans in excess of 70 years are not exceptional, and indeed become the norm when extrinsic causes of death are reduced (see Box 5.3).

5.2 General Overview of Life-history Theory

Assembling data points across a wide array of mammalian species illustrates not only variation but also some underlying similarities in the linkage between traits such as growth rate, body size, and lifespan. These patterns in turn beg deeper questions about the biological processes which underpin them. What are the differences in biological strategy which distinguish a mouse from a human or an elephant?

A central tenet of life-history theory is that organisms vary, in large part, because of differences in the ways that they harness and use the finite energy and nutritional resources at their disposal. Functions such as maturation, growth, reproduction, and the intrinsic repair processes that slow ageing and extend lifespan each require energy and commitment of resources. There are constraints on the size of the pool of resources that a species has available to it, and this unavoidably leads to trade-offs between functions. This concept of finite resources provides important insights into how physiological processes have evolved to balance the trade-offs that necessarily ensue. Before exploring these trade-offs, let us first look at the forces which limit resources.

That an organism must have some upper limit to its energy use is intuitive and has long been appreciated among biologists. Resting metabolic rate can be described as a power function of mass in mammals (Figure 5.1); this is sometimes called Kleiber’s law (Box 5.1). Not surprisingly, as one moves from small to large organisms, total energy expenditure
Box 5.1 The Power of Quarters

Early proposals to explain Kleiber’s law pointed to the fact that an organism’s surface area, which radiates heat, increases more slowly than does its volume or mass, which produces it. The geometry of that arrangement implies a scaling component of 2/3 or 0.67, close but not close enough to the empirically determined 0.75. Newer concepts look to the geometry of the circulatory system for an explanation. Organisms face an important challenge in distributing resources within their bodies. As blood circulates around the body, it must service all the body’s cells. An optimal solution to this problem involves a system of branches upon branches upon branches, beginning with the aorta and branching all the way down to very small capillaries. This type of nested branching network is an example of what is termed fractal geometry. Although the mathematical proof is beyond the scope of our discussion, because of this geometry it turns out that the distribution of resources will tend to scale to body mass with an exponent of 0.75, or 3/4.

Intriguingly, there are many other relationships within life that scale at the power of a quarter. For instance, animal lifespan is proportional to the quarter power of body mass and heart rate varies inversely with the quarter power of body mass. Taking out the body-mass component from those relationships implies that all animals will have a similar number of heartbeats during their lifetime, which is roughly true: about a billion. (Humans, of course, with their anomalously long lifespans, have more: 2–3 billion!) At the ecological level, population density of a species scales inversely to the 3/4 power of body mass. There are few species of elephant but many species of insects; indeed, the relationship between body size and species diversity follows another quarter-power law. The foraging area of human hunter-gatherer bands is proportional to the 3/4 power of group size. And finally, returning to the metabolic level, calculation of drug dose according to body mass (for instance, in adapting an adult dose for a child) is often improved by using a 3/4 power relationship.
increases. What is notable, however, is that energy expenditure increases at a rate that is slower than the increase in body mass. As a result, an elephant may expend more joules per hour in total than a mouse, but a kilogram of elephant requires less energy than a kilogram of mouse. Larger organisms tend to be more efficient than smaller ones. As can be seen in Figure 5.1, this relationship is quite regular, with metabolic rate scaling to body mass as a power function with an exponent of 0.75.

This mathematical relationship linking metabolic rate with body size reveals that there are limits to the energy available for use within a body of a given mass. Between-species variations in traits such as growth rate, body size, and lifespan result in part from different strategies of energy partitioning (Figure 5.2). Although there is much leeway in the types of allocation strategies that an organism can evolve, it is important to keep in mind the fundamental law of physics that energy must be conserved: a molecule of ATP used in one cell to service one function cannot be used again elsewhere. The simple fact that the pool of resources at any point in an organism’s life is finite, and can only be used once, has profound implications for understanding the forces that have shaped the evolution of life-history variation. It is also central to understanding how members of a species adapt to environmental challenges such as energetic limitations. Because the body faces trade-offs when allocating resources, species evolved to balance those trade-offs in a fashion that optimizes fitness.

5.2.1 Key Trade-offs in Life Histories

The life-history traits that species manipulate (another metaphorical term) to achieve optimal fitness include: the number, size, and sex ratio of offspring; size and maturity at birth; the pattern of growth; age and size at maturity; parental investment in offspring; and the investment in cellular repair and maintenance which ultimately affects intrinsic mortality and lifespan. There is an extraordinary variation in the solutions that different mammalian species have evolved, each providing an adaptive means of maintaining gene flow within their usual environment(s). Despite this variation, there are trade-offs that constrain the possible solutions, some of which are described in the following sections.

5.2.1.1 Number versus Quality of Offspring

Some species give birth to many offspring, of which few survive. This strategy is typical of rodent species. Parental investment in these offspring is low, if not non-existent, once weaning has occurred. Other species such as humans and other large mammals give birth to a single offspring and invest heavily to nourish and protect that offspring at least until it reaches the juvenile stage. These examples represent two extremes of a range of potential approaches.

5.2.1.2 Current versus Future Reproduction

Another reproductive trade-off is between reproducing now and reproducing in the future. At one extreme are species such as the salmon and the males of some marsupial species, which reproduce once before dying (termed semelparous breeding). These species invest enormous effort in producing a large number of offspring in a single mating event that exhausts their energetic capacity, but is a successful fitness strategy for them. At the other extreme are species that invest in reproduction across multiple reproductive events and have long reproductive lifespans. Human females limit their energetic investment in an individual pregnancy by generally only conceiving a single fetus, but nevertheless that investment is very high given the long gestation and the energetic demands of lactation. Human females have a potential reproductive life of about 15–35 years after puberty, depending on their extrinsic mortality risk. However, over that period of time relatively few children are born and nurtured, and fitness therefore depends on a high proportion (in comparative terms) of these offspring surviving. Female hunter-gatherers generally have five to six children over their life time with between two and four surviving to adulthood (Pennington 2001).

Females limit their investment in each fetus by mechanisms which restrict nutrient flow to the fetus (a phenomenon termed maternal constraint; see Sections 8.9.5 and 9.4.3.2). The first-born child in humans is on average about 150 g lighter than subsequent children. The proximate explanation
may be that the uterine arteries cannot dilate as well in the first compared with subsequent pregnancies because the elastin in the arterioles and arteries breaks down during the initial pregnancy, leading to a lower vascular resistance (the uterine arteries have to dilate considerably in pregnancy and fetal oxygen delivery is limited by blood flow in the uterine arteries; Wang and Zhao 2010). But there is also a life-history perspective: primates may have evolved with a restriction on energy utilization in the first
pregnancy, which tends to have lower maternal and fetal survivability because of maternal immaturity. This has the benefit of conserving energy for subsequent pregnancies.

5.2.1.3 Age versus size at maturity

A defining life-history trait is the timing of sexual maturation, and this is discussed in detail in Section 5.4.2.2. Depending on the species, growth may stop or slow at or soon after sexual maturity. The trade-off here is between investing energy in continued growth and investing available energy into reproduction. Models of mammalian life histories generally assume that there are two competing factors determining the optimal timing of this transition. The first is the importance of adult size for reproductive success. To the extent that body size influences reproductive fitness (by allowing females greater investment in larger, more resilient offspring with lower mortality, and by allowing males greater success in competitive mating), this will favor a delay in the onset of reproduction. This reproductive benefit must be balanced against a second factor, namely the excess risk of dying, which increases as maturation is delayed. Thus where organisms face a high risk of mortality, particularly in the juvenile phase, there is a generally a shorter time to maturation.

5.2.1.4 Fecundity versus Lifespan

Fecundity is the measure of the total lifetime reproductive performance of an organism. In female mammals, fecundity is measured by the cumulative number of offspring from a lifetime of multiple pregnancies. But pregnancy and lactation are energetically expensive for the mother and divert resources from maintenance functions. This trade-off is one reason why there is a reciprocal relationship between lifespan and fecundity across species; short-lived animals tend to have more offspring. Lifespan is, by definition, shorter where there is a high extrinsic mortality rate, and in species where this occurs one fitness-enhancing strategy is to have high fecundity (Box 5.2).

5.2.2 Extrinsic and Intrinsic Mortality

A key determinant of the life-history strategy adopted by a species is its mortality risk profile. Allocating resources to maintenance functions such as tissue repair can be interpreted in a metaphorical sense as representing an “expression of optimism” from the social and economic contexts of their times. Does the relationship still hold for a “normal” (i.e., socioeconomically heterogeneous) population? A similar analysis of demographic data from northern Germany in the eighteenth and nineteenth centuries found an increasingly strong negative relationship between longevity and fecundity with increasing poverty (Lycett et al. 2000). This is what would be expected if the trade-off is mediated by resource availability. A recent study has examined the potential genetic basis of the parity–longevity trade-off, this time in participants in the Framingham Heart Study, a large, ongoing multi-generational cohort of Massachusetts residents that was initiated in 1948. Negative correlations were detected between parity and lifespan, and a GWAS approach identified several SNPs associated with this trade-off, although inclusion of covariates abolished their statistical significance (Wang et al. 2013).

Box 5.2 A Royal Trade-off

Does the trade-off between longevity and fecundity operate in humans? To investigate this would require an extensive data set of the lives of individuals under natural fertility (pre-conception) conditions. Historical reality means that such records have generally only been kept for the wealthy, and indeed one data set used by gerontologists to study this question involves the genealogical records of the British aristocracy, which go back some 1200 years and record information for over 33,000 individuals. As predicted by life-history theory, in this population there was a negative correlation between longevity in women and the number of children that they had (the effects of mortality in childbirth were accounted for by considering only women who reached their post-reproductive years) (Doblhammer and Oeppen 2003).

The individuals represented in these records were of course privileged, in that they were presumably insulated from the social and economic contexts of their times. Does the relationship still hold for a “normal” (i.e., socioeconomically heterogeneous) population? A similar analysis of demographic data from northern Germany in the eighteenth and nineteenth centuries found an increasingly strong negative relationship between longevity and fecundity with increasing poverty (Lycett et al. 2000). This is what would be expected if the trade-off is mediated by resource availability. A recent study has examined the potential genetic basis of the parity–longevity trade-off, this time in participants in the Framingham Heart Study, a large, ongoing multi-generational cohort of Massachusetts residents that was initiated in 1948. Negative correlations were detected between parity and lifespan, and a GWAS approach identified several SNPs associated with this trade-off, although inclusion of covariates abolished their statistical significance (Wang et al. 2013).
that the organism is likely to live into the future, and thus reap the reproductive rewards of investing in keeping its body functioning and healthy. Thus the rate of unavoidable mortality sets how optimistic and “forward-looking” a species can afford to be in its life-history strategy.

A simple thought experiment illustrates how the optimal life-history strategy for a species depends upon the local mortality risk. Imagine a species that inhabits a small island and faces a source of mortality that is beyond its control, such as lightning strikes. On any given day, there is some small (but not trivial) risk of being struck by lightning. Now, as a result of climate change, the occurrence of lightning increases markedly and the risk of being struck increases dramatically. Note that this ecological change is experienced by each individual organism as a greater risk that, on any given day, it will meet an untimely death. This increase in risk will have several important implications for the species and the life-course strategies available to it. The period of growth and development is one in which these risks are greatest from the perspective of fitness optimization. This is not because of the simple fact that the young organism is small and physically vulnerable, but rather because from a genetic perspective it has not yet had a chance to reproduce. A juvenile struck by lightning is an individual who will not be represented in the gene pool of the next generation. It therefore follows that, as the risk of death on any given day increases, the genes of earlier-reproducing individuals will become more common in the gene pools of subsequent generations. This will lead to a decline in the age of maturity in future generations of offspring.

This evolutionary shift in the age at maturity will have broad and cascading effects on the rest of the species’ life history. Assuming that other parameters such as nutrition or growth rate are not affected by the increase in lightning, this reduction in the age at maturity means that there will be less time for growth, and thus future generations of adults of this species will also evolve to a smaller final size. This reduction in adult size will, in turn, have additional effects. All else being equal, species that are smaller as adults have fewer resources to invest in reproduction; that is, in support of fetal and infant growth. Thus, offspring size will also decline as adult size declines.

As the size of young decreases, this has additional cascading implications, for it influences how vulnerable each offspring will be. Small newborns typically have fewer nutritional stores to draw upon, for example for heat generation, in the event of an energy shortfall. They also tend to be more vulnerable to predation. Thus, a by-product of this reduction in offspring size is an increase in offspring mortality. As juvenile mortality increases, it would be too risky for mothers of that species to give birth to a single offspring, for the chances of that offspring making a contribution to the gene pool in the next generation will be very small. As a result, smaller species not only have higher juvenile mortality but they also tend to “hedge their bets” by giving birth to large litters. Having large litters will decrease the size of each offspring further, and thus further increase the risk of fewer offspring reaching adulthood. But this strategy may still yield more surviving offspring than investing in a single larger offspring.

Thus we see how a simple change in the local environment—a change in the risk of unavoidable mortality—will tend to have cascading effects on the entire life history. Higher mortality will favor earlier maturation, which will yield smaller adults with fewer metabolic resources and thus a lower capacity to invest in reproduction. This in turn will lead to smaller offspring with higher mortality. As that mortality risk increases, we expect to see an evolutionary shift from singleton births to litters of smaller and more vulnerable offspring.

So far we have seen how a change in local externally induced mortality will favor earlier reproductive maturation, and how this will in turn have subsequent effects on adult size, offspring size, offspring mortality, and litter size. These influences are not the only pathway by which a change in mortality will lead to an evolutionary shift in a species’ life history. As unavoidable adult mortality increases, investing scarce resources in processes which extend lifespan by maintaining and repairing the integrity of tissues becomes a gamble which is decreasingly likely to have a fitness benefit. As the likelihood of living into the future declines, we expect individuals who reproduce...
earlier and invest fewer of their scarce resources in their own future to be better represented in the gene pool of the next generation. Thus, species faced with higher mortality tend not only to mature earlier and at a smaller body size, but they also invest a larger percentage of their available resources in immediate reproduction at the expense of faster senescence and reduced lifespan (that is, they have higher rates of intrinsic mortality). Thus, from a 1-um Staphylococcus aureus bacterium with a generation time of 27 min, to the 80-m giant sequoia in which each generation spans 60 years, the generation time across species is generally proportionate to mature body size (Bonner 1965).

This hypothetical example is very much analogous to the types of selection pressure that have shaped life-history variation in the animal kingdom. A more common source of extrinsic mortality is predation. From this perspective, one important factor that has favored a strategy of rapid growth, early maturity, small size, large litters, and short lifespan is the high unavoidable mortality from predation faced, for example, by mice. Long-lived species with slow life histories, such as humans or elephants, have evolved ways to avoid predation and reduce their risk of unavoidable mortality. In humans, substantial reduction in extrinsic mortality has been achieved via improvements in public health care, leading to a dramatic increase in the average lifespan in most societies.

Mortality can apparently be a driver of evolution, as in the cascading effects of lightning on the life history in our thought experiment, but it can also be an outcome of a species’ strategy of allocating resources, exemplified in this case by the evolution of a faster pace of functional decline and shorter lifespan as the risk from lightning strikes increases. What distinguishes these types of mortality, and their role in a species’ evolution, is the extent to which an organism can modify them by changing its strategy of allocating energy within the body. Lightning and predation are examples of extrinsic mortality because they are not directly related to how an organism allocates its resources: rather, the risk for an individual is stochastic, dependent on factors such as the size of the population and the number of predators. This is in contrast to intrinsic mortality, which is directly related to how the body allocates its resources across biological functions. An example would be the risk of becoming infected with a life-threatening pathogen. While this too has an element of chance, there are biological decisions within the body that can modify how fatal this infection is likely to be. The effectiveness of the immune system depends on the level of resources allocated to it. Additionally, expending energy in raising body temperature may reduce the effectiveness and impact of a viral or bacterial infection. Intrinsic mortality appears to be largely related to the level of investment made in repairing DNA and other cellular constituents, and modern theories of ageing are largely based on this concept, as will be discussed in Section 5.2.4.

5.2.3 Lifespan and Ageing

Humans have evolved with a strategy of waiting over a decade before beginning their reproductive careers, suggesting that unavoidable mortality (in particular, predation) has been relatively low during hominin evolution. While perhaps only 50% of children in the Paleolithic reached puberty, this is a high proportion compared with most other species. This relatively low rate of extrinsic mortality allowed our species to invest in growing to a larger body size and obtaining more skills and capacities as an adult by extending the period of growth and learning. As discussed earlier, this has other effects on the life cycle: it favors enhanced investment in somatic maintenance and repair, a prolonged reproductive phase, and a slower pace of senescence, and thus a reduced rate of intrinsic mortality.

Ageing and senescence are not fully interchangeable terms. Ageing is the simple fact of time—getting older. Senescence is a biological process (or processes) in which, following a phase of development and attainment of maximal reproductive potential, there is an age-associated physiological degeneration which adversely affects vitality and function and leads to an increased risk of death. Senescence is a universal phenomenon of sexually reproducing species and is associated with a number of biochemical and cellular changes. Lifespan (the age at death of a population), longevity (the lifespan of an individual), and life expectancy (which depends on
the age from which it is specified) are other related terms with specific meanings.

Different species have very different lifespans, and these are finite even where extrinsic causes are largely absent. It is estimated that the average lifespan of Homo erectus was about 15–20 years, and that of humans in the Paleolithic was about 25 years. Then, humans died primarily in childbirth, in infancy, or from trauma; infection became more important in more recent times following the development of agriculture and population settlement (see Sections 6.3.10 and 9.3.2). While low compared with that in many other species, pre-reproductive deaths would probably have approached 50% in the Paleolithic due to high infant and childhood mortality. It is common to adjust estimates of life expectancy by excluding children under the age of 5, giving an adjusted life expectancy of 35–40 years for those in the Paleolithic who survived infancy. In modern high-income countries, pre-reproductive mortality is now generally very low (1–2%). But sadly even now there are human populations in some of the lowest-income countries where infant and child mortality approaches 30%.

The maximum achievable lifespan is generally inferred from the singular example of the longest-living member of a species or population, and is generally much greater than the average lifespan. The longevity of an individual is affected by its environment: zoo-housed animals live longer than those in the wild, as extrinsic mortality risk is reduced. Similarly, human longevity is also greatly influenced by our cultural evolution, as we have developed technologies to deal with many extrinsic causes of mortality, particularly in recent decades through advances in public health and medicine. Thus, while our rate of senescence has likely not changed over the last few millennia, the average lifespan has increased (Figure 5.3). While there is some uncertainty as to the average lifespan in pre-historical times (Box 5.3), historical data show relatively short average lifespans until the last 200 years. Since then lifespan has risen exponentially, and coincidently we have seen a rise in the prevalence of degenerative disease, in particular the dementias. The person who had the longest verified lifespan was a French woman, Jeanne Calment, who died in 1997 at the age of 122 years; female life expectancy in France at the time of her birth was 46 years and at the time of her death it was 82 years.

5.2.4 Evolutionary Theories of Senescence

Evolutionary theory must address the question of senescence (Rose 1991; Crews 2003). Why do we not

---

**Figure 5.3** Life expectancy at birth in Europe, 1725–1990. Plotted from data in Fogel (2004).
live and reproduce indefinitely? The initial proposition put forward by August Weismann in 1891 was that death was an adaptation for the good of the species because old and potentially decrepit immortals would otherwise take resources from those who are healthy. This was recognized as a flawed argument based on group selection. But nevertheless Weismann did recognize that there are links between the pattern of growth, the pattern of reproduction, and the timing of senescence. From this there gradually emerged a body of knowledge that forms the basis of current understanding.

In 1952 Sir Peter Medawar demonstrated that the force of natural selection decreases with age. This means that if negative effects on fitness only appear late in life there will be less selection against the determinants of the fitness-impairing traits. Selection for life-history trade-offs will favor investment in reproduction soon after reaching puberty because of the cumulative risks of extrinsic mortality, so there will have been selection for those who successfully reproduce early. This can also be proved mathematically, and such calculations show that the point at which the force of natural selection starts to decline in human females is approximately 20 years. Further, women show marked reproductive decline from their early thirties (see Section 8.9.7)—a problem now confronting many women whose lifestyles and choices lead them to delay the decision to reproduce. From this follows the proposition that mutant alleles with late-acting effects, but which had no effect on fitness, can therefore accumulate in the gene pool. This may play a role in diseases which appear primarily in middle and old age.

Pleiotropy is the phenomenon whereby the product of one gene locus can have multiple actions on physiology and phenotype. In 1957 George
Williams, building on the work of Medawar, proposed the concept of antagonistic pleiotropy, whereby genes might have been selected which exhibit beneficial effects in early life but then have detrimental effects later (Williams 1957). He proposed this as the basis for senescence, namely a trade-off has been made for advantage in early life with adverse consequences later. One example may be determinants of testosterone secretion and action. High levels of testosterone promote fitness in males in early life by increasing body mass and promoting aggressive behavior, but such individuals are more prone to prostate cancer and heart disease in later life (see Section 12.5.2). Another example relates to insulin-like growth factor 1 (IGF1) and ageing (Box 5.4). Nutritional factors acting at multiple levels regulate the secretion and activity of IGF1 and IGF1 promotes fetal growth and growth of the skeleton and muscle during childhood and adolescence; high levels would increase fitness in early reproductive life. But high levels of IGF1 in later life are associated with an increased risk of breast or prostate cancer, which are usually problems of middle and older age. Given the higher rates of extrinsic mortality and shorter lifespans in humans until recently, the negative effects in later life of such antagonistic pleiotropic selection would have been largely hidden, further favoring selection for the beneficial early life effects. Another example of these pleiotropic effects is provided by the fecundity-enhancing effect of the breast cancer-associated BRCA1 mutation (see Box 12.3).

While the evolutionary origin of senescence is still an area of debate, the concepts of antagonistic pleiotropy remain the basis of our current understanding (Figure 5.4). A derivative theory is the disposable soma theory (Kirkwood 2008), which proposes that the energy available to an organism may be used either to preserve the soma or to reproduce the germline. Because of the cumulative risks of extrinsic mortality, natural selection favors investment in reproduction over somatic repair once sexual maturation has been achieved. Prior to that, there is a fitness advantage in early investment in maintenance and repair of somatic tissues to promote survival until reproduction. The capacity to maintain the soma is linked in turn to the reproductive strategy of the species and the number of redundant subunits at different levels of organization (e.g., nephrons, neurons, mitochondria) in somatic tissues as well as the ability to repair chromosomal and mitochondrial DNA and cellular and structural proteins in the face of cumulative damage, much of which is mediated by reactive oxygen species. Other sources of damage, often in turn mediated at least in part by reactive oxygen species, include exposure to infection, low-level radiation, toxins, or mild trauma. Hence the disposable soma theory proposed that senescence is a function of an inability to maintain defensive mechanisms and to repair tissues in later life. Such a theory explains why ageing is a period of co-morbidities: it probably reflects the accumulated changes of senescence leading to multisystem disorders which manifest as the limits for maintaining function of each organ or system are reached.

Not all age-related changes are necessarily the result of senescence: they may simply be the stochastic result of living longer, although the distinction may be more theoretical than real. Whereas lowered immune function, atrophy of muscle, and decreased permeability of cell membranes are likely to be manifestations of senescence, the risk of neoplasia may simply rise progressively with age: the longer one lives, the greater the chance of exposure to an environmental mutagen or for a spontaneous error in cell replication to lead to a somatic mutation (Chapter 12).

5.3 Body Size and Shape

There are a number of interrelated determinants of body size and shape, including gravity, the pattern of locomotion, and nutrient sources. Humans are bipedal whereas all other primates are quadrupedal. The size of a terrestrial mammal is reflected in its skeleton, which makes up 18% of the weight of an adult human but less than 10% of the weight of a sparrow or a mouse. Similarly, the shape of trees is affected by their size: larger trees have thicker trunks as trunk strength is related to its cross-sectional area. In addition, the height and shape of trees is affected by competition for sunlight and by the needs of their root systems. Thus the body composition of an organism must change as its size changes. Further, the proportions of the body

5.3 BODY SIZE AND SHAPE 107
Foxo transcription factors may be key in the integration of metabolism, reproduction, and lifespan. FoxO activity is modulated in part by the sirtuins, which are histone deacetylases. Sirtuin activity is positively correlated with lifespan in several species, and sirtuin 1 (SIRT1) is the target of the putative lifespan-extending polyphenol resveratrol: it shifts FoxO-dependent responses away from cell death towards cell survival. CR also induces SIRT1 activity. Studies in populations of Japanese and of German ancestry have shown that genetic variation within the FOXO3 gene is associated with human longevity. Indeed, in the population of Japanese origin, the frequency of one particular allele of the gene was linearly associated with age at death, such that homozygosity for the allele tripled the chance of living until nearly 100 years (Willcox et al. 2008).

Can caloric restriction also extend lifespan in humans? To date this remains an open question, although there are indirect hints that it may be the case. At a clinical level, moderate, non-malnutritive CR has been associated with marked improvements in blood pressure, lipid profile, glucose homeostasis, and other parameters of metabolic and cardiovascular health, suggesting decreased risk of obesity and metabolic diseases that incur major morbidity and mortality costs (Fontana et al. 2004). At the endocrinological level, CR has also consistently reduced circulating levels of growth factors, adipokines, oxidative stress markers, and inflammatory cytokines associated with multiple forms of cancer.

From a mechanistic perspective, CR could induce alterations in the transcriptional profile of skeletal muscle to more closely resemble that of younger individuals, with affected molecular pathways shifting cellular metabolism from growth to maintenance and repair; this pattern of change was also mirrored in corresponding experiments on rats (Mercken et al. 2013). Transcriptional regulation of multiple components of the IGF1/insulin/FoxO pathway was modified, in accord with previous work involving genetic manipulation in other organisms. Although much more remains to be understood about the underlying mechanisms promoting CR-induced longevity in humans, and the inherent difficulties in conducting long-term randomized controlled trials in this area, there is ongoing work tracking the health of individuals who have adopted a CR diet. Comparison of their longevity with that of close relatives may yield additional clues on whether humans, too, can enjoy the benefits of longevity by eating less.

Box 5.4 Eat Less, Live Longer

A common observation in ageing-related research is the lifespan-increasing effect of calorie restriction (CR) in organisms ranging from yeast to worms to rodents. But nutritional cues can have quite distinct effects depending on when in the life cycle they act, and in mammals it is important to distinguish between nutritional cues acting in fetal or early neonatal life and those acting later after weaning. For example, in rodents post-maturational CR increases longevity. In marked contrast, pre-natal undernutrition can lead to a range of adverse consequences in later life, including metabolic dysfunction and shortened lifespan. Similarly, humans who are born smaller have shorter lifespans. Such effects can be placed in the context of the adaptive significance of responses in early life to external cues (see Chapter 4). Limited studies suggest that so-called catch-up growth—that is, accelerated growth after an earlier period of nutritional limitation—appears deleterious, and its prevention by post-natal CR prevents the adverse effects of earlier undernutrition (Dai et al. 2012).

Post-weaning undernutrition in many species prolongs lifespan, although generally this is associated with a reduction in reproductive performance, again highlighting the complexities of the interaction between life-history traits. Although it is tempting to attribute the effects of post-weaning CR to a concurrent reduction in metabolic rate causing a corresponding reduction in the production of tissue-damaging oxygen free radicals, this interpretation appears to be too simplistic. Rather, the complex process of ageing and its modulation by nutrition-related factors may represent the balance of numerous disease-inducing and longevity-promoting factors.

In knockout and transgenic experiments in organisms ranging from flatworms to mice, manipulation of individual genes involved in the regulation of growth, nutrient sensing, and cellular metabolism can have profound effects on lifespan, providing a molecular basis for the longevity-increasing potential of CR. These genes include those coding for proteins involved in the insulin/IGF1 signaling pathway, such as the insulin receptor substrate (IRS) 2, and its downstream targets such as the FoxO transcription factors which regulate gene expression associated with cell growth, survival, and metabolism. For example, knockout of IRS2 increases lifespan and alters nutrient partitioning (Taguchi et al. 2007). FoxO transcription factors may be key in the integration of metabolism, reproduction, and lifespan.
change as an animal grows. Figure 5.5 shows how the relative size of the head changes during growth, being 33% of body length at birth but only 13% in an adult.

Larger objects have a greater surface area, but this does not necessarily increase to the same degree as volume. In a spherical object, surface area is a function of volume$^{2/3}$, and thus complex shapes are needed to have sufficient surface area for diffusion in larger organisms. Placental and intestinal villi, the tortuosity of the gastrointestinal tract, and the alveolar structure of the lungs are all similar ways of increasing the area for diffusion within a constrained space. Similarly, the cortical sulci and gyri of the primate brain are a solution to generating a sufficiently large cortical surface without requiring a gravitationally impossible head size.

There are often potential adaptive advantages in having a larger body size, such as providing defense against predators. But there are important constraints including gravity, cardiovascular pump capacity, limb size, diffusion capacity, and nutrient availability. In a stable environment there will inevitably be species of many different sizes, reflecting the multiple ecological niches available. Because there is always a potential niche for a larger organism that can gain advantage through predation of smaller animals or by being more effective at capturing resources (e.g., by being the highest tree in a forest), there is an impression that over evolutionary
time organisms have become larger. However, this is not directional evolution; it is simply a reflection of the selective environment, and increasing size is not inevitable. In our hypothetical lightning-plagued island, we described how a change in extrinsic mortality may well cause a lineage to evolve to a smaller size. Indeed “island dwarfism” is well described for several mammalian species, including elephants, where a subpopulation has become isolated on a resource-limited island. It has also been evoked as one possible, but controversial, explanation for the recent discovery of skeletal remains of a small hominin on the island of Flores in Indonesia (see Box 6.2).

5.3.1 Allometry

As mentioned at the beginning of this chapter, many functions and components of the body are not simply a linear function of body size, and *allometry* is the study of these relationships. For example, the overall weight of an animal is generally a function of the cube of its linear dimensions. However, the strength of bones rises in proportion to the square of their linear dimensions, effectively putting a limit on body size for a land-based animal. The supportive action of water allows animals to grow larger in aquatic environments—the blue whale being the largest mammal. These mathematically regular relationships between components of body size and functions across species are known as allometric, and result from generalizable trade-offs between life-history traits. A deviation from an expected allometric relationship suggests natural selection for a particular trait. The most common application of allometric analysis is in the examination of relationships between organ size and body size. A comparative evaluation of brain size in primates and hominins provides an extensive illustration of how allometric analysis can be used (Box 5.5).

5.3.2 Variation in Growth and Development

There is considerable variation in human size and shape and in the tempo of maturation both within and across populations. Some of this will represent the influences of the developing and concurrent environments. But there are also genetic variations in the patterns of growth and maturation apparent across populations (e.g., the Watusi are very tall compared with many other African populations), and this may represent further microevolution.

Figure 5.5 Relative head size during human growth. Adapted from Wells et al. (1931).
The size and complexity of the human brain allows a sophistication of human social and cultural achievement. We rely upon cognitive ability where other species rely upon traits like brute force, speed, or camouflage. But is the human brain really that special? We think so, but how do we go about evaluating this scientifically? As there are limits on assessing functionality in other species, most comparative studies have focused on brain size, despite its limitations. Figure 5.6(a) shows the size of a human brain relative to that of other apes, including chimpanzees, gorilla, orangutan, and the lesser apes, siamangs and gibbons. Also plotted in the same chart is the brain size of a 5000-kg elephant. Looking at raw data is deceptive, because elephants should have a large brain given their far greater overall body size.

Figure 5.6 Allometric relationships in brain size. Primate brain weights in comparison with that of an elephant (a). The data are also plotted on linear (b) and logarithmic (c) scales. Plotted using data from Harvey et al. (1987). Figure courtesy of Dr. Chris Kuzawa, Northwestern University.
And the same is true for the other comparisons, even though the differences in body size are more subtle. What is needed is a comparison of a species’ brain size relative to its body size. We could simply calculate the ratio of brain to body size, but this is not ideal. If the ratio of brain to body size is not constant as we move from small to larger species, this might make some species appear relatively “smarter” when in fact they are merely small or large. The allometric approach is a tool to disentangle such relationships.

Figure 5.6(b) shows the same brain-size data along with data from a much larger set of primates, but this time plotted in relation to body size. The broken line best describes the relationship between brain and body size in the primates included in this data set (the elephant was left out of this calculation). Several features of this figure stand out. First, the line is not straight: it has a slight curve to it, and is best described by a curvilinear power function of body size. Note that the slope of the line diminishes as one moves to larger-sized species. Thus, brain size increases as one moves from small to larger-bodied species but at a slower rate than body size itself; that is, the exponent for the power function between the variables is less than 1.0. Perhaps certain efficiencies are gained in the architecture of the brain as body size increases. The second feature of the chart that stands out is that it is very difficult to read! The elephant is so much larger than the remaining species in the data set that both the y and x axes extend to much larger sizes than any of the primates achieve. In fact, although it is not apparent, there are about 150 primate data points compressed into the very tip of the graph! This is a common problem when performing comparative analyses of this sort: there are generally very many small species and relatively few large species, and the graphs are quite challenging to interpret as a result. The solution is to plot the data points—brain and body weights—as log-transformed values. This stretches out the smaller values while it crunches in the very large ones.

The effect of this is seen in Figure 5.6(c). The data used are identical, as is the best-fitting equation describing their relationship. All that has changed are the units on the graphs and thus the visual representation. Not only are the data easier to see and interpret, but when plotted on log-transformed units, the power function is linear. This assists the interpretation, and we can finally evaluate the question that we set out to address: do humans really have larger brains than expected? Human adults have an average body weight of 60 kg. Plugging this into the formula reveals that a typical primate of human adult body weight should have a brain around 500 g, when our brains in fact weigh 1250 g. Thus, humans have a brain size about 2.5 times what would be expected for other primates if they reached an equivalent body size to a human.

Within particular environments or may simply reflect random drift. For example, adaptive arguments have been used to explain the short stature and absence of a pubertal growth spurt in pygmies (Box 5.6). Biogeographical perspectives have been applied to explain some variations in the developing and adult phenotype, and there are several rules that have been used to describe these. Bergmann’s rule is the observation that, in homeothermic species,
subpopulations in colder climates are generally larger than those in warm climates: a larger mass has relatively less surface area in relation to volume than a smaller one (assuming that they are the same shape) and thus loses less heat from its surface. Allen’s rule posits that humans in colder climates were selected for shorter limbs relative to the trunk to reduce the surface area radiating heat. Thomson’s rule argues that the nose becomes narrower and more protuberant as the climate becomes colder and drier: it was suggested that this allows the nose to warm and humidify the air before it reaches the lungs. All these are simple and rather sweeping generalizations, but they do highlight the role that microevolution in different geographical contexts may have played in defining variation in human form.

5.4 Growth in Humans

5.4.1 Phases of Growth

Most primates have relatively few and generally singleton offspring which require a long period of post-natal suckling and parental care and protection. All the non-human species of the Hominidae (orang-utan, gorilla, bonobo, and chimpanzee) have a comparatively late puberty and a potential lifespan of several decades. Humans fit this general pattern but are further distinguished by their disproportionately large brains (Box 5.5). Together with our upright posture and associated changes in pelvic dimensions, this means that the human infant has evolved to be relatively more immature neurologically at birth than other primates (see Chapter 8) and a greater proportion of brain development occurs after birth. The life history of humans is thus both described and further defined by these factors.

There are distinct phases to the human life cycle which can be defined by our distinct pattern of growth, our maturational state at birth, our age of reproductive competence, and our energetic sources and needs at different stages through life (Table 5.1). Prior to maturity, the requirements for nutrients can be divided into those required for maintenance and those required for growth. The energetic needs of the developing brain are particularly important in infancy and childhood. The newborn infant uses about 85% of its resting metabolic rate for brain growth and function and this is still 45% in the 5-year-old, falling to about 20% in the

Box 5.6 Why are Pygmies Short?

Several human populations are characterized by small body size—anthropologists generally define pygmies as having average adult male height of less than 155 cm. The word pygmy brings to mind the hunter-gatherers of the Central African tropical forests, but similar groups are found in Southeast Asia and South America. Various adaptive hypotheses have been proposed to explain the small stature of pygmies, and the selective pressures invoked have included difficulty of movement through a dense forest environment, the greater need for heat loss in the tropics, vitamin D insufficiency because of reduced sunlight exposure, or nutritional stress. However, many populations facing similar environmental challenges have not evolved small stature, and the reason why pygmies have done so has remained unclear.

Analysis of the life histories of women in several pygmy populations provides one clue (Migliano et al. 2007). Recall the discussion earlier in this chapter about why organisms trade off age at maturity versus size at maturity, and how this trade-off is affected by levels of extrinsic mortality. Comparison of growth curves reveals that pygmies grow at the same rate as other nutritionally limited populations in childhood, but cease growth at a much earlier age—about 12 years old—and have virtually no pubertal growth spurt. This premature cessation of growth is accompanied by earlier age at first reproduction and earlier peak fertility. This indicates that pygmies employ a “fast” life-history strategy characterized by early cessation of growth, early maturation, and early reproduction, and in turn predicts that they will experience high early mortality. Indeed, life expectancy at birth in pygmies is very low: an average of 19 years has been reported in the populations studied. The growth cessation in pygmies occurs well after human brain growth is completed (at about 7 years; Section 5.4.1) and therefore their brain size is unaffected by their small stature. In this respect, pygmies are distinct from the small-brained fossils of the putative dwarfed Homo floresiensis (see Box 6.2).
The energetic dependence plays a significant role in explaining the pattern of human growth, as illustrated by the height-velocity curve (Figure 5.8), has several distinct characteristics.

Growth is rapid in fetal life and in infancy but slows rapidly in the second and third years (Figure 5.8). This infant phase is associated with maternal milk as the primary source of nutrition. Lactation in hunter-gatherers lasts for around 24–36 months, although from about 6 months the energy available from milk alone is not adequate and additional energy from supplementary feeding is required for optimal growth and brain development (Figure 5.9). Our immature dentition and small gastrointestinal system require that children receive a diet that is readily chewable, digestible, and low in volume, yet dense in energy, amino acids, and proteins. It is also a period in which the deciduous teeth erupt and there is rapid locomotor and cognitive development.

Following this phase, childhood is a period of relatively constant growth until about 7 years of age. It is associated with a continued need for some adult assistance for feeding, and with the eruption of the first permanent teeth and the completion of brain growth, which in terms of mass occurs by 7 years (although changes in neuronal number, axonal organization, and synaptic density continue until well after puberty). The continued dependence on adult assistance during this phase is reflected in the general observation that orphaned “street” children cannot survive unaided and alone until about 7 years of age.

At about this age, the onset of adrenal androgen production, known as adrenarche, occurs. There may be a small transient alteration in growth rate associated with this steroid secretion before it again declines until the start of the pubertal growth spurt which starts earlier in females than in males.

Juveniles can be defined as pre-pubertal individuals who are no longer necessarily dependent on their mothers for survival. During this juvenile

<table>
<thead>
<tr>
<th>Stage</th>
<th>Duration</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>First trimester of pregnancy</td>
<td>Fertilization to week 12</td>
<td>Embryogenesis</td>
</tr>
<tr>
<td>Second trimester of pregnancy</td>
<td>Months 4–6</td>
<td>Rapid growth in length</td>
</tr>
<tr>
<td>Third trimester of pregnancy</td>
<td>Month 7 to Birth</td>
<td>Rapid growth in weight and organ maturation</td>
</tr>
<tr>
<td>Neonatal period</td>
<td>Birth to 28 days</td>
<td>Extraterine adaptation, most rapid rate of post-natal growth and maturation</td>
</tr>
<tr>
<td>Infancy</td>
<td>Month 2 to end of lactation (usually by 36 months)</td>
<td>Rapid growth velocity with steep deceleration in velocity with time, feeding by lactation, deciduous tooth eruption, many developmental milestones in physiology, behavior, and cognition</td>
</tr>
<tr>
<td>Childhood</td>
<td>3–7 years</td>
<td>Moderate growth rate, dependency for feeding, mid-growth spurt, eruption of first permanent molar and incisor, cessation of brain growth by end of stage</td>
</tr>
<tr>
<td>Juvenile</td>
<td>7–10 (girls) or 7–12 (boys) years</td>
<td>Slower growth rate, capable of self-feeding, cognitive transition leading to learning of economic and social skills</td>
</tr>
<tr>
<td>Puberty</td>
<td>At end of juvenile stage</td>
<td>Reactivation of central nervous system mechanism for sexual development, drastic increase in sex hormone secretion</td>
</tr>
<tr>
<td>Adolescence</td>
<td>The 5–8 years following onset of puberty</td>
<td>Adolescent growth spurt in height and weight, permanent tooth eruption virtually complete, development of secondary sexual characteristics, socio-sexual maturation, intensification of interest and practice in adult social, economic, and sexual activities</td>
</tr>
<tr>
<td>Adulthood</td>
<td>From 20 years to end of child-bearing</td>
<td>Homeostasis in physiology, behavior, and cognition, menopause for women by age 50</td>
</tr>
<tr>
<td>Old age and senescence</td>
<td>From end of child-bearing years to death</td>
<td>Decline in the function of many body tissues or systems</td>
</tr>
</tbody>
</table>
phase, individuals are capable of feeding themselves and are in a period of cognitive transition, allowing the learning of a number of useful skills including social and economic skills.

Biological maturation at puberty is described in Section 5.4.2. Adolescence can be defined as the period between the onset of biological maturation and full acceptance as an adult. The obvious features of this phase are the pubertal growth spurt, the development of sexual characteristics, the maturation of reproductive competence, and changes in cognitive, psychosocial, emotional, and social skills.

Figure 5.7 Proportion of the resting metabolic rate (RMR) allocated to the brain during human growth and development. Brain metabolism represents more than 60% of RMR in infancy, and about 20% in adulthood. From Leonard et al. (2003), with permission.

**Figure 5.8** Human height-velocity curve (growth per year) for boys (blue line) and girls (black line). Note the growth spurts at around 7 years and at puberty (10–15 years), and the earlier pubertal growth spurt in girls. Adapted from Bogin (2001), with permission.
and skills leading to the individual being able to make an economic contribution to the community.

5.4.2 Puberty

Puberty is the physical manifestation of adolescence and is characterized by a transient acceleration in growth for 2–3 years followed by a deceleration which ultimately leads to termination of linear growth with epiphyseal fusion of the long bones. The pubertal growth spurt adds about 25 cm in height in females and 27 cm in males. As will be discussed later, the pubertal growth spurt is a unique characteristic of *H. sapiens* not seen in other primates: it is thought to have evolved late in the hominin lineage. There are also changes in body proportions as the different components of the skeleton do not all grow synchronously. Thus the span to height ratio changes, with the limbs getting longer relative to the trunk. The relationship between the pubertal growth spurt and reproductive maturation differs between the two sexes (Figure 5.10). The initiation of spermatogenesis in males occurs relatively early in puberty at a stage when development of the genitalia and pubic hair is only partially progressed and well before peak height velocity is achieved. In contrast, menarche occurs late in the progression of pubertal development in the female, when development of the breasts and pubic hair is almost complete and well after pubertal growth has peaked. There is no event in males equivalent to menarche that allows their sexual maturation to be readily assessed; thus most of the discussion will focus on the female, where clearer statements can be made.

Body composition also changes. Whereas females continuously accumulate fat through puberty and adolescence, males have a greater increase in muscle mass and no overall increase in fat mass (although this is now changing as children have an increased rate of obesity). The possible selective advantage of these different patterns of body composition between sexes could be explained in terms of the role of muscle mass in generating fitness in males (through its potential to affect competition for mates) and of energy stores in doing so in females (i.e., supporting several pregnancies and lactation periods).

5.4.2.1 Endocrine Control of Puberty

Soon after birth, the fetal activation of the sex hormone system which is necessary for pre-natal sexual differentiation in the male (but is also active
in the female fetus and infant) is suppressed by hypothalamic mechanisms. The precise nature of this “gonadostat” is still uncertain but the hypothalamic–pituitary–gonadal axis then stays quiescent until puberty. In adrenarche, a poorly understood process that precedes puberty, the adrenal cortex markedly increases its production of adrenal androgens, starting in mid to late childhood (6–8 years). Adrenarche is generally covert but it can induce early and slight production of pubic hair in both sexes. Onset of adrenarche is confined to humans and higher-order primates, suggesting that it has evolved relatively recently. Humans are distinct in having a pubertal growth spurt; in most other primates growth is largely complete by the age of sexual maturation. Thus adrenarche may in some way be an evolutionary echo of different controls on growth and maturation which operate across species. Still, the evolutionary significance of adrenarche, and indeed the mechanistic factors

Figure 5.10 The relationship between the pubertal growth spurt and timing of events of sexual maturation in girls (top) and boys (bottom). Note that menarche follows peak height velocity (PHV) in girls, whereas sperm production precedes PHV in boys. CNS puberty refers to maturation-related events in the hypothalamus and other brain regions. Modified from Bogin (2001), with permission.
underlying its regulation, remains unclear, although recent clinical data suggest an important regulatory role of the adrenocorticotropic hormone, produced by the pituitary gland (Boettcher et al. 2013).

At puberty, the gonadostat mechanism is reversed and release of gonadotropin is reactivated. This leads to the Leydig cells of the testis secreting testosterone and to the Sertoli cells supporting active spermatogenesis. In the female, ovarian follicles reactivate and start to secrete estrogen and progesterone. Eventually this develops into a cyclical pattern due to changes in the feedback mechanism of estradiol on the release of luteinizing hormone. Estradiol generally acts as a negative-feedback controller of release of luteinizing hormone via its actions on hypothalamic control of the secretion of gonadotropin-releasing hormone, but in mid cycle it starts to exert positive feedback, leading to the pre-ovulatory surge of luteinizing hormone and thus to ovulation. Follicular development recommences and the cyclical process, in which generally one oocyte fully matures and ovulates in each cycle, is established.

Testosterone and estrogens determine the development of secondary sexual characteristics. In males this causes pubic, facial, and axillary hair to develop, the larynx to change shape and shift its position downwards leading to the voice deepening, and skeletal and muscle growth. The phallus grows and the scrotum expands to accommodate the growth of the testis associated with the onset of spermatogenesis. In females, development of the breasts and pubic hair are the most obvious secondary sexual characteristics. There are sex differences in both the age at onset of puberty, which is earlier in the female, and the time at which reproductive competence occurs during the pubertal process. The entire process of pubertal maturation takes several years. The first sign of puberty in girls is breast development, and in boys testicular enlargement. Progress to completion of puberty takes 3–5 years.

In humans it is not until the end of puberty that the epiphyseal growth plates of the long bones fuse and linear growth is completed. Boys can produce sperm relatively early in puberty: much of their pubertal growth spurt thus follows development of their biological capacity to reproduce. Girls, on the other hand, stop growing soon after their first menstruation but fertility may take up to 2 years after menarche to be fully established. This can be interpreted as a result of evolutionary processes, because females who had not reached pelvic dimensions large enough to allow a successful birth would have failed to reproduce successfully. Indeed, the pelvis reaches maximal size only about 4 years after menarche, and is the last skeletal structure to do so. Many of the initial cycles of a newly matured female are associated with failure of ovulation of a mature egg; monthly periods may be irregular for a year or so, and fertility is low for about 1–2 years before full reproductive competence is reached. Further, while the potential mother is young there is a competition between her and her fetus for nutrients as she is still laying down soft tissue for some years after menarche. Thus infants born to young mothers are smaller not only because they are first-born but also because of nutritional limitations. The constraining role of pelvic size in influencing the evolution of the human life history is discussed further in Box 9.11.

5.4.2.2 The Timing of Puberty

The age at onset of puberty varies quite considerably within and across populations. The average age at onset of puberty is earlier in the female than in the male, and, as this subsection will demonstrate, a complex hierarchy of life-history controls operates to influence the age at pubertal onset and sexual maturation.

One factor is the nutritional state of the girl both in utero and in childhood (Gluckman and Hanson 2006a; Sloboda et al. 2007). Some years ago it was suggested that puberty was triggered in girls when they reached a critical percentage of body fat. This so-called Frisch hypothesis did not stand up to critical analysis, but it is true that in general better weight gain in childhood is associated with an earlier menarche. The precise mechanisms involved are unclear, but the neuroendocrine control of gonadotropin release interconnects with the hypothalamic regulation of satiety and metabolism. Conversely, childhood energy deficits can delay puberty and the onset of menses: this is not uncommonly seen in children who exercise heavily, such as competition gymnasts or ballet dancers, and is a feature of anorexia nervosa. This relationship can be interpreted
as an override: irrespective of other drivers of maturation, there is an energetic and fitness logic to delaying puberty when nutrients are scarce and energy expenditure levels are high, and permitting an earlier puberty when nutrients are in abundance and require little energy expenditure to obtain.

As historically pregnancy generally followed soon after maturation, the variable age at menarche can be interpreted in relation to the energetic costs and fitness effects of pregnancy at a young age. A pregnancy in conditions of famine is more likely to be associated with a poor outcome, and will have energetic costs for the mother that may compromise her own survival and future fitness; thus, to delay maturation in the expectation of a potentially better nutritional future is adaptive. Conversely, to be well nourished may allow an earlier puberty, with a longer potential reproductive life and enhanced fitness.

The age at sexual maturation is a central feature of a species’ successful life history. Therefore, controls on its timing in the female probably evolved to match the age of reproductive competence to the capacity of the female to support her offspring. In turn this must have roughly matched the physiological and psychosocial demands of being an adult and the social ecology of our ancestral lineage (Gluckman and Hanson 2006a). Hence it is to be expected that puberty will be linked to both energetic and social conditions, and may be influenced by the extrinsic risks of mortality. Reproduction is energetically costly and is a major risk to the female, and hence to her previous but still-dependent offspring. All else being equal, evolutionary processes will favor earlier maturity because it reduces the cumulative risks of extrinsic mortality on fitness, and because it increases the potential for live offspring by increasing the reproductive lifespan. But from the life-history perspective, fitness would be compromised if a strategy evolved in which reproductive competence preceded the capacity to function adequately as a mother. Thus, the age at menarche must have related to the age at which a degree of psychosocial maturity and adult competencies is reached; we shall further explore this concept later (Figure 5.13). A similar argument applies to explaining why the presence of the mother’s mother to assist with childrearing enhances childhood survival, and this is the basis of the “grandmother hypothesis” to explain the evolution of the menopause (see Section 8.9.7).

Certain ecological conditions may have led to the selection of mechanisms which influence pubertal timing and in turn link it to environmental conditions; those conditions may be very different from the ones now observed in modern societies. In other words, the mechanisms controlling the timing of puberty evolved to adjust puberty within an anticipated range of environmental inputs, but the very novel environments of modern society may be having maladaptive effects even though they operate through mechanisms that evolved for potential adaptive advantage. It is not known when menarche occurred in the Paleolithic; most have assumed at relatively late ages while others have posited that it may have occurred at a relatively young age. Indeed there are modern forager groups who have menarche at an average age of 12–14 years (Walker et al. 2006). Larger apes progress through puberty between 6 and 9 years and the major auxological difference between apes and humans is the 3–4-year delay in maturation induced by the additional childhood growth phase (Section 5.5.1).

The controls on the age of pubertal maturation are clearly multigenic, and there is clear evidence of plasticity in relation to this age, depending on ecological circumstances. Indeed there are marked differences across populations in this age. Monozygotic twins reared together have a minor difference in age at menarche, whereas the difference in those reared apart is almost as large as that in dizygotic twins (Segal and Stohs 2007). While twin studies in general have problems of interpretation, such studies imply a considerable environmental effect as well as inherited factors (the correlation in age at menarche between mother and daughter ranges from 0.15 to 0.4).

Nutritional and psychosocial influences in both early development and childhood have significant effects on the timing of reproductive competence. However, there are complexities in the relationships depending on the age at which cues are experienced. Pre-natal factors may accelerate maturation through effects which shift the overall life-history strategy in anticipation of adverse post-natal circumstances. Yet, childhood influences tend to operate in the
opposite direction: a delay in maturation when environmental circumstances are poor in the immediate peri-pubertal period may be a safety override in a long-living species which can delay maturation until conditions improve without putting maternal health at risk. Pre-natal adjustments are likely to reflect the state of the environment over a longer time frame, whereas childhood influences can respond to the immediate energetic realities.

Females who are born small or are exposed to intrauterine or infant stress advance their age at menarche by up to 18 months (Sloboda et al. 2007). Evolutionary arguments would suggest this to be the appropriate response for an individual anticipating a challenging environment. There is commonly a trade-off between age at maturation and investment in growth and longevity. If the fetus interprets the signals from the mother as representing a threatening future, then based on the principles of developmental plasticity it becomes an appropriate response to accelerate the tempo of maturation in expectation of a shorter life expectancy. Fitness is maintained in threatening circumstances by earlier maturation.

However, the effect of low birthweight on the advancement of menarche is modified by the effect of post-natal nutrition. If the child is well nourished, then maturation is accelerated and the pre-natal effect is fully demonstrated. On the other hand, if the child is undernourished the pre-natal influence can be overridden, as fitness would be compromised by her undergoing menarche and potential reproduction if she could not energetically support her fetus through pregnancy and lactation. The earliest menarche is therefore seen in girls who were born small and then become relatively fat as children (Figure 5.11). This may be the explanation of the marked acceleration in the age at puberty seen in girls adopted as infants from refugee camps into higher-social-class homes (Box 5.7). Indeed, there is experimental evidence to suggest that animals born small have a changed regulation of appetite and metabolism so as to gain fat before puberty to support earlier reproduction, and the human data also show that those born smaller tend to gain fat in childhood relative to those who were born of normal size. Thus metabolic adaptation in such circumstances may serve to support the strategy of accelerated sexual maturation.

![Figure 5.11](image.png)

**Figure 5.11** Complex interplay of fetal and childhood nutrition in determining the age at menarche in Australian girls. Girls who were born small but were overly nourished in childhood were most likely to experience menarche at an earlier age. In contrast, those born larger but who did not gain excess weight in childhood had a lower chance of menarche at an early age. Data plotted from Sloboda et al. (2007).
There is good evidence that poor weight gain in childhood is associated with delayed menarche and more rapid weight gain with earlier puberty. Indeed there is increasing evidence of an interaction between the neuroendocrine systems controlling gonadotropin release and those controlling metabolic regulation. Childhood psychological stress may also affect puberty, with severe family stress appearing to accelerate puberty (Ellis and Essex 2007). Several evolutionary psychosocial arguments have been advanced to explain these observations. It is suggested that girls who experience negative and coercive family relationships accelerate maturation to independence. Similarly girls from homes with an absent father accelerate their maturation. In contrast, a warm, nurturing environment is one in which there may be a fitness advantage in remaining immature for longer to reach a larger adult size with time to accumulate more life skills. Neuroendocrine interplay between systems controlling mood (influencing the release of glucocorticoids) and those controlling gonadotropin release provides a mechanistic basis for these effects.

In Europe during the late eighteenth century the age at menarche ranged from 16 to 18 years, and it has fallen dramatically in the last 200 years (Figure 5.12). This so-called “secular trend” is a secondary consequence of improvements in maternal and child health, such as better nutrition and reduced infection. The decrease seems paradoxical, given that the concomitant fall in extrinsic mortality would be expected to ameliorate or reverse pubertal acceleration, since a better prognosis for survival should obviate the need for early reproduction. The resolution of this paradox lies in understanding that the age at puberty is the outcome of a hierarchy of influences and controls, depending on whether they operate across the population and over generations or whether they are acting within the life course. But fitness has to be sustained in the face of variable environments, and this adds another layer of control.

The secular trend in falling age at menarche can therefore be understood in terms of recent ecological history. It is now recognized that agriculture and settlement increased the risk of undernutrition and disease. The development of settlement is associated with a reduction in skeletal size and increased disease (as evidenced by skeletal remains). Figure 5.13 shows the hypothetical relationship for Europeans between age at biological maturation (menarche) and at psychosocial maturation.

The assumption is made that in the Paleolithic menarche and psychosocial maturation would have been roughly concurrent. This was probably necessary for functioning as an adult in Paleolithic society, which was based on small groups of hunter-gatherers. But with settlement, disease and post-natal undernutrition became more likely and thus the mean age at menarche was delayed. Nonetheless, this matched the increasing complexity of society following the invention of agriculture which led to settlement and aggregation of populations. This in turn led to the differentiation of
Figure 5.12  Secular trend in age at menarche in several European countries and the USA. Plotted from data in INSERM (2007), de Muinck Keizer-Schrama and Mul (2001), and Morris et al. (2011).

Figure 5.13  Hypothetical relationship for Europeans between the age of biological maturation (menarche) and the age of psychosocial maturation. Adapted from Gluckman and Hanson (2006), with permission.
tasks within the population and the development of social hierarchies. As hygiene deteriorated with increasing population density in Europe, the age at menarche was delayed still further.

But since the Industrial Revolution and improvements in public health in the nineteenth century, these constraints on puberty have been removed and the age at menarche has again fallen to its evolutionarily determined range. However, the complexity of society has continued to increase and it appears to take longer to be accepted as an adult. This creates an evolutionarily novel life phase of a prolonged period of biological maturation without full psychosocial maturation. There is evidence that adolescents with the earliest onset of puberty have the highest rates of psychological disturbance, and even attempted suicide (Michaud et al. 2006), pointing to the costs of this greater discrepancy in these two components of maturation (see Sections 7.4.2 and 11.4).

### 5.4.3 Final Height

Under limiting conditions, growth must be sacrificed to protect the maintenance of critical body functions, and in particular brain and cardiac function. The result is a reduction in growth rate, and if the nutritional limitations persist for a long period then adult height is affected. Stunting remains a major concern in the developing world (Figure 5.14). It reduces productivity both through reduced physical capacity and also, because it is an index of infant and childhood nutritional limitations, through the associated effects on schooling, cognitive ability, and educational achievement.

A well-recognized feature of migration from poor circumstances to nutritionally enriched circumstances is an intergenerational increase in height, which can be quite substantial (Figure 5.15). There is a secular trend in adult height in most populations, representing an improvement in child health. Both linear growth and the incidence of stunting are used in comparing public health status across populations.

While this discussion has focused on linear growth, it is important to note that growth is but one element of development described by these phases (Table 5.1). There are parallel changes in many aspects of behavior and physical and cognitive abilities.

---

**Figure 5.14** Global prevalence of stunting among pre-school children using latest country estimates up to 2010. The numbers of stunted children in low- to middle-income countries has been estimated at 167.2 million, compared with 4.2 million in high-income countries. Adapted from de Onis et al. (2012), with permission.
5.5 Evolutionary Analysis of the Distinct Features of Human Growth

When phenotypic change evolves in a lineage it may be that a new structure or control is introduced by way of mutations which lead to a different characteristic, or it may be that during development the timing of the appearance of a feature is altered in relation to others (a process known as heterochrony), either because of the slowing down of the rate of development (neoteny) or because there may be a prolongation of time for development of a feature leading to its relative magnification (hypermorphosis).

Neoteny leads to the retention of juvenile characteristics, and Stephen Jay Gould suggested that humans evolved by neoteny: that is, they retain childlike characteristics of our ancestor species into maturity (Gould 1977). Features that might be explained by this process would include small jaws and a relatively flat face. But there are strong arguments against this explanation. Humans do not grow particularly slowly compared with other primates, but rather humans appear to have evolved by extending each developmental phase (i.e., hypermorphosis) relative to their ancestor species and other primates. This can be seen with respect to both the pattern of growth and cognitive and behavioral development.

These models, while influential, are oversimplified in seeking a unitary explanation for the evolution of the distinct pattern of human growth and development. Careful examination of the patterns of human growth and their relationship to other species of extant primate and extinct hominin species shows that there is an extraordinary complexity, with some features being magnified, others being reduced, and new features appearing (e.g., the pattern of sulci and gyri in the human cerebral cortex is very different from that in other primates). This demonstrates that the pattern of human evolution cannot be explained by any single process, or indeed by any simple theory, given the multiplicity of extrinsic influences and consequent trade-offs between traits that would have occurred over our evolutionary history.

![Growth curves for Mayan children living in rural Guatemala and for those living in the USA after migration of their parents, compared with US standard growth curves (labeled NHANES). The Mayan children living in the USA are over 10 cm taller than their genetic counterparts living in Guatemala. Adapted from Smith et al. (2003), with permission.](image-url)
Human development involves the appearance of new characteristics such as the distinctive childhood phase, the pubertal growth spurt, and the menopause, and we will discuss possible evolutionary explanations for the appearance of these characteristics.

5.5.1 The Childhood Phase

In many mammalian species the period between weaning and sexual maturation is short relative to the lifespan. For example, in female rats the time between weaning and sexual maturation is about 20 days in a lifespan of about 500 days (an interval of less than 5% of the lifespan). In contrast, most primates and many other social mammals (e.g., cheetahs, lions, elephants, and whales) have a rather prolonged period between weaning and sexual maturation. In humans the weaning to prepubertal phase may have been as much as 20–30% of the lifespan in the Paleolithic, and even now it is about 12% of the lifespan in high-income countries.

Any evolutionary explanation for this prolonged juvenile period must explain how advantage is achieved in delaying reproduction when there is a potential fitness cost arising from pre-reproductive mortality, given that this is a particularly vulnerable age and death during this period means that individuals never contribute to the next generation. Further, a prolonged childhood phase may interfere with the capacity of a mother to support another pregnancy or infant as older offspring will compete for food or attention. Thus there must be a considerable fitness advantage created by prolonging the pre-pubertal period.

Several possible explanations have been put forward. First, a long phase may be needed to allow a female child to reach a size sufficient to be able later to support a pregnancy with a large fetal mass. Second, it may allow time for the young offspring to learn how to live within the social group and to learn complex skills such as foraging and hunting, and how to survive in variable environments. A third explanation is that social groups form where the risk of predation is high, but this creates intra-group competition for food. Newly weaned animals may be at a particular disadvantage in this competition as they have less developed foraging skills. One way to overcome this is to grow more slowly and have lower demands: modeling work suggests that the closer a juvenile’s metabolic rate is to the mean energy supply rate in its home range, the more likely it is to starve (Janson and van Schaik 1993).

The human biologist Barry Bogin has synthesized data from extant humans and chimpanzees and extrapolated from skeletal and dentition data available for early hominin species to suggest that the childhood phase of growth and development, which is not present in the chimpanzee, emerged in early Homo species (Bogin 1999). He has argued that this is likely to have been a necessary accompaniment of larger brain growth (Figure 5.16). This is a period in which there is a continued need for parental support for nutrition even though the offspring are weaned (weaning is defined as when the mother’s lactation stops, not when complementary feeding starts). In industrialized societies the age of weaning has been reduced, but it is probable that humans evolved with a pattern of being breastfed until about 36 months, which is typical of
many foraging societies. Even at this age, humans are weaned while they still have deciduous teeth. This means that during the following childhood phase the child must be fed a different diet from adults, and thus human children are dependent on adults for food preparation and assistance (see also Chapter 6). Thus humans evolved to have a weaning strategy which enhanced maternal fitness by allowing her to be pregnant again, but this required that adults provided prolonged post-weaning nutritional support to the offspring. The special infant/childhood diet is of further value in meeting the large energetic demands of brain growth which, given the relatively short gut of humans, requires a predominance of energy-dense foods. The short gut itself may be a trade-off against a large brain, as both the gut and brain are energetically demanding. The cultural evolution of practices such as cooking improved the capacity to provide children with digestible high-energy foods to support brain growth. Thus the pattern of maternal care, growth, nutrition, and brain development evolved in an interdependent manner. This prolonged pre-pubertal period allows a greater time for brain growth, acquisition of technical skills, and socialization. The net outcome is that the mother is able to reproduce again, having weaned early, but is able to continue to protect and care for the weaned offspring. Additionally, others in the social group can increasingly assist and there is a longer period for the child to learn the survival skills they will need as an adult. Thus a fitness advantage is conferred on both the mother and her offspring.

5.5.2 Pubertal Growth Spurt

The presence of a linear growth spurt during puberty is unique to Homo sapiens and some ancestral hominins. A change in growth velocity during sexual maturation is not observed in any other primate, or indeed any other species, although in some species such as the gorilla the males may show a pubertal spurt in weight gain due to accumulation of muscle mass. Several possible evolutionary explanations can be offered for why the pubertal growth spurt evolved. One hypothesis is that a large component of growth is delayed until after brain growth is mostly complete (about 90% of the adult brain volume is reached by 6 years of age; Stiles and Jernigan 2010), that is, a trade-off to protect the high energetic demands of the brain. A related argument is that a delayed pattern of growth maintains the individual as a less competitive phenotype within the social group while it is learning the skills of language and socialization required to survive in the group as an adult. This may have been particularly important in males because, as we will discuss, this cultural practice may have evolved as a way of reducing sexual competition from younger males. A further argument is that there is survival advantage in being taller, and achieving this requires a delay in growth because of the energetic demands of the growing brain.

Yet another possibility is that there was selection for females who grew during adolescence and gave birth to larger babies who were more able to survive: given the centrality of brain growth in hominin evolution this may have been accelerated by selection for the impact of larger brain size on fitness. A relevant observation may be the evidence from prehistoric Native American populations that the estimated age of death is correlated with pelvic inlet size (Tague 1994), suggesting that complications of pregnancy were a leading cause of death for prehistoric females (just as they still are today in parts of the developing world), and this would have exerted a selection pressure for larger pelvic inlets and tall stature.

An alternative and not mutually exclusive set of speculative arguments is based on sexual selection. Perhaps tall stature was a surrogate for the capacity to be a “good” mother and/or a dominant male, and so sexual selection acted to favor a pubertal growth spurt given the limitations discussed previously on childhood growth.

Males are fertile early in puberty, long before they reach adult height. In females immediately after menarche, which is late in the pubertal growth spurt, only a low percentage of cycles are associated with ovulation. It is not until about 2 years after menarche that ovulation rates reach adult levels. This prolonged period in the female allows a delay in the likelihood of pregnancy until other aspects of her biology are mature; in particular until her pelvic inlet has reached its maximum size and she has had time to be included as an adult female in the social
group. The slower appearance of mature physical features in the male may reflect the lesser likelihood of a young incompletely mature male competing for mating advantage in the social group, a battle he is likely to lose, and indeed there may have been survival advantage in delaying maturation while additional skills were learned. These differences may be reflected in the very different nature of pubertal rites in foraging societies. Female rites are generally short and followed by early marriage. Male pubertal rites often involve years of seclusion from females, and involvement in male-only activities including warrior duties before marriage is allowed.

5.5.3 Reproductive Decline and the Menopause

The menopause is a further feature of the human life cycle effectively not found in other mammals in the wild (Figure 5.17). Female fertility starts to decline well before the menopause, which usually occurs between 45 and 55 years of age. In contrast to menarche, the menopause is relatively uninfluenced by environmental factors. Oocytes only form during fetal life and oocyte numbers rapidly decline through childhood, and progressively thereafter until the time of menopause (Section 8.9.7). Most of the oocytes that a female starts with in her life die by apoptosis, with only a very small fraction ever reaching maturity and being ovulated. This suggests that oocytes have an intrinsic lifespan which is less than that of the female’s own potential lifespan, and that there is a decline in fertility from the mid-thirties, reflecting a loss of oocyte quality. Given that humans had a lifespan that averaged only about 35–40 years through the bulk of existence of our species, even if they survived infancy, this decline in fertility may simply represent the absence of any selective pressure for longer oocyte survival. However, this argument is becoming unsatisfactory in light of the increasing consensus that a significant proportion of women lived well into their seventh and eighth decades of life even in the Paleolithic (Box 5.3).

The evolutionary origin of the menopause has been subject to intense debate. At one extreme there are those who consider it an accidental consequence of now living for much longer on average than when humans evolved. However, some humans have always lived to an advanced age in every society. Others have argued that the menopause gives a fitness advantage in one of two ways: either the presence of a maternal grandmother can aid the daughter in supporting more offspring, or a decline in fertility as ultimately reflected in the menopause allows a mother whose lifespan on average was relatively short to have a greater opportunity to support her youngest offspring to reach independence before she dies. These concepts are discussed at length in Section 8.9.7.

5.6 Conclusion

This chapter highlights the role of trade-offs between life-history traits in defining the evolved strategies of a species and the interdependence of different components of the phenotype. Ultimately the life history is the result of different pressures operating in different ways, as exemplified by the various influences on the age at menarche.

_Homo sapiens_ has evolved with a very distinct life-history strategy with a unique pattern of growth, maturation, and reproduction, a distinct social existence, and a dependence on a large brain.
Life-history theory shows how interdependent these various features are. In turn, energy availability and utilization has been a selective driver in the evolution of the human life-history strategy. Differences in energy sources across the life cycle are linked to the patterns of growth, and the consequences of disturbed energy intakes can be understood in light of the trade-offs that must follow.

The outcome of these various evolutionary forces has been the rapid emergence of a primate species with a particularly large brain, whose fitness strategy is defined by a long life, parenting very few offspring who have a delayed and distinct pattern of maturation, and parents who invest heavily in these few offspring ensuring a high rate of survival to adulthood.

But there is much evidence that there are also important environmental influences on individual components of the life history (Box 5.8, Figure 5.18). These influences can affect our health, the risks of disease, and elements of our subconscious behaviors.

**Box 5.8 The Tempo of Contemporary Human Life Histories**

It is well documented in many animals that a threatening early life environment accelerates sexual maturation, thus maximizing the chance of successful reproduction before death intervenes. Does a similar phenomenon occur in humans? Earlier in this chapter we discussed how pubertal timing in humans is influenced by a complex hierarchy of pre- and post-natal factors. A substantial body of work undertaken by behavioral biologist Daniel Nettle now supports the notion that under conditions of deprivation there are evolutionarily embedded drivers that prompt behavioral and social changes in humans to speed up various life-history parameters.

Using data from the Millennium Cohort Study, a large, longitudinal investigation of 8660 British families with babies born during 2000–01, Nettle determined that, compared with those living in more affluent neighborhoods, relatively deprived families had a younger age at first birth (AFB), lower birthweight, shorter duration of breastfeeding, and slightly larger family size (Nettle 2010a). Furthermore, affluence was also associated with increased life expectancy (Figure 5.18).

Nettle has interpreted the earlier (and perhaps more frequent) pattern of reproduction as a manifestation of a “dying young and living fast” life-history strategy; if social

![Figure 5.18](image) Age at first birth (AFB) and life expectancy for women in English neighborhoods of varying qualities (1 = most deprived, 10 = most affluent). Plotted from data in Nettle (2011).
and environmental indicators point towards hardship (and therefore morbidity and mortality are more likely), the corresponding behavioral adaptive response—enabled through phenotypic plasticity—is to accelerate the life history accordingly. The same study also revealed strong positive correlations between AFB and female life expectancy, and AFB and female annual income (Nettle 2011). These relationships held when equivalent data were taken from across 116 countries.

The “live fast” life-history strategy has further public health implications. The link between socioeconomic position (SEP) and health behavior is well established: those with a lower SEP are in general more likely to be heavy smokers and drinkers, to consume a poor diet, and to be more sedentary. What could explain this socioeconomic health gradient, especially when some of the behaviors, such as purchasing highly taxed cigarettes and alcohol, present a financial disincentive to the very people to whom it should matter most? Furthermore, it appears that low-SEP individuals are much less amenable to behavioral changes associated with public health information campaigns than are those of high SEP.

Nettle’s previous modeling work has linked increased extrinsic mortality risk to decreased investment in protective health behavior (Nettle 2010b). Based on this work, his group have proposed an ultimate explanation for the effect of SEP gradients on health behavior—specifically, they drew upon an adaptive framework from behavioral ecology that considers both intrinsic (controlled) and extrinsic (uncontrolled) components of mortality risk (Pepper and Nettle 2014). A large group of volunteers were surveyed on several aspects: thoughts on their financial status, to formulate a subjective measure of SEP; effort made in ensuring personal health and safety; and their opinion on their chances of reaching old age with and without personal effort, to determine their perceived risk of intrinsic and extrinsic mortality. It was found that, in agreement with modeling data, a high perception of extrinsic mortality risk showed a strong negative correlation with the reported uptake of preventative health measures. In other words, believing that there is little one can do to live a long life prompts individuals to live for the present rather than invest in the future, by diverting time and energetic costs potentially devoted to investments in health to other areas of greater perceived value. Perhaps this life-history approach can inform public health interventions through identifying and modifying factors that influence perceived extrinsic mortality risk.

Key Points

- Life-history theory describes why species have particular patterns of growth, development, reproduction, and mortality.
- The life-history strategy of a species is determined by an evolutionarily optimized allocation of limited resources between growth, reproduction, and tissue repair to maximize reproductive success. This requires trade-offs among life-history traits, such as age versus size at maturity, number versus quality of offspring, current versus future reproduction, and fecundity versus lifespan.
- Humans are characterized by large brains and long lives. Their development is typified by a singleton pregnancy, a long phase of post-natal nutritional dependency, a prolonged juvenile period, delayed sexual maturity, and only modest sexual dimorphism. Females terminate reproduction before the end of their intrinsic lifespan. Humans produce very few offspring, which benefit from high parental investment, ensuring a comparatively high rate of survival to adulthood.
- The timing of puberty is influenced by developmental factors, and there is a potential disconnection between the age of biological puberty and acceptance as an adult; this is reflected in the problems of adolescence.
- Evolutionary considerations can offer explanations for the unusual characteristics of human life history, including the relatively long childhood phase, the pubertal growth spurt, and the menopause.
- Environmental influences within the life course can affect key elements of the life history.
CHAPTER 6

Human Evolution and the Origins of Human Diversity

6.1 Introduction

Human biology can only be understood fully in terms of our evolutionary history. Our biology is determined by and constrained by this history. Humans are upright apes with particular forms of locomotion and communication, cognitive capacity, capacities for intentionality and prescience, and the ability to develop and use technologies. We are a species that lives in social groups and has uniquely developed cultural capacities (Section 2.3.3). Our biology shows much evidence of our deep evolutionary past; for example, rudimentary “vestigial” organs such as the appendix and anatomical arrangements such as the long pathway of the recurrent laryngeal nerve. The evolution of bipedalism from a quadrupedal ancestor has many consequences, such as the risk of lower back pain (Section 6.3.2). The first sections of this chapter will focus on the key elements of hominin and human evolution. This is important for understanding the origin of our species, and thus the origin of common features which may be reflected in disease. Later in the chapter we will consider how humans have adapted differently in different environments, and how this contributes to the considerable diversity in some aspects of our phenotype. Finally we will briefly turn to a controversial and speculative question: what is the evolutionary future for the human species?

6.2 The Hominoid Clade

The evolutionary history of Homo sapiens has been considerably clarified by the use of molecular biology approaches. These have led to a revision of our understanding of the evolutionary relationships between the various hominoid species and the order of separation from last common ancestors. First some definitions are in order: among the highly diverse species that form the order Primates, hominoids comprise the superfamily of apes and humans, hominids refers to the great apes (excluding the gibbons or lesser apes), and the term hominins is restricted to humans and their direct ancestors after division from the ancestors of the other great apes in the evolutionary tree. The hominid superfamily comprises the currently living great apes (orang-utans, bonobos, chimpanzees, gorillas, and humans), their extinct ancestors, and other extinct species that had evolved from the last common ancestor of these five species (see Figure 1.4).

Molecular evidence shows that the orang-utan lineage was the first to split from the primitive hominid lineage some 12–15 Mya, and that the gorilla lineage split perhaps 9–13 Mya. We shared a last common ancestor with the chimpanzee and the bonobo as recently as 7–10 Mya. These dates are based on measurements of the mutation rate between human parents and children together with observed generation times of modern-day hominid species, and are rather older than previously calculated from rates of mutational change calibrated from the fossil record (Langergraber et al. 2012). It is now clear from fossil and genomic data that H. sapiens originated only in Africa. The alternative theory of multiple origins in diverse geographies from H. erectus is no longer tenable, although the position has recently been complicated by molecular evidence of admixture or
introgression between modern and, until relatively recently, coexisting archaic Homo species (Box 6.1).

There are approximately 300 species of primates existing today. They are typically arboreal species living in tropical and subtropical environments with common features of grasping hands and feet, often with opposable thumbs, locomotion which tends to be hind-limb-dominated, enhanced vision and a diminution in the olfactory sense compared with many other mammals, a specific pattern of dentition, and a flattened face. The brains of primates are large relative to their body size (Box 5.5), and as a result there are associated changes in life-history traits generally reflected in longevity, late onset of reproduction, having few offspring, and monotocous pregnancies.

The first primates appeared approximately 50 Mya. Primate evolution is characterized by a series of adaptive radiations (Box 2.1) leading to the prosimians (tarsiers and lemurs), New World monkeys, and Old World monkeys, from which primates later diverged. An intriguing question concerns how the New World monkeys reached the Americas from Africa. It is generally thought that this early “out-of-Africa” event occurred some 25–40 Mya by serendipitous (and of course passive) “rafting” on mats of vegetation at a time when the positions of the continents were such that the Atlantic Ocean was much narrower.

### 6.3 Hominin Evolution

#### 6.3.1 Timeline and Species

Hominin evolution also shows adaptive radiation. The drivers of this radiation are speculative but are likely to reflect changes in the arboreal environment, perhaps reflecting climatic change. The earliest paleoarcheological evidence of the hominin clade is fragmentary and subject to frequent claims and revision as individual fossil hominins and proto-hominins are discovered in Africa. We shall restrict our discussion to the generally accepted understandings.

Perhaps the earliest fossil for which there is reasonable evidence suggesting that it was an ancestral hominin is a specimen of *Sahelanthropus tchadensis*, which is dated to at least 6 Mya and was found in the region of Lake Chad. A more recent hominin, *Ardipithecus*, was discovered in the Afar region of Ethiopia and is dated to around 6–4 Mya. *Ardipithecus* is of particular interest as a transitional species because it has features suggesting that it was partially arboreal (tree living) but was also an effective upright walker when required (facultative bipedality) without the knuckle-walking that is characteristic of the modern great apes. The Australopithecines comprise several species of early hominin found in both eastern and southern Africa and which date to between 4 and 1 Mya. Their appearance and radiation has been related to a change in global temperatures, which fell during that time and led to a more open habitat in eastern Africa. While there is debate over the classification of hominin species, it is clear that the hominin clade has included many species, some of which will have coexisted. Few of these species have a direct ancestral relationship to *H. sapiens*, but rather represent terminal branches in the evolution of the clade (Figure 6.1; see also Lewin and Foley 2004).

Partial bipedalism was present in the earliest hominins such as *Ardipithecus*, and the Australopithecines were clearly bipedal. The evolutionary explanations for bipedalism and its consequences are discussed in Section 6.3.2. It seems that bipedalism preceded the appearance of a larger neocortex. Other novelties which evolved within the hominin clade include a change in dentition and jaw shape, and the development of technology and culture.

The early Australopithecine species were bipedal, sexually dimorphic, and weighed up to about 45 kg with a brain size of about 400–500 cm³, giving them a slightly larger brain relative to body size than the modern chimpanzee. They were probably exclusively vegetarian and lived in a woodland habitat. The fossil skeleton known as Lucy, dated to about 3 Mya, is from an Australopithecine. Skeletal analysis suggests that Lucy was adapted for bipedal walking but was not able to run efficiently. By 2.5 Mya so-called robust Australopithecine species appeared which had greater encephalization. Studies of their teeth suggest they were primarily plant eaters but did include some meat in their diet.

From about 2.5 Mya a further radiation occurred, and species of hominin appeared with smaller jaws.
and teeth but with a clear increase in brain size. There was probably more meat in their diet although it is not clear if this was hunted—more likely it was scavenged. Tool use appears then for the first time in our ancestral history. This cluster of features marks the appearance of the first species of the genus *Homo*.

These classifications into genera and species are somewhat arbitrary, and there are authorities who lump and others who split the scanty evidence to proclaim either fewer or more species. The difficulty is that variation is a characteristic within any species. When relying solely on skeletal data without access to a large number of specimens or to molecular or reproductive behavior, it is not possible to distinguish unequivocally between variation within a species and variation between species.

The earliest member of the genus *Homo* was *H. habilis*, which had a brain size of more than 600 cm³ and a body weight of about 45 kg, showing greater encephalization. These early species of *Homo* gave rise to a species with a larger body and larger brain, *H. erectus*. The best preserved example of *H. erectus*, the “Nariokotome boy” who died about 1.6 Mya, is estimated to have been destined for an adult height of 1.59–1.68 m and had a cranial capacity of 880 cm³. The nomenclature regarding our forebears can be confusing: some authorities refer to the earliest forms of *H. erectus* as *H. ergaster*. The issues surrounding these distinctions are not relevant to this book and so we will generally use *H. erectus* to include *H. ergaster*. The position in the human lineage of *Homo naledi*, a so far undated hominin species sharing characteristics of both Australopithecines and *Homo*, remains unclear.

Skeletal remains of *H. erectus* have been found not only in Africa but also throughout the Middle East and southern Asia (the first fossils were found in Indonesia), and in southern Europe. While *H. erectus* evolved about 2 Mya, it migrated out of Africa only about 1 Mya. *Homo erectus* appears to have had a prolonged childhood and a life-history pattern intermediate between that of modern great apes and humans. This estimate is based on calculations of tooth eruption patterns and of pelvic size to estimate brain size at birth and mature brain size. Hand axes appeared as a new kind of tool, and the archeological evidence has been interpreted to

---

**Figure 6.1** Timeline and relationships of hominin species. Query marks indicate uncertain relationships. Adapted from a figure kindly provided by Jacquie Bay, Liggins Institute.
suggest that *H. erectus* engaged in hunting and had more meat in its diet.

*Homo erectus* was succeeded by *H. heidelbergensis*, fossils of which have been found in Africa, the Middle East, and southern Europe. *Homo heidelbergensis* is considered the precursor of modern humans, as well as of two other varieties of hominin: Neanderthals and Denisovans. Evidence from their genomes (Box 6.1) suggests that the latter two groups diverged from the direct lineage to modern humans around 600,000 years ago, and Neanderthals and Denisovans themselves diverged around 400,000 years ago. Whether Neanderthals (and, presumably, Denisovans) should be classified as separate species (e.g., *H. neanderthalensis*) or as subspecies of *H. sapiens* (e.g., *H. sapiens ssp. neanderthalensis* versus *H. sapiens ssp. sapiens* for modern humans) is the subject of some debate. Anatomical studies have tended to favor separate species (e.g., Máñquez et al. 2014); the case of the “lumpers” was of course advanced by the discovery of interbreeding among archaic and modern humans (see Box 6.1). We will avoid the controversy and simply use the descriptors “Neanderthals” and “Denisovans.”

Neanderthals have been well characterized from skeletal remains and associated artifacts discovered over the past 150 years across Europe. They may have evolved as early as 300,000 years ago and were restricted to Europe, the Middle East, and Central Asia. The most recent Neanderthal remains appear to date from 25,000–30,000 years ago. They were a highly successful species that used tools similar to those of early *H. sapiens* and had a complex culture. Compared with modern humans, Neanderthals were more robust (stocky) with relatively shorter limbs; their skulls were characterized by sloping

---

**Box 6.1 Out of Africa, Making Friends on the Way**

The appealingly simple model of modern *H. sapiens* leaving Africa to colonize the world, completely replacing archaic hominins without any gene flow, has recently been complicated by the availability of the complete nuclear and mitochondrial genomes of a number of Neanderthal and Denisovan individuals (Meyer et al. 2012; Prufer et al. 2014). This has revealed the presence of a small proportion of Neanderthal and Denisovan alleles within the genomes of modern humans. Although the distribution of those introgressing sequences is consistent with the out-of-Africa model, their existence has revealed unexpected complexities in the evolution of current human populations.

Specifically, Neanderthal sequences constitute 1.5–2.1% of the DNA of all modern non-African humans, whereas Denisovan sequences form up to 6% of the genomes of present-day Papuans and Australian Aborigines, with a smaller proportion in other island populations of Southeast Asia and Oceania (Reich et al. 2011). Modern East Asians (e.g., from China, Japan, and Mongolia) have little (up to 0.2%) Denisovan admixture, although Section 13.11.2 discusses how a Denisovan sequence may be contributing to the adaptation of modern Tibetan populations to altitude.

How can we interpret these genomic data? First, they suggest that modern humans interbred with Neanderthals shortly after the initial dispersal from Africa, probably in the Levant where modern humans and Neanderthals are known to have coexisted around 55,000 years ago (Hershkovitz et al. 2015). This would account for the absence of Neanderthal sequences from modern Africans, and the similar level of Neanderthal admixture in all modern non-Africans. Second, modern humans who formed the ancestors of modern Papuans and Australians must have interbred with Denisovans in Southeast Asia after separation from the ancestors of East Asians, explaining the relatively high Denisovan admixture in Melanesia and Oceania but its low levels in modern East Asians (see Figure 6.4). The Denisovan introgression event in Southeast Asia, together with the location of the physical remains in Siberia, suggests a considerable geographical range across several altitudinal and/or climatic zones for the Denisovans despite the low genetic diversity inferred from the sequence data (implying a small population size).

Further complexity is added by the suggestion of gene flow from an unknown early hominin (possibly *H. erectus*) into the Denisovan genome, and by the discovery that the mitochondrial genome of a nearly 400,000-year-old hominin found in a Spanish cave resembles more closely the Denisovan sequence than the Neanderthal sequence that the location would imply (Meyer et al. 2014). The evolutionary pathway of hominins clearly involves much more intergroup gene flow than previously anticipated, and will no doubt yield further surprises.
brows and a protuberant face, with a brain volume 10% greater than that of modern humans. While their capacities and capabilities can only be indirectly inferred, there is some anatomical evidence, such as the size of the hypoglossal foramen and the shape of the hyoid bone, which suggests some ability for vocalization. There is also some scanty and controversial evidence for carving and artistic representation which may have been copied from coexisting *H. sapiens*. There is evidence that at least 70,000 years ago Neanderthals buried their dead and that this was associated with defleshment, which might represent some form of ritual.

Much less is known about the Denisovans. Their physical remains consist only of a few small bones and teeth found in the Denisova Cave of the Altai Mountains in southwestern Siberia, dating from around 50,000 years ago, although it has been suggested that the Dali and Maba skulls found in China could also be of Denisovan origin. The size and shape of the skeletal fragments indicate that the Denisovans, like the Neanderthals, were robust in form. Our knowledge of the Denisovans comes mostly from genomic studies (Box 6.1), which revealed that the Denisovan individual sequenced probably had dark skin, brown hair, and brown eyes.

The evolution of anatomically modern *H. sapiens* from the *H. erectus* and *H. heidelbergensis* lineage involved a decrease in robustness of the face and skeleton and a progressive change in functional capacities associated with locomotion, behavior, social organization, and culture. It is now clear from molecular studies that this change occurred only in Africa, with other populations of *H. erectus* and their successor species eventually becoming extinct. Anatomically modern humans appeared in southern and eastern Africa approximately 160,000–200,000 years ago and remained localized to that continent until about 60,000 years ago (although there is evidence for an unsuccessful expansion into the Levant, eventually defeated by changes in climate, occurring 80,000–130,000 years ago). The use of mitochondrial DNA and Y-chromosomal mapping has allowed the reconstruction of subsequent migrations of humans from Africa to be studied in considerable detail (see Box 6.8; Wells 2002). That dispersion plays a major role in helping us to understand the origins of human diversity.

Climate change leading to ecological changes and progressive deforestation in eastern Africa is likely to have played a major role in our development from an arboreal ancestor into a bipedal terrestrial ape. The evidence suggests that over the last 5 million years there have been a number of cycles of warming and cooling. Of particular relevance are major cooling events about 5 Mya and again between 3.5 and 2.5 Mya, which led to a major build-up of polar ice. At about this time the Panamanian isthmus rose and joined the Americas, altering patterns of circulation in the oceans. Many other major climate events leading to changing sea levels and environments occurred in the subsequent 2.5 million years and played a role in allowing humans to migrate around the globe. The range of the Neanderthals was restricted to the north by the glaciations that occupied northern and middle Europe from approximately 100,000 to 10,000 years ago, and that cold period may have delayed the expansion of modern humans out of Africa. The late stage of the last glacial period 12,000 years ago may have allowed grasses to grow in the Middle East, promoting the development of agriculture. The changing nature of food supply meant a shift from an exclusively herbaceous diet to a mixed diet, first by scavenging and then by hunting. Clearly there was selective advantage in the development and use of technology: group living, communication, and cultural capacities both reinforced and were reinforced by these factors.

6.3.2 Bipedalism

By about 4 Mya, *Australopithecus* was a generally bipedal hominoid, and facultative bipedalism was present in even earlier species such as *Ardipithecus*. Fossil hallmarks of bipedalism include the positioning of the foramen magnum, which becomes more central as the posture becomes more upright and the skull has to be supported by the spinal column. There are also changes in the shape of the pelvis, hip socket, and femora. There have been many hypotheses put forward, and entire books written, regarding the adaptive origins of bipedalism, some compromised by an anthropocentric and even teleological tendency to think of humans as special.
Human bipedalism has features which are distinct from the occasional bipedalism of other apes. The knee is very different, allowing the leg to straighten and the knee to lock, minimizing energy expenditure in supporting the body when standing upright. Humans have a pelvic and muscular structure which allows the center of gravity to shift only slightly with each step, whereas species such as the chimpanzee waddle because an enormous shift in the center of gravity occurs with each step. Being able to walk this way has meant the evolution of a differently shaped pelvis, an angled femur, and a change in the limb-moving muscles such as the gluteal abductors. Humans can run—a development which first became anatomically possible in early Homo species. Humans are thought to be unique in being able to undertake endurance running as well as sprinting, which perhaps was an adaptation that allowed the hunting down of prey.

One hypothesis is that bipedalism may have evolved in Australopithecines in association with a shift in land cover from forest to a more open grassland environment in East Africa about 4 Mya (Kingdon 2003). As this environmental shift was associated with a greater dispersal of food resources, bipedal movement would have been energetically more effective than quadrupedal movement, particularly at walking speeds. The paleoanthropologist Robert Foley has calculated that provided early hominins spent the majority of their time in terrestrial (as opposed to arboreal) foraging, then bipedalism would be adaptive in such environments (Foley and Elton 1998).

The discovery of *Ardipithecus*, however, has called into question this open grassland origin of bipedalism. *Ardipithecus*, which probably lived in a mixed woodland/forest environment and was omnivorous rather than fruit-eating, appears to have been both an efficient climber and capable of upright walking. Nevertheless, its pelvic anatomy suggests that, unlike Australopithecines, *Ardipithecus* derived little energy advantage from an upright posture. Bipedality in *Ardipithecus* may have evolved to allow the carrying and sharing of food as part of a more cooperative social structure (Lovejoy 2009).

Other arguments that have been put forward for the origin of bipedalism include better thermoregulatory efficiency—being upright exposes less of the body to the sun. This would significantly reduce the demand for fluids. The loss of body hair has also been suggested as a component of this thermoregulatory adaptation as it would allow more effective evaporation of sweat. Both these putative adaptations would allow foraging to extend through the heat of the day. Others have suggested that being upright allows greater capacity for predator avoidance by allowing the individual to see further, just as the meerkat stands upright when on guard. While these arguments have much popular appeal, they are more likely to be secondary exaptations (Section 2.4.4) which arose from, and amplified, the initial adaptation of standing upright and adopting a bipedal gait.

However, bipedalism has costs, the most obvious being lower back pain, which is one of the most common reasons for modern humans to consult a doctor. The shift from a quadrupedal to an upright posture came late in the evolutionary history of primates, and the accommodation reached in terms of a lumbar lordosis creates pressure on the intervertebral discs and the sacroiliac joints. Prolapse of a disc, with injury to the lumbar nerves innervating the lower limbs, leads to pain, paraesthesia, and sometimes motor dysfunction. Such problems are aggravated by being overweight, which puts even greater pressure on the intervertebral discs, and by osteoarthritis and osteoporosis, problems of advanced age characteristic of the post-industrial modern human. The upright posture makes injurious falls more likely as people age, because their gait and balance become more unsteady and they have failing vision. As they develop osteopenia the risk of femoral neck fractures becomes very high; the mortality associated with femoral neck fracture is a major concern in the older population. The shape of the pelvis is changed by an upright posture, necessitating a flattening of the pelvis and a change in the shape of the pelvic canal. Indeed, many of the potential complications of human childbirth are consequential upon this anatomical change (see Section 8.9.5).
6.3.3 Body Size

The Australopithecines were small, ranging in weight from 18 to 45 kg, whereas Homo species were larger—their size reached a peak with Neanderthals. It is not clear why body size increased over the course of hominin evolution. Clearly a number of competing evolutionary pressures were at work. Perhaps living in a more open environment exposed early hominins to a greater risk of predation and so there was selection for larger individuals. There are costs to this increase in body size: the resulting increase in metabolic rate (see Section 5.2) required hominids with their mixed diets to expand their home range for foraging and to shift from a purely herbivorous diet. The alternative approach is to become adapted to eating large quantities of low-caloric foods: this is the strategy of the gorilla, which has a small foraging range but requires a large intestinal system to digest the considerable amounts of plant material it eats. The increase in body size would have caused greater problems in thermoregulation, which may have been ameliorated by an upright posture and loss of body hair. Further, the time to reach adult maturity took longer, and living in social groups thus became a key feature for care and protection. The presence of larger social groups allowed a greater feeding range to be occupied. Energy requirements would have been further exacerbated by the high energetic demands of a relatively large brain (see Section 6.3.7.2 and Box 9.7). In contrast, the selective pressures that may have caused a Homo species to evolve towards a decreased body size remain the subject of speculation (Box 6.2).

6.3.4 Face, Jaw, and Dentition

There were major changes in the face, jaw, and dentition as the hominoid line evolved. In general, hominins evolved a flatter face with a less protuberant jaw and the teeth tucked under the face relative to modern apes. Thus there was a progressive change in mandibular shape to the modern L-shape. Humans have much less developed masseter muscles compared with their early forebears, and this is associated with other skeletal changes such as a smaller zygomatic arch. These changes suggest that a less powerful chewing action was needed by later-evolving hominin species and this became even more so once fire was used to cook meat. The first solid evidence for controlled use of fire, such as hearths, earth ovens, and charred animal bones, dates to about 500,000 years ago.

Box 6.2 Homo floresiensis?

In September 2003, an archeological expedition to the Liang Bua cave on the island of Flores, Indonesia, uncovered the skull and partial skeleton of an 18,000-year-old hominin. Several other specimens have subsequently been found in the cave. Various primitive and derived features led the discoverers to denote the remains as a species distinct from H. sapiens. Named Homo floresiensis in recognition of its island of discovery, the skeleton was determined to be that of an adult female who was just 1 m tall. This, the researchers postulated, was a consequence of dwarfing of an ancestral H. erectus population on Flores; the phenomenon of “insular dwarfism,” in which a restricted ecological range and/or resource limitation cause a reduction in body size over generations, is well known in other animals. Inevitably, the popular press dubbed these hominins “hobbits.” Besides its extremely small physical size, what stood out about the skeleton was its very small brain volume (about 380 cm³ compared to about 980 cm³ for late H. erectus and 1350 cm³ for modern humans).

Given the presence of sophisticated stone tools at the site, other researchers believe that the specimen is a small-bodied, microcephalic modern human. Yet others have argued that the remains are of H. sapiens affected by endemic cretinism because of low iodine levels, or by Laron syndrome (growth-hormone insensitivity) or Down syndrome (trisomy 21). Supporters of H. floresiensis cite unusual anatomical features of the skeletons, traditional stories of small-bodied people in the area, and evidence for the persistence of H. erectus in island Southeast Asia until less than 30,000 years ago. Unfortunately, it is unlikely that DNA can be recovered from the putative H. floresiensis remains. Controversy still rages regarding the true nature of H. floresiensis (Callaway 2014), although the majority of evidence supports a distinct Homo species.
The change in jaw shape allowed for a more grinding type of mastication, and this was associated with a change in dentition. The incisors became less prominent compared with those of other primates and the molars became larger. These changes reflect a difference in diet from the purely fruit-based diet of apes to a more omnivorous diet. There is evidence that when in recent centuries human infants began to be fed on particularly soft foods, the need for grinding movements became less, and the risk of dental malocclusion increased (see Section 13.8.2; Rose and Roblee 2009). The flattening of the face and the change in posture also changed the positioning of the eustachian tubes. This contributes to the risk of otitis media. The changes in the position of the larynx associated with vocalization are associated with increased risk of obstructive apnea, especially during sleep.

6.3.5 The Gastrointestinal Tract

The earliest primates were insectivores; as they became larger they became herbivores. There are some modern primates such as the langur that have large stomachs, analogous to the ruminant stomach, to allow bacterial digestion of cellulose in the upper gut. But most herbivores rely on slow intestinal transit and ileocecal digestion. A large cecum is typical of such an intestinal system that needs to digest high-fiber food. As the diet shifts from mainly frugivorous to more omnivorous diet, a very long gastrointestinal tract and a large cecum become energetically inefficient.

The earliest hominins were exclusive herbivores, but they relied primarily on fruits and tubers rather than on the high-cellulose diets typical of leaf and grass eaters. The later Australopithecines probably also scavenged for protein- and energy-rich meat, and meat became a larger component of the *Homo* diet, particularly once fire was adopted for cooking. As noted earlier, these changes are reflected in the jaw and dentition as well as in the gastrointestinal tract. The appendix is an atavistic reminder of our evolutionary origin as frugivores (as is our requirement for ascorbic acid in the diet; see Box 6.3). Indeed, Darwin recognized the appendix to be a rudimentary or vestigial organ, there being no evidence that it has unique functions, even though it does contain some lymphoid tissue. Appendicitis is the outcome of this vestigial organ becoming infected: it is prone to inflammation and rupture. Appendicitis can be fatal, and its peak incidence occurs in adolescence and early adulthood, so it is interesting to speculate why there has not been more active selection against persistence of the appendix. The evidence is that appendicitis was rare until recently: its prevalence appears to have risen in the nineteenth and early twentieth centuries in high-income countries, where its incidence is now declining again; the incidence is much lower in traditional and less developed societies. Thus there would be little selection pressure operating on this vestigial organ. While the reasons for the association between modern development and appendicitis are unclear, they are likely to be related to changes in the intestinal microflora or the presence (or lack) of intestinal parasites (Laurin et al. 2011).

6.3.6 The Hairless Ape

All other primates, including the other apes, are fully covered with hair apart from on the palms of their hands and feet. It is not known when hominins lost this characteristic. The most generally held explanation is that the hair loss aided thermoregulation, even though all other primates living in tropical climates are fully covered in hair. An alternative thesis would hold that hairlessness does not have an adaptive origin through natural selection but rather may have evolved by sexual selection. When a feature appears rapidly and uniquely in one species within a clade, sexual selection needs to be considered as a possible explanation. The retention of pubic hair in both sexes and of facial hair in males might support this argument. Pubic hair only appears at puberty, and in both sexes the full pattern of sexual hair would be a sign of sexual maturity. Similarly, facial hair may serve this purpose in the male.

What was the color of the skin exposed by the loss of body hair coverage? Chimpanzees have generally pale skin under their fur, suggesting that the first hairless hominins may have had the same. The sun-resistant (dark skin-conferring) variant of the human melanocortin 1 receptor (*MC1R*) gene appears to have evolved around 1.2 Mya, after
the time early hominins first inhabited the open grasslands of Africa. Protection against ultraviolet-induced skin cancer may have been the selective pressure responsible for the emergence of this allele (Greaves 2014). An alternative proposal is protection against folate deficiency (Jablonski 2012), as this vitamin is photosensitive and can be destroyed by exposure to sunlight. The mechanisms and consequences of the reversion to lighter skin as humans migrated to more northerly latitudes are discussed in Section 13.4.1.

### 6.3.7 The Hominin Brain

Figure 6.2 shows the change in brain volume calculated from the dimensions of fossil skulls across the hominin lineage. Brain size, in both absolute terms and relative to body size, did not change dramatically in the earliest members of the lineage. Australopithecines had absolute and relative brain sizes that were not greatly different from those of the modern apes. Expansion of brain size started with the appearance of Homo species and showed
an exponential increase from *H. habilis*, through *H. erectus* to Neanderthals, and then a slight decrease to *H. sapiens*. During this period, average brain volume increased from about 400 to 1250 cm$^3$. This can be expressed as the encephalization quotient (EQ), which is an allometric calculation of the relationship between brain mass and body mass. In mammals the most widely used formula is \( \text{EQ} = \frac{\text{brain weight}}{[0.12 \times (\text{body weight})^{2/3}]} \).

The common chimpanzee has an EQ of 2.0, Australopithecines had an EQ of about 2.5, *H. erectus* an EQ of about 3.3, and modern *H. sapiens* has an EQ of 5.8. As hominoid body size has generally increased through this lineage, the enhanced encephalization suggests positive selection for brain size. But in addition to brain volume there are changes in brain structure: the brains of non-human primates have relatively smaller parietal and temporal lobes and simpler frontal lobes.

It is important to recall at this point that there is no inherent direction in evolution (Section 2.4.1), and in particular no inherent progression in brain size—the Neanderthal brain had a larger volume than the modern human brain. Indeed, brains are energetically expensive and there are many large animals with highly successful evolutionary histories yet relatively small brains (witness the dinosaurs prior to their asteroid-induced extinction). Thus the answer to the question of why hominins evolved large brains and *H. sapiens* is endowed with its unique set of capacities is not self-evident.

6.3.7.1 The Social Brain and Theory of Mind

There are several inter-related theories about the evolutionary origin of brain expansion. Effectively they either place emphasis on *Homo* species finding adaptive advantage in social interactions within
Humans are social animals, and within the group individual fitness is advanced by maintaining stable social alliances that promote potential mating opportunities. But if this strategy is to be successful then an “evolutionary ratchet” may be created where societal pressures build, requiring even greater intellectual complexity to maintain reproductive opportunity. Thus a series of feed-forward loops may be created in which rising social complexity requires greater intelligence. In turn this allows the development of more sophisticated technology, more complex ways of living, and more sophisticated social structures, and these in turn generate the need for still higher cognitive function. Communication becomes a critical component of such a feed-forward system, and the development of language (Section 6.3.9) was clearly permitted and expedited by this increasing neural complexity.

Group structures in mammals, including humans, have hierarchies. In some species such as the hyena these are very apparent and determine mating rights. Alliances are needed, and in Chapter 11 we will discuss how concepts of reciprocal altruism assist in maintaining the integrity of a group. Grooming in primates, and perhaps language in humans, served the purpose of maintaining group cohesiveness, protecting friendships and alliances, and repairing injury to that cohesion. But in human groups there are frequently attempts by individuals to deceive others or to manipulate the group dynamics for individual advantage. This can also be seen in chimpanzee society. Such deception must remain within limits otherwise group cohesion would be lost. However, the capacity to behave in this way requires an ability to interpret the intentions of other individuals. This requires an appreciation that others of the same species have minds with thoughts, beliefs, and the capacity to interpret one’s own actions in return. This concept is called the theory of mind and it is categorized in terms of orders of intentionality (Box 6.4).

There are demonstrable relationships between brain size and the extent of interaction in social groups. For example, Dunbar (2003, 2007) has extrapolated from measurements of brain size to estimate the levels of intentionality of earlier hominins (Table 6.1), concluding that intentionality above fourth order did not emerge until anatomically modern humans (although some Neanderthals may have been close). Brain size can also tell us something about group size within the primates (Figure 6.3). This relationship suggests that pre-agricultural human groups comprised between 50

---

**Box 6.4 Theory of Mind**

This construct is used to describe how members of a species interact when interpreting each others’ behaviors and thoughts. It is directly linked to neocortical capacity and function. There is a close relationship between these concepts and that of consciousness: the ability to assess one’s own state of mind and that of others. The basis of theory of mind is a hierarchy of orders of intentionality.

First-order intentionality is awareness of self (plants have zero-order intentionality). Second-order intentionality is awareness that others are also aware of themselves. Third-order intentionality is awareness that another individual is not only self-aware, but is aware that the first individual can think too.

Levels of intentionality extend beyond these three levels, and this reciprocal understanding is important for successful social interactions in humans. Most adult humans can handle five or six layers of intentionality: for example, I know (first order) what you are thinking (second order) about my thoughts (third order) about what will happen to you (fourth order) if you do not respond to my overture (fifth order) in the way I would like you to do (sixth order). At this point we usually get tangled up.

Although most other primates have no capacity beyond the first level of intentionality, some Old World socially advanced species such as baboons and chimpanzees clearly practice deception, particularly in relation to sexual matters and sometimes food access. This requires second-order intentionality. It is thought that it was the capacity to develop advanced levels of intentionality during human evolution that gave rise to art, music, science, and religion.
and 150 individuals (Dunbar 2008), a group size that has persisted in some aspects into modern society. Section 6.3.9 extends consideration of social group interactions to the evolution of language.

### 6.3.7.2 Other Consequences of Greater Encephalization

Of all the organs in the body, the brain is particularly energetically expensive. Even though in the adult human the brain comprises only 2% of body weight, it consumes 20% of the body’s energy (and much more in infancy; Figure 5.7). This puts limits on how much the brain can grow, and in turn relates brain evolution to ecological history. The capacity to evolve a larger brain may also be related to the capacity to hunt, cook, and digest meat, which would have provided higher intakes of fat and protein to support brain growth and energetics.

The development of the large brain had other consequences. While other extant apes give birth to neurologically precocial babies, evolutionary pressures on *Homo* had to address the problem of passing a large head through a pelvic canal which had

---

**Table 6.1** Calculated achievable levels of intentionality for hominin species. Second-order intentionality represents minimal theory of mind, possibly achievable by some modern-day social primates. Data adapted from (Cole 2011)

<table>
<thead>
<tr>
<th>Species</th>
<th>Level of intentionality</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sahelanthropus tchadensis</em></td>
<td>First order</td>
</tr>
<tr>
<td><em>Orrorin tugenensis</em></td>
<td>First order</td>
</tr>
<tr>
<td><em>Ardipithecus ramidus</em></td>
<td>First order</td>
</tr>
<tr>
<td><em>Australopithecus anamensis</em></td>
<td>First order</td>
</tr>
<tr>
<td><em>A. afarensis</em></td>
<td>First order</td>
</tr>
<tr>
<td><em>A. africanus</em></td>
<td>First order</td>
</tr>
<tr>
<td><em>Paranthropus aethiopicus</em></td>
<td>First–second order</td>
</tr>
<tr>
<td><em>P. boisei</em></td>
<td>Second order</td>
</tr>
<tr>
<td><em>P. robustus</em></td>
<td>Second order</td>
</tr>
<tr>
<td><em>Homo habilis</em></td>
<td>Second order</td>
</tr>
<tr>
<td><em>H. erectus</em></td>
<td>Second–third order</td>
</tr>
<tr>
<td><em>H. heidelbergensis</em></td>
<td>Third–fourth order</td>
</tr>
<tr>
<td><em>H. floresiensis</em></td>
<td>First–second order?</td>
</tr>
<tr>
<td>Neanderthals</td>
<td>Fourth (fifth?) order</td>
</tr>
<tr>
<td>Denisovans</td>
<td>Fourth (fifth?) order?</td>
</tr>
<tr>
<td>Modern <em>H. sapiens</em></td>
<td>Fifth order</td>
</tr>
</tbody>
</table>

---

**Figure 6.3** Mean social group size for different species of primate plotted against relative neocortex volume (neocortex volume divided by the volume of the rest of the brain). Ape species are shown as blue symbols; the human point is predicted from the ape regression equation. From Dunbar (2008), with permission.
had its shape affected by the adoption of an upright posture. The solution was to evolve with a secondary altricial strategy by limiting fetal life to an earlier stage of development. This makes the human infant entirely dependent on its mother for many months after birth (Chapter 8), and it is not independently fully mobile (i.e., able to run) for several years. If humans were born at the same stage of maturity as other primates, then pregnancy would last about 21 months, and this would require a pelvic canal so wide that it would be impractical for efficient bipedal locomotion. This secondary altriciality in turn determines human social structure: if the mother is to support the infant, she needs to be confident of support from the father. Thus while early hominins showed great sexual dimorphism, with the males much larger than the females, implying a harem-type mating system with fighting between males for mating rights, \textit{H. sapiens} has a much lesser degree of sexual dimorphism, suggesting that in general females were able to have continued support from one male (see Sections 8.6 and 8.7 for further discussion).

### 6.3.8 Tool Making and Art

Hominins are not the only animals to use tools. Some birds such as the New Caledonian crow are quite adept at making tools from twigs and using these to catch insects. Similarly, chimpanzees use sticks as tools and even as weapons, and there are quite different cultures of tool use between the chimps of western and eastern Africa, highlighting the importance of cultural transmission in non-human primates.

However, humans as a species are characterized by a remarkable capacity to make tools and exploit technology. Because stone survives the ravages of time, much more is known about stone-tool making than the use of wood and other perishable materials such as those needed for clothing (Box 6.5). The earliest stone tools date to about 2.5 Mya; they are shaped to make scrapers, choppers, and cutters. Experiments show that while these seem simple constructions, it takes considerable skill to make them, requiring the capacity to strike the stone at a precise angle at the appropriate point. It is generally thought that \textit{H. habilis} was the first species to make tools, although there is some limited support for the view that the later Australopithecines could also make simple tools. After a million years in which tool technology seemed static, about 1.6 Mya new forms emerged with the appearance of rounded tools and two-sided handaxes and cleavers. Such technologies, associated with \textit{H. erectus}, have been found not only in Africa but also in the Asian sites.

---

**Box 6.5 A Lousy Way of Determining When Humans Started to Wear Clothing**

Use of clothing by hominins was prompted by two events: loss of insulating body hair, which is believed to have evolved over 1 Mya, and migration resulting in exposure to cooler climates. The archeological evidence for the origin of clothing is scant, since furs and fibers do not preserve well. The oldest tools associated with the manufacture of clothing—bone needles—date from about 40,000 years ago, although there are many older examples of scrapers that might have been used to clean an animal pelt.

An innovative approach to finding an answer came from studies on the molecular genetics of the human louse (\textit{Pediculus humanus}). There are two subspecies of this parasite: the head louse feeds exclusively on the scalp and attaches its eggs to hairs, while the body louse feeds on the body and attaches its eggs to clothing. It is a reasonable assumption that the two subspecies began to diverge at the time humans began to wear clothing. Molecular clock analysis of DNA sequences from head and body lice suggests a divergence date of around 80,000 years ago, and possibly as long ago as 130,000 years ago, which would place one origin of clothing in anatomically modern humans before their dispersal from Africa (Toups et al. 2011). Short of finding fossilized lice associated with archaic humans, this approach cannot tell us whether Neanderthals and Denisovans also wore clothing—although the archeological evidence suggests that Neanderthals did. Intriguingly, before genome sequencing confirmed gene flow between modern and archaic humans (Box 6.1), an earlier study on louse genetics had suggested direct contact between these groups (Reed et al. 2004); we now know how intimate that contact really was.
where remains of *H. erectus* have been found. Tool design remained remarkably unchanged until perhaps 250,000 years ago, after which there is a rapid expansion in the range of tool types identified.

In Africa there is some evidence for barbed harpoons dating from 100,000 years ago, and that long-distance exchange of goods and bone tools appeared about that time. Modern humans arrived in Australia as early as 50,000 years ago; even though sea levels were lower, this required the crossing of a number of channels between islands, suggesting that these humans had advanced technologies. Unfortunately, the Eurocentric nature of much of science has given greater emphasis to findings in Europe. About 40,000 years ago there was a further expansion in the range of stone tools made in Europe. This is the time when modern humans displaced Neanderthals, and is termed the *Upper Paleolithic revolution*.

The same period provides strong evidence of other forms of cultural development including representative art, carvings, musical instruments, and decorative items such as beads. Although much older objects associated with *H. erectus* and dated to 300,000–500,000 years ago have been claimed as the first examples of symbolic or representative art, their human provenance has been questioned. Rudimentary markings in the Blombos Cave in South Africa are estimated to be about 70,000 years old, but it was not until 40,000–50,000 years ago that humans began to produce artistic objects such as the cave and rock shelter paintings found in Europe and Australia. Even so, some have argued that the utilitarian objects from the 50,000 years before that began to show a preference for the aesthetic in the form of symmetry or shaping beyond that necessary for their practical function.

6.3.9 Language

One of the most contentious debates in human evolutionary biology is over the timing of the evolution of the capacity for language. Informed views date this anywhere from early *Homo* some 400,000 years ago to as recently as 50,000 years ago. One difficulty is the limited anatomical substrates that can be used to infer language, since the soft tissues of the larynx are not preserved in fossils. The scanty anatomical evidence suggests that some form of vocalization was possible in early *Homo*, and the hyoid bone from the Neanderthal vocal tract appears to be very similar to that of modern humans. Molecular evidence can also be used to infer the timing of language evolution (Box 6.6).

We can set a most recent boundary for the emergence of the capacity for language from the fact that all modern human populations have a complex language, and therefore that capacity for language must have been present in ancestral humans before the expansion from Africa 50,000–60,000 years ago. The other major form of evidence has been inference from the study of materials such as tools and development in areas thought to be associated with vocalization in mammals and birds, and also changes behavioral traits such as learning and sound production (Scharff and Petri 2011). The high rate of divergence in *FOXP2* after the human/chimpanzee split is suggestive of a role for *FOXP2* in the evolution of language in humans. Neanderthal and Denisovan *FOXP2* also possesses the same two mutations as modern humans, although there are changes in regulatory regions around *FOXP2* in modern humans that are not found in the archaic genomes. It remains uncertain whether Neanderthals had the capacity for language, although the extent to which they used tools suggests a developed capacity for communication.

---

**Box 6.6 Can the Genetic Basis for Speech and Grammar Tell Us About When Language Evolved?**

Studies of a large family with severe impairment of language ability identified a causative point mutation in the gene *FOXP2*. This gene codes for a transcription factor able to regulate the expression of multiple target genes. Although language is complex and clearly not reliant on any single gene, investigation of *FOXP2* has provided some evidence to suggest that the development of language was selected for during human evolution.

The protein coded by the human *FOXP2* gene differs from that of mice by three amino acids and from that of chimpanzees, gorillas, and rhesus macaques by two. Introduction of the human *FOXP2* sequence into transgenic mice affects brain development in areas thought to be associated with vocalization in mammals and birds, and also changes behavioral traits such as learning and sound production (Scharff and Petri 2011).
artifacts, and of social structure and behavior. It has been suggested, for example, that relatively consistent patterns of tool-making, some of which involved quite complex but arbitrary patterns of design, imply the use of language for their transmission. A stronger argument can be derived from studies of the development of art, rituals, and social rules. But unequivocal representative art in the form of cave paintings or carved figures dates to only about 32,000 years ago in Europe and perhaps as early as 50,000 years ago in Australia. Some of this art, particularly in Australia, clearly involves use of abstraction and suggests the capacity for thought, which is intimately related to the capacity for language.

Language is generally considered to have evolved as a way of assisting communication within the social group. Cooperative ventures such as hunting would be aided by such communication, although many other species such as wolves can cooperate in hunting without requiring advanced language. Indeed, there has been a shift in emphasis towards viewing the evolution of language in a different context: namely to aid the capacity to be conscious and to analyse the perceived world. It is widely believed that it is not possible to build a construct of the world beyond the immediate present without language in some form. Dunbar (2003, 2008) has gone further in arguing that language was key to the maintenance of larger stable social groups of the order of 150 individuals which characterized our ancestral social organization, and to some extent still does (Table 6.2, Figure 6.3). Whereas grooming is used to achieve social cohesion in other organized primate groups, language would have been a more effective and efficient means of doing so as social groups became larger and the capacity for grooming across the full community became limited by time. There is thus a loose interaction between social group structure, brain size, and language, and while it is not possible to be definitive, the weight of evidence suggests that language evolved relatively recently, perhaps 70,000 years ago. In turn this supported and reinforced our social organization and allowed the development of a more complex mental world which could support the development of art, music, belief, and political systems.

<table>
<thead>
<tr>
<th>Grouping</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neolithic villages (Middle East, 7500–6500 years before present)</td>
<td>150–200</td>
</tr>
<tr>
<td>Roman army maniples “double century” (2000 years before present)</td>
<td>120–130</td>
</tr>
<tr>
<td>Average English village size (Domesday Book, 1086)</td>
<td>150</td>
</tr>
<tr>
<td>Average English village size (eighteenth century)</td>
<td>160</td>
</tr>
<tr>
<td>Tribal societies (mean and range of nine communities)</td>
<td>148 (90–222)</td>
</tr>
<tr>
<td>Hunter-gatherer societies (mean of 213 clans)</td>
<td>165</td>
</tr>
<tr>
<td>Hutterite farming communities in Canada (mean of 51)</td>
<td>107</td>
</tr>
<tr>
<td>Amish parishes in Nebraska (mean of 8)</td>
<td>113</td>
</tr>
<tr>
<td>Church congregations (recommended ideal size, 1974)</td>
<td>200</td>
</tr>
<tr>
<td>East Tennessee rural mountain community</td>
<td>197</td>
</tr>
<tr>
<td>Social network size (mean of two)</td>
<td>134</td>
</tr>
<tr>
<td>Clothing factory unit size</td>
<td>150</td>
</tr>
<tr>
<td>Military company (mean and range for 10 armies in World War II)</td>
<td>180 (124–223)</td>
</tr>
<tr>
<td>Christmas card distribution lists (mean of 43)</td>
<td>154</td>
</tr>
<tr>
<td>Research specialties (sciences and humanities)</td>
<td>100–200</td>
</tr>
<tr>
<td>Facebook friends</td>
<td>130 (mean), 200 (median)</td>
</tr>
</tbody>
</table>

Equally contentious in the area of evolutionary psychology is the substrate for language evolution. There are basically two schools of thought: one holds that language is an incidental outcome of developing a large brain and is largely derived from cultural evolution; the other maintains that language evolved specifically through selection (for a flavor of this debate, see Pinker 2013). Some notable theorists, including Stephen Jay Gould, have adhered to the first school, arguing that the substrates for language did not evolve through natural selection. Gould considered that language evolved because of the cultural possibilities arising from possessing something like a biological general-purpose computer. This view is not generally accepted, but seems
to have been a response to both the enthusiasm for finding an adaptationist explanation for every phenomenon and because of the position that Gould took up in the sociobiology debate (see Box 11.1), namely to argue against evolutionary explanations for much of human behavior.

The leading theorist for the other camp—a selective origin for language—is Steven Pinker. There is little doubt that hominins evolved with specific features enabling them to speak and use language (e.g., the shape of the larynx, Broca’s and Wernicke’s areas in the neocortex). There is evidence for these features even in early species of *Homo*. Thus the “design” argument would maintain that active selection for language occurred. Pinker would argue that language is the key reason why hominins developed a large brain, and this implies that protolanguage developed in early *Homo* species. It is easy to envisage adaptive advantage with even the minimal capacity for vocalization and then symbolic language. Further there are genetic abnormalities that affect speech and grammar (see Box 6.6), and this is an argument for genetic determinants and thus selection on the capacity to use language. Young children acquire language very easily, suggesting an innate understanding of the rules of language arising from a genetic and therefore presumably selected substrate. In this regard it is important to recognize that infants learn fluent and grammatically correct language by passive exposure rather than by being formally taught grammatical rules and syntax. This in turn leads to another debate, beyond the scope of this book, as to whether there is a universal grammar with a genetic basis underpinning it: it is a debate which keeps linguistic research lively.

### 6.3.10 Evolution of Human Society

“Culture” is an amalgam of knowledge, behavior, and tradition within a particular community or population. It can be manifest in technology and tools, in art and music, in belief, myth, stories, and tradition, in behavior, and in social structure and organization. There is an intimate relationship between our biological and cultural evolution (Section 2.3.3; Richerson and Boyd 2005). There are, however, two cultural factors which have played a major role in defining how modern humans live and its impact on our health: the development of agriculture and the development of towns.

The earliest modern *H. sapiens* lived as small mobile bands of foragers. They moved to find new food supplies if there were ecological changes that affected their environment. As populations grew they would have dispersed across wider and wider geographical areas and the environments to which they had to adapt would have become more diverse. But *H. sapiens* achieved a capacity to control its environment through fire, tools, clothing, and building shelters. It is this innovative capacity which made, and still makes, humans such an effective generalist species.

Agriculture developed independently in several geographical areas. Its earliest appearance was in the Levant of the Middle East about 12,000 years ago, and independent developments appeared about 5000 years ago in sub-Saharan Africa, in parts of North and South America, and more recently in New Guinea. The adoption of agriculture was a gradual process, and whether it preceded or followed the development of the first villages is unclear (Kelly 1992). There is, for example, evidence of a village in northern Syria which was settled some 13,000 years ago, prior to evidence of agriculture; this may have been a traditional forager clan which, while living in one place, was able to forage locally for abundant wild cereals and hunt game. There were similar historical forager societies in British Columbia which lived as sedentary groups in villages because food was plentiful locally.

Climate change may have played a role in the development of agriculture. The cooler conditions in Mesopotamia about 12,000 years ago would have promoted the growth of wild cereals and attracted game, and this would have allowed the adoption of a sedentary lifestyle. The development of an agricultural lifestyle implies a degree of social complexity to allow the organization of both pastoralism and crop care. So there is a direct link between the evolution of agriculture and the shift to living in villages.

These changes had profound implications for the health of our species (Cohen and Armelagos 2013). First, there was a progressive change in diet: as plants were artificially selected to allow
for planting and cultivation, the balance of foods changed. Herding, at least in some societies, allowed the collection of milk as a food source as well as access to meat on a more consistent basis. The domestication of animals brought humans into much closer contact with them, with a consequent increase in the risk of infectious disease. Many viral, bacterial, and parasitic diseases such as influenza, tuberculosis, and toxoplasmosis have their origin in domestic animal hosts (see Section 10.5). Further, as populations became sedentary and lost the capacity to forage, they were more at the mercy of drought, floods, and crop failure. Reliance on a limited variety of crop plants may have caused vitamin and mineral deficiencies. Malnutrition and disease thus paradoxically become a greater risk with the development of agriculture. Larger cities created problems of hygiene, which had become a catastrophic problem in Europe by the nineteenth century.

Second, settlement brought a drastic change in the way we live. Groups of people began to live in a common place and, as these groups got larger, skills started to be separated in society, and a group structure with traditions, rules, and beliefs was required. Political structures had to be developed. The tendency to urbanization, complex social structures, and differentiation of social roles has continued unabated since that time. As we shall discuss in Chapter 11, this in turn creates pressures on our capacity to cope and may play a major role in the origin of mental disorders as well as impacting on our physical health.

As society became more complex and formalized, political and formal religious systems evolved. Religion probably first evolved informally as a way for our prescient species to explain natural phenomena, and perhaps to propitiate against possible future disasters (Boyer 2001). As belief became a central component of social organization and identity for a group, enshrined in ritual and in special knowledge held by shamans, the capacity for religion to become a force for social control became apparent. Thus religion and political power became linked. Indeed much of history over the last 4000 years has been dominated by religion defining an in-group and violence towards those who are seen to be non-compliant or in out-groups. The beginnings of medicine are linked very much to the evolution of religious belief, with the power of healing being linked to shamanism, special knowledge, and invocation of the supernatural. Healing gods became a focus for, and their priests the providers of, primitive medicines.

### 6.4 Genomic Changes That Make Us Human

The anatomical and behavioral changes that distinguish *H. sapiens* from other hominins must be underpinned by genomic differences (O’Bleness et al. 2012). Indeed, there are a number of genomic regions that have shown particular change since the most recent common ancestor of humans and chimpanzees, which probably lived about 7 Mya, suggesting that these “human-specific” sequences contribute to some of the unique features of *H. sapiens*.

The human and chimpanzee genome sequences vary by about 1% in single nucleotide differences, or approximately 35 million bases. However, insertions and deletions of longer DNA sequences (about another 90 million bases) bring the total difference up to around 4%. By comparison, the genome sequences of modern humans and Neanderthals differ by less than 0.2%. The most obvious difference between chimpanzees and humans is the karyotype: what are two separate chromosomes in early hominids have fused to form human chromosome 2, so that chimpanzees, gorillas, and orangutans have 24 chromosome pairs while humans (including Neanderthals and Denisovans) have 23. This fusion is estimated to have occurred around 4–5 Mya.

Research at a more detailed sequence level has revealed the existence of DNA segments that changed little during most of mammalian evolution but have shown accelerated rates of base substitution since the divergence of humans and chimpanzees. Most of these human-accelerated regions (HARs) are in non-coding but regulatory segments of the genome, proximal to developmental genes. Comparisons of *in vitro* gene expression under the control of HARs versus their equivalent primate regions show human-specific changes in gene expression in the brain, eyes, and limbs. For example, the human version of the *HACNS1* sequence is particularly strongly...
active in developing limb joint regions, leading to speculation that the human-specific sequence changes in \textit{HACNS1} might contribute to unique human digit and limb patterning, such as the fully opposable thumb and the features of the foot that allow permanent bipedalism (Prabhakar et al. 2008).

Glycans are the carbohydrate components of the glycoproteins that coat cell surfaces. Glycans act as recognition sites for various molecules involved in the regulation of cellular interactions, and also as binding sites for a number of pathogens and parasites, facilitating their entry into the cell. A mutation that inactivates the \textit{CMAH} gene, which codes for the enzyme cytidine monophosphate-\textit{N}-acetylneuraminic acid hydroxylase, means that all human cell-surface glycans are terminated by \textit{N}-acetylneuraminic acid instead of the \textit{N}-glycolyneuraminic acid found in nearly all other mammals, including the great apes. This mutation is specific to the human lineage and is estimated to have occurred about 2.5 Mya. In parallel, human-specific adaptations have occurred in a number of glycan-binding proteins (collectively known as Sigles) involved in regulation of the innate and adaptive immune systems. This change in cell-surface properties has a number of consequences for the susceptibility of humans to infectious disease and other immune-associated pathologies (Varki 2009).

Another case where loss of gene function may have contributed to evolution of human-specific anatomy is a deletion after the chimpanzee–human split affecting a myosin gene (\textit{MYH16}) specific for jaw muscles (Stedman et al. 2004). The resulting reduction in size of the jaw muscles may have contributed to the evolution of the hominin face towards a more human-like shape and, some have suggested, to the increase in human brain size (see also Section 13.8.2).

The copy number of sequences coding for the DUF1220 protein domain varies widely across mammals, with humans having nearly 300 copies, chimpanzees half that number, and mice only one. DUF1220 domain copy number correlates with brain size in interspecies comparisons and in humans with developmental brain disorders (Dumas and Sikela 2009). DUF1220 is located mainly on a genomically unstable region of chromosome 1q, in which copy number variation has been implicated in various human diseases affecting cognitive function. Despite these suggestive associations, it remains unclear whether DUF1220 copy number actually underlies the expansion in human brain size.

Epigenetic regulation is also more complex in the human brain. In particular, RNA editing is two orders of magnitude more complex than in mice, and humans have higher levels of edited and multi-edited RNAs than any other species, including other primates, particularly in the brain. RNA editing involves the APOBEC and ADAR enzyme systems, and leads to a greater complexity of regulatory RNAs. APOBEC1 has been shown to be under positive selection in humans. It is particularly apparent in \textit{Alu} elements (see Box 3.2), which are distinct to primate brains; humans have the highest concentrations of \textit{Alu} elements in the 5′ untranslated regions of non-coding genes. There is evidence that RNA editing is environmentally sensitive, and it seems likely that this complexity of RNA regulation plays a role in memory and learning (Mattick and Mehler 2008).

Apart from these and a few similar examples it has in general been difficult to link genomic differences with changes in phenotype between other primates, early hominins, and modern humans. At this stage the only conclusion we can draw with certainty is that there are no large changes in the genome that could account for these phenotypic differences. Human-specific features are likely to have arisen as a series of small changes within the hominin genome, often in regulatory rather than protein-coding sequences.

6.5 Human Adaptation to Local Selection Pressures

6.5.1 Hominin Origins and Migrations: Out of Africa Again

Archeological, molecular genetic, and linguistic data are now sufficient to document with reasonable certainty the origin and dispersal of modern humans. It is clear that there have been two major dispersals of \textit{Homo} species from Africa within the
last million years. As described in Section 6.3.1, *H. erectus* spread out of Africa into southern Europe and Asia about 1 Mya, transitioning to their successor species *H. heidelbergensis* and in turn into the Neanderthals who populated Europe from about 300,000 years ago up to about 25,000 years ago. We can also infer from fossil and genomic data that the Denisovans, probably also derived from *H. erectus*, were present in parts of Asia at least as recently as 50,000 years ago. Anatomically modern humans, *H. sapiens*, evolved from *H. heidelbergensis* in East Africa between 200,000 and 160,000 years ago, but remained largely restricted to that continent for many millennia until about 50,000–60,000 years ago, when a relatively small group of those people departed from eastern Africa across the mouth of the Red Sea. They became the founder population for all modern humans living outside Africa.

Genetic and other evidence points to a sub-Saharan African origin for anatomically modern humans (Box 6.7), and indeed the genetic lineages found in the Khoisan people of southwestern Africa (also known as the !Kung san: we will meet them again in Section 9.3.1.2 as prototypical hunter-gatherers) provide evidence that they are among the oldest extant human populations. Human mitochondrial DNA lineages coalesce about 140,000 years ago, or in other words they can be traced to a single woman who lived at that time, a finding that prompted much media comment about “mitochondrial Eve.” Naturally, this does not mean that she was the only woman living at the time. Women who only have male offspring do not pass on their mitochondrial DNA and there will have been many other women living at the time of “Eve,” although their genetic legacy, at least in the form of mitochondrial DNA,

---

**Box 6.7 Mapping Human Migration**

The use of mitochondrial DNA and Y-chromosomal DNA to map human origins and migrations has caught the popular imagination, resulting in wide publicity for mitochondrial Eve (the most recent female ancestor of all humans, who lived about 140,000 years ago) and for “Genghis Khan’s Y chromosome” (the discovery that about 8% of men now living in Central Asia carry a Y-chromosome haplotype that can be traced back to a specific man who lived in Mongolia about 1000 years ago; Zerjal et al. 2003). Indeed, several companies now offer to report your genetic ancestry on receipt of a cheek-swab sample (and your credit card details). The value of these genetic markers is two-fold: they are inherited only from a single parent, and they are not subject to recombination during gametogenesis. Consequently, DNA sequence patterns (haplotypes) in these types of DNA are passed unchanged to the offspring, and it is possible to re-create the history of human migration by tracking genetic markers in mitochondrial DNA (for the female lineage, or *matriline*) and Y-chromosomal DNA (for the male lineage, or *patriline*) among different human populations.

Another way to look at relationships between populations is to measure the sequence diversity of their whole genomes. In practice, markers such as microsatellites or SNPs are used rather than whole genome sequences. Diversity is calculated using Wright’s $F$ statistics (named after the influential mathematical geneticist Sewall Wright, who developed the approach). This method derives a coefficient—usually called $F_{ST}$, where $S$ stands for subpopulation and $T$ for total population—that shows the relative level of genetic variation between subpopulations of a species. Greater variation suggests a longer history of population subdivision, resulting from a correspondingly longer time for selection and drift to operate. Calculated $F_{ST}$ values for human populations correlate well with geographical distance between them, especially when corrected for historical constraints on migration such as the need to travel over land. For instance, $F_{ST}$ is greatest for African versus South American populations, as would be expected from the patterns of migration shown in Figure 6.4.

Overall, human genetic diversity supports the model of a serial founder effect (in other words, migration and settlement followed by further onward migration of part of the population—originating in Africa, as also predicted by the studies on specific mitochondrial DNA and Y-chromosomal markers (Henn et al. 2012).

Other lines of evidence support a serial founder effect. Genetic variation in microorganisms associated with humans, such as *Plasmodium* (the malarial parasite) and *Helicobacter* (a bacterium found in the digestive tract) also shows a pattern of declining diversity with distance from Africa. More controversially, human language may show a similar pattern, with high diversity among phonemes (the building blocks of language) in Africa, and reduced diversity elsewhere (Henn et al. 2012).
has been lost. Studies of Y-chromosome variation point to a somewhat later coalescence; the apparent discrepancy may be the result of a smaller effective population size (Section 3.4.1) in males possibly arising from the socio-cultural phenomenon whereby, in many traditional societies, a few dominant men father most of the children.

During the period of about 100,000 years between the emergence of modern humans and the beginning of their worldwide expansion, the *H. sapiens* population in Africa appears to have lived in small and geographically isolated groups that only began to mix towards the end of that time. The event that precipitated the out-of-Africa migration will probably always remain unknown, but displacement by some form of intergroup rivalry is a possibility. Certainly the departing group was relatively few in number, because they carried only a small proportion of African genetic diversity. Remarkably, of the 40–50 human matrilineal lineages (as assessed by analysis of mitochondrial DNA) that existed in Africa at the time of the out-of-Africa event, only two—known as L3M and L3N—actually contribute to current worldwide non-African diversity. Similarly, all non-African Y chromosomes carry the mutation M168. This founder effect means that modern African populations carry the highest amount of human genetic diversity, and people in the regions that were last to be populated by modern humans (western Europe, the Americas, and Oceania) the least.

Where did the African emigrants go? The world was colder 50,000 years ago, the icefields of the Northern Hemisphere had sequestered large amounts of water, sea levels were perhaps 100 m lower than they are today, and the coastlines of the continents stretched much farther out to sea than they do now, with some present-day islands forming single landmasses with the adjacent coasts. These conditions allowed human expansion around the tropical beaches of southern and southeastern Asia, and there is archeological evidence for the presence of humans in Australia by around 40,000–50,000 years ago. Another group appears to have headed northward through the Arabian Peninsula to the Levant and Mesopotamia, thence spreading eastwards across the steppes of Central Asia to populate Siberia and, for those who traveled south of the Central Asian massif, India and eventually China. There is also evidence that some of the earlier coastal migrants moved northwards through Southeast Asia. Westward migration of modern humans into Europe appears to have come from two sources: via a Mediterranean route from the Levant, and from Central Asia via a route north of the Black Sea.

Thus by 40,000 years ago modern humans had populated essentially all of Africa, Europe, and Asia (Figure 6.4). We know that these modern humans encountered, and eventually replaced, some remnant populations of descendant species derived from the earlier migration out of Africa by *H. erectus* over a million years previously. There is good archeological evidence that humans and Neanderthals coexisted in the Levant around 55,000 years ago, some populations of Neanderthals persisted in the Iberian Peninsula up to 25,000–30,000 years ago, and *H. erectus* fossils in Java may date from as little as 27,000 years ago. Molecular data show that some of the encounters between these groups have left traces of the archaic genomes in our own species (see Box 6.1).

The Americas were populated relatively recently by humans, probably no earlier than 14,000 years ago. The most commonly accepted model is that of one or more migrations by relatively small groups of Siberian hunters over the landmass called Beringia, joining present-day Siberia and Alaska, which existed from about 20,000 to 8000 years ago because of the fall in sea level during the last glacial period. The North American glaciers began to melt about 15,000 years ago, opening the way for a southward migration that reached the tip of South America only 1000 years later. The original founding population was very small. The low genetic diversity of native Americans has caused some investigators to propose an effective population size for the migrating group of no more than 70 individuals (but remember that, for the reasons explained in Section 3.4.1, effective population size always underestimates true population size). An example of this lack of diversity is the low frequency of blood type A, and the virtual absence of blood type B, in native American populations—a classic example of a founder effect.

The requirement for advanced technologies to enable humans to cross oceans meant that Oceania...
was the last part of the world to be populated by *H. sapiens*. Current theories of the peopling of Oceania suggest a southward and eastward migration of the Taiwanese Austronesian culture, beginning around 5500 years ago and reaching the last significant landmass, New Zealand, around 800 years ago.

### 6.5.2 Variation Caused by Migration

The migration of humans around the world created two sources of variation. Variations created by genetic drift, founder effects, and population bottlenecks (Chapter 3) are, by definition, non-adaptive in origin. But as *H. sapiens* spread into different ecological situations, the potential for selective pressures to create variation both within and between populations became present. Thus local gene pools came to vary for both adaptive and non-adaptive reasons (Lachance and Tishkoff 2013). Selective pressures on humans during their migration (Table 6.3; see also Figure 13.1) would have included changes in climate, for example in temperature and sun exposure, as well as changes in food sources and availability, which would themselves have been affected by climate.

When species spread over a geographical range their phenotype may vary because of the differing environments faced by the different populations. Rabbits (which were only introduced into Australia as a small founder population some 200 years ago) in northern Australia have longer ears than their conspecifics from southern Australia, presumably reflecting selection for longer ears in the warmer latitudes to give better heat loss. The “rules” reflecting such biogeographical effects, such as Bergmann’s rule and Allen’s rule (see Section 5.3.2), play a role in human variation. For example, the Nilotic peoples of eastern Africa are tall and thin, whereas the Inuits of the North American Arctic are stocky and short.

Other clear examples of biogeographical variation involve skin color. Human skin color shows a strong geographical correlation with the intensity of ultraviolet light in a person’s region of ancestry, and the possible reasons for this are discussed in Section 13.4.1. The African origin of modern humans means that dark skin is their ancestral condition (although as mentioned in Section 6.3.6 earlier hominins probably had lighter skin), but genetic studies show that lighter skins have subsequently evolved independently, each involving changes in a few genes, in European and East Asian populations. Sequencing of
DNA fragments recovered from a Neanderthal skeleton suggested that that individual at least had red hair and fair skin, while the Denisovan individual sequenced probably had brown hair and dark skin.

Differences in hair morphology between human populations, with East Asians having a coarser hair texture, appear to result from a single amino acid change that increases the effectiveness of a particular cell-surface receptor protein called EDAR. The mutation probably occurred during human migration through Central Asia, since it occurs at high frequency in East Asian and Native American populations but is virtually absent in Europeans and Africans. The adaptive significance, if any, of coarser hair is unclear, but the receptor also has effects on the development of teeth and sweat glands, suggesting that the change in hair morphology may be a spandrel (Section 2.4.4) arising incidentally from an adaptive change in some other trait.

Migration from equatorial Africa may also have caused selection in physiological traits related to cold tolerance, such as energy metabolism, reactivity of peripheral blood vessels, salt retention, and a tendency to deposit fat. Climate-related variations in the allele frequencies of numerous genes relevant to these physiological processes have been demonstrated, for example in a keratin gene specifically expressed in sweat glands, underlining the importance of climate as a selective agent during human evolution (Hancock et al. 2011).

### Table 6.3 Examples of local adaptation in human populations

<table>
<thead>
<tr>
<th>Selective pressure</th>
<th>Physiology affected</th>
<th>Population</th>
<th>Gene</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geography/climate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude</td>
<td>Oxygen transport and/or response to hypoxia</td>
<td>Tibetan, Andean</td>
<td>Various (different across populations)</td>
<td>Beall (2007)</td>
</tr>
<tr>
<td>Temperature (and/or sexual selection?)</td>
<td>Ectodermal development (pleiotropic for sweat glands, mammary glands, tooth and hair morphology)</td>
<td>East Asian</td>
<td>EDAR</td>
<td>Kamberov et al. (2013)</td>
</tr>
<tr>
<td>Low sunlight</td>
<td>Decreased skin pigmentation</td>
<td>European, East Asian</td>
<td>Various (different across populations)</td>
<td>Sturm and Duffy (2012)</td>
</tr>
<tr>
<td></td>
<td>Increased vitamin D biosynthesis</td>
<td>Northern populations in Europe and Asia</td>
<td>DHCR7</td>
<td>Kuan et al. (2013)</td>
</tr>
<tr>
<td><strong>Infectious disease</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaria</td>
<td>Erythrocyte function (Duffy antigen, hemoglobins, glucose-6-phosphate dehydrogenase)</td>
<td>Sub-Saharan Africa, Middle East, Southeast Asia</td>
<td>DARC, HBA, HBB, G6PD</td>
<td>Gomez et al. (2014)</td>
</tr>
<tr>
<td>Trypanosomiasis (sleeping sickness)</td>
<td>Apolipoprotein L1</td>
<td>Western Africa</td>
<td>APOL1</td>
<td>Pays et al. (2014)</td>
</tr>
<tr>
<td>Unknown</td>
<td>Zinc transport</td>
<td>Sub-Saharan Africa</td>
<td>Zinc transport</td>
<td>Engelken et al. (2014)</td>
</tr>
<tr>
<td><strong>Nutrition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-fat diet</td>
<td>Mitochondrial long-chain fatty-acid oxidation</td>
<td>Inuit, Siberians</td>
<td>CPT1A</td>
<td>Clemente et al. (2014)</td>
</tr>
<tr>
<td>Milk consumption</td>
<td>Lactase persistence</td>
<td>Pastoral populations in Europe, Middle East, Africa</td>
<td>LCT</td>
<td>Tishkoff et al. (2007)</td>
</tr>
<tr>
<td>Starch consumption</td>
<td>Salivary amylase</td>
<td>Populations consuming high-starch diets</td>
<td>AMY1 (copy number)</td>
<td>Perry et al. (2007)</td>
</tr>
<tr>
<td><strong>Toxins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arsenic</td>
<td>Arsenic metabolism</td>
<td>Argentine Andean</td>
<td>AS3MT</td>
<td>Schlebusch et al. (2015)</td>
</tr>
<tr>
<td>Plant toxins</td>
<td>Cytochrome P450 diversity</td>
<td>Various</td>
<td>Multiple, e.g., CYP2D6</td>
<td>Ingelman-Sundberg (2005)</td>
</tr>
</tbody>
</table>
6.5.3 Variation Caused by Change in Lifestyles

As we describe in Section 9.3.2, perhaps the largest change in the human lifestyle was the adoption of sedentism and agriculture—the so-called Neolithic Revolution—some 12,000–10,000 years ago. That innovation introduced additional selective pressures on humans, and resulted in changes in several traits. The first involved adaptation to new food sources and to the increased vulnerability to famine as the possibility of crop failure increased. Part of this adaptation may have been cultural, such as the introduction of new means of food preparation and storage. Part would have involved gene–culture coevolution, such as the human association with pastoralism leading to the spread of lactase persistence in several populations. Part would have involved physiological adaptation to new foods, such as cereals with their high levels of starch. Salivary amylase is responsible for starch digestion in humans: the number of copies of the amylase gene varies between individuals, and the more copies of the gene in an individual, the higher their salivary amylase activity. Human populations with starch-rich diets, such as East Asians who have a rice-based diet, have a higher amylase copy number than do populations with diets low in starch, such as high-latitude hunters who consume a largely meat-based diet (Figure 3.2). This adaptation appears to be specific to the *H. sapiens* lineage, as amylase gene duplication was not observed in the Neanderthal and Denisovan genomes (Perry et al. 2015).

The second major selective pressure arising from the move into settlements was the increased prevalence of infectious disease as populations became large and dense enough to support transmission, both directly and via invertebrate vectors. Elsewhere we discuss human adaptation to infectious disease (Chapter 10) and how it explains population differences in the frequency of diseases such as beta-thalassemia (Section 13.9.1).

6.6 Are Humans Still Evolving?

Has human evolution stopped, or at least slowed, because our cultural and technological adaptations now insulate our biology from the selective pressures of the environment? It could be argued that as humans gained control over their environment they increasingly overcame energetic limitations on reproduction. More recently, because social and medical advances enable most people in a society to reproduce, the selective advantages or disadvantages of particular phenotypes (and genotypes) for reproduction have become less important. Genetic isolation is becoming less common as populations are highly mobile and gene flow between previously dispersed groups is facilitated, removing founder effects and bottlenecks as sources of human variation. Perhaps it is only in a calamitous situation such as a worldwide pandemic that selection would occur for those with genotypes which confer relative protection. Such scenarios cannot be ruled out.

But, as discussed in Section 6.5, there is also clear genetic evidence of positive selection for individual genes in humans over the past few thousand years. Even without a pandemic, the continued emergence of new infectious diseases means that the ability to respond to pathogens is likely to remain one of the strongest targets of selection in humans. The persistence of the hemoglobinopathies, sickle cell anemia and thalassemia, is a reminder of the strength of selection caused by malaria. The widespread occurrence in European populations of the deletion in the CCR5 gene that causes partial resistance to HIV infection may be another example of selection caused by some previous infectious disease (Novembre and Han 2012). Changes in nutrition are occurring worldwide, and it is worth remembering that the mutation allowing lactase persistence in East African farmers only occurred within the last 2000 or 3000 years.

Even so, and despite repeated claims to the contrary, the likelihood that humans will evolve to tolerate a nutritionally rich environment without incurring metabolic disease is fanciful. While nutritional mismatch (Chapter 9) may lead to physiological disturbance even in young people, in general the effects within the peak period of reproduction are not sufficient to affect reproduction and thus no selection pressure is created. Advances in medical care and family planning make those who are obese (a condition often associated with infertility) and those who are not obese largely equal in terms of
reproductive capacity and survival through peak reproduction. Assortative mating, whereby people tend to mate with people with similar characteristics, is also a complicating factor. For example, if relatively obese people marry other relatively obese people, genetic as well as cultural factors involved in driving obesity may be consolidated within families.

From the perspective of classical population genetics, there is no reason why humans should not still be evolving. Humans are probably the most abundant of any animal of similar size that has ever lived on the Earth, with a current population of about 7.3 billion. A larger population means that a larger number of potentially adaptive mutations will occur in each generation, providing an increased amount of variation for selection to work on. Recall also from Chapter 3 that genetic drift can overcome the effects of selection in situations of small population size and limited gene flow, leading to the notion of an effective population size. Historically, human effective population sizes have been small. But humans now have a very large actual population size with less reproductive isolation, thanks to modern travel and migration patterns. Further, social factors have reduced the effect of assortative mating. These conditions result in a higher effective population size, increasing the efficiency of selection. The large population size also means that in aggregate there is a higher stochastic probability that the majority of potentially adaptive base changes in the human genome will occur and will, over relatively few generations, be tested against the current environment.

It is thus difficult to find the balance between those issues which constrain ongoing selection from those which might drive it. Has our capacity to overcome the selective pressures of the environment effectively abolished positive selection other than through catastrophe? Will culture rather than environment provide the major selective pressures for further human evolution?

Gene–culture coevolution (Section 2.3.3) occurs when cultural and behavioral factors drive biological evolution, as illustrated by the following thought experiment. Because of advances in public health that have nearly eliminated infant mortality, nearly all children born in high- and middle-income countries survive to reproductive age. The possibility for selection on survival in such societies has all but disappeared, and any selection must be based on differences in fertility. There is a recent tendency for women in high-income countries to delay their pregnancies as they increasingly contribute to the workforce (Figure 6.5). Older women

Figure 6.5 Cultural evolution. The peak age of first birth in Canada increased by 5 years from 1976 to 1996. From Lochhead (2005), with permission.
Box 6.8 Demographic Studies Reveal Contemporary Human Evolution

One way of studying how societal changes might impact on human evolution is to look at the so-called “demographic transition.” This is the change in low- and middle-income countries from high birth and mortality rates (and hence survival-driven selection) to low birth and mortality rates (and hence fertility-driven selection) as they advance economically, leading eventually to stable population levels. The demographic transition can also be traced in the history of modern high-income countries (Moorad 2013). The transition is brought about by a number of factors, including improved nutrition and public health, increased access to technology, and enhanced educational opportunities (particularly for women).

One study looked at the relationship between female phenotype (height and BMI) and fitness across the demographic transition in rural Gambia (Cortiol et al. 2013). It confirmed that changes in relative fitness during the transition were predominantly related to improved survival of children, but also showed fitness differences between particular female phenotypes: before the transition “selection favoured short females with high BMI values,” whereas after the transition “selection now favours tall females with low BMI values.” Whatever the cultural or physiological reasons behind the change in selection pressure, the authors pointed out that height and BMI are highly heritable traits, and the phenotypic changes are likely to be reflected in corresponding genetic changes.

Evolutionary change has also been demonstrated in a contemporary population from a high-income country (Byars et al. 2010). Analysis of female participants in the multi-generational Framingham Heart Study in the USA indicated that selection is acting to increase reproductive span via earlier first birth and later menopause, to decrease height and increase weight, and to decrease blood cholesterol and blood pressure. However, it is difficult to be certain about the extent to which these effects are biological or are confounded by cultural evolution changing reproductive practices over that time.

Natural selection favoring earlier age at first birth has also been demonstrated in a nineteenth- and twentieth-century rural population from Canada (Milot et al. 2011). It is worth noting that these biological changes towards earlier first birth, which may reflect the secular trends for a fall in the age at menarche (Section 5.4.2.2), occurred before the cultural changes of the 1980s onwards that have led many women to delay their first birth (Figure 6.5).

6.7 Social and Medical Implications of Human Diversity

Modern humans evolved in Africa and spread across the world, adapting locally to the selective pressures of the climates, food sources, and pathogens that they encountered. That adaptation has resulted in the visible differences in phenotype—body shape, skin color,
hair color, and hair texture—that provide clues to the geographical ancestry of an individual, as well as many invisible differences, for instance in digestive ability and in disease resistance. Although humans are at least 99.9% identical in the sequence of their genomes, 2–3 million bases will still be different between any two unrelated individuals. These polymorphisms will have arisen through random and mostly neutral mutations accumulating in family lineages and will be distributed differently over the world, so that statistical analysis of the pattern of polymorphisms in any one person usually allows identification of their geographical ancestry (Elhaik et al. 2014).

It is a convenient administrative shorthand to classify people into a few groups depending on their appearance, and call those groups “races”. Most of us will be familiar with census forms asking our “race” and/or “ethnicity.” (Note that “race” refers to physical appearance; “ethnicity” is a more nuanced categorization that includes not only geographical and family ancestry but also factors such as religion, language, and cultural traditions.) But does the concept of “race” correspond to any biological reality?

If we fly from, say, Scandinavia to West Africa the people whom we encounter at departure and arrival will look very different. But if we were to take that journey southward in a more leisurely fashion, overland and by boat, we would recognize a gradual and progressive change in the skin color of the people we meet. But nowhere would there be a sudden change which would allow us to distinguish those people currently north of us as “race A” and those people south of us as “race B.” So it is with many if not most human characteristics: there is a continuous geographical gradient of a trait rather than discontinuous variation. These gradients are called clines. Clines are discordant in direction. The cline of skin color in our example follows a north–south orientation, but the cline of distribution of blood group B in Europe follows an east–west orientation. As would be predicted, genetic analysis of the alleles underlying clinal traits demonstrates corresponding clinal variation in allele frequencies. Therefore, our genotype, and consequently our phenotype, is determined by the position of our geographical origin on a multitude of intersecting clines: somebody from Central Europe is more likely to have lighter skin color and lactase persistence than a person who originated 1000 km south, but less likely to have blood group B than a person who originated 1000 km east. This continuous variation in human traits on gradients which vary in orientation means that there are no boundaries, geographical or genetic, that can be used to assign individuals to particular races.

From a medical perspective, does the social construct of “race” correlate with clusters of genetically determined traits, and hence with disease risk, in any way that is useful for diagnosis, treatment, or prevention? Superficial consideration of some of the examples in this book would suggest that to be the case, but we caution against extrapolating from populations to individuals. For example, in Chapters 3 and 13 we use sickle cell disease, highly prevalent in West Africa and in individuals originating from there, to illustrate how balancing selection can cause a deleterious trait to be maintained in the population. But a clinician treating someone presenting with pain and anemia should not rule out sickle cell disease if her patient does not have recent West African origin, because hemoglobinopathies with similar clinical presentation also occur in individuals from other areas, such as the Mediterranean and the Middle East, where humans have had to adapt to the malarial parasite.

In Chapter 9 we discuss how people from South Asia appear to be more susceptible to certain metabolic disorders at a lower level of obesity than others; this has led to a call for population-specific algorithms for predicting diabetes risk. Although population differences in patterns of fat deposition have been interpreted by some as the outcomes of “thrifty” alleles selected by ancestral famine, the effect of obesity on metabolic risk varies widely among individuals across all continents, with genetic, developmental, and environmental determinants. The higher susceptibility in the South Asian population is perhaps best interpreted in terms of a steeper reaction norm (see Section 4.1) underpinned by polygenic variation.

Polymorphisms in drug-metabolizing enzymes across populations lead to differences in the pharmacokinetics of certain medications, with potential implications for treatment success or adverse effects. An example is dosage adjustments of
warfarin (Jorgensen et al. 2012). Even so, genotypes vary enough within populations that information on individual genotype and phenotype, which may also be affected by diet and other medications, is required to adequately predict treatment outcome (Yasuda et al. 2008).

The important point we wish to underline from these examples is that while broad-brush examples of population differences in disease risk are useful for illustrative purposes, interindividual variation, both genetic and environmentally determined, is much more significant than variation across populations. Individual diagnostic and treatment decisions need to be based on the history of the individual patient in the context of their social and economic environment, rather than any notion that geographical ancestry alone affects disease risk.

Our critique of the perception of “race” as a biological determinant of disease risk does not deny the importance of population stratification as a social determinant of health. Ethnic minorities in most countries have consistently poorer health and life expectancy. Economic, social, and public health efforts to correct such disparities have had some limited success; the few clinical interventions targeting so-called racial differences in biology have been more controversial. An example is the launch in the USA of a drug for congestive heart failure aimed at African-Americans, who have a higher prevalence of the disorder and of one of its underlying conditions, hypertension. The drug’s biological rationale is to target population-specific differences in the metabolism of the vasodilator nitric oxide, and a clinical trial that recruited only African-Americans showed marked benefit. However, critics of the drug point to the lack of a direct comparison of its efficacy across population groups and to the likelihood that the differences in disease prevalence between African-Americans and other US ethnic groups are more likely to be a result of socioeconomic disparities than biological differences (Brody and Hunt 2006).

6.8 Conclusion

Molecular studies have identified that the hominin clade diverged from other apes some 7–10 Mya. The limited nature of available specimens makes the study of hominin evolution itself more uncertain, but there have been many hominin species since that point of divergence, most of which were not our direct ancestors. Anatomically modern H. sapiens appeared first in Africa about 160,000 years ago from a precursor species that shared a common ancestor with the Neanderthal over half a million years ago. Cogent but speculative adaptive arguments can be put forward linking ecological change to the emergence of bipedalism. All hominins, and particularly humans, are characterized by a large brain, which is likely to have evolved because of the fitness value of living in cohesive social groups. There remains much uncertainty as to when language evolved.

Recent molecular studies show that the frequencies of particular alleles have changed within human populations in the last few thousand years, suggesting positive selection: an example is the emergence of lactose tolerance. There is some evidence for the continuing evolution of the human phenotype in modern populations, but the nature of evolutionary processes means that the potential for and extent of further biological evolution of our species is inherently impossible to forecast.

Biological evolution in humans cannot be separated from cultural evolution. More recent human history is characterized by the development of agriculture and large population clusters. Throughout this book we illustrate how this cultural evolution has changed our biotic, physical, and social environments, with consequent impacts on human health.

There is much genetic and phenotypic diversity in our species, which arose as humans migrated from Africa around the globe in the last 60,000 years. Some variations represent adaptations to local climates, others the effects of population bottlenecks and founder effects. Despite this phenotypic diversity, genetic variation within a population exceeds that between populations. There is no biological basis for the concept of race. Clinicians should base diagnostic and treatment decisions on individual genotype and phenotype rather than on perceived notions of broad population differences in disease susceptibility.
Key Points

• Humans shared a last common ancestor with the chimpanzee and the bonobo at least 7 Mya.
• The hominin clade has included many species, some of which will have coexisted. Few are in direct ancestral relationship to *H. sapiens*.
• Humans are distinguished from the other great apes by specific physical (bipedalism, brain size, hairlessness) and cultural (tool making, language, society) traits.
• Modern humans evolved in Africa and spread across the world, adapting locally to the selective pressures of the climates, food sources, and pathogens that they encountered.
• Recent human cultural evolution has to a large extent insulated our biology from the selective pressures of the environment, although pandemic diseases may represent an exception. Gene-culture coevolution is likely to be a major driver of future human adaptation.
• Human diversity is best explained by continuous rather than discontinuous variation. There is no biological basis for the concept of race.
PART 2

Evolution in Health and Disease
CHAPTER 7

An Evolutionary Framework for Understanding Human Health and Disease

7.1 Introduction

Nothing in medicine makes sense except in the light of evolution.

The above dictum is a slight reworking of a famous quotation from Theodosius Dobzhansky, one of the most influential evolutionary biologists of the middle of the twentieth century and a key player in the Modern Synthesis (see Chapter 2 and Section 14.5). We have replaced his word “biology” with the word “medicine,” but the essence of the message is the same: evolutionary processes provide the unifying and integrating principle of all biology. As we have discussed in the preceding chapters, evolution has shaped all aspects of our biology at every organizational level from our genome to our social structure and function.

This chapter is a bridge from the first part of the book, which focused on medically relevant aspects of evolutionary theory, to conceptual and practical applications to clinical medicine and public health. Our goal is to demonstrate how knowledge of evolutionary processes can extend our understanding of human biology, health, and disease. Generally the practice of medicine focuses on the issues of proximate causation—namely the actual physiological and anatomical disruptions that lead to disease, because it is these pathological processes that inform most diagnostic and therapeutic choices. That is why pathology textbooks classify disease according to proximate mechanisms: traumatic, inflammatory, infective, autoimmune, neoplastic, and so forth. But what this book aims to demonstrate is that a more holistic understanding of health and disease also requires a consideration of ultimate, that is evolutionary, pathways that affect the risk of developing disease or promote resilience to it. While the value of evolutionary considerations is primarily that of changing perspective and understanding, we will demonstrate in this and subsequent chapters how evolutionary explanations do impact on clinical decision making and on the practice of public health.

Further, because the evolutionary framework requires a consideration of the ecological context of the individual and the population, be it the physical, biotic, or social environment, it can provide greater understanding of why individuals may present with illness or problems, or alternatively may be resilient. Indeed, much of public health and medicine in effect relies on evolutionary principles, even though this is often not appreciated. For example, antibiotic resistance is simply the demonstration of natural selection in rapidly reproducing bacteria within a toxic (to the bacteria) environment created by the use of antibiotics.

In this chapter we provide a brief categorical framework of the various ways in which evolutionary considerations can affect the patterns of health and disease, and individual risk. This framework will then be revisited in Chapter 13, which will both summarize the clinical section of the book and provide further case studies. In general, we are not arguing that evolutionary processes cause disease, rather that they have important effects on the relative risk of developing symptoms or
disease. While we will describe a number of distinct pathways, any categorization is somewhat artificial, and more than one evolutionary process may be involved in the generation of individual risk. Nevertheless, just as a pathogenetic classification is useful in describing proximate pathways to disease, it is valuable for the physician and public health practitioner to have an understanding of the various pathways by which our evolutionary history affects disease risk.

7.2 Fundamental Principles of Evolutionary Medicine

There are a number of general principles that we have reiterated throughout the first part of this book and which provide the basic conceptual framework for evolutionary medicine (Gluckman et al. 2011a). These are summarized in Box 7.1.

Central to an evolutionary understanding of disease is the recognition that selective processes operate to enhance inclusive reproductive fitness. This does not necessarily mean selection for health or longevity. There is generally no direct fitness advantage to living a healthy or a very long life; there is only a substantial selective advantage in reproducing successfully, although that may require parental health until the offspring are mature and thus create selective pressure for life beyond the immediate reproductive period. Furthermore there are some situations where fitness can be advanced indirectly through kin selection: this is the most generally held hypothesis for the origin of the menopause in human females (see Section 8.9.7).

Every species evolved with distinct life-history characteristics that were appropriate for the environmental and ecological niches that it occupied; the resultant phenotype is a well-integrated outcome. The human life-history strategy has allowed, and to a large extent was driven by, the evolution of a large complex cerebral neocortex, which requires a prolonged period of postnatal maturation dependent on a high level of parental investment (Chapter 5). Thus each human female can support only a small number of progeny through her life. Homo sapiens evolved to live in relatively small, highly cooperative, groups and much of our social behavior is influenced by the need to ensure group cohesiveness (Chapter 11). Our species evolved the essentially unique capacity to progressively develop an extensive material culture, for example fire, tools, clothing, dwellings, and technology. Biological evolution, cultural evolution, and their interaction have allowed humans to live successfully as a generalist species across a broad range of environments. For example, the development of clothing from fibers and animal skins has allowed humans to live in cold climates, such as those found

Box 7.1 Fundamental Principles of Evolutionary Medicine

- The phenotype, at any age including in maturity and in any individual, is the result of genomic and non-genomic inheritance (of various types) and the cumulative effects of the developmental environment, and is further defined by the interaction with the current environment.
- Our history as a species through our particular lineage, and then our developmental history, influence our susceptibility to disease.
- Selection operates to enhance fitness, not primarily to enhance health or longevity.
- Humans now live in very different ways and in different environments from those where the majority of selective processes affecting the modern human phenotype operated. In this respect we are biologically “mismatched” to many aspects of our current environment.
- Constraints on evolutionary processes (the speed, substrate, or direction of selection, or the scope of plasticity) in the presence of environmental novelty and rapidly changing environments can lead to ill-health.
- Definitions of normality, abnormality, and disease are not absolute and are influenced by the environmental context of the individual and the individual variation in phenotype.
- In applying evolutionary medicine it is essential to avoid the trap of teleology: there is no pre-ordained plan, purpose, or design to evolution.
at high altitudes, but selective processes have also operated to allow more effective tissue oxygenation despite low oxygen pressure, as observed in the Tibetan population (Section 13.11.2). In turn, the Tibetan adaptation may have been possible because a particular allele was present in that population as a result of past introgression from another hominin species.

The human body contains considerable evidence of organizational compromise and exaptation, reflecting our evolutionary history, as well as novelty. But, as we have emphasized, the mature phenotype is not solely a function of the genome. We have evolved a series of mechanisms which allow the environment in early life to mold our phenotype (most broadly defined to include molecular, functional, and structural components) through the processes of developmental plasticity (Chapter 4). Furthermore, inheritance involves more than genes: it may involve epigenetic marks that persist across generations, and it certainly comprises cultural inheritance. The phenotype at any age is thus a construct of inheritance, including genomic and non-genomic inheritance of various forms, and the result of developmental plasticity, itself informed by environmental history. It is defined further by the interaction with the current social, physical, and biotic environments.

With respect to all that we consider in this book, it is important to think in terms of variations or changes in disease risk rather than viewing ultimate mechanisms as leading directly to causation of disease. Being exposed to the influenza virus only increases the risk of contracting influenza; it does not cause influenza, which is the result of the virus invading, avoiding immune defenses, and compromising the cells lining the respiratory tract. Similarly, in considering evolutionary pathways to altered disease risk, we are focusing on how the phenotype confers lesser or greater risk in a particular context.

Often, increased disease risk results from the constraints on the rate of, or the potential for, evolutionary change in the face of a changing or challenging environment. The context of human existence has changed in many ways: many more of us now live much longer lives because of the impact of public health and medicine. Longer lifespans expose phenomena such as degenerative diseases that were previously not seen (Figure 7.1). We live in different physical, social, and energetic environments from those of our ancestors and can use technologies to deal with many challenges, for example immunization against bacteria and viruses. Technological and social innovation continues, and we can only speculate on the ultimate impact of some of these changes. But already there is evidence that the move to virtual forms of communication and interaction may be affecting brain development and in turn might create selective pressures. As discussed in Chapter 6, there is evidence that selective pressures continue to affect human evolution.

The consideration of health and disease from an evolutionary perspective raises some interesting challenges. It can challenge some widely held definitions of disease. In Chapter 1 we questioned whether a person with lactose intolerance has a disease or not: most humans are lactose intolerant and most, until recently, have not been exposed to lactose after they have been weaned from human milk so they have had no health problem. Similarly, does the person who is adapted to a low plane of nutrition (Chapter 9) have a disorder? He or she may remain healthy unless exposed to an obesogenic environment, when they will be at higher risk of obesity than someone with a differently adapted metabolic capacity. Conversely, an infant better adapted to a high-nutrition environment is more likely to develop the more serious syndrome of kwashiorkor if faced with famine, because their metabolic physiology has not adapted by the processes of developmental plasticity to cope with severe undernutrition (Box 7.2). In both cases, the potential proximate cause of ill-health is the concurrent nutritional environment. But understanding why there is such variation in susceptibility requires an understanding of ultimate processes as well as the proximate mechanisms, and that is the sphere of evolutionary medicine. And as we shall see in both cases this has potential implications for public health interventions.

Before considering the evolutionary pathways to increased disease risk in a systematic framework there are some important caveats that need to be reiterated. Because of the pervasive use of
metaphor in evolutionary biology (often seen as necessary to describe something complex or abstract) it is easy to confuse evolutionary analysis with sloppy teleological descriptions. Nothing in biology was “designed” for a higher or future purpose; many traits are present because they were selected for some other adaptive advantage in the past. That is, a trait may have originally been selected for one function but it now serves another function and indeed may have undergone further selection on quite a different basis (i.e., exaptation). The bones of the inner ear have their origin in the jawbones of the teleost fish: quite different selection pressures acted on the fish to first evolve the jawbone and then, as terrestrial mammals and ultimately primates evolved from early fish, the selective pressures built on the earlier development of the jawbone and further evolved these bones.

**Figure 7.1** Ten projected leading causes of morbidity in 2030 worldwide (top) and in high-income countries (bottom). Note the high morbidity arising from depression in both settings. COPD, chronic obstructive pulmonary disease. Plotted from data in Mathers and Loncar (2006).
The predictive adaptive response model (Section 4.5.2.2) posits that early life signals induce developmentally plastic responses that have long-term phenotypic implications. Central to the hypothesis is that these longer-term changes will offer a fitness advantage later in life, particularly in the post-weaning period through to reproduction, but only if the developmental cues predict and match the post-natal phenotype to the actual environment. The hypothesis is supported by studies in rodents (Vickers et al. 2005), but it has been controversial due to a lack of supportive clinical evidence; however, recent studies in infant malnutrition provide strong support for it.

Sustained and severe undernutrition in childhood results in one of two distinct clinical syndromes: marasmus and kwashiorkor. While marasmus is characterized by general wasting, patients with kwashiorkor also present with edema, the extensive accumulation of fluid in bodily tissues. The latter also have a higher mortality rate because of the breakdown leading to a high rate of infection, and they cannot effectively mobilize their own stores of protein and fat. Studies of children who have recovered from these two syndromes show that even after nutritional rehabilitation they have very different intermediary metabolic physiologies (Jahoor et al. 2006). In particular, the survivors of marasmus appear better suited to a low-energy environment, and, as adults, are more likely to be stunted and to develop glucose intolerance (Patrice et al. 2014).

It was hypothesized that the survivors of marasmus and kwashiorkor had developmentally induced differences in metabolic physiology that were present prior to the original presentation of either syndrome in infancy, and that these differences determined how the individual infant responded to the severe malnutrition. To test this hypothesis, a study in Jamaica examined records of 1336 children hospitalized with severe acute malnutrition between 1962 and 1992. It was found that children who presented with marasmus had a significantly lower birthweight (more than 300 g lower) than their peers who presented with kwashiorkor (Forrester et al. 2012).

These findings suggest that children with lower birthweight developed in utero with a metabolic phenotype that was pre-adapted—through the induction of predictive adaptive responses—to a lower post-natal nutritional plane. They were thus able to cope better with post-weaning exposure to undernutrition, developing the lower-mortality syndrome of marasmus. In contrast, those fed more plentifully pre-natally “expected” a higher nutritional plane after weaning and thus, in conditions of severe malnutrition, suffered from kwashiorkor. Given that the mortality rate of untreated kwashiorkor is much higher than that of marasmus, it suggests that there is a fitness advantage of using signals in utero from the mother to make the immediate adaptive response of reducing fetal growth, and also to predict a greater post-weaning nutritional risk. Here the predictive response has led to a survival advantage in infancy, and although it may have costs in later life of stunting and developing insulin resistance, neither has the negative effect on fitness that death in infancy does. This observation may be the most direct evidence to support the predictive adaptive response model in humans, in that it uses survival to infer a fitness effect.

The significance for those who predict a low-nutrition environment during development, but then find themselves developmentally mismatched to a later nutritionally enriched environment, is explored in Chapter 9.
several possible frameworks. Our scheme is largely derived from our further consideration of these various iterations: for example, we have added balancing selection, drift, and founder effects, because these too reflect our evolutionary history. Accordingly, Box 7.3 lists eight possible pathways by which biological evolutionary processes can affect disease risk.

The fundamental question is: why have the various evolutionary processes, and in particular the processes of natural selection, left our bodies vulnerable to disease? At the most integrated level, there are three possible answers. First, human evolution is too slow and cannot cope with either the coevolutionary contest with micro-organisms or with novel environments, especially those of human making such as the obesogenic environment of highly processed, calorie-dense foods. We have, however, evolved with a range of defensive responses to environmental challenges: from stress and immunity to behavioral responses. There are situations in which these responses may become overactive or inappropriately activated and then have health consequences. Secondly, there are many constraints (e.g., speed of evolution, constraints on the magnitude of change that is possible, and constraint resulting from the foundational anatomy and physiology created by the history of the hominin lineage) on what selection can do. Often there have been inevitable trade-offs between competing selective pressures, which themselves have consequences as not all systems can operate optimally under all conditions. Thirdly, there may be consequences from what selection has shaped, because selection is about fitness and not primarily about health, particularly later in life.

Any categorical division of integrated biological processes must necessarily be somewhat artificial and arbitrary. For example in creating a category of life-history-related processes we are conflating some quite distinct concepts ranging from antagonistic pleiotropy to an altered tempo of development. Antagonistic pleiotropy, a concept that describes traits that have been selected for because they are advantageous in promoting reproductive fitness in early life but then may incur costs later (Section 5.2.4), has been implicated in some cases of breast cancer. There is evidence suggesting that the BRCA gene, associated with an increased risk of breast cancer later in life, is also linked with greater reproductive success (Smith et al. 2012; Box 12.3). Yet many other cases of breast cancer have been associated with a novel pattern of human reproduction, arising from the use of contraception and smaller family size (Section 12.5.2). Longer lifespans in themselves reflect technologically driven alterations in our life history. There is a linear relationship between the risk of cancer and rising age (DePinho 2000), so ageing itself, and its associated risk of disease, is a further example of a life-history factor.

Increased disease risk may well involve more than one of these processes, and indeed the boundaries between them are blurred; in many cases detailed evolutionary arguments would lead to further

---

**Box 7.3 Pathways by Which Evolutionary Processes Can Affect Disease Risk**

1. An evolutionarily mismatched or novel environment: the individual has been exposed to an environment beyond their evolved capacity to adapt, or to an entirely novel environment or challenge.
2. Life-history-associated factors (including antagonistic pleiotropy, altered reproductive behavior, and changes in life-history phasing).
3. Excessive or uncontrolled defense mechanisms: evolved processes which are normally adaptive become inappropriate and harm the individual.
4. The consequences of coevolution with microbes.
5. Results of evolutionary (“design”) constraints.
6. An apparently harmful allele is maintained by balancing selection.
7. Sexual selection and competition, and their consequences.
8. The outcomes of cladal and demographic histories.
subdivision. Nevertheless, an ordered analysis of the problem of disease causation using these eight pathways provides a useful systematic way of analyzing a medical problem from an evolutionary perspective.

There are diseases for which an evolutionary perspective may add nothing to their management. Traumatic injury is one such example. But even here an evolutionary perspective can aid in understanding the origins of a problem—for example why young men are more likely to be involved in knife crime. Cars, guns, and other weapons are of course the consequences of our cultural evolution; yet, greater risk-taking in adolescence is an evolved part of human behavior.

7.4 An Evolutionary Classification of Ultimate Mechanisms Affecting Disease Risk

7.4.1 Pathway 1: An Evolutionarily Mismatched or Novel Environment

Evolutionary change is slow and our social and physical environments have changed very fast through the broader processes of cultural evolution (Chapter 2). If we consider how rapidly we are changing the nature of our social communication and the way many societies are constructed, we can speculate about, and in some cases demonstrate, the effect on mental health (Section 11.4). The biological processes that determined our present structure and function largely evolved in very different environments from those we now inhabit. Thus the most common way in which evolutionary pathways are associated with ill-health is through the consequential mismatch that can arise. Many of the examples we will consider in subsequent chapters highlight the health impact that can ensue when an individual lives in an environment which is evolutionarily novel or where the individual’s evolved capacity to adapt is exceeded. One example, which we will expand upon later, is obesity and its associated morbidities. Another example is the mismatch between the evolved reproductive decline in women, starting in the fourth decade of life, and the pattern of reproduction shaped by cultural evolution and widely practiced in modern societies, with later pregnancies and resulting demand for fertility services.

Scurvy represents a historical example of mismatch, but is also a consequence of the evolutionary history of the primate clade (Buklijas et al. 2011). The history of Western sea exploration is filled with stories of the ravages of scurvy, which is caused by deficiency of vitamin C, the only significant natural source of which is fruit. Humans, unlike most other species, cannot synthesize vitamin C and are entirely dependent on an environmental source (see Box 6.3). The British surgeon James Lind observed in 1747 that citrus fruit prevented scurvy, but the disease only disappeared from the Royal Navy 50 years later when a regular daily ration of fresh citrus juice was introduced. During human evolution there was continual access to fruit, and thus the mutation which led to our inability to synthesize vitamin C was “neutral” until exposed by an evolutionary novelty produced by cultural evolution (i.e., boats capable of long sea voyages during which a dietary source of vitamin C was absent).

Humans are highly dependent on their social environment. As we will discuss in Chapter 11 there have been enormous changes in this environment, which for some people may have exceeded their adaptive capacities. The twenty-first-century social environment has many elements of novelty: the way we communicate electronically rather than in person, the high volume of interpersonal interactions we have, the complexities of changed societal structures of control, and increased autonomy and choice. Our dramatically changed ability to control reproduction also impacts on our choices and interpersonal behaviors. The changing nature of our society means that the rules of the social game have changed; the nature of family structures, peer perceptions, and group dependences has changed too. Our evolved means of identifying and dealing with freeloaders and cheaters are no longer adequate (see Chapter 11). The probable contribution of these evolutionary mismatches to mental health disorders and to behaviors that are considered antisocial, such as acting-out or delinquent behavior, is growing (see Section 11.4). The World Health Organization’s Global Burden of Disease study estimates that, by 2030, depressive disorders will be the largest cause of morbidity in high-income countries, and the second largest worldwide (Figure 7.1).
Mismatch can be shown to affect structure as well as function. We evolved, as did other species, to have a good match between the upper and lower jaw. Malocclusion, in which the bite is mismatched, was rare until the last few hundred years (Rose and Roblee 2009). With the advent of softer foods for children, the need for well-developed jaw muscles declined and in turn this led to more malocclusion (Section 13.8.2). Orthodontics as a profession is heavily dependent on this evolutionary novelty of well-cooked soft food being fed to children.

Myopia is a further example of an interaction between a genetic predisposition (which may be population-specific and originated because of genetic drift or possibly unknown selective pressures) and exposure to a novel environment (Box 3.7). There is a strong familial and a probable genetic basis to juvenile-onset myopia (Cordain et al. 2002a). Being indoors in artificial light may be a factor in the development of myopia, and focusing on close objects, such as text, from an early age for long periods of time is likely to contribute. Recent research suggests that children need to have their peripheral vision exposed to the horizon for some time every day to ensure optimal growth of the length of the eyeball and thus an appropriate focal length (Lin et al. 2004). The result of a childhood largely spent indoors is an eyeball which is misshaped for short-distance focusing, with the consequent need for refractive correction. Some health authorities in Asia are now responding by using regular outdoor experience for children as a prophylactic approach.

We have evolved mechanisms to deal with many toxic substances to which we were commonly exposed, but we cannot rapidly evolve mechanisms to deal with novel toxins. Mesothelioma was until recently a very rare tumor of the membrane that lines the pleural cavity. The use of asbestos in buildings in the 1950s to 1970s led to a large amount of asbestos dust. This dust can enter the lungs, penetrate to the pleural membranes, and cause an oncogenic change that, after a long latent period, is reflected in mesothelioma. Sadly, there has been an epidemic of mesothelioma and asbestosis as a result of such exposure. There is also much concern about a wide range of chemicals in plasticizers, such as bisphenol A, and insecticides, many of which persist in the environment for a long time and are now found in low levels even in the cleanest environment on Earth, the Antarctic.

In Chapter 9 we will discuss the rapidly rising incidence of gestational diabetes mellitus (GDM), which in turn has the potential to drive a further rise in the prevalence of type 2 diabetes in both women and their children. Unlike other nutrients such as amino acids, there is no limit to the amount of glucose that can cross the placenta (Lager and Powell 2012). Thus while fetal growth is normally constrained by limiting nutrient transfer to the fetus (see Section 8.9.5), elevated maternal glucose levels have deleterious effects on the fetus with both short-term (e.g., heart defects) and long-term (e.g., obesity and type 2 diabetes mellitus) consequences. As the short-term consequences also include fetal macrosomia and obstructed labor, which would have led to fetal and maternal death, this suggests that the need for the fetus to cope with the high levels of maternal glucose seen in GDM is an evolutionary novelty.

Most human populations have only been exposed to the toxins found in tobacco smoke for a few generations. Nicotine and other drugs of abuse also highlight another health concern arising from a novel environment: that of addiction. Addiction is essentially a product of the hedonistic sense of reward induced by neurochemical pathways normally associated with satiety from, for example, food or sex. Whereas the intermittent use of such substances, including alcohol, was and is common in foraging societies, their extensive use is novel. These substances override the normal mechanisms of emotional and behavioral control which ensure full functionality as a member of the social group.

### 7.4.2 Pathway 2: Life-history-associated Factors

As mentioned above we have concatenated several interrelated concepts within this category. Each relates in some way to the consequences of the particular nature of the evolved human life history and how that can create risk, given the changed way in which humans now live. Additionally the human life-history strategy can be influenced by events early in life generating trade-offs with costs that may be reflected in ill-health later in life.
The human life-course strategy is one of deployment of resources in the period up to peak reproductive performance, while trading-off that investment against the associated loss of reparative function in the post-reproductive period later in life, when in general a fitness advantage does not operate. Thus evolutionary processes select for optimal development to, and functioning through, peak reproduction, and the primary investment in cellular and molecular maintenance and repair occurs before peak reproductive age and declines in later life.

There has been a marked increase in the average human lifespan in most societies (Section 5.2.3) because of considerable reductions in extrinsic mortality following improvements in public health and, to a lesser extent, medical care. But this reduction in extrinsic mortality is associated with a longer period of post-peak reproductive life and the consequent greater importance of biological senescence in middle and old age. These later periods in our lives are characterized by many co-morbidities leading to a considerable burden of chronic non-communicable disease. These diseases of middle to older age are characterized by a loss of repair, a reduced ability to meet challenges such as infection, exercise, or trauma, and accumulating damage from oxidative stress and exposure to other environmental toxins. The results include post-menopausal osteoporosis, atherosclerosis, sarcopenia, arthritis, neurodegeneration, and cognitive decline. These are the inevitable consequences of longevity, and thus they will become an increasing burden on medical services in many societies. However, it may be that taking an evolutionary perspective will help to provide tests to predict those who are most vulnerable (those who have made the greatest trade-offs), and to devise novel interventions based on an understanding of this variability.

The related concept of antagonistic pleiotropy (Section 5.2.4) is exemplified by the persistence of stem cells in tissues: this feature is of adaptive value during the period of growth and reproduction in promoting tissue maintenance and repair, but it can enhance the risk of neoplasia in later life (see Section 12.4). The example of BRCA has already been discussed. Trade-offs during development are central to adjusting life-history strategies to variable circumstances. Some trade-offs are made during early life to meet immediate extrinsic challenges and allow survival, but these may have irreversible consequences that may compromise later fitness. One example is the child exposed to placental insufficiency in utero: that child will trade off intrauterine growth and is more likely to be born small and prematurely. This trade-off of fetal growth and gestational length has allowed survival within a threatening intrauterine environment, but at the expense of greater infant morbidity and mortality, and longer-term consequences including impaired cognition and a greater risk of metabolic disorder in adult life (Section 13.5.2). Such children, however, may be better able, through predictive adaptation, to cope with famine (see Box 7.2).

Another developmental trade-off is to accelerate the age of puberty to achieve reproduction sooner (Gluckman and Hanson 2006a), but it may have subsequent costs. Both early undernutrition and exposure to high levels of severe stress, such as abuse in early childhood, can accelerate the age of puberty, but this accelerated biological maturation may be at a cost of shorter final height (Section 5.4.3).

In some situations the response of the individual to a challenge is less directly linked to a trade-off than to an inappropriate realignment of components of their life-history strategy. As discussed above, the consequences of adverse early life may lead to a trade-off in terms of an earlier timing of reproduction at the expense of linear growth. But the adaptive strategy of advancing puberty brings with it a mismatch, because other aspects of neural maturation are not advanced, in particular those associated with function of the prefrontal cortex. The result of this loss of synchrony between biological and psychosocial maturation includes the effects on risk-taking behavior and emotional resilience in adolescence (Gluckman et al. 2011b) (Section 5.4.2.2). The consequences can be acting-out behaviors which themselves are harmful (e.g., the effects of drugs and alcohol), but the consequences can also exceed a person’s emotional adaptive capacity and be associated with pathological depression. Recent data show that the incidence of suicide in boys who went through puberty early is several times that of their peers who went through puberty at an older age (Michaud et al. 2006).
Other trade-offs are not for immediate advantage but rather for future advantage, and the adaptive value of these predictive responses depends on the fidelity of the prediction (Section 4.5.2.2). In Chapter 9, we will discuss the consequences of errors in prediction which can contribute to a greater risk of obesity and metabolic disease in later life. The prediction may be erroneous because of faulty signals about the environment in early life, or because of a change in the environment over the life course. Thus an adult may be at greater risk of metabolic disease because maternal or placental ill-health in pregnancy implied a nutrient environment poorer than that to which the individual would be later exposed. Such individuals are more sensitive to developing insulin resistance in an obesogenic environment. This might explain why people originating from the Indian subcontinent develop type 2 diabetes at much lower levels of central adiposity than Europeans (Chiu et al. 2011): in general they are born much smaller due to the shorter stature and poorer nutritional states of their mothers.

Similarly, in high-income countries there is considerable evidence that those who have a poorer start to life are more likely to have earlier puberty and develop obesity and insulin resistance. They invest less in repair and maintenance, as reflected in a lower nephron number, and are more likely to develop chronic non-communicable diseases such as hypertension in middle age. In life-history terms such individuals can be interpreted as having forecast a threatening existence. They have altered their life-history strategy to invest for early reproduction—that is, to create a greater chance of reproducing because they anticipate a threatening environment. Alternatively, faulty prediction may occur because the individual lives in an environment undergoing very rapid change or moves to an energetically richer environment; thus populations undergoing rapid nutritional transition and migrant families may be at particular risk.

Knowledge of such trade-offs and alterations in life-history strategies can also be very important in public health. Among children born in low- and middle-income countries, where nutrition during development has been poor or unbalanced, stunting will be relatively common (Section 5.4.3). This has consequences for their intellectual development, their ability to withstand infection, and their risk of obesity. Such effects can be predicted on the basis of evolutionary principles, and agencies or governments need to take account of them in planning interventions or dispersing resources.

7.4.3 Pathway 3: Excessive and Uncontrolled Defense Mechanisms

Processes which have developed for advantage in providing defense against infection or toxins may become excessively or inappropriately manifested and cause symptoms which are themselves harmful (Section 10.7.4). Alternatively, symptoms which may be annoying or distressing may arise as part of evolved defense mechanisms. It is not certain that suppressing those symptoms is helpful to the disease process, although their alleviation may comfort the patient.

This latter point is highlighted by a consideration of fever. Pyrexia evolved as a component of the defense against infection, and thus as a protective function. Lizards, when infected, move to a warmer situation, suggesting that in some way pyrexia is part of the defense against infection rather than simply being a consequence of it, even though some bacteria do secrete fever-inducing toxins (Rockel and Hartung 2012). Clinically, elevated temperature after admission to hospital is associated with an improved outcome in patients with infections, and suppression of fever with antipyretic drugs in patients with influenza also worsens the clinical outcome (Kushimoto et al. 2013), underlining the importance of fever in response to infection (Section 13.6.1). On the other hand extreme pyrexia itself can be harmful. In children, pyrexia associated with infection can induce so-called febrile seizures which are generally benign but occasionally can have long-term morbidity.

Diseases of autoimmunity, allergic–atopic diseases such as asthma and anaphylaxis (Section 10.7.5), and a number of mood disorders such as phobia (Section 11.6.2) can similarly be considered as situations where the normal evolved processes of defense are inappropriately and/or excessively activated and disease ensues.
7.4.4 Pathway 4: The Consequences of Coevolution with Microbes

Humans are in a constant coevolutionary relationship with viruses, bacteria, and parasitic organisms (see Section 10.3). In some cases, such as for many classes of gut microbiota, these are mutualistic relationships, but in many other cases, particularly for pathogenic organisms, there is an evolutionary arms race under way. The short generation time of these organisms compared with that of humans provides opportunities for microorganisms to out-compete our defense systems. New strains of influenza and the emergence of antibiotic-resistant bacterial strains both represent situations where our defenses have been overcome by the pace of microbial evolution.

Some organisms have successfully managed to evolve ways to co-opt human biology to their advantage; for example, HIV and other retroviruses have successfully co-opted the machinery of human macrophages and lymphocytes to enter those cells. Other organisms use mimicry to try and outwit human defense systems. For example, some organisms have evolved epitopes that are very similar to the human ones. This trait may help them evade attack from the immune system but the outcome can be harmful to the host, who then forms autoantibodies which can cause disease such as post-streptococcal glomerulonephritis or rheumatic fever.

The emergence of both antibiotic and chemotherapeutic resistance represents the challenges of coevolution. The inevitability of microorganisms to evolve resistance to new agents is increasingly a public health concern and has led to new strategies in the use of antibiotics. For example, does the tradition of insisting on the completion of a course of antibiotics once the infection has been treated lead to a greater risk of infection and should we therefore use shorter courses? The use of combination therapy that combines drugs with distinct mechanisms in treating HIV is a direct consequence of developing a strategy which makes the evolution of resistance less likely. Similar thinking underlies the use of multiple chemotherapeutic agents together in treating cancer—the goal is to try and prevent the development of resistant clones of cells (Section 12.6.2).

7.4.5 Pathway 5: Results of Evolutionary Constraints

There are many ways in which our evolutionary history and the associated compromises of evolutionary “design” express themselves in clinical medicine.

Appendicitis is caused by infection in a small blind cecal pouch that has no function in human digestion (although there has been a suggestion that it is a refuge for gut microbiota), but which was part of a much extended gastrointestinal tract needed when earlier members of the primate clade were exclusively vegetarian and required a longer digestion time to adequately extract nutrients from the grasses and leaves that made up their diet. There are still some primate species which are exclusively herbivorous, and one or two species have maintained or evolved expanded gastrointestinal tracts more akin to those of ruminants, in which bacterial breakdown of the cellulose is essential to sustain nutrition. The high frequency of appendicitis (lifetime risk of 7–8%) in high-income countries highlights this imperfection in our gastrointestinal design.

A detached retina occurs when the fibrovascular outer layer of the retina separates from the underlying retinal pigment epithelium. The resulting hemorrhage and scar can lead to a rapid loss of vision if not treated with laser surgery to re-fuse the retina. Because of the arrangement of the layers of the retina, humans also have a blind spot where the optic nerve radiates its fibers to innervate the photoreceptor cells. Both the blind spot and the detached retina are consequences of how the mammalian eye evolved (see Section 2.3.1.3). This is not the case in all species: the eye evolved independently at least 40 times during animal evolution, and eye arrangements are very different between taxa. Notably, the eyes of cephalopods (squid and octopus) are generally very similar to vertebrate eyes but the retinal layers are reversed, with the photoreceptor cells facing forwards and the nerves and blood vessels entering from behind, eliminating a blind spot (see Figure 2.3). This multiple evolutionary origin of eyes is a classic example of convergent evolution, but the particular arrangement of the mammalian eye with its consequent faults highlights the simple fact that evolution is not logical and has no direction.
Obstructed labor or dystocia is a common problem. One estimate is that in the Paleolithic between 10 and 20% of pregnancies ended in maternal death, many from obstruction or hemorrhage (Hassan and Sengel 1973). This is the outcome of the evolutionary compromise between a large brain and a bipedal posture which necessitated changes in the orientation and shape of the pelvis. Gestational length is set to try and achieve maximal intrauterine neural development without compromising the potential for delivery. The consequences are a risk of dystocia if the fetus is large or the mother is small or has a misshapen pelvis. Intriguingly, dystocia is also common in domesticated monotocous species such as the cow and sheep where the domesticated species tend to be smaller than the wild ones.

Back pain is one of the most common presentations to general practitioners and represents the skeletal compromise resulting from quadrupedal precursors adopting a bipedal posture several million years ago. Our skeletal structure is one of gradual accommodation from that of our quadrupedal ancestors, and has been compromised further by the development of the ability to walk and run. With a large head and considerable truncal weight, the potential for lower back pain due to mechanical injury to the spinal discs is high.

Humans choke, sometimes fatally, when there is obstruction of breathing by food entering the airway. The lungs evolved as a diverticulum of the gastrointestinal tract in the first terrestrial air-breathing vertebrates. Thus we have a situation where a common tube is used at the rostral end of both the gastrointestinal tract and the respiratory tract. This creates some risk, and while the epiglottis is an evolved mechanism to try and prevent choking, occasionally this fundamental design flaw reveals itself.

A further example is that of obstructive sleep apnea, in which the soft palate and the epiglottis obstruct the upper airway during the phase of rapid-eye-movement sleep. This is the phase of sleep when the tone of many postural muscles is diminished. The reduced muscle tone is not a problem except in the muscles which maintain airway patency and are not found in other animals (with the exception of the bulldog, which has been bred to have a very peculiar anatomy of the face and neck). In humans the problem seems to have arisen because of the change in shape and position of the larynx necessary for speech. The problem is further exacerbated by obesity, where fat deposition distorts the shape of the pharynx and larynx.

About 1% of all breast cancer occurs in males and it has a similar prognosis to that in females. Men have no need for breasts, but while ovarian hormones are needed for full breast development, the underlying architecture of the breast is not gender-specific. As the breast probably had no significance in the male, it is an atavistic feature. But the male breast has not regressed through neutral mutations to the point where it has disappeared altogether because the genes underpinning it are essential in the female for breast development and lactation and thus survival of the human lineage.

7.4.6 Pathway 6: An Apparently Harmful Allele is Maintained by Balancing Selection

Normally, harmful mutations that manifest in disease before middle age are rapidly selected out of a population, but for some recessive diseases the underlying alleles are maintained at relatively high frequency within a population. For this to be the case there must be positive selection for persistence of the allele and this implies that the heterozygous state confers a fitness advantage. While balancing selection is often suggested as a reason for the persistence of a potentially deleterious allele in a population, the only examples for which the evidence is compelling relate to resistance to malaria.

In sickle cell disease, homozygosity for the mutant allele of the beta-globin gene leads to hemoglobin with a lower oxygen carrying capacity and this is associated with a sickled appearance of the red cells. The resulting chronic tissue hypoxia produces myocardial, cerebral, and other effects, and reduces life expectancy. While the condition is present to a lesser degree in heterozygotes, the sickled red cells do confer advantage in that they cause resistance to the malarial parasite. As a result, the sickle cell allele was positively selected in parts of the world such as sub-Saharan Africa where malaria was endemic. A similar heterozygote advantage exists with the beta-thalassemias, hemoglobinopathies that produce
more of a graded change in globin synthesis; they are commonest in people of Mediterranean descent.

Glucose-6-phosphate dehydrogenase (G6PD) is a key enzyme necessary for the production of NADPH. In red cells, reduced G6PD activity confers protection against infection by *Plasmodium falciparum*, the most lethal of the malarial parasites. The enzyme deficiency is very common and is carried by 6% of the global population, particularly those of African, Asian, and Mediterranean descent. As the affected gene is on the X chromosome (Section 3.5.1), symptomatic disease is predominantly seen in males. There are many different forms of mutation showing that it originated multiple times and its persistence can be linked to areas where *Plasmodium falciparum* malaria was common. Normally, relative G6PD deficiency is covert but can manifest as an acute hemolytic anemia when the individual is exposed to fava (broad) beans (a condition known as favism) or to certain drugs such as primaquine. Additionally, it may lead to prolonged jaundice in the neonate. Note that these examples can also be seen as consequences of coevolution with microbes (Pathway 4).

### 7.4.7 Pathway 7: Sexual Selection and Competition and Their Consequences

Sexual selection and competition between males for access to females and/or between females for the most desirable mates has been a central part of the evolution of all sexually reproducing species, including humans. There are still possible reflections of this history in the way we now live. While we have evolved emotional and cognitive behaviors to sublimate these traits within a group structure, echoes of sexual selection can still manifest with significant health outcomes. The sexual dimorphism in human characteristics and behavior is reflected in the higher death rate of males in every age group. In nearly all societies males have a shorter life expectancy than females (see Figure 8.4). By far the biggest contribution to this differential is the greater tendency of males to undertake risky activity and to resort to violence. Until recently, males would duel with each other to the death over a matter of “honor.” Males drive cars faster and are more likely to be involved in a road accident. Males are more aggressive in trying to acquire assets, so are more likely to commit a violent crime or be the victim of a gunshot or knife wound.

These characteristics of the male brain can be understood as manifestations of the different evolved roles of males and females and their different approaches to achieving maximal fitness. In evolutionary terms males evolved to effectively compete with other males or to try to demonstrate features that will convince a female that they are the optimal mate. Another possible example of these differences is the greater likelihood of a stepfather murdering a stepchild than of a father murdering his biological child (Harris et al. 2007).

The popular and academic literature is full of claims relating to the evolution of human behavior. Some of this has been overstated and much misinterpreted, and most is highly speculative. The application of evolutionary principles to human behavior has a long but controversial history which will be discussed in Chapter 11. By definition, evolution has made a contribution to human behavior—to deny that is to deny biological reality. However, it is important to point out that while behavior has an evolutionary origin, humans have also evolved to live within stable societies, with both culturally driven and evolved expectations about behavior appropriate to living within the group. These group rules have required a level of sublimation of certain behaviors, and in turn this has created an environment which will have affected our biological evolution.

### 7.4.8 Pathway 8: The Outcomes of Cladal and Demographic Histories

As discussed under Pathway 5, our cladal history has created a number of constraints that can have disease consequences. But our more recent history also has consequences. For example, it is now apparent that some populations have different degrees of introgression of alleles from other hominin species and this can have physiological consequences (e.g., Section 13.11.2). As humans spread out of Africa and across the globe in the last 60,000 years (Section 6.5.1) they passed through tight population bottlenecks which decreased genetic diversity and thus generated drift. Others lived in small isolated communities where otherwise rare mutations could be magnified in prevalence because of a founder effect.
Assortative mating within a population can also magnify any founder effect (Section 3.4.1). Each of these processes can influence the pattern of disease, and clinical genetics offers many examples.

Genetic drift can be considered a form of random sampling and is often invoked to explain differences across populations. If a very small proportion of a population becomes isolated, for example by migration, then the distribution of alleles within that derived population need not be representative of that in the parent population. Alleles that were present in the originating population may not be present in the derived population, or an allele that is rare in the founder population may, by sampling (random drift), become more common within the derivative population. Thus a deleterious mutation may become fixed in a small population before it can be eliminated by selection. Within a large population there may also be partially isolated subpopulations (e.g., the Ashkenazi Jews), called demes, which will also favors the effects of drift.

The high incidence of aldehyde dehydrogenase type 1 deficiency in East Asia, which leads to a low tolerance for alcohol, is thought to have originated through drift when the small founder population migrated to East Asia. The modern Finnish population is derived from a very small population who passed through a bottleneck when they migrated from Europe across the Baltic and who then lived in relative isolation because of geography and climate. Partially as a result of this, the pattern of disease in Finland is rather different from that of the rest of Europe (Box 3.4). There is a lower incidence of Huntington’s disease, cystic fibrosis, and phenylketonuria, but a higher incidence of many rare metabolic diseases with a genetic basis. The incidence of type 2 diabetes and cardiovascular disease is very high in Finland, but while there has been much searching for genetic determinants, the particular diet and lifestyle of Finns may be an explanatory factor.

There are many rare diseases that have been mapped back to a founder effect and are maintained by assortative mating. A classic example is the pattern of hemophilia in the European royal families. The more closely related individuals are, the greater the risk of each carrying a deleterious allele from a common ancestor. Different societies have very different views of couplings between close relatives. Sib-based couplings were apparently common in ancient Egyptian royal families, and even today cousin marriage is closely linked to protecting inheritance and concentrating wealth and power in some societies.

Many rare genetic diseases are found in small populations. For example, growth hormone resistance (Laron syndrome) related to one specific mutation in the growth hormone receptor has been found

---

**Box 7.4 Formulating and Testing Hypotheses for Evolutionary Medicine**

The historical nature of evolution and the long human generation time make study of human evolution a difficult—but not impossible—task. This list is based on the framework proposed by the co-founder of evolutionary medicine, Randolph Nesse (Nesse 2011).

1. Precise definition of the object of enquiry.
   (a) Type of object: a human trait, facultative or general; pathogen trait; gene.
   (b) Representation in the population. Explaining the existence in the entire species or variation between different populations.

2. Precise definition of the explanation sought. Phylogeny? Selective pressures? Evolutionary forces that shaped its distribution?


4. What method is used/should be used?
   (a) Modeling using quantitative methods.
   (b) Comparative methods (be it among species, populations in a species, or individuals varying in a trait, e.g., including clinical trials and epidemiological studies).
   (c) Experimental methods (extirpation or disruption; augmentation; examining regulation of a facultative trait, observing evolutionary changes in the lab or the field).
Key Points

• Medical practice and public health are primarily concerned with preventing and managing ill-health. In turn, this requires an understanding of the determinants of individual phenotypes and the factors that lead to variation in disease risk.

• Understanding why one person develops disease while another does not is improved by considering the evolutionary pathways to altered risk.

• The human phenotype at any age is a construct of inheritance and the consequences of developmental plasticity, itself informed by environmental history, and is defined further by the interaction with the current physical, biological, and social environments.

• Selection operates to enhance inclusive reproductive fitness, not necessarily health or longevity.

• Nothing in the human phenotype was selected for a higher purpose. Many traits are present because they were selected for some adaptive advantage in the past.

• Several non-mutually exclusive pathways can be identified by which ultimate causation affects disease risk, and they provide an important heuristic for medicine and public health.

7.5 Testing Hypotheses in Evolutionary Medicine

A practical problem that emerges in evolutionary medicine is that while it is relatively easy to speculate regarding ultimate causation, in many cases testing the derivative hypothesis is difficult. Microevolutionary processes in general occur over many generations, yet in medicine we deal with individuals and even in public health only with a group covering at most three to four generations. This creates the problem that apparently credible but not always evolutionary sound “just-so stories” may be presented as evolutionary arguments. However, academic evolutionary medicine has emerged with a distinct set of approaches to testing evolutionary hypotheses related to health and disease. Some of the most important approaches are listed in Box 7.4. In the chapters that follow we will limit our examples to those where there is reasonable evidence to support the evolutionary explanations proposed or indicate where plausible explanations exist but require more evaluation.

In a province of Ecuador (Laron 2004). A very rare form of intersex due to a mutated 5-alpha reductase enzyme which converts testosterone to the more active dihydrotestosterone in peripheral tissues is found in the Dominican Republic. It causes an unusual syndrome in which male children are born with ambiguous genitalia and are raised as girls to puberty when they masculinize (Andersson et al. 1991). Such founder effects likely explain why Tay–Sachs disease, a rare and devastating neurodegenerative disorder affecting children, is found essentially only in Ashkenazi Jews—a population which, while now dispersed around the world, has for cultural reasons remained reproductively isolated from other populations.

There are additional possible explanations for the geographical distributions of diseases with an evolutionary component. We have discussed the relationship between the geographical distribution of the hemoglobinopathies beta-thalassemia and sickle cell disease and the past distribution of malaria. Humans in different parts of the world developed phenotypic variation which can affect their disease risk; for example, fair-skinned people are more likely to develop skin cancer.
CHAPTER 8

Reproduction

8.1 Introduction

While the unit of selection and differential survival is the organism, at the more fundamental level it is the genes that encode successful traits that survive over time. Extant lineages are those that have reproduced successfully in each generation whereas, by definition, extinct lineages are those in which reproduction failed. Successful reproduction is thus the crucial outcome of evolutionary processes, and the specific reproduction-associated traits of a species have evolved to optimize fitness in the environment it inhabits. This does not mean that the strategy of simply producing more offspring is always going to produce higher fitness. Because that strategy has energetic costs (Chapter 5), the alternative of producing very few offspring and investing greatly to ensure their survival has proved very successful for some species, especially humans. Across the biotic kingdoms an extraordinary range of reproductive strategies has evolved, allowing each successful species to reproduce by matching its life-history strategy and its reproductive pattern to its ecological situation.

Selection operates to maximize fitness, which is in turn a function of reproductive success over a lifetime. Therefore, for species that have recurrent reproductive episodes this need not mean producing the potential maximum number of offspring per episode. Humans can have multiple births, but singleton pregnancies have evolved as the norm because this strategy is most likely to lead to successful survival of the few offspring a woman can support energetically, deliver successfully, and nurture over her reproductive life. Indeed, infanticide of one of a pair of twins, the norm in some species such as the giant panda, has also been recorded in some forager societies (Hrdy 1992). In some species of bird, clutch size is reduced by the parents to an optimal size for the resources available. These energetic and nurturing limitations are reflected in the much higher infant mortality in human societies when the inter-birth interval is reduced (Figure 8.1). Thus optimal fitness considerations include spacing reproductive events to match resource availability. These resources include parental investment and the availability of food.

A successful reproductive strategy therefore requires balancing the investment of the parents: from the costs of mating, reproduction, and parental care, to the life-history-associated energetic constraints imposed on the organism. In other words, organisms must cope with the trade-off between fecundity and the energetic costs associated with reproduction and support of the ensuing offspring. The exact outcome of the trade-off depends on a range of variables, from cultural practices to the environmental niche: the !Kung san (Khoisan) population, a traditional foraging society in southwest Africa, has an average fertility of 4.7 live births per woman across the life course, lower than the Hadza, an East African foraging society with a fertility rate of more than six. The difference has been variously explained by the easier availability of food in the bush, greater efficiency at foraging, and lower age at weaning among the Hadza (Blurton Jones et al. 1992). As suggested by life-history considerations, excess fecundity may have a cost for the parent. One well-supported hypothesis for the age-related decline in fertility in women in the fourth decade of life is that it evolved to limit the risk of dying before their youngest child reached the age of independence (Section 8.9.7). There are few sets of records
which allow these ideas to be explored in humans but, as described in Box 5.2, one study of the British monarchy provides some evidence for a reciprocal relationship between fecundity and the lifespan of the female members.

### 8.2 Sexual Reproduction

Reproduction does not always require sex; that is, two organisms combining genetic material in their joint offspring. There are many plant and animal species that either always reproduce asexually or do so under certain circumstances. Exclusively asexual organisms do not have multiple genders. Parthenogenesis (a type of asexual reproduction) has been described in reptiles as large as the Komodo dragon when housed in isolation in a zoo, although this species normally propagates by sexual reproduction. Some other lizards such as the New Mexico whiptail always reproduce asexually. Perhaps the most studied family of species showing successful asexuality over millions of years is the bdelloid rotifer, a microscopic aquatic metazoan. There are over 500 species of this rotifer and all individuals are exclusively female, meaning they produce eggs. Indeed, sexual reproduction is nonexistent in this entire metazoan phylum, although several species of rotifer do show at least intermittent sexual reproduction. Propagation by budding is common in plants and bacteria. Other species, including some grasses and some insects, may have many generations of asexual reproduction and then a generation of sexual reproduction before another long series of cycles of asexual reproduction.

If fitness is approximated by the number of progeny possessing a parent’s genes, then asexual reproduction could be viewed as twice as efficient as sexual reproduction. This is because all of an individual parent’s genes are transferred to its progeny in asexual reproduction, whereas only half of that parent’s genes are transferred in sexual reproduction. Assuming that two offspring survive per generation, after three generations an individual reproducing asexually would have eight great-grandchildren carrying all of its genes; a sexually reproducing equivalent would have eight great-grandchildren each carrying a random one-eighth of its genes. Viewed from this angle, what would be the fitness advantage of sexual reproduction? After all, sexual reproduction is obviously a highly successful strategy for many species of plants and animals alive now. Thus a fundamental set of questions can be posed: why has sexual reproduction evolved, how has it conferred apparently greater fitness for those species that employ it, and why does sexual reproduction require more than one gender?

Many features of our reproductive physiology, anatomical structures, and components of our behavior which ensure successful reproduction and have considerable influence on human health and the risk of disease are different in the two sexes. These biological differences underlie the relationship between and the roles of the sexes, the social construct of gender, and, more broadly, the structure of human societies. In turn, and as shown throughout this book, socio-cultural aspects of sex and reproduction have influenced the direction of human evolution via cultural evolution.

### 8.3 Why Did Sex Evolve?

Sexual reproduction is based on the fusion of genetic material. This involves two steps: the first is meiosis, in which the chromosomes of a diploid organism align thus allowing recombination where...
the homologous portions of the paired chromosomes may interchange (Figure 3.3). Recombination is a key event in creating genetic diversity from one generation to the next, because in recombination neither of the two original maternal or paternal chromosomes are passed on unchanged to the haploid gamete. What was originally the paternal chromosome will contain material from the maternal chromosome derived from recombination events, and vice versa. But recombination is not the only source of genetic novelty in the next generation. In the second step, at least in all birds and mammals, the two haploid genomes (one from the maternal line and one from the paternal line) are fused into a new single diploid organism at zygosis. Where alleles are heterozygous in a parent, only one is passed on to an offspring. Thus sexual reproduction is ultimately about developing new admixtures of alleles from one generation to the next.

The male is defined as the sex that produces a large number of small mobile gametes, whereas the female is characterized by large immobile gametes; this difference is termed anisogamy. Why gametes diverge so considerably in size is open to speculation, but it may be to make the risk of self-fertilization remote. The male gamete carries nuclear genes but, in contrast to the ovum, contains essentially no cytoplasm and no mitochondria or other extranuclear organelles. Mitochondrial chromosomes code for 37 genes in humans, and these genes are critical for cellular metabolic function. Mitochondria and chloroplasts are derived from bacterial cells that eventually became engulfed by proto-eukaryotic cells in distant evolutionary time. The exclusion of male-derived organelles and mitochondrial genes from the fertilized zygote may have evolved as a way of excluding viruses and bacteria from the cell. There is some evidence from algae that mitochondria of different genetic origin existing within the same cell can be mutually destructive. Perhaps this reflects ancestral bacterial conflict, and suggests that the gender difference in gamete structure may have evolved to avoid such conflict: anisogamy ensures that there is only one line of mitochondria (or bacteria) within the fertilized zygote and its descendant cells. Parenthetically we should also note that although the sperm do not contain major organelles, they contain multiple classes of non-coding small RNAs that can affect gene expression in the early stages of zygote development (Krawetz et al. 2011). This is a potential source of epigenetic control (see Sections 3.7 and 3.8.3).

Sex is basically about genetic trading. The question is why did it evolve as a reproductive strategy; that is, how does it confer a fitness advantage? This issue has been a major component of evolutionary discourse for many decades. An early hypothesis was that sex generates genetic variation and, because it provides a greater range of phenotypes on which selection might act if environmental conditions change, this greater genetic variation might assist the survival of the species (Maynard Smith 1978). But this is essentially a group-selectionist argument, in that it suggests that sex evolved for the good of the species rather than the organism or gene, and thus it is less compelling. A related argument is that it could be advantageous for the lineage, because in variable conditions a greater variety of genotypes/phenotypes in the progeny gives a greater probability of at least some offspring surviving to pass the parent’s genes to subsequent generations. But there is no empirical evidence to support this hypothesis.

Ecological models may provide a better explanation. It turns out that asexual reproduction is more common in highly variable environments, especially in small short-lived animals, whereas sexual reproduction is a feature of longer-lived species inhabiting more stable environments (Bell 1982). Bacteria propagate by parthenogenic division and lineages show very successful mutations, for example those conferring resistance to antibiotics. The high rate of reproduction, very short generation time, and thus high rate of mutation per unit time ensure a high probability that some lineages will survive. Bacteria can also interchange genetic material through conjugation, a process involving direct contact and passage of plasmids through a tube-like structure called a pilus. The exchange of genetic material between bacteria is not reciprocal, and only some strains containing genes allowing conjugation can act as donors of self-replicating elements of DNA that become incorporated into the bacterial genome. The R factor involved in one form of antibiotic resistance is an example of such
an element. A further consequence of the genetic variation argument is that mutations can be better dispersed across a population by sexual reproduction. However, most mutations are deleterious, and those that are beneficial do not need to be transferred by sex to be maintained since they can be positively selected in asexual progeny.

A second set of theories argue that sex is about chromatin repair (Bernstein et al. 1981). DNA is continually subject to environmental injury, and there are elaborate DNA repair mechanisms that can correct damage to both single- and double-stranded DNA using enzymes that specifically target damaged areas. One form of repair involves DNA crossover, which is similar to meiosis. With single-strand damage the uninjured complementary DNA strand is used as a template to repair the injured strand. But if reproduction is asexual and both copies are damaged at the same locus there would be no template for repair. Thus sexual reproduction provides a template for repair by providing an alternative copy of the gene from the chromosome of a second individual. There is some evidence for this theory because asexual organisms do accumulate disadvantageous mutations over time, just as mitotically replicating cells may do in the pathway to cancer (Chapter 12) and those regions of the genomes of sexually reproducing organisms that do not undergo recombination. This phenomenon is known as Muller’s ratchet (Felsenstein 1974) and may be relevant to understanding the Y chromosome: except for the human pseudoautosomal region 1 (see Section 8.4), the Y chromosome is not subject to recombination so tends to accumulate mutations (Charlesworth and Charlesworth 2000) and contains few active genes (though there may be several mechanisms at play; see later). It has thus been argued that the purpose of intermittent sexual reproduction in some species such as *Daphnia*, which has generations of asexual reproduction followed by an intermittent sexually reproducing phase, is genetic rejuvenation made possible by the provision of a new repair template (Hebert 1978).

The third and most favored set of arguments for the evolutionary success of sex relates to the need to cope with parasites. When a species becomes extinct bears no relation to how long that species has existed, suggesting that the challenges to a species’ survival are persistent: at any moment in that species’ history it may face a something that can overwhelm its capacity to adapt. The major challenges to the survival of a species are generally not physical (despite the great extinctions associated with asteroid collisions) but due to the relationship between that species and its biotic environment. This includes the part played by other species through predation or competition for the same ecological niche (Lively and Morran 2014) (e.g., humans have destroyed many species by hunting them to extinction or by destroying their environment and displacing them from it). But by far the greatest biotic challenge to any species is its parasite burden. All mammalian species carry a plethora of parasites, from metazoans to viruses. Parasites and hosts have complex interrelationships (see Chapter 10). Parasites are smaller than their hosts and have faster generation times and thus a greater potential rate of evolution. The generally used analogy is that of an arms race. If one army has spears, the other side will soon have shields. If one side then has guns the other will perish or invent armor plates and, as has shown, eventually ballistic missiles and missile defenses follow. Some form of unstable equilibrium is reached until one side or the other has a technological breakthrough. In the same way, while a host is continually trying to evolve better defenses against parasites, the parasite is constantly attempting to overcome these very defenses.

Mammals have two related primary defenses against parasites: their immune system and genetic change. Sexual recombination can, in a single generation, create a novel combination of defense mechanisms against parasites, for example by creating new allele combinations in the genes for transmembrane proteins called histocompatibility complexes. These genetically determined complexes play a key role in self-recognition to avoid the development of autoimmune diseases, but they also must recognize foreign substances so are highly variable among individuals (Section 10.7.2). Sexual reproduction leads to a major change in their pattern in the next generation, allowing the recognition as “foreign” of parasites that have adapted to evade the host’s immune response by displaying epitopes on their cell surface that are hidden from the host’s immune system. In turn the parasite is itself evolving, and to
survive must overcome the change in host defenses or move to a new host in which it can still survive. Generally, the level of change in one generation is sufficient to compensate for the parasite’s much faster reproduction time and potential for evolution. This interspecific coevolutionary competition, always ongoing and seldom with a clear winner or loser, has been called the Red Queen effect after the character in Lewis Carroll’s *Through the Looking Glass* who had to keep running to stay in the same place.

If this Red Queen hypothesis is true, then how do we explain the survival of ancient asexually reproducing species such as the bdelloid rotifer and the New Mexico whiptail lizard? Why didn’t they succumb to parasites ages ago? Two ideas have emerged. First, when the pools in which the rotifers live dry up, they survive by becoming extremely dehydrated—dehydration may kill off their parasitic loads (Wilson and Sherman 2010). Interestingly, in *Potamopyrgus*, a New Zealand snail, the oldest clonal lineages (500,000 years old) come from environments where parasites were rare (Neiman et al. 2009). Such examples suggest that asexual reproduction persists in the absence of selection imposed by parasites. Second, both the bdelloids and the whiptails are polyploidic organisms, with multiple copies of most genes. This provides additional self-repair templates and greater potential for variation and mutation during parthenogenic reproduction.

### 8.4 Sex Determination

In mammals, sex is defined chromosomally (in contrast to gender, which is socially constructed), with two homologous sex chromosomes defining a female (XX) and two different sex chromosomes (XY) defining a male. In eutherian (i.e., non-marsupial) mammals the X chromosome is large with many genes expressed, while the Y chromosome is small with very few genes expressed, and these are most directly associated with reproduction. In females the second X chromosome is almost entirely inactivated by a form of chromosome-wide epigenetic suppression (Section 3.7). This suppression has evolved to maintain an equivalent gene dosage in both males and females. However, some genes in the second X chromosome are biallelically expressed and are also found in the pseudoautosomal region (PAR) of the Y chromosome (Helena Mangs and Morris 2007). For example, the *SHOX* gene (short stature homeobox-containing) resides in a PAR on the X chromosome. Monoallelic instead of biallelic expression of this gene contributes to the skeletal dysplasia of Turner syndrome which affects girls who have only one sex chromosome (45 XO karyotype). Symptoms include short stature, cubitus valgus (forearm angled away from the body more than usual), and short fourth metacarpals.

It is generally believed that the X and Y chromosomes evolved from an identical ancestral pair of chromosomes, but that there was gradual differentiation of the two chromosomes such that the Y chromosome did not undergo recombination with the X chromosome during meiosis in male germ cells. This led to a Y chromosome that only had the small PARs showing homology to the X chromosome, the remaining loci being found only on either the X or the Y chromosome. In contrast to the X chromosome, the Y chromosome is never paired and is therefore more vulnerable to DNA damage and mutations that are not repaired (Charlesworth and Charlesworth 2000). The Y chromosome has low levels of genetic diversity, which is explained by the action of negative (purifying) selection (Wilson Sayres et al. 2014). Because of this rather low variability, Y-chromosomal analysis is a way of studying male-line inheritance and population divergence (Box 6.7), allowing fine distinctions between populations.

Birds also have sex chromosomes, called Z and W, but the heterochromosomal sex (ZW) is the female and male is ZZ. Gene-dosage compensation is not complete as the second Z chromosome in the male is not inactivated, although individual loci are hypermethylated (Section 3.7). However, in many other taxa, sex is not defined chromosomally but by the developmental environment. In some reptiles sex is chromosomally determined, but in others sex determination is solely defined by the temperature at which the egg is incubated. To complicate matters still further, in some species the warmer egg becomes the male and in some it becomes the female, and there is at least one species of lizard that can use both chromosome- and temperature-dependent sex differentiation under different circumstances (Bull 2015).
Across these different systems, the processes of differentiation of the testes and male genitalia and suppression of structures derived from the Müllerian duct involve similar genetic cascades, but the process by which this cascade is activated varies. In mammals testicular differentiation is initiated by the expression of \textit{SRY}, a locus on the Y chromosome whose function is still poorly understood. There are a few mammals, such as the spiny rat, that do not have a functional \textit{SRY} locus and the alternative mechanism for initiating testicular differentiation in these species is not known. In temperature-dependent sexual differentiation, temperature influences the expression of genes that regulate gonadal differentiation, such as \textit{DMRT1} in the genital ridge. In birds and some reptiles that have chromosomally determined sex, \textit{DMRT1} is expressed at a higher dose in male embryos. \textit{DMRT1} is downstream of \textit{SRY} and plays a role in mammalian testicular determination, suggesting extensive parallels in the processes of sex determination across very large evolutionary distances.

### 8.5 Reproductive Strategies

Some species, including most fish, reptiles, and amphibians, produce many offspring and generally have no parental involvement in their care; few of these offspring survive to reproduce. At the other extreme are species such as whales and elephants that have few offspring and high parental investment in each, such that a high proportion of the offspring survive. There are species such as the salmon where the male mates only once and then dies, and species that have very short lives or very long lives. Each of these strategies has an evolutionary logic, as discussed in Chapter 5.

Humans are towards one end of the spectrum of life-history patterns. Our lives are long, we have very few offspring, and as parents we invest greatly in their care and development. Fitness requires that a high percentage of offspring survive, and humans live successfully in wide range of environments. Relative to their lifespan, humans have a very late puberty—that is, the age at which reproduction becomes possible. This is a trade-off for having a large brain and a strong reliance on cognitive and other neural capacities (Chapter 11). Because of the compromise needed to achieve a large brain while allowing birth through a birth canal in a narrow pelvis restructured by the demands of bipedalism, humans have evolved as a secondarily altricial species (see Figure 8.9). That is, compared with other primates humans are born at a relatively immature neurological stage and depend totally on parental support for mobility, protection, and feeding for a considerable time while brain development proceeds. In turn this requires a high level of maternal involvement, adequate spacing between pregnancies, and a clan-like family structure to assist the mother, thus driving the evolution of social structures and the allocation of roles between the sexes.

These evolutionary compromises have made human childbirth a precarious event: it is thought that in the Paleolithic the chance of maternal death in pregnancy or childbirth was about 15%. The inter-birth interval is thought to have been about 3–4 years based on what is known from the few remaining forager societies. Thus if menarche occurred at about 12–14 years and the first birth about 5 years later (Allal et al. 2004), then the average woman might be expected to have given birth to about five or six children, of whom about half might have survived. Thus, selection pressure will have given advantage to those traits that enhanced the capacity to nurture these few offspring.

In mammals there is a fundamental difference in the cost of reproduction for the male and female. Females invest heavily in pregnancy and lactation, and their fitness is determined by the survival of the few offspring that they have. Males, however, mate but do not necessarily invest in the care of their offspring. The only potential exception is lactation observed the male Dayak fruitbat, but not enough is known about its biology and the environmental influences on this species to understand if it is a normal occurrence or a pathological finding; the small amount of milk makes it unlikely that offspring can be solely nourished by the father (Kunz and Hosken 2009). So, equilibrium must be reached in this sexual disparity for each population, and across species very different strategies exist. Optimal fitness in males may be achieved by mating with multiple females, in a “harem”: this is the nature of the social structure of, for example, gorillas. Here a dominant male has several females in his harem with
which he mates, and other males are excluded from mating until the dominant male is overthrown (or his rivals succeed in a surreptitious opportunity). Males in such systems thus expend a good deal of energy fighting to maintain dominance and mating rights. For example, in the red deer we can view the evolution of the antler, a costly organ in energetic terms, as a tool for establishing dominance. In some species females may choose from several males, as is the case with the peacock (Section 2.3.2.3). In yet others, such as the chimpanzee, females may mate with several males in a single estrous. Or monogamous relationships may be the basis of resolution of the potential conflict, as in gibbons and albatrosses. In species where multiple males mate with a female, there is advantage in producing large amounts of sperm to reduce the chance of success of other males’ sperm. This is true even in primates, where ejaculate volumes are greater in species such as the chimpanzee in which multiple matings occur (Box 8.1). Or they may produce a plug that blocks access to the vagina by sperm from other males. Females may develop sperm storage systems so they can select the father despite multiple matings. In some species, particularly in invertebrates, extraordinary genital anatomies have thus evolved.

8.6 Mate Choice

It is a fundamental and self-evident principle that in any given population the total number of females must produce the same number of progeny as the
The way in which mothers and fathers invest in their offspring is very different, reflecting their different strategies for ensuring high fitness levels. Energetically, mothers invest from before pregnancy through to lactation, whereas the direct energetic investment of the father is minimal. However, both parents will invest considerable effort in providing resources, protection, and education for their offspring.

A number of evolutionary arguments have been used to try and explain variations in the sex ratio at birth by considering parental investment. Under normal circumstances the sex ratio of most species that produce offspring is around 1:1 (but see Section 8.8). The evolutionary theorist Ronald Fisher explained this ratio in the following way: a skewed ratio would lead to better mating prospects for the rarer sex, leading to more offspring in the next generation, and the spread of the genes favoring the birth of the rarer sex. Eventually that would lead to a 1:1 ratio and the advantage that the rarer sex had would disappear. While Fisher’s theory applies to a population in equilibrium, there are circumstances under which groups might gain advantage by altering the sex ratio of their offspring transiently. There is evidence that the number of live-born offspring in many mammals can be influenced by the condition of the mother. For example, in many species under conditions of very severe undernutrition the sex ratio becomes skewed to more females being born. There is some limited evidence to suggest that this may also be the case in humans (see later). The physiological mechanisms involved are unknown and are likely to vary between species.

The evolutionary biologist Robert Trivers and mathematician Dan Willard have formulated a hypothesis to account for this. They proposed that parental investment is preferentially directed towards the sex that provides a higher reproductive pay-off relative to the investment made (Trivers and Willard 1973). So, because female offspring are more certain to reproduce than male offspring in poor condition, mothers in poor condition are more likely to give birth to daughters. Conversely, as males can gain more mates, mothers in good condition who are able to invest heavily in offspring growth and thus produce high-quality sons will achieve greater inclusive fitness if they have male offspring with greater reproductive success. For example, there is some evidence from zoos that the sex ratio in some species can be affected by changing the maternal diet during the period around conception. Note that good and poor conditions need not relate to nutrition; in red deer social rank has a similar effect, with high social rank leading to more male offspring.

Furthermore, there is limited evidence for such biased sex ratios at birth in some human populations. In famine conditions, such as in China during the Cultural Revolution, the sex ratio was reportedly skewed towards female offspring. It has also been suggested that there is a tendency for humans of high social rank to produce more male offspring. A study in the UK shows that the percentage of male offspring is highest in nulliparous women who had the highest dietary energy intake in the pre-conception period (Mathews et al. 2008). The Trivers–Willard model has been used to explain the origin of infanticide in some populations and of neglect of one sex (not always the female) in some cultures, and the longer nursing of male children in high-social-class individuals and the reverse in low-income families.

It might be argued that under conditions of high nutrition and low threat, having more male children may enhance fitness as most of those males are likely to have progeny, whereas under limiting conditions favoring female nutrition is more likely to ensure successful mating and progeny for these offspring.

In mammals the cost of reproduction is much higher for the female, as it entails not only the production of ova but also support of both the pregnancy and lactation. For a male mammal, fitness is generally increased by maximizing the number of females he mates with provided that he can ensure the survival of his offspring. To do so may require pair bonding with a female, and this must be traded off against the fitness opportunities of having multiple partners. So while a polygynous mating system may at first appear to be advantageous for males, depending on the species and ecology, monogamy

---

**Box 8.2 Parental Investment Strategies**

The total number of males. But within the population some individuals may produce more and others fewer. For males, the number of offspring generally increases with the number of mates; but for females that is not the case, because the cost of female reproduction—with the single ovum being more costly to produce than sperm cells—is the limiting factor. The observation that the sex for which the cost of reproduction is higher (usually the female; Box 8.2) is likely to be more selective about potential mates is known as Bateman’s principle (Bateman 1948).
can evolve as the most effective fitness strategy. For females, fitness is maximized by selecting a mate who can contribute most to sustaining viable offspring. This may be manifest in evidence of health or strength or, in some species, of the capacity to assist the mother in parental care. In humans the latter is of particular importance because our secondarily altricial nature requires a long post-natal period of nurturing and education.

This interplay between the sexes in obtaining a fitness advantage is reflected in a number of sex-specific behaviors, and is the focus of disciplines such as evolutionary psychology and social anthropology in which researchers seek evolutionary explanations for behavior (Chapter 11). While this is an interesting field of research, and potentially allows for a greater understanding of human behavior, we must urge caution and recognize the limits of our capacity to separate those components of behavior that have a direct evolutionary origin from those that may have an entirely different explanation, albeit based on an evolutionarily derived framework for human behavior.

For example, giving a gift such as a diamond ring has been argued to be an echo of an evolved behavior whereby a female, knowing the investment that would need to be made in their potential progeny, had to be assured that the male had the ability to obtain sufficient resources to support her and her offspring. Yet the association of diamond rings with romance and love in our modern culture may have more to do with the successful advertising campaign that the De Beers diamond cartel first launched in the 1930s to combat the decline in the price of diamonds and the numbers sold (Otnes and Scott 1996).

The tendency for males to undertake risky behaviors such as racing fast cars has similarly been suggested to be the result of sexual selection, whereby the female seeks the strongest male to father her offspring. Some scholars, however, have argued that women see risk-taking as attractive in a short-term sexual partner rather than long-term mates; others distinguish among different types of risk-taking activities, with those in social and recreational areas deemed attractive and other kinds such as health risks and gambling unattractive (Wilke et al. 2006). Note that taking risks in social and recreational domains may imply qualities that are rather different: altruism versus physical strength and agility.

In many societies older males mate with younger women; is this the outcome of a preference for a mate who is likely to live long enough to fully support his offspring’s growth and maturation? At the same time, females may select older or richer males as mates because of the greater resources they have accumulated. These examples illustrate the storytelling inherent in such interpretations and the caution needed in thinking about them.

Sexual dimorphism in body size is taken as evidence that humans evolved with a polygynous mating system. However, the relatively small degree of sexual dimorphism in body size, and comparative considerations of testis size across the greater ape species (Box 8.1), suggest that pair bonding—rather than extreme polygyny—was more often the norm for humans. Nevertheless, the ethnographic data are mixed. Many cultures permit polygamy, serial monogamy is a common human behavior, and mating outside the pair bond is not infrequent. Pair bonding has obvious advantages in ensuring resources for the offspring, and polygamy is less common where the male does not have adequate resources to support all his mates and offspring. Pair bonding may also be a form of mate guarding to ensure paternity: why would a male invest resources in an offspring he cannot be certain is his? But perhaps cultural rather than strictly biological elements are the major drivers of monogamy to encourage successful living in family and social groups. Otherwise, the complex taboos which exist in many cultures may not have been required to reinforce it.

Individuals who are raised together up to the age of 5 years tend to have little sexual interest in each other. Similar effects can be seen in some other species. This phenomenon is known as the Westermarck effect (Westermarck 1891). In most cases such closely reared infants are kin and so the risks of incest are reduced. The usual assumption has been that this process evolved to reduce the risks of inbreeding with consequent loss of genetic variation and accumulation of harmful mutations (inbreeding depression). In many species, such as the Malaysian tree shrew, matings will only occur between animals with distinct major histocompatibility (MHC) antigens. Many societies have taboos
against incest, which can be extremely complex. But exceptions are sometimes made, particularly for rulers, and this may be related to safeguarding inheritance and as a way to prevent wealth accumulating outside the dominant family (Box 8.3). Another argument is that these taboos encouraged marriage and alliances between clans, which may have had advantages in ensuring access to key resources or reducing the risk of warfare.

8.7 Sexual Differences in the Human

Males and females are anatomically, physiologically, and behaviorally different. These sexually dimorphic characters are the result of both natural and sexual selection. One obvious difference between males and females is adult body size and muscle mass. The human male has a greater growth spurt at puberty and develops considerably more muscle mass. Humans demonstrate some sexual dimorphism in body size, although to a much lesser degree than the gorilla, suggesting that in our evolutionary history a partial-harem mating system was the norm. In some hunter-gatherer societies, mating rights are largely restricted to older men who take a number of younger female partners.

Differences in the age of mating may reflect the age at puberty (Section 5.4.2). Males tend to have delayed entry into puberty compared with females, there being no advantage to a male in having earlier puberty. Indeed, it may create a disadvantage in that the young male is immediately a potential reproductive threat to the older males. In many forager societies, females who have reached puberty are viewed as reproductively competent and are married soon after. In contrast, male pubertal rites may involve long periods of exclusion or the development of rituals where the young males cannot mate for some years after sexual maturation. For

---

**Box 8.3 Inbreeding and the Fall of a European Dynasty**

In 1700, the last descendant of the Spanish branch of the Habsburg dynasty, Charles II, died without an heir, despite having been married twice. His death opened the door for dynastic succession. Charles was disfigured and mentally and physical disabled: among many other features, he had an extreme version of the so-called “Habsburg jaw” or “Habsburg lip,” also known as mandibular prognathism, where the lower jaw outgrows the upper. This trait was shared by many of Charles’ relatives and ancestors, as family portraits show. Historians have speculated that the Spanish Habsburgs’ preference for consanguineous marriages caused the dynastic downfall and finally extinction. To preserve their heritage within the family, the Habsburgs began to intermarry more and more, leading to an accumulation of deleterious mutations. Several years ago, this hypothesis was subjected to a test (Alvarez et al. 2009). Charles’s pedigree was constructed going back 16 generations and including over 3000 individuals, and the inbreeding coefficient computed.

The inbreeding coefficient is a measure of the “pedigree collapse,” or the degree to which the family tree of the offspring of two individuals with shared ancestors is smaller than the family tree of the offspring of unrelated individuals. Over 200 years since the establishment of the Spanish branch by Philip I (crowned in 1504), 9 out of 11 marriages were consanguineous to the degree of third cousin or closer. Indeed the unfortunate Charles II was the culmination of inbreeding among the Spanish Habsburgs: not only was his father his maternal uncle, but his paternal grandmother, Margaret of Austria, was also his maternal great-grandmother—and the family connections did not stop there, as the pedigree in Figure 8.3 shows. Indeed, relatedness through multiple ancestors contributed to Charles’s inbreeding coefficient of 0.254, exceeding the coefficient expected in parent–child or brother–sister unions (0.25).

It is known from modern studies of first-cousin marriage that the degree of increased risk of congenital malformation/disease depends to a large extent on the mean coefficient of inbreeding for the population studied (Paul and Spencer 2008). For example, while standard calculations of around 3% of additional risk assume a pedigree in which the ancestors are unrelated, in certain populations, such as in Pakistan, more than 50% of marriages are between cousins so the inbreeding coefficient—and possibly the additional risk—will be much larger. While for the Spanish Habsburgs it is impossible to assess the association between inbreeding and congenital disease, a statistically significant association between the parents’ inbreeding coefficient and the childhood mortality of their offspring was found.
example, the Masai tradition was for young adult males to become the soldiers for the tribe, and marriage was not possible until that period of effective sexual separation and exclusion had been completed. Even in Western society there are still echoes of these differences in our social structures.

Many species signal sexual receptivity by marked physical change. Female baboons, for example, exhibit a dramatic change in perineal coloration. By contrast, human females have covert ovulation. It is speculated that covert ovulation evolved as a strategy to ensure that the male must
attend to the female almost constantly to ensure other males do not gain access. This would help reinforce the maternal–paternal–infant bond and thus ensure infant survival. But other explanations have also been put forward. If a male can detect when a female can ovulate then paternity may be more certain and this may lead to behaviors that are not in the mother’s interests. In some species, such as the lion, a male may kill the offspring of a potential mate and another male. In humans, step-fathers are more likely to murder their step-children than their biological children (Daly and Wilson 1988).

Males and females have evolved a number of physical and behavioral characteristics that may reflect sexual selection. For example, human females have evolved large hips and pendulous breasts, and males appear to have a sexual preference for women with an “hour-glass” body shape. It is suggested that large breasts and pelvis may have been interpreted as surrogate markers of a capacity to give birth and to then nourish the offspring successfully. On the other hand there are clearly major cultural overlays which influence perception of body shape. Sexual selection has also been suggested as one possible reason why humans have lost most of their body hair.

8.8 Sex Differences in Morbidity and Mortality

Differences in mortality between males and females begin before birth. The slight net excess of male births was first recorded in the seventeenth century, with about 1.05 boys for every girl registered fairly consistently across time periods, societies, and ethnicities, but the trajectory of the human sex ratio from conception to birth remained unknown until very recently. A study combining data from early embryos conceived with assisted reproduction technologies, induced abortions, chorionic villus sampling (around 12 weeks), amniocentesis (around 15 weeks), and fetal deaths and live births, showed that each sex has periods of excess mortality over the course of gestation (Orzack et al. 2015). Male embryos died more frequently in early and late pregnancy, and females in the middle, with an overall excess female mortality. The reason for this curious trend is not currently known.

In nearly all societies, women have a longer life expectancy than men (Figure 8.4). In high-income countries, two-thirds of those over the age of 80 are female. In part this difference is due to males being more vulnerable to extrinsic causes of death. Men are more likely to die from violent causes, and in the past from diseases related to smoking or alcohol consumption (although women have been catching up in this respect). In societies where violence is very much part of life, whether it be a war which can affect a generation, continual skirmishes with neighboring tribes, or in urban situations in which rival gangs fight over turf or two neighbors have a dispute, it is usually the males who fight and are killed. In many animal species the battle between rival males for reproductive supremacy is also a constant feature of life. Males from the age of puberty are more likely to engage in risk-taking behavior, but males also have a higher intrinsic mortality rate. As shown in Figure 8.5, males are more likely to die at every age from birth. The male/female ratio in a Western society (the USA) at birth to 10 years is 105%, at 50 years it is 96%, and at 80 it is 65%. The physiological reasons why males are more vulnerable to disease are complex, and some non-human primates share this trend, but an evolutionary explanation is possible. Assuming humans evolved with a mildly polygynous mating system, then the female fitness strategy depends on a longer, healthier life, whereas males depend on a period of sexual dominance, which may be relatively transient, to maximize fitness.

An example of sexual dimorphism in disease risk is cardiovascular disease. Hypertension is a feature of younger men and post-menopausal women. Lifestyle, dietary, and exercise patterns do not explain this difference, and nor do patterns of smoking. It appears that estrogens are protective, and in women the disease process is enhanced by a reduction in estrogen leading to an increasing risk after the menopause. Estrogen acts via the estrogen receptor alpha on vasculature to enhance vasodilatation, which reduces remodeling of vascular smooth muscle and counteracts the effects of oxidative stress.
Figure 8.4 Women live longer than men in high-income countries. Figure shows life expectancy at birth for men and women in 2011. From OECD (2013), with permission.

Figure 8.5 Risk of death by age and sex in England and Wales. At all ages, the risk of death is higher in men than in women. Note in particular the divergence of the curves, disfavoring men, during the reproductive ages of 15–44 years. Data plotted from Office for National Statistics (2014).
8.9 Human Reproductive Life Cycle

In Chapter 5 we discussed the physiology of and the evolutionary explanations for changes in the time of puberty. This section examines reproduction from an evolutionary perspective, from the onset of menstruation in females to the post-reproductive period.

8.9.1 Why Menstruate?

There are various competing theories about why humans and some primates such as apes, as well as the elephant shrew and some species of bats, have evolved menstruation; none of these is entirely satisfactory, although some are more plausible than others. Menstruation is the shedding of the endometrium—which undergoes progesterone-induced decidualization (differentiation) after ovulation—and the associated bleeding when progesterone levels fall at the end of the reproductive cycle. In non-menstruating species, the differentiation of the endometrium only takes place if the embryo implants, while the fall in progesterone is not followed by the breakdown of the tissue. But why is it that some mammals menstruate and others don’t? Is there an adaptive value to menstruation?

One hypothesis relates to the non-seasonal nature of ovulation in humans. Having evolved in tropical regions where there was no marked seasonality in day length, humans can conceive all year round. In comparison, many mammals, particularly those which evolved at higher latitudes, have a seasonal pattern of reproduction which allows offspring to be born in the spring and grow in climatic conditions which are likely to be favorable. Thus human females have evolved to ovulate throughout the year, although there is an echo of seasonality in the annual pattern of prolactin secretion which is influenced by day length. Humans therefore maintain the endometrium in a receptive state, that is differentiated repeatedly throughout the year, in contrast to seasonal breeders. The energetic cost of maintaining the endometrium in a receptive state for implantation might be reduced by having a cyclical pattern (Strassmann 1996). However, this hypothesis does not consider the absence of menstruation in many other species in tropical regions.

Several hypotheses focus on the decidualization of the endometrium in the post-ovulatory phase of the menstrual cycle—that is the processes involved in endometrial remodeling in preparation for pregnancy, which includes secretory transformation of the uterine glands, influx of specialized immune cells, and vascular remodeling. One hypothesis has argued that spontaneous decidualization is an outcome of maternal–fetal conflict (Box 8.4) as it evolved for advance protection against the invasive fetal tissue of the human placenta (Emera et al. 2012; Section 8.9.3). Menstruation is then just the necessary consequence of decidualization. A related suggestion posits that spontaneous decidualization may have evolved as a system for sensing the quality of an embryo, as it was found that endometrial stromal cells have a stronger response to an impaired than to a healthy embryo (Teklenburg et al. 2010; Section 8.9.2). Genome-wide screening of embryos obtained through in vitro fertilization have found that more than 70% of high-quality early embryos contain cells with large-scale chromosomal abnormalities, many of which have never been recorded in clinical miscarriage samples. Cyclical decidualization (and menstruation) could then be explained as an adaptive response to the high incidence of chromosomally abnormal embryos. Another theory links menstruation to the type of placenta that humans (and to some extent apes) have, suggesting that endometrial shedding and associated bleeding have a role in “pre-conditioning” the uterus in order to protect its tissues from the profound hyper-inflammation and oxidative stress caused by deep placentation (Brosens et al. 2009).

A well popularized but less convincing theory arising from a different perspective posits that menstruation may help to the uterus shed harmful pathogens, perhaps introduced with sperm (Profet 1993). Shedding the entire lining rather than developing very active cell-mediated (macrophage) antibody responses, as in other parts of the body such as the gut and respiratory tract, again seems unnecessarily costly. A refinement of this hypothesis is that menstruation ensures that any residual sperm are removed. Yet menstruation occurs in some cases weeks after copulation. If this is so important, however, why is menstruation not a more frequent mammalian phenomenon?
Metaphors of struggle, competition, and conflict have been used in evolutionary biology from its early days to describe implications for individuals when resources are limited, and when the genetic make-up of some individuals is facing elimination by natural and sexual selection. In the 1970s, the evolutionary biologist and theoretician of parental investment (see Box 8.2) and parent–offspring relationships, Robert Trivers, drew on William Hamilton’s work on kin selection and altruism (see Chapter 11) to explain the discrepancy between the amount of resources that the offspring demands and what the parent is willing or able to give (Trivers 1974). Because the parent and the child share only 50% of their genes, their evolutionary interests differ. Trivers constructed his model from data on the social behavior of young animals, in particular, conflict around the time of weaning. The model was immediately accepted by sociobiologists and evolutionary psychologists who described it as “inherent to the human condition” (Pinker and Bloom 1992), yet empirical studies of the behavior of young mammals gave only limited support for it. Conflict around weaning is far from universally present: in many species the offspring wean themselves, while in others mothers monitor and respond to the progress of the offspring, with respect to environmental conditions (Bateson 1994).

Then, in the 1990s, David Haig moved the focus of parent–offspring conflict discourse from post-natal behavior to pregnancy, arguing that the mother and the fetus (carrying paternal and maternal genes) are engaged in conflict over nutritional resources (Haig 1993). He distinguished three interrelated sources of conflict: (1) between genes expressed in the mother and genes expressed in the fetus and the placenta (maternal–fetal conflict); (2) conflict between maternally derived and paternally derived genes in the fetal genome (parental conflict); and (3) conflict between maternal alleles that recognize themselves in the offspring and the rest of the maternal genome. Of course, conflict can

**Figure 8.6** Nutritional interaction between a mother and her future offspring. Growth of the fetus—promoted by, for example, *IGF2*, an imprinted gene that is paternally expressed (shown in blue) in some but not all species—may signal increased demand to the placenta. *IGF2* also increases the placenta’s nutrient-transport capacity. Several nutrient transporter genes (*Slc*) are also imprinted. Genes that are in some species maternally expressed (such as *IGF2* receptor (*IGF2R*) and cyclin-dependent kinase inhibitor 1c (*CDKN1C*), a negative growth regulator (shown in dark gray), may reduce nutrient supply or demand. The block arrow (in blue) indicates the growth promoting effect of *IGF2* on the fetus. Adapted from Constância et al. (2004), with permission.
only occur if these genes somehow retain a memory of their origin. This is the case with genomic imprinting, a phenomenon where either the maternal or the paternal copy of the gene is expressed while the other copy is silenced (see Section 4.6).

Haig’s argument that imprinting evolved to solve the conflict over resources was largely based on observations in the mouse where the insulin-like growth factor 2 gene IGF2 is paternally expressed in the fetus, while the IGF2 receptor gene (IGF2R) is maternally expressed in the placenta. IGF2 has anabolic and pleiotropic actions via the IGF1 receptor, while the IGF2 receptor acts as a clearance receptor. Thus, high levels of IGF2 will promote fetal growth and nutrient utilization by the fetus, and high levels of the IGF2 receptor will lower effective IGF2 levels and impede fetal growth.

The patterns of menstruation and, more broadly, hormonal cycling, may have clinical significance. Most Western women now experience 450–500 menses across their reproductive lives. Yet in cultures in which no contraception is used, and where family size is generally of the order of four to eight, women might only experience 100–150 menses (Eaton et al. 1994; Strassmann 1999). Thus the patterns of exposure to estrogen and progesterone throughout the life course have changed dramatically, and this shift is thought to explain partially the greater incidence of cancers of the reproductive tract in women in developed societies (Chapter 12).

8.9.2 Conception and Implantation

Gametogenesis, fertilization, and implantation each provide opportunities for selection on viability and quality. Of the many eggs within an ovary, only one is generally released in each menstrual cycle. Indeed, of the many million oogonia originally formed, only a few hundred at most may ever ovulate. Is selection operating in this filtering process? Similarly, the normal ejaculate contains at least 40 million spermatozoa. Of these, not more than several thousand reach the ampulla, the middle portion of the Fallopian tube, where, in humans, fertilization normally occurs. Others are stopped at one of the several junctions, most importantly the cervix and then the utero-tubal junction. In particular the latter appears to impose a selective process by monitoring sperm surface antigens (Hunter 2012), as it has been shown that sperm cannot cross the utero-tubal junction when one of the sperm proteins is modified (Avilés et al. 2015). The Fallopian tube presents a major filtering site: here gametes need to survive until fertilization, oocytes mature, capacitation of sperm cells occurs, and transport of the immotile oocyte and embryo is enabled. The oviduct furthermore seems to be directing events outside its immediate reach, for instance by downregulating immune-related genes affecting the uterus even before the arrival of the embryo. While genetic quality must be the basis of the selection process, it is not clear what signals, and thus what traits, might be the basis of this selection.
Yet even with these filters in place, human pregnancy has a high rate of early failure. About 15% of clinically recognized pregnancies miscarry, but when combined with pre-clinical losses (“bio-chemical pregnancies”) the incidence is closer to 50% (Rai and Regan 2006). It is estimated that even more are lost prior to implantation, which generally occurs 6–12 days after ovulation (Wilcox et al. 1999). Indeed, the time of implantation seems to present an important checkpoint. It appears that developmentally impaired human embryos—for instance those containing chromosomal errors—elicit a stress response in differentiated endometrial stromal cells, known as decidual cells. In particular HSPA8, a gene encoding the ubiquitously expressed molecular chaperone HSC70, is sensitive to signals from impaired embryos and its downregulation is followed by a stress response (Brosens et al. 2014). It also appears that aberrant differentiation of endometrial stromal cells (causing them to lack the capacity to sense an embryo) is associated with recurrent pregnancy loss. The nature of the embryonic signals that trigger the decidual response is not known.

8.9.3 The Placenta

Different taxa have adopted very different strategies to care for offspring from conception to maturity. In insects the only parental involvement may be the choice of where to lay the eggs. For example, parasitic wasps lay their eggs in the larvae of the host, and this provides a good food source. In many species of fish, once eggs have been laid and fertilized there is no further parental involvement, although in other species such as the tilapia the parents guard their young. Some amphibians, some fish (e.g., the hammerhead shark), and some reptiles exhibit viviparity whereby the egg is incubated within the parent and the offspring are live-born. In some the male incubates the egg in a pouch (e.g., the seahorse) or his mouth (e.g., the sea catfish). However, in these cases the nutritional support for the developing embryo is independent of the parent, and the yolk sac is the only nutrient source. Birds develop very elaborate nesting, incubation, and post-hatching behaviors, and in some cases, such as the wandering albatross and emperor penguin, the level of parental commitment appears extraordinary when viewed from an anthropomorphic perspective.

Primitive placentae, by which nutritional support is provided by the mother to the developing embryo, have evolved more than once (Wildman et al. 2006). For example, placental structures are found in some fish, including the poeciliids, and some lizards, such as the Australian scincid lizard. In contrast the placenta of the marsupial provides only limited and transient support and the fetus is born at a very immature stage. However, eutherian mammals have evolved a particular way of nurturing their young. The yolk sac plays a nutritional role in early gestation, but then the embryo and fetus are nourished by nutrients from the mother that pass through a specialized and metabolically active organ—the placenta. This brings the fetal and maternal circulations functionally close to each other although they remain anatomically separate. The placenta also acts along with the corpus luteum to suppress maternal ovulation so that the uterus generally only contains one embryo at any stage of development. Otherwise, the uterus would expel the second immature fetus at delivery, an energetically wasteful outcome. In humans the placenta is the major site for the production of progesterone, a key hormone maintaining the quiescence of uterine muscle to avoid premature labor. In other species progesterone is variously made by the fetal gonad or adrenal gland, or the maternal corpus luteum. Progesterone withdrawal is a common feature of the induction of labor.

Whereas the overall primary function of the placenta is similar in all mammals, the details of its anatomy differ substantially. Sheep and cattle have discrete placentomes or caruncles (sites at which the placenta will develop) that are discernible even in the non-pregnant uterus; there are usually about 100 of them in the sheep, for example. Horses and camels have microcotyledons, much smaller structures which develop over a much wider area of the endometrial lining. In these species there are five layers of cells between the maternal and the fetal blood. Nutrients, hormones, and gases must pass across these layers by diffusion, or active or facilitated transport. This type of placenta is called epitheliochorial.

Humans have a single site for placentation, although the placenta itself is much larger and is
REPRODUCTION

divided into cotyledons which receive branches of the uterine blood supply. The human placenta is termed hemochorial because the villi, which are of fetal origin, project into a maternal blood space, and it is across these villi that nutrient and gas exchange occurs. There are only two layers of cells through which nutrients must pass: the microvillus membrane and the capillary endothelium. After the invasion of the trophoblast and the erosion of the uterine blood vessels, the placental microvilli are bathed directly with maternal blood from sinusoids which open directly into intervillous spaces. Pulsatile maternal arterial blood is squirted over the villi in these spaces rather like rain falling on a tree. Clearly this gives an efficient exchange process, but it also raises the stakes: if the maternal vasculature is not eroded sufficiently, then the rising cardiac output and blood volume in the maternal circulation as pregnancy progresses will lead to hypertension and pre-eclampsia.

The reason for the spectrum of variation in mammalian placentation is a matter of debate. The traditional view, first proposed in the early twentieth century, was that shallow placentae with greater barriers between maternal and fetal blood arose early in evolutionary time; the deeply invasive and eroded human placenta, which brings the fetal and maternal blood into closer contact, evolved later. More recently, conflict theory (Box 8.4) was invoked to explain the evolution of the placenta towards the more invasive form because it would have been an advantage for fetal genes to increase invasiveness and access to maternal resources, and an advantage for maternal genes to restrict invasiveness (Crespi and Semeniuk 2004). However, it appears that the epitheliochorial placenta is not the ancestral state; that combinations of placental shape, invasiveness, and structure are diverse across mammalian species; and that placentae evolved multiple times (Sterner et al. 2013).

There is considerable commonality in the patterns of genes expressed across all mammalian placentae early in post-implantation development, although in late gestation species-specific patterns of gene expression emerge. For example, there are considerable differences between the patterns of placental gene expression in late gestation in humans and other primates. This suggests that after a common early phase of development there has been divergence leading to the particular placental pattern which in turn relates to life-history traits. Thus more invasive hemochorial placentae are found in species with longer gestation periods carrying individual offspring, while less invasive placentae evolved in species characterized by short gestations and multiple offspring (Garratt et al. 2013).

There is tremendous variation in the length of gestation as well as the degree of maturation at birth across species. Marsupials give birth to very immature embryos which then mature postnatally in the pouch. In the wallaby, oxygen transfer at the time of birth must occur across the skin as the lungs are not yet mature. As soon as the joey is born it is possible for the mother to implant another embryo; indeed, she may carry several unimplanted fertilized embryos in her uterus, a state known as reproductive diapause. Reproductive diapause is also used in seasonally reproducing mammals such as the brown bear, to match birth to the climatic conditions so that offspring are born in the spring. Humans evolved in the tropics, where births need not be seasonal. There is no evidence that human embryos can delay implantation while awaiting the right maternal cue. Eutherian mammals have a placenta which supports the fetus to a relatively greater degree of maturation. However, some species (altricial species) such as the mouse or cat are born at a very immature stage, and there is total dependence on the mother for care until the animal develops motor, thermoregulatory, sensory, and other capacities. Other more precocial species such as the guinea pig can be virtually independent from their mothers from birth. Primates generally have long pregnancies with offspring born in a neurologically mature state. Humans are distinct from other primate species in having regressed to a secondarily altricial state because of the need to terminate pregnancy relatively early to allow the relatively large fetal head to pass through the pelvic canal. The implications of this were discussed in Section 6.3.7.2.

8.9.4 Maternal–Fetal Interactions

There are important consequences of the evolution of a placenta as part of the strategy for supporting the next generation. Because the mother has become
a host to a developing organism which is not genetically identical to her there is a risk of an immunological conflict between them. The high degree of exchange between the mother and fetus can lead to the passage of antigens, such as blood cells, from the fetus to the mother in late gestation and immunoglobulin G antibodies from her to the fetus. As a result, fetal DNA can be detected in maternal blood and this can be used in antenatal diagnosis. Under these circumstances the limitations of the placenta as an effective immunological barrier can be demonstrated. Rhesus (Rh) incompatibility provides a good example. The Rh blood-group antigens are cell-surface antigens that are extremely polymorphic, and as a result antigenic differences may arise between mother and fetus. The function of the Rh cell-surface proteins is unknown, but they are molecularly related to ammonia transporters present in yeasts. There are two major classes of Rh antigen which formed from gene duplication and occur adjacent to each other on the short arm of chromosome 1. One site codes for the so-called D antigen and, depending on molecular evolutionary events including recombinations, deletions, and polymorphic mutations, the site may be functionally deleted—this state is termed Rh-negative, or D-negative. There is enormous ethnic variation in the distribution of these alleles in the population. If blood cells from a D-positive fetus reach the maternal circulation, a D-negative mother will form antibodies which can then cross back to the fetal circulation and lead to fetal hemolytic anemia. This is known as rhesus isoimmunization disease. The condition is likely to be worse in subsequent pregnancies if the mother is re-exposed to a D-positive fetus, leading to further antibody production, hydrops fetalis, and fetal death. Given the serious consequence of hemolytic disease of the newborn, one could expect that negative selection would have acted against the D-negative phenotype. Yet, it appears that at least in European populations the D-negative phenotype is typically present in 15% of the population and sometimes more. It has been hypothesized that D-negative phenotype might have conferred some sort of selective advantage, but no compelling argument has been made and no evidence of positive selection has been found on the RHD deletion allele (Perry et al. 2012).

Pre-eclampsia is a common disorder of pregnancy in which the woman develops high blood pressure, proteinuria, and edema in about the twentieth week of pregnancy. These symptoms may progress to severe maternal renal disease and severe hypertension leading to seizures, a life-threatening syndrome known as eclampsia. In addition, the fetus may develop severe growth restriction and die in utero. The cause is unknown but is believed to relate to immunological difference between mother and fetus. Pre-eclampsia involves disordered trophoblastic invasion of the uterine vasculature, thus leading to inadequate erosion of the spiral arteries and lack of a fall in vascular resistance. As a result, blood flow to the placenta is compromised, leading to fetal growth retardation, which may be fatal for the fetus.

Pre-eclampsia affects between 5 and 10% of pregnancies, yet it appears to be a syndrome unique to humans. With such a high incidence, and it being a clearly disadvantageous syndrome, an evolutionary perspective may provide some additional insights. The incidence of pre-eclampsia decreases with successive pregnancies by the same partner. If a woman changes her sexual partner, the risk of pre-eclampsia in pregnancy increases again to become the same as her first (primiparous) pregnancy (Li and Wi 2000). The risk is also higher in pregnancies from matings between close relatives, and there are clearly some male-associated factors: pre-eclampsia in a pregnancy is more likely if the man had previously fathered a pregnancy in another woman that was associated with pre-eclampsia.

These factors, together with the probable immunological basis of the disease, raise the possibility that humans evolved a process which to some extent favors the production of offspring from stable relationships. A recent study looking at the post-natal health of offspring whose mothers had hypertension at some stage during their pregnancy found that hypertension in early pregnancy is associated with reduced risk of disease in the offspring in later life, while hypertension in the second half of pregnancy had the opposite outcome (Hollegaard et al. 2013). These findings were explained as an expression of parent–offspring conflict, where in early pregnancy investments in fetal growth are associated with relatively low maternal risks, while
in late pregnancy risks posed to the mother exceed the benefits to the child. However, there will be many other factors operating in such pregnancies, so interpreting the findings in terms of conflict must be done with caution.

Pregnancy sickness is another example of a complication during pregnancy for which evolutionary arguments have been made, although debate on whether it has adaptive value or not still continues (Box 8.5).

### 8.9.5 Regulation of Fetal Growth

In monotonous species, fetuses generally grow to a relatively large body size compared with offspring of similarly sized polytocous species. Indeed, in humans the mean birth size of multiple pregnancies is less than that of singleton pregnancies. This suggests limits on the potential for fetal growth. These limits are defined by the capacity of the utero-placental unit to supply nutrients to the fetus, and the processes underpinning these limitations are generally referred to as maternal constraint. Maternal constraint implies that fetal growth is not solely determined by genetic drivers; indeed, estimates based on birth-size correlations in humans suggest that perhaps only 40% of the variance in fetal size can be attributed to genetic factors, with the remainder being due to the intrauterine environment. An examination of birthweights between first-degree relatives shows that there is a higher interpair correlation between half-siblings who share the same mother than those who share the same father (Gluckman and Hanson 2005, p. 42). Thus paternal genetic factors appear to play little direct role in regulating fetal growth. Adult body size, by contrast, is highly correlated with maternal and paternal size (as well as being influenced by gender and developmental history).

Fetal growth is directly regulated by the supply of nutrients across the placenta, leading to altered levels of the fetal anabolic hormones insulin and insulin-like growth factor 1 (IGF1), whereas postnatal growth is largely regulated through growth hormone. The origin of this difference in growth regulation has probably been to separate fetal growth from paternal genetic control and to ensure a close linkage between fetal growth and nutrient availability, which in turn is linked to maternal size. Growth hormone-dependent growth develops after birth in all altricial and precocial mammals studied so far.

The greater part played by maternal versus paternal factors in the development of offspring phenotype is a general principle which appears to operate across a wide range of animal and plant species. From an adaptive perspective it makes sense because the female is more likely to give accurate cues of local environmental conditions than the male: this may be true in plants where the male gamete in pollen can be carried a long way by wind or insects and in mammals where the integration of environmental cues through pregnancy and lactation may be more representative of conditions than the production of sperm at one point in time.

The classic demonstration of maternal constraint was made by Walton and Hammond in the 1930s. They crossed the very large Shire draught horses with the much smaller Shetland pony breed, and showed that the size of the offspring at birth depended on the mother's breed: Shetland mothers gave birth to small foals while Shire mothers had large foals (Walton and Hammond 1938; Figure 8.8). Importantly, this study showed that the Shetland fetuses were capable of growing to a larger size than they did in their usual constraining uterine environment where both parents were Shetland ponies. Subsequent experiments using embryo-transfer techniques confirmed these results and showed the importance of non-genome-mediated regulation of fetal growth. Maternal constraint operates through incompletely understood mechanisms, including determinants of uterine–placental function which in turn are linked to maternal size. Thus size at birth correlates with maternal stature—and her birthweight—but not with paternal stature or birthweight.

The evolutionary importance of this mechanism is illustrated by considering what would otherwise happen if a large male mated with a small female. This would have been a common scenario, particularly in sexually dimorphic species, and would have led to a high risk of dystocia and death of mother and fetus. Accordingly, mammals must find an equilibrium between maternal size and fetal growth which allows the fetus to be born as large as possible.
Pregnancy sickness, or a combination of symptoms including food and odor aversions, nausea, and vomiting, is experienced by about two-thirds of pregnant women, mostly in the first half of pregnancy but in some cases until birth. Yet despite its widespread nature and the associated debilitating discomfort, the reasons for its occurrence are still a matter of speculation. Some investigators, using evidence concerning proximate mechanisms, argue that it is a non-functional by-product of hormonal changes in pregnancy. Intense pregnancy sickness is associated with a healthy viable fetus, so the observation that mothers with morning sickness are less likely to experience spontaneous abortion (which is mostly due to chromosomal abnormalities or low-quality embryos), has been used as supportive evidence of the by-product theory. Other scholars stress that nausea and vomiting are in general functional adaptations so pregnancy sickness should be regarded as such too.

One popular theory of the basis for pregnancy sickness is the “prophylaxis” hypothesis, which posits that it evolved as a mechanism for causing pregnant women to expel potentially harmful foods. The negative reinforcement would then lead to future avoidance of such foods. Thus, the fetus is protected from substances that may be harmful to its development; this viewpoint received much publicity in the 1990s (Profet 1992). These substances included pungent and bitter vegetables such as broccoli, cabbage, and Brussels sprouts that are thought to contain high levels of naturally produced toxins that ward off insect attacks; foods with a burnt odor (suggesting the presence of mutagenic compounds); and foods that smell spoilt (indicating the presence of bacteria). This theory is purely speculative and has provoked much controversy.

Other authors pointed out that different aspects of the maternal diet may be more powerful triggers of pregnancy sickness. These include meats (which are historically more likely to contain pathogens), alcohol, and caffeinated drinks. A cross-cultural meta-analysis has indeed demonstrated correlations between morning sickness and high intake of these foods in different countries (Flaxman and Sherman 2000). It has been suggested that the variable incidence of morning sickness across cultures could be accounted for by cultural differences in dietary intake. Building on these studies, and using the ethnographic record of food taboos in pregnancy across a number of societies, it has been suggested that pathogen-contaminated meat was probably the key target of nausea and aversions (Fessler 2002; Figure 8.7). The early months of pregnancy are characterized by strong reproductive suppression: following implantation, the trophoblast

![Figure 8.7](image-url)
to maximize neonatal viability without compromising the mother. Maternal constraint mechanisms are particularly important in monotocous species, especially humans where the shape of the maternal pelvic canal creates additional risks of dystocia (Hanson and Godfrey 2008).

A distinctive feature of humans is the very large brain, which creates a delicate equilibrium: while the fetal head readily passes through the pelvic canal in chimpanzees, the head of a human fetus cannot pass through the comparatively narrower pelvis (a necessary consequence of bipedalism; Section 6.3.2) without considerable twisting and turning (Figure 8.9). As a result, and in contrast to other primates, the head is delivered facing the dorsal aspect of the mother. Indeed, it has been suggested that this led to a fitness advantage arising from living in a clan where a birth attendant was needed to help deliver the fetus and clear the airway of the neonate.

The importance of maternal constraint processes operating even in normal, uncomplicated pregnancies is shown by a recent very large population study in the Netherlands where it was found that mean birthweight (that is, the result of maternal constraint and the optimal weight for the mother) was significantly less than the optimal birthweight for fetal and neonatal survival (Vasak et al. 2015).
8.9.6 Lactation and Post-natal Care

The offspring of all mammals are nutritionally dependent on their mother’s milk for some time after birth. Mothers must invest considerable energy in their offspring in this phase. If huge resources have been devoted to a long gestation, it does not make sense to abandon the energetic investment made during that time. Earlier in this chapter we discussed the potential ways in which human behavior has evolved to promote survival during this phase. There is evidence in rodents of a considerable interplay between maternal and offspring behavior which is influenced by imprinted genes. Again, those who are advocates of conflict theory have argued that this is an interaction where maternal genes might restrain infant nutrition and paternally imprinted genes might enhance it (Haig 2014). But the failure of the mother to suppress lactation even in famine conditions suggests that the conflict metaphor is inappropriate when considering the biology of human lactation.

While traditionally humans have breastfed for several years, exclusive breastfeeding is much shorter: current guidelines recommend a duration of around 6 months. Breast milk provides all the nutrients required at this early period of intense development yet high immune vulnerability, in a form that is hygienic and easy to digest. Its protein, carbohydrate, and fat profiles are unique to humans and differ in many ways from other animals: the diluted,
low energy content of human milk matches the protracted “slow lane” human life history (Milligan 2013). Breast milk also contains bioactive components such as digestive enzymes, hormones, growth factors, and specific binding molecules that increase the bioavailability of micronutrients. Antimicrobial agents such as white blood cells, immunoglobulins, lysozyme, complement, and others are essential at the time when the infant immune system is still undeveloped. Indeed, breast milk plays a key role in the development of the infant’s immune system and gut microflora (Box 10.5).

Breastfeeding is a life-saver for children in low- and middle-income countries, where sterile water and bottles may not be available. There have been many tragic cases of death from infant diarrhea as a result of the feeding of formula aggressively marketed by food companies in conditions where it could not be properly prepared. There are many beneficial effects of breastfeeding for all children. They include reduced risk of infection, especially of the gastrointestinal tract, lung, and ear; more appropriate growth patterns, with lower risk of childhood obesity; reduced risk of type 2 diabetes and cardiovascular disease; better cognitive function and neurodevelopment; reduced incidence of allergy and atopic disease; improved bone health; and a reduced risk of breast cancer in the mother. Despite these obvious benefits not all women can or do breastfeed their infants for the minimum 6-month period recommended. Cultural attitudes to breastfeeding and the lack of appropriate facilities in the workplace contribute to this, and social class and education play a major role.

It is obvious that parents continue to care about their children throughout life. Many view this from a cultural perspective as a uniquely human characteristic, but there are also other species such as the elephant where the offspring remain in the protective environment of the family group after weaning. Humans evolved as a cooperative species, and while there is still room for debate regarding the role of higher levels of selection (Section 2.3.2.4), such processes may have largely evolved through kin selection, which drives behaviors including maternal and family care that assist in the preservation of genetically closely related individuals and groups.

But there are limits on nurture: not all interests are shared between offspring and parent, and other types of “conflict” will emerge as a child gets older. During adolescence the conflict between parental and children’s interests can be very acute, and every family knows the kind of arguments which ensue (Section 11.4).

8.9.7 The Menopause

The menopause reflects the end of the capacity to ovulate and is accompanied by a decline in secretion of estrogen and progesterone. The lack of these ovarian hormones has a number of consequences for the post-menopausal woman, including thinner skin, reduced vaginal secretions, and loss of bone mineral which can lead to osteoporosis. The menopause is essentially unique to human females, although there is some evidence to suggest that it may exist in pilot whales. Human males, who do experience reproductive senescence, do not become infertile with age. Why the menopause evolved in humans has been a subject of considerable debate.

The menopause can be considered as evidence of a highly successful life-course strategy, in the sense that a woman has lived long enough to terminate her reproductive phase of life naturally. While many other species show a reproductive decline as they get older, only humans show complete cessation of ovulation regularly and consistently. For example, African elephants show a 50% decline in reproduction beyond the age of 50 years, but only about 1 in 20 female elephants lives longer than this in the wild. Female rhesus monkeys also show a reproductive decline as they age beyond 20 years, but again this approximates to their lifespan and they do not show a complete cessation of reproductive capacity. The only other primates in which menopause is reported are large species such as gorillas kept in zoos, where lifespan is well beyond that generally seen in the wild. Thus there may be an inherent tendency towards ovarian failure in animals that is not seen under natural conditions, but which is exposed if they live longer when housed in artificial conditions in zoos. It is of interest therefore to consider that human lifespan is now on average much longer than when our species first evolved (see Chapter 1 and Box 5.3).
While the menopause is defined by the termination of menstruation, fertility starts to decline well before then. The progressive loss of oocytes actually starts before birth. From the millions of ova formed in the primitive ovary, most are lost by apoptosis and only a few hundred by ovulation (Figure 8.10). The menopause is not the reverse of puberty but is the result of ovarian failure while control of gonadotropin release remains functional. Indeed, it is the continued secretion of hormones that causes some of the symptoms of the post-menopausal state, such as hot flushes.

We can only speculate about the timing of menopause in Paleolithic times, but modern hunter-gatherers generally enter menopause by the age of 40–45. It is closer to 50–55 in developed societies (World Health Organization 1981). While there is some evidence that environmental factors such as smoking can influence the timing of the menopause, the effect size is small. There is no evidence that developmental factors such as the age at menarche play a significant role in determining the age of menopause.

There are considerable difficulties of interpretation, and varying viewpoints, on the evolutionary origins of the menopause. They can be broadly divided into epiphenomenal and adaptationist explanations. Epiphenomenal and indirect selection models focus on aspects of reproductive function such as the number and attrition rate of oocytes and hormone-producing follicles over the lifespan. They argue that menopause is a by-product of a highly conserved mammalian pattern of pre-natal oogenesis with atresia over the lifetime. Because deleterious mutations accumulate in oocytes, females enter menopause whereas sperm quality is somewhat protected by the continuous nature of spermatogenesis. With extension of the lifespan, this scenario would lead to uncoupling of menopause and post-reproductive life (Sievert 2011). In another model, female longevity after menopause has been postulated to be a epiphenomenon of selection for longer life in men (whose reproductive ability does not disappear with age), allowing alleles associated with longevity to spread in both sexes (Bribiescas 2001; Tuljapurkar et al. 2007).

Figure 8.10 Changes with age in the total number of follicles in the human ovaries. Plotted from data in Timiras et al. (1995).
The alternative argument is that menopause has a specific adaptive advantage and was the direct result of selection. To be accepted, this argument must demonstrate a fitness advantage, and this is difficult for a trait which only appears after reproduction is complete. To advance a fitness argument we therefore have to think in terms of inclusive fitness and seek indirect advantage for women exhibiting menopause.

Women show a decline in fertility from well before menopause, starting at around 35 years of age. This may simply reflect the effect of ageing eggs noted above, but there may well also be an adaptive advantage in having evolved such a pattern. Mothers needed to nurture their children well into childhood for them to have a higher probability of surviving to reproduce. Modeling can show that the probability of having more surviving offspring can be enhanced if a woman stops having children in time to support the development of her youngest child. For a species that reproduces slowly this may be a better strategy for maximizing fitness. It balances the cumulative and rising risk of death from the next pregnancy and childbirth against the survival of the last child, which might otherwise perish if the mother were to die. Thus in evolutionary terms it may pay to stop having children and to invest in those who have already been born. This suggestion was the first adaptationist theory of the evolution of menopause and is known as the “mother hypothesis” (Williams 1957).

A second but non-mutually exclusive adaptive argument relies on the concept of inclusive fitness. A mother with many children may be limited in her capacity to support all of them. Thus, as she ages, her youngest may be compromised. But if she is helped by her own mother she can learn mothering skills at a younger age and more of her children are likely to survive. In turn, the grandmother assists her own fitness by ensuring more of her own grandchildren live until adulthood. This is the so-called grandmother hypothesis (Hill and Hurtado 1991). Modeling again shows that this can confer a fitness advantage. It has empirical support in studies of child survival in Africa where child mortality is lower in families where the maternal grandmother lives in the same village. The presence of older women, and indeed older men, might also confer other advantages through accumulated knowledge, which can help a clan in times of stress (Kaplan et al. 2000). Analogous arguments have been used to explain the role of the matriarch in an elephant herd. This might explain why human societies in general have maintained elderly people as an integral and respected component of the clan. Recent modeling studies suggest that a combination of these two adaptive arguments confers a greater fitness advantage than either alone (Box 8.6).

Thus the current consensus is that menopause is an evolved rather than an accidental phenomenon.

---

**Box 8.6 Using Modeling Studies to Explain the Menopause**

Modeling can be a useful approach for investigating complex phenomena. It involves mathematical and computational analysis of theories for which parameters of known variables have been established. Modeling can then predict the outcomes when the variables are changed.

Evolution of the menopause in humans has been studied this way using survival and fecundity data from a Taiwanese pastoral population (Shanley and Kirkwood 2001). To investigate the effects of human altriciality and grandmother assistance on the age of menopause, two models were generated to look at each factor in isolation, and a third model was developed to look at a composite of the two previous models. Many variables were incorporated into the analyses, such as childbirth-induced maternal death, juvenile survival, fecundity, and the availability of a grandmother’s assistance. Various combinations of the variables were examined under fixed or changing conditions. Essentially, it was shown that the first two models were insufficient to explain age at menopause. Only when the presence of the grandmother was combined with maternal survival for nurture of her youngest child did the model fit with the available data. Indeed, this theoretical life-history model was subsequently verified empirically using a comprehensive data set from The Gambia (Shanley et al. 2007).
8.10 Conclusion

Given that the net effect of evolutionary forces is eventually expressed through successful reproduction, it is to be expected that there is a broad diversity of reproductive strategies across the animal kingdom. This is a simple reflection of the diversity of biotic and abiotic environments that organisms have evolved to live within. Changes in reproductive strategy underlie, or are reflected in, key evolutionary steps such as the transition from unicellular to multicellular organisms. There is evidence from a species of green alga (*Volvox*) that cell differentiation into nutritionally supporting and reproductive cells was an important early component in the origin of complexity. The development of sexual reproduction appears to have given a major advantage in allowing organisms to survive in complex biotic environments and to sustain defenses against parasites. Placentation emerged several times across taxa and, combined with lactation, has created the basic mammalian reproductive strategy. It requires considerable pre- and post-natal maternal investment, which in turn creates a number of gender-related interactions. The evolution of mammalian imprinting appears to be directly linked to the evolution of eutherian (placental) mammals.

The human life history has a number of distinct elements. These include delayed maturation, predominantly singleton pregnancies, a long inter-birth interval, high post-partum parental investment, and, in females, complete reproductive decline before the end of the lifespan, at least in the modern world. Both the social and physical environment prior to maturity can affect age at puberty. The changing relationship between the age of physical maturation and the age of acceptance or performance as an adult in developed societies has consequences for mental and societal health. Equally, the rapid change in female reproductive behavior associated with women’s empowerment and widespread use of effective contraception has led to a major change in when women choose to conceive their first child (Figure 8.11). This changing reproductive behavior may be at odds with our evolved biology, by which ovarian function declines from the age of about 35. This mismatch between expectation and biology is partially addressed through assisted reproductive technology such as *in vitro* fertilization.
Key Points

- Sexual reproduction is not universal across species, and the favored hypothesis is that it evolved as a strategy to address the threat from parasites.
- There is a broad range of reproductive strategies across species, which have evolved to maximize fitness for each species and each sex according to the physical, biotic, and social environment.
- The investment of males and females in reproduction is very different. Each species has evolved strategies in which an equilibrium of interests between the sexes is reached. This is reflected in patterns of parental investment, mate choice, and social structure.
- The mammalian reproductive strategy relies on placentation and lactation to provide nutritional support and immune and physical protection to the offspring.
- The human reproductive strategy includes relatively delayed puberty and few, generally singleton, pregnancies with a high parental investment in the offspring.
- The menopause is a distinctly human characteristic with several potential evolutionary explanations.
- Sexual dimorphism in body shape, behavior, and physiology is reflected in the higher risks of morbidity and shorter lifespan of human males.
- There is a growing mismatch between the evolved rate of ovarian ageing and the sociological trend for women to delay the age of first pregnancy.
CHAPTER 9

Nutritional and Metabolic Adaptation

9.1 Introduction

How an organism adapts physiologically to the available food supplies and adjusts its behavior, metabolism, and energy expenditure accordingly are defining characteristics of a species and major points on which selection has acted. Indeed, life-history theory, as discussed in Chapter 5, focuses on how an organism’s allocation of finite energy during different phases of its life course, including growth and the reproductive period, is fundamental to its evolved strategy as a species and is thus under strong selective pressure. Humans have evolved as a generalist species able to derive energy from a range of food sources and, uniquely among animals, to use technology to gather food and alter energy expenditure. These characteristics of human evolution are important in understanding how modern environments influence the patterns of health and disease.

Obesity and the metabolic disorders associated with it, such as type 2 diabetes mellitus and cardiovascular disease, are today often described as a global “epidemic” (Figure 9.1). The epidemic is well established in high-income countries, but is also of rapidly growing importance in low- and middle-income countries. The simple explanation that humans are living longer, and so becoming more susceptible to these diseases as they age, is inadequate given the increasing incidence of obesity and metabolic disorders among younger people. For example, in the UK about 30% of children aged 2–15 are overweight or obese (Public Health England 2015); childhood obesity is associated with multiple co-morbidities that affect all major organ systems and psychosocial well-being. In this chapter we use evolutionary perspectives to explore the changing patterns of obesity and its associated metabolic disorders that are sweeping the globe today.

The balance between energy consumption and expenditure has changed substantially in modern times. Modern city-dwelling humans consume large amounts of carbohydrates with a high glycemic index and excess amounts of refined sugars, particularly in beverages, as well as fatty foods, but generally have low levels of physical activity. How does this compare with the food and activity patterns of ancestral hominins? The information we have about such aspects of the lives of our ancestors is inevitably limited and somewhat speculative, but it suggests that humans are now in an environment of “evolutionary novelty” that involves exposures beyond the levels previously experienced in our evolution, and that this environment exceeds our evolved capacity to efficiently regulate metabolism, appetite, and food preferences. In other words, human biology, which is the product of evolution over millions of years, is metabolically “mismatched” to the novel and rapidly changing environments that we now inhabit.

In this chapter we will explore evidence that evolution has indeed resulted in a human metabolism that is not equipped for modern environments, and which now contributes to our patterns of susceptibility to non-communicable diseases. We will also show how thinking about the evolution of these conditions is rapidly moving to focus on a more complex interaction: not merely between our changing lifestyle and our inherited “ancient” genome, but also potentially adaptive developmental adjustments which our bodies and metabolism make to local environments very early in life.
9.2 Strategies for Energy Storage

The deposition of energy reserves primarily in the form of fat is not simply the passive outcome of excessive consumption of food but is an integral component of the adaptive strategy for many species, from the fat body located in the abdomen of insects to the subcutaneous fat of hibernating mammals such as the grizzly bear. Fat deposits also provide thermal insulation in cold-adapted species such as marine mammals. Fat is the major way in which the body stores excess energy—the net difference between energy intake and that expended in maintenance, growth, reproduction, repair, and physical exercise. Each aspect of this balance is under regulatory control.

Not all fat stores have the same biology or physiological significance (Box 9.1). While excess fat can have adverse consequences in any region of the body, it is increased deposition of ectopic fat in particular, including visceral, hepatic, and intramuscular fat, which is associated with the development of insulin resistance, vascular dysfunction, and eventually diabetes (Box 9.2; Shulman 2014).

A significant proportion of the energy consumed by an adult organism is directed to maintenance and repair of its structure (Chapter 5). This includes ensuring function of vital organs such as the brain, heart, and kidneys. Indeed, the adult human brain expends of the order of 25% of all calories consumed at normal levels of intake (Holliday 1986). There is also an essential component expended in non-exercise activities such as standing and fidgeting—this is known as non-exercise activity thermogenesis, or NEAT. Beyond these basal maintenance requirements is the energy required for physical activity—walking and running, hunting and gathering, agriculture or work—and the energetically costly reproductive processes of pregnancy and lactation. In addition, young animals require energy for growth.
Adipose tissue was long considered a passive organ for energy storage, simply accumulating triglycerides in its lipid droplets and releasing them according to the body’s changing requirements. Within the last few years, that view has been overturned by the discovery of numerous hormones and other cytokines (termed adipokines) secreted by adipocytes, the fat-storing cells of adipose tissue. These secreted factors serve as signals, and have wide-ranging effects in the control of appetite, energy metabolism, immune function, and even reproduction. Importantly for metabolic disease, it is now becoming apparent that superficial subcutaneous adipose tissue is biologically different from visceral adipose tissue (the fat stores surrounding abdominal organs, apparent as “truncal” or “central” obesity), and perhaps also from deep subcutaneous tissue, in terms of metabolic and endocrine characteristics. These differences may underlie some of the metabolic disturbances associated with obesity and the metabolic syndrome (Lebovitz and Banerji 2005). Visceral adipose tissue is a labile depot, meaning that it is more sensitive to the hormones that stimulate the release of fatty acids during a fast or in response to a stressful challenge. Release of fatty acids into the hepatic portal venous system can induce insulin resistance. Improvements in insulin sensitivity in response to weight loss are more closely related to the loss of visceral adipose tissue than to the reduction in total body fat or subcutaneous fat.

Indeed, recent observations with magnetic resonance imaging demonstrate that visceral obesity, with its pathological consequences, may occur hidden from view in otherwise apparently lean individuals. This has been termed the “thin outside—fat inside” syndrome, and it has been suggested that some populations may be at particular risk of this condition owing to a combination of genetic and developmental factors.

The pathophysiology of type 2 diabetes is complex: it is generally associated with visceral obesity, supporting the view that the insulin resistance is related to some degree of abnormality in fatty acid metabolism that leads to lipid deposition in muscle and liver and loss of insulin sensitivity in these tissues. For this reason, lack of exercise, leading to reduction in muscle mass, is an additional risk factor. Increased insulin resistance requires increased insulin secretion to maintain normal blood glucose levels, and thus a typical feature of early disease is higher levels of insulin than anticipated for a given level of glucose. It is likely that this state of enhanced insulin secretion acting over a number of years in the pre-diabetic phase eventually exhausts the capacity of the beta cells to secrete sufficient insulin, causing hyperglycemia and symptomatic type 2 diabetes (Yki-Järvinen 2011). Type 2 diabetes can often be controlled by drugs that mimic or stimulate the actions of insulin rather than needing treatment with insulin itself. For example, the sulfonylureas stimulate insulin release from beta cells, metformin acts to decrease blood glucose levels by inhibiting gluconeogenesis and increasing insulin sensitivity, and the thiazolidinedione drugs such as rosiglitazone increase insulin sensitivity by acting on transcriptional mechanisms within the insulin signaling pathway.

Historically, type 2 diabetes typically appeared in middle age, but in recent decades, as obesity has become more common at younger ages, type 2 diabetes is increasingly appearing in the second or third decade of life. Very rarely, specific genetic defects in the pathways to insulin action lead to monogenic causes of type 2 diabetes. MODY, or maturity onset diabetes of the young, refers to a mixed group of monogenic causes of diabetes—the most common being mutations of glucokinase (MODY type 2) and hepatic nuclear factor 1α (MODY type 3).

Although type 2 diabetes has a strong familial element, the search for causative genes has in general been rather unproductive. For example, meta-analyses of GWAS have shown that 44 independent loci significantly associated with type 2 diabetes could, collectively, only explain about 10% of observed familial clustering in Europeans (Wheeler and Barroso 2011).

The metabolic syndrome is a constellation of abnormalities frequently associated with type 2 diabetes, including central obesity, hypertension, hyperglycemia, and dyslipidemia. Diagnosis of the metabolic syndrome implies increased risk for cardiovascular disease. Insulin resistance is thought to be a key underlying defect in the metabolic syndrome, although some degree of systemic inflammation, possibly associated with release of inflammatory factors from visceral adipose tissue and endothelium, is likely to also contribute to the hypertension associated with the syndrome (Goldberg 2009).
The main sources of energy in the diet are carbohydrates and fats, whereas protein is needed to provide amino acids for growth and support the turnover of tissues. These food constituents of fat, carbohydrate, and protein are referred to as macronutrients. Additionally, mammals require micronutrients in the form of vitamins (that function as co-factors in enzyme reactions, as hormones, and as antioxidants) and minerals (e.g., calcium for bone structure, iodine for thyroid hormone synthesis, and sodium/potassium to maintain the ionic composition of bodily fluids). Although mammalian intermediary metabolism allows interconversion between macronutrients, certain essential amino acids and fatty acids cannot be synthesized by the body, or only to a limited extent, and must thus be obtained from the diet. It is also important to consider the role of the gastrointestinal microbiota (all the microorganisms within a particular environment), which itself utilizes some nutrients and generates others (Box 9.3). What is absorbed from the gut is therefore a function not only of what is eaten but also of gastrointestinal function and the gut microbiome (i.e., the microbiota, their genomes, and products generated by the host and microbiota).

A typical adult human male consumes about 1 million calories each year. If weight is stable, this means that he is also expending 1 million calories, suggesting that humans balance their energy intake and energy expenditure with remarkable accuracy. Only a small imbalance is required to lead to progressive weight gain. So, faced with the choice of foods available to us in a modern diet, how do we choose what and how much to eat in light of the constantly changing demands for energy that we face in our daily lives?

Neuroendocrine mechanisms have evolved to help ensure that energy balance is generally maintained, other than in chronic scarcity or starvation or in disease-induced cachexia. Various inputs from the periphery—signaling blood glucose levels, adiposity, nutrient availability, and filling of the gastrointestinal tract—are integrated centrally in the hypothalamus (Lustig 2010). The long-term control of energy balance involves the adipoinisular axis: as adiposity increases, rising plasma leptin concentrations act on the hypothalamus to decrease food intake and on the pancreatic islet to lower circulating insulin levels, thereby reducing adipogenesis. On the other hand, when adipose stores decrease,

---

**Box 9.3 Energy Balance and the Microbiota**

The metabolic activities of the gut microbiota (see also Section 10.3) have come into focus in the context of increased rates of obesity and type 2 diabetes. Recent studies have revealed that the coevolved gut microbiota is involved in the digestion of food and in the production of peptides and fatty acids that can influence metabolism, appetite, and satiety. In animal models, manipulation of the microbiota can influence the development of dietary-induced obesity, and early clinical studies in humans show alterations in the microbiota linked to disrupted metabolic homeostasis (Rosenbaum et al. 2015).

A series of elegant studies showed that germ-free mice brought up in sterile conditions ate more than conventionally raised animals, but paradoxically were leaner (Bäckhed et al. 2004). Recolonization of the germ-free mice with the gut contents of the normal animals led to restoration of the “conventional” phenotype—the animals ate less but became fatter, and also developed insulin resistance. These metabolic effects of the gut microbiota were mediated by a number of mechanisms, including increased energy harvesting from the gut by breaking down otherwise indigestible complex carbohydrates into absorbable monosaccharides, accompanied by an increased capacity of the host intestine for monosaccharide transport. Additionally, the gut microbiota appeared to directly stimulate peripheral fat storage in the host.

In humans, alterations in the balance of the major phyla comprising the gut microbiota are associated with changes in energy balance, for example as seen in twin studies (Turnbaugh et al. 2009) and in individuals who have undergone bariatric surgery (Tremaroli et al. 2015). Generally, weight loss is associated with a decrease in the ratio of Firmicutes to Bacteroidetes; there is some evidence that the Firmicutes may be more efficient at energy extraction. Nevertheless, it remains unclear whether changes in the microbiome are a causative factor in the development of obesity or merely a secondary effect of the condition (Harley and Karp 2012).
falling plasma leptin concentrations increase feeding activity and permit increased insulin production, resulting in the deposition of additional fat (Kieffer and Habener 2000). Central leptin resistance is a feature of the metabolic syndrome. Other adipokines such as adiponectin also play a role by affecting the sensitivity of the tissues to insulin.

Short-term mechanisms acting on an hour-to-hour basis involve insulin, which itself also acts centrally to decrease food intake, and a variety of gut-related peptides which signal hunger or repletion following a meal. The latter are affected by therapeutic techniques to reduce food consumption, such as gastric bypass surgery, which not only reduces stomach size but also affects the release from the stomach of appetite-controlling hormones such as ghrelin. There are also interactions with emotional states. For example, ghrelin signaling mediates reward-eating behaviors induced by chronic stress/depression (Chuang et al. 2011). In phylogenetic terms this is perhaps not surprising as predation-related fear and hunger would often coexist in nutritionally limiting circumstances.

Because hunger and thirst are critical survival mechanisms, it has been suggested that the default position is to favor excess food intake over energy expenditure, and that this is why humans, and indeed other animals in captivity, are so prone to obesity. This argument is perhaps supported by the large number of polymorphisms that have been associated with morbid adiposity, and that generally involve disturbance of appetite control (Box 9.4).

Glucose is the primary metabolic fuel for most tissues, so homeostatic mechanisms have evolved to keep its concentration in the blood reasonably constant despite variations in nutritional supply caused, for example, by overnight fasting. Brain

---

**Box 9.4 Candidate Genes Associated with Severe Obesity**

A number of SNPs have been associated with obesity (see also Section 3.6). In general, they have been identified by studying familial severe obesity. A broader number of genes have been implicated in obesity by GWAS, although their overall contribution to explaining the variance in body fat found in Western populations is less than 10% (Speliotes et al. 2010), indicating that non-genetic factors are more important and/or that current methods of GWAS that deal poorly with epistatic mechanisms are inadequate. Some of the mutations associated with severe obesity syndromes include FTO and MC4R.

**FTO**

Haplotypes in the fat mass and obesity-associated gene, FTO, are associated with increased risk of type 2 diabetes and obesity. Homozygosity for the risk allele (SNP rs9939609) increases the risk of obesity by 67%, equating to an extra 3 kg of body weight (Frayling et al. 2007), and individuals with these haplotypes tend to have greater appetite and a preference for energy-dense food even before they become overweight. Polymorphisms in the first intron of FTO have been associated with obesity-related traits in European, Asian, and African populations. There is marked variation in frequency of the risk allele, which is present in 45% of Europeans but in only 14% of East Asians.

**MC4R**

Mutations in the melanocortin receptor 4 gene, MC4R, are one of the most frequently identified causes of monogenic obesity in humans, explaining up to 4% of early onset and morbid childhood obesity. Heterozygosity is sufficient to confer a substantial risk of obesity, although homozygous carriers present with a more severe phenotype. Carriers are also characterized by severe hyperinsulinemia and binge-eating activities. MC4R encodes a transmembrane receptor with a central role in energy homeostasis and somatic growth, probably mediated by the sympathetic nervous system. The greater the impact of the mutation on the signal transduction function of MC4R, the more severe the degree of obesity.
function is adversely affected if blood glucose drops too low, yet excessive levels of blood glucose are toxic, primarily by glycation of many constitutive proteins and glycoproteins, affecting their function. Insulin efficiently promotes the uptake of glucose by tissues after a meal. Some of this glucose is metabolized, some is converted into glycogen, which acts as a short-term energy store in muscle and liver, and some is converted by the liver into triglycerides for longer-term storage as fat in adipose tissue. Conversely, as circulating glucose is depleted in the fasting state, other hormones (e.g., glucagon, growth hormone, glucocorticoids, and catecholamines) mobilize the energy stores for conversion into glucose.

For any animal faced with day-to-day or season-to-season variability in food intake the ability to store energy as fat, and to mobilize that energy when required, is a capacity fundamental to its survival. This required the evolution of mechanisms to regulate and promote tissue uptake of glucose from the circulation, as well as maintain readily available storage depots for long-term backup energy reserves in the form of adipose tissue.

9.3 Human Diet: An Evolutionary History

9.3.1 Pre-agricultural Hominins

The earliest members of genus Homo appeared about 2 million years ago (see Chapter 6). The first archeological evidence for agriculture, in contrast, dates from a mere 12,000 years ago in the Middle East. Given that humans and our hominin ancestors survived as hunters and foragers for 99% of our existence, it can be argued that selection has driven our biology and metabolism to be better matched to the physical activity and diet that characterized the foraging way of life. From this perspective, the current global epidemic of metabolic disease can be understood, in part, as a result of a mismatch between our “ancient” foraging-adapted genome and our rapidly changing modern diet and lifestyle. Thus, one way to gain insights into contemporary patterns of disease is to understand how our ancestors lived, and how modern culture might conflict with our biological adaptations to our previous lifestyle (Gluckman and Hanson 2006b). This is no simple task, and the answers to these questions are not straightforward. Indeed despite the hype of marketeers, there is no single “paleo diet,” and in fact what is marketed as a paleo diet bears little relationship to the probable realities of our ancestors’ lives (Zuk 2013).

9.3.1.1 Anatomical Evidence for Diet Quality in Early Humans

Clues about the evolution of the human diet come from our anatomy and digestive biology, which differ from that of our closest living relatives, the great apes, in important ways. Most great apes have a gut biology that is well suited to foods that require more extensive processing and digestion, such as leaves and roots. In contrast, gut size is greatly reduced in modern humans, showing that at some point since the split between us and the great ape lineage about 7 Mya the need to maintain a large gut was reduced. The evolution of a shorter gut would have accompanied the adoption of a more digestible diet, which was calorifically dense and with less cellulose bulk.

One likely possibility is that this reduction in gut size accompanied an increased reliance on eating meat (Stanford and Bunn 2001). Mammalian bones found in association with early H. habilis sites, such as Olduvai Gorge, have cut marks and other evidence of butchery. Early Homo populations would probably have acquired meat by scavenging remains left by savannah carnivores. Assessment of modern carnivore kill sites shows that carnivores routinely leave the brain, bone marrow, and other scraps of meat. These would have provided a rich source of dietary fat, including essential fatty acids such as docosahexaenoic acid that are required in abundance in humans to support early brain growth. Human requirements for the longer-chain (C<sub>20</sub> and C<sub>22</sub>) fatty acids are similar to those of carnivores, reflecting low metabolic capacity for elongating and desaturating shorter plant-derived (mostly C<sub>18</sub>) fatty acids and implying a significant contribution of animal-derived lipids to the early hominin diet. Evidence for reliance upon game, whether hunted or scavenged, increases with the appearance of H. erectus, and becomes quite prominent with
the emergence of fully modern humans around 150,000 years ago. The increase in dietary quality that came with eating meat probably relaxed selection for maintaining a larger gut, leading to the reduction in gut size (Section 6.3.5).

### 9.3.1.2 Modern Foraging Populations: What do They Teach Us?

One important source of information on ancestral human diets comes from modern or historical human populations who also subsist as foragers (hunter-gatherers) (Jordan 2014). Before considering what we can learn from such modern populations, it is important to bear several points in mind. It would be a mistake to view such populations as carry-overs from an earlier time in human evolution. There are probably no populations today that have not been influenced, either directly or indirectly, by the flow of ideas, technologies, and cultural practices from other societies. The idea of a “lost tribe” roaming the backwoods untouched by the global economic system, or the earlier institutions of colonialism, is little more than a myth. All modern foragers have been shaped by the passage of trade goods, infectious disease, cultural practices, and in myriad other ways that are not always obvious.

Although the environments that modern foragers inhabit are remarkably varied, what they share in common is that they tend not to be suitable for more intensive resource-extraction methods such as farming. The most productive environments, which today are the seat of the densest human settlements supported by intensive agriculture, would have been the preferred habitats of our distant foraging ancestors; yet the environments of modern forager societies have been markedly changed. Because of this, we have no record of what foraging diets might have looked like in these once-common settings. In many cases modern foragers live in marginal environments to which they have been displaced by colonizing populations. Indeed, some populations that forage today were not foragers in the past: some are the modern descendants of agriculturalists who were pushed, often by force, into carving out a new living in a marginal environment. For all these reasons and more, we should not view modern foragers as evolutionary relics.

However, despite these caveats, modern foragers do provide important insights into the types of diets that our distant ancestors probably consumed. A quick thought experiment illustrates why. Imagine that you and 20 of your closest friends and family were transported to a region far from roads, communication, or human settlement. If your group survived, it would necessarily do so by subsisting on dietary resources that occur naturally in the local environment, such as plants, roots, small prey, berries, and fruits. Through trial and error, you would gradually build up a sophisticated repertoire of knowledge of which resources taste good, which are not digestible, and which are dangerous. You would refine methods for trapping or hunting prey species, and you would develop ways of processing and cooking the meat harvested from them. Within several generations, your group would develop a local knowledge, culture, and technology of food extraction, processing, and preparation that was sustainable within the confines of what was available in your local habitat.

In any given ecology or habitat, there are finite ways to subsist by hunting and gathering. For instance, there are likely to be only a handful of important potential sources of calories, a handful of easily captured game species, and so on. As a result, different groups left to survive in the same environment would probably converge on similar strategies and diets, and perhaps even develop similar migratory patterns to follow seasonal gradients of resource availability between food patches. While your hypothetical extended group of friends would clearly not be an evolutionary link with the past, in all likelihood you would settle into a diet and style of living not all that different from what ancestral hominins must have developed when faced with a similar palette of natural resources. For the purposes of this discussion, modern foraging societies can therefore be viewed as the equivalent of this thought experiment: not as evolutionary relics or stepping stones to industrial society, but as a source of insight into the types of strategies that are sustainable by humans living solely, or primarily, from gathered and hunted food sources and without access to a staple crop.

So what do modern foraging populations eat? One defining feature of forager diets is the sheer
diversity in what different populations in different environments consume—this is an important consideration in addressing the naïve belief that Paleolithic societies had a stereotypic way of life. In most, hunting is an important component of the diet, and often contributes half or more of the total calories consumed (Cordain et al. 2000). Recently, it has also been shown that these populations have very different gut microbiomes from that of industrialized populations, in that they display more complexity and biodiversity (Obregon-Tito et al. 2015). This highlights the interplay between us, our microbiome, and our environment.

However, the proportion of calories derived from meat varies immensely, from 99% among the Arctic-dwelling Inuit to a majority of vegetable calories in groups like the !Kung san (or Khoisan) of Botswana, who have access to a plentiful source of non-animal calories (Box 9.5). No known group of foragers lives solely as vegetarians, and meat constitutes at least 25% of the calories consumed by all modern or historical foraging populations for which detailed information is available.

Studies of foraging populations show that they also acquire a substantial percentage of their calories from vegetables and fruits that may be gathered in wild settings. Typically, these plant sources would provide high levels of vitamins and fiber compared with the diet of modern urban humans, and their carbohydrate content would have a low glycemic index (a measure of the ability to raise blood glucose levels and therefore promote insulin secretion). Another important difference between a foraging diet and a modern “Western” diet is intake of sodium: modern foragers without access to commercially produced salt consume much less sodium than is found in a modern Western diet, and are generally free from arterial hypertension in consequence.

Two more recent subsistence transitions—the Neolithic Revolution and the “nutrition transition”—have radically transformed the human diet and lifestyle. The modern scourge of metabolic disease can be partly understood as a result of these transitions, which have brought our Paleolithic biology into conflict with our modern lifestyle and diet.

9.3.2 The Neolithic Revolution

In a process that began around 12,000 years ago in the Middle East, two fundamental changes occurred in the human way of life: first, in some regions the mobility of hunter-gatherer bands was gradually replaced by settled village life; and second, humans began to domesticate and exploit for food the plant and animal species that had become associated with their settlements.

At the time of the Neolithic Revolution, the human population of the planet was estimated to be about 5 million. The ice of the Last Glacial Maximum had retreated thousands of years earlier, but it was a time of marked climate change, with the cool and dry period of the Younger Dryas giving

---

**Box 9.5 Are we Primarily a Vegetarian or Meat-eating Species?**

There are many food fads in Western society that often invoke trite and generally uninformed arguments about our evolutionary past. The role of hunting in human evolution is a long-standing topic of debate. By the 1950s, the human fossil record was revealing evidence for hunting among early hominins, and the notion of “man the hunter” began to capture the imagination of anthropologists and the public alike.

This was later challenged by one of the first detailed ethnographic studies of a foraging population. Starting in the late 1960s and carrying on into the 1970s, anthropologists conducted a detailed study of the ecology, diet, and demography of the !Kung san of Botswana, who managed to carve out a living in what would seem to be a very challenging environment in the scrublands of rural Botswana. What they found in their studies surprised them. The !Kung spent around 3 h each day in food procurement and processing, leaving much of the day for leisure, crafts, child care, naps, and relaxation. What’s more, the assumption that hunting was their dietary mainstay was brought into question by the finding that they relied to a great extent on mongongo nuts as a source of calories, which they harvested in large quantities. Men did hunt, which supplemented this
nut-based diet, but this contributed a relatively modest proportion of the !Kung’s daily calories (Truswell 1977).

The !Kung project remains one of the most detailed studies of a foraging population ever conducted, and at the time it was the first in-depth study of forager ecology. It is no wonder then that the !Kung became something of a living icon of our foraging hominin past, and they were viewed by many as a model for how our distant ancestors must have lived.

The problem is that the !Kung, though fascinating, are by many measures not typical human foragers. More recent assessments of the subsistence of modern and historical foraging populations reveal that the !Kung are outliers. Most foragers do, it turns out, rely to a greater extent on hunting for their calories. Whereas meat constitutes around 33% of the calories consumed by the !Kung, the other foraging populations for which data are available places the estimate closer to 75% (Figure 9.2).

Although debates remain regarding the importance of meat eating in human evolution, there is little question that hunting was a component of human subsistence for much or all of modern human evolution. What is intriguing is that chronic degenerative diseases were traditionally not present in foraging societies, even among those, like the Arctic Inuit, who procured 99% of their calories from animal meat and fat (Hegele 2001). Some of the health problems associated with meat eating in modern industrialized societies can be traced not to the quantity but the composition of the meat that we now consume. These qualities, which include a higher fat content and a changed fatty acid and cholesterol composition, among others, are not intrinsic to meat but have resulted from alterations in meat composition from historically recent domestication and industrial herding practices.

![Figure 9.2](image_url)

**Figure 9.2** Foraging populations vary in their diets: the contribution of plant and animal calories to the diets of modern or historical foraging populations. Note that two studies of the !Kung are shown. Plotted from data in Kaplan et al. (2000).

way to the warming of the early Holocene a little more than 10,000 years ago. Most explanations for the inception of agriculture include some component of resource or population pressure, whether caused by the extinction by hunting of the large animal populations that survived the Last Glacial Maximum or by the growing human population that had already colonized the most easily exploited regions. Whether sedentism preceded or was followed by agriculture may have been dependent
on the region in question (Rocek 1998). Whatever the cause of sedentism—perhaps protection from other foraging groups, or clustering at a favored site (oasis) in a drier climate—it would by itself lead to population pressure as fecundity increased from the low levels typical of nomadic groups. Nomads are often forced to limit the number of dependent children by ensuring that one child can walk before a subsequent one is born (and can be carried), resulting in inter-birth intervals of children that survive of up to 4 or 5 years.

The early sedentary groups would have exploited a wide range of animal and plant species, and it is easy to imagine how familiarity with, and utilization of, wild cereals would have extended to the deliberate collection and planting of their seeds, and how later association with animals such as sheep and goats led to their domestication. By about 11,500 years ago in the Middle East the crop range, or package, of cereals (barley and wheat) and pulses (lentils and peas) cultivated by settled farmers was well established, and by 9000 years ago the major animal species (sheep, goats, cattle, and pigs) had been added. Export of this package or its components to Europe, North Africa, and western Asia, whether by movement of individual farmers or by acculturation, followed. Farming later arose independently elsewhere in the world, with distinct crop packages, in East Asia (predominantly rice), Central America (maize), South America (potatoes), West Africa (millet and sorghum), and New Guinea (yams), and biogeographical patterns of early farming may have determined later patterns of economic development. Whatever its original stimulus, the increase in population that accompanied agriculture ensured that the process was irreversible. By the time of the Roman Empire 9000 or 10,000 years later, the global population was about 130 million.

Was farming a positive development in terms of its effects on a population’s food stability and health? It brings population growth and, importantly, the opportunity for a society to develop specialists—potters, metalworkers, soldiers, philosophers, and so on—which has been so important for the constitution of settlements and modern society. But it also brings less welcome effects: the beginning of a rigid social hierarchy with its attendant stressors, the opportunity for human pathogens and parasites to flourish as their hosts move to crowded communities (Chapter 10), the nutritional consequences of reliance on a restricted range of food sources, and the possibility of famine as crops fail because of climate change or infestation (Cohen 1989). There is evidence from skeletal and dental remains that child growth, adult stature, rates of infection and anemia, and lifespan were all adversely affected by the consequences of the shift from foraging to farming. For example, skeletal remains provide evidence of an increased prevalence of iron-deficiency anemia and infection in sedentary farmers (Cohen and Armelagos 2013).

Famine as a result of crop failure, and often exacerbated by political disputes or conflict, is a recurring theme in recent human history, from the famines of ancient Egypt when the Nile failed to flood, to the Irish famine which started with disease of the potato crop in 1845, to the present crises in the Horn of Africa. The seasonality of agricultural production causes a pre-harvest “hungry season,” even among modern subsistence farmers in West Africa. Conversely, as we will discuss later, it is far from certain that famine was a significant factor for human existence prior to the Neolithic Revolution.

9.3.3 The Modern Nutrition Transition

We see then that the advent of agriculture should not be equated with dietary security or abundance; in fact, it was often associated with the opposite. Agricultural populations were often more prone to shortfalls in both energy and specific micronutrients than were their foraging ancestors. The nutritional abundance that is now the hallmark of many modern agricultural societies, and which many of us take for granted, is in fact a far more recent—and in many regions, ongoing—development. It is a product of a set of changes in diet and lifestyle that have been described as the nutrition transition (Popkin 2001). In contrast food security remains a critical issue for many lower-income countries with rapidly growing populations.

There are many factors that contribute to the nutrition transition. One relates to the cheaper production of refined carbohydrates and fats, in
particular fats of plant origin. Government subsidies for certain crops such as corn, introduced for political reasons, further incentivize farmers and food producers. Greater trans-national trade, the development of intensive industrial farming practices that reduce the cost of crop production, and the burgeoning market of populations moving from rural to urban environments all make a contribution. Economic considerations show how the processing and ultra-processing of foods increase the value (and thus the retail price) of foods as well as their preservation. Additional profits lead food companies to focus increasingly on such production rather than on the provision of minimally processed, usually less energy dense and healthier, foods. Advertising and communication also play a part in making these products desirable: for evidence we need only look to the success of fast food chains and of the marketing of sweetened beverages. But we must also recognize that these trends occurred against the backdrop of human societies who were willing to make the transition. It is unfortunate, for example, that the traditional scarcity of fats for cooking in poor communities has encouraged overconsumption of these items as their cost has declined and their availability increased. In many such populations, as income rises so does the intake of additional high-fat foods and sweetened beverages.

China provides a good example of this transition. During the 1970s, in the period after the Cultural Revolution, food insecurity was a major problem. There was no television, limited bus and other mass transportation, and little food trade with other countries. Very little processed food existed, and most rural and urban occupations were very labor-intensive: oxen were used for plowing and factories still required huge amounts of human labor to move stock and equipment about. However, work and life in China have changed dramatically in recent decades. Tractors are now available on many farms and fork-lift trucks in factories; the internet has arrived in offices along with printers, fax machines, and modern telephone systems. By 2000, soft drinks and processed foods were consumed everywhere and nearly 90% of all homes had a television. Hong Kong- and Western-based advertising have been on the increase and may be received on many televisions as well as seen on billboards and in magazines. Use of the bicycle has declined and public transport and cars increased. In the course of one generation, the lives of millions of Chinese people were transformed from a subsistence agriculture-based economy to a modern, industrialized one. Data on the incidence of obesity in Chinese children demonstrates the consequences of this. The overall prevalence of overweight and obesity in childhood increased from 2.2% in 1981–5 to 20.6% in 2006–10 (Yu et al. 2012). The increase largely occurred in urban areas, in which children were nearly twice as likely to be obese compared with their rural counterparts. Similar changes are also seen in other low- to middle-income countries, for example in Oceania, Latin America, elsewhere in Asia, and parts of the Middle East (Ng et al. 2014). Refugees or migrants who were born in conditions of poor nutrition but who then moved to nutritionally abundant environments are likewise at greater risk of diabetes, obesity, and cardiovascular disease (Box 9.6).

9.3.4 Well Fed but Poorly Nourished

The foraging populations we describe were generally considered to be free from chronic diseases like diabetes and obesity until the adoption of more Westernized dietary and lifestyle practices in recent generations. In fact, the relatively high meat consumption of foragers may cause some readers to pause for thought. The amount of animal products in the diet of foragers appears to far exceed the intake in most Western industrialized nations. In the USA, where typical per-capita food consumption is of the order of 2700 kcal/day, 15% of energy is from meat, 10% from dairy, and the remaining 75% from flour, sweeteners, oils, fruits, vegetables, and other items. Americans appear to consume far fewer calories from meat than most foragers. If excessive meat consumption is indeed bad for our health, how then do we explain the rise of chronic metabolic disease in societies like the USA and the foragers’ comparatively disease-free existence?

One important point is that not all meat is alike. Wild meat of the sort that foragers consume is lean (protein dense) and typically provides about half as much of its energy from fat as does meat from
grain-fed domestic cattle (Figure 9.3). Through time, human herding cultures have selectively bred the animals that produce the fattiest, tastiest meat. Then in the post-war period, as meat has been marketed to a consumer society, the trend towards taste and textural qualities has rapidly grown. This has maximized factors that improve flavor, such as marbling, which involves the deposition of fat droplets (triglycerides) within and between the muscle cells. In addition, industrial herding practices in places like the USA often involve feeding animals grain, such as corn, which is grown in abundance with generous federal subsidies, rather than the grasses that are their natural food resource. Compared with grass-fed cattle, grain-fed beef is higher in fat and has more saturated and monounsaturated fats, and is also very low in beneficial polyunsaturated fats (Figure 9.4).

As a result of these changes, what many of us call meat today is very different from the meat that has probably been an important component of the hominin diet for several million years.

*Figure 9.3* The high fat content of supermarket beef: fat as a percentage of dietary calories from muscle in various domestic and wild game species. Plotted from data in Cordain et al. (2002b).

---

**Box 9.6 Migration and Health**

The Jewish population of Ethiopia, the Falasha (or properly the Beta Israel), claim to trace their origin as far back as the time of Moses. In the last quarter of the twentieth century, the Israeli government, increasingly alarmed at the growing political and economic instability of Ethiopia and its possible effects on the Jewish people there, began to encourage and facilitate their emigration to Israel. The most notable event of the emigration was the mass airlift of Operation Solomon in 1991, during which over 14,000 people, 1122 of them on a single flight, were transported covertly to Israel within 36 h.

In Ethiopia, the Beta Israel had been subsistence farmers and many were suffering the effects of famine, but overnight they found themselves in a high-income country with access to refined foods with a high energy density. Inevitably, with the speed of this nutrition transition, many of them have developed metabolic dysfunction. Among a sample of young adult immigrants tested some 4 years after their arrival in Israel, nearly 20% had developed either frank diabetes or impaired glucose tolerance. Other studies have found increased rates of deposition of abdominal fat and atherogenic blood lipid profiles (Jaffe et al. 2001).
These changes have consequences for our health. Not only are we eating more calories as a result of the greater energy density of domesticated meats, but we now consume higher levels of the saturated fatty acids that elevate cholesterol, which is implicated in cardiovascular disease. In addition to the well-known deleterious effects of an excessive intake of saturated fatty acids, recent work is revealing with higher resolution the importance of specific fatty acids, and especially the ratios of fatty acids, as key to conditions such as atherosclerosis. The omega-3 fatty acids are particularly important. These are not produced within the body to a great extent and thus are essential. Omega-3 fatty acids have been shown to have a wide range of effects such as reducing clotting and inflammation. Omega-6 fatty acids, which are also essential, can tip the balance in the other direction, encouraging inflammation. It appears that the effects of these fatty acids on health depend critically on their balance within the body, and that inappropriately high omega-6/omega-3 ratios in the diet can be harmful. It is notable, then, that grain-fed beef is significantly lower in omega-3 fatty acids than grass-fed beef (Daley et al. 2010). The ratio of omega-6 to omega-3 fatty acids has increased substantially in human breast milk over the past 20 years, demonstrating the subtleties of how influences from these changing food behaviors could affect our biology.

The public health push to reduce consumption of meat and meat products in many countries has encouraged consumption of vegetable oils. Vegetable oils in general tend to be higher in monounsaturated and polyunsaturated fats, which have more favorable effects on risk factors such as cholesterol profiles when compared with the saturated fats present in meat or poultry. But some commonly used vegetable oils have fatty acid profiles and other properties that are also not beneficial. For instance, soybean and corn oils, which are mass produced and widely utilized in processed foods, have very high omega-6/omega-3 ratios. Palm and coconut oil, the most widely used vegetable oils, have very high levels of saturated fats.

Our diets have changed in other ways that also influence our health and risk of metabolic disease. Modern foragers acquire most of their carbohydrates from fruits and starchy tubers. Western societies increasingly rely on refined sugars and sweeteners, which differ from naturally occurring

---

**Figure 9.4** The poor fat quality of supermarket beef: fatty acid composition of muscle from wild game and domestic cattle. Plotted from data in Cordain et al. (2002b).
sugars and starches in that the body metabolizes them more rapidly, giving them a high glycemic index with consequent effects on homeostatic processes such as insulin release. The process of refining involves extracting the sugars from the surrounding food matrix, which is often the source of important micronutrients and of fiber which also reduces the glycemic index.

Even as our diets increasingly focus on refined energy-dense foods, our level of physical activity has decreased in recent decades in response to several important global economic and social trends. As nomadic foragers, our ancestors spent much of their time on the move, and our modern life increasingly deviates from this in fundamental ways. Physical activity can be broken down into that engaged in at work, during leisure time, and in locomotion, and all have seen dramatic changes in the past few generations. The percentage of individuals who engage in physical activity at work has greatly declined with the rising importance of the service sector and the mechanization of manufacturing processes. Leisure time is also increasingly sedentary, focusing on activities like watching television or playing video games rather than physically active pursuits. With the advent and wide distribution of the internal combustion engine, we now rely to a great degree upon the hydrocarbons in fossil fuels, rather than our own dietary energy, to move from place to place. These trends, the so-called nutrition transition, are sweeping the globe.

Although there clearly is no single Paleolithic-type diet or pattern of activities that will keep us healthy, we see that studies of our evolutionary past and of modern foragers do give us important insights into the aspects of our lifestyle and diet that have changed the most, and that consequently are most likely to be involved with the modern epidemic of metabolic disease. The Western diet and lifestyle, with its chronic imbalance of intake and expenditure, appears to have exceeded our homeostatic capacity to cope metabolically, helping drive the global pattern of weight gain and its associated problems.

The modern epidemic of metabolic disease is not merely a function of calories, but also of the composition of what we eat. Modern dietary change has led to other, no less important, imbalances and sources of novelty, such as our increasing consumption of non-protein calories like processed sugars and alcohol, and foods that are not only high in fat but also have an imbalanced fatty acid composition. All of these changes have been rapid, leading to unprecedented global changes in human metabolism and health.

It is also important to remember that even in high-income countries many members of the population, especially those of low socioeconomic status and living in parts of urban environments sometimes called “food deserts,” have a diet deficient in micronutrients and high-quality protein, and have food insecurity.

9.4 How Can Change in the Environment Increase Disease Risk?

These lifestyle transitions have been implicated as causal in the changing patterns of obesity and diseases such as type 2 diabetes and cardiovascular disease. The evolutionary question is, why has this change in environment increased the risk of disease? Many of us are now living in environments that are novel in an evolutionary sense, and the underlying consensus is that the disease syndromes are the outcome of a mismatch between the human capacity to maintain metabolic homeostasis and the modern energetic and nutritional environments.

Two sets of possibilities exist to explain how this mismatch may arise, and they are not mutually exclusive (Low and Gluckman 2016). First, because the contemporary environment is novel in evolutionary terms, there has never been selection to match our genotype to it; rather it was primarily selected and adapted for the range of Paleolithic environments. But there are additional questions: was the genotype selected because it allowed the individual to survive bottlenecks produced by famine and now has deleterious consequences manifest in obesity and its complications, or was it that the rate of change in the environment outstripped the capacity of selection to match the organism’s physiology to its environment in a more general sense, and obesity and metabolic disease are the outcomes of this mismatch?

The second possibility is that direct selection on these traits of metabolic regulation may not be
the only way in which an organism can evolve to be adapted to its environment. Selection may also act on the processes of developmental plasticity and use these mechanisms to adjust its phenotypic development to its actual or anticipated environment (Section 4.5.2). Inappropriate developmental responses in response to developmental cues could provide an alternative mechanism by which mismatch between the organism and its environment can occur. There is evidence to suggest that each of these possibilities—evolutionary mismatch and developmental mismatch—plays a role.

9.4.1 A “Thrifty” Genotype?

The hypotheses developed to explain the rise of metabolic and related diseases in recent history were among the earliest attempts to use evolutionary principles to understand human health and disease. In 1962 the geneticist James Neel proposed that such diseases arose through patterns of inherited genes that may have been previously advantageous in human evolution but were now disadvantageous. He referred to these genes as “thrifty genes” because they were proposed to be associated with energy-saving characteristics such as insulin resistance in muscle or the propensity to deposit fat (Neel 1962). He reasoned that such genes would have conferred an adaptive advantage in times of famine and perhaps would have conferred little disadvantage in times of plenty, because, as we saw for many other chronic diseases, there would have been little selection operating to maintain health into later life, as in general these diseases do not inhibit reproduction. Neel proposed that these genetically mediated thrifty traits that were useful to our ancestors would produce inappropriate effects, including obesity, in a modern world. He argued that this was particularly manifest in populations in parts of the world where famine had been frequent. Then, when such populations underwent a nutrition transition from a hunter-gatherer lifestyle to a modern lifestyle of chronic dietary abundance they would be at a particular disadvantage metabolically. The “thrifty genotype hypothesis” led to an early focus on predominantly genetic explanations for the high prevalence of metabolic disorders in some populations.

What properties would we expect of a thrifty gene in the sense originally proposed by Neel? First, we would expect that the gene would demonstrate (and indeed would probably have been identified by) significant linkage to some trait which could represent thriftiness: perhaps BMI or glucose tolerance. Second, we might expect plausibility of gene function, i.e., that it would code for a protein or regulatory RNA associated with some aspect of control of metabolic partitioning (e.g., insulin secretion or action, a key metabolic regulatory enzyme, or some facet of adipocyte biology). Third, and ideally, we might expect population genetic studies to show some signal of selection on the gene, perhaps for a “risk” allele in populations particularly exposed to cycles of feast and famine in historical times, or even for a “protective” modern allele where the ancestral allele is now detrimental.

Clearly, susceptibility to obesity and type 2 diabetes varies between individuals in the population, and there is evidence for heritability of these traits. Over 800 candidate genes associated with obesity or with type 2 diabetes have been identified (Dai et al. 2013). However, of these only a small number have been confirmed by replicate studies in different populations with large enough samples to give assurance that the associations are indeed real. Moreover, even the strongest of these associations typically explains only a very small amount of the variation in susceptibility to obesity or diabetes in a population (of the order of 1–3%). Moreover, the mere demonstration that variation in obesity is related to genes does not prove that these genes were selected in the human gene pool due to advantages during famine, as Neel originally proposed. Thus, the thrifty gene hypothesis remains just that, a hypothesis without any direct support.

In fact, putting Neel’s ideas under greater scrutiny gives one reason to pause when considering the role of famine as a source of selection on human metabolism. Contrary to Neel’s central assumption, there is little anthropological evidence that famine is a regular event for modern hunter-gatherers. The foraging lifestyle is marked by its flexibility and by the wide range of resources that are exploited. This provides the forager with far greater opportunities to adapt to changing ecological conditions, such as drought. As discussed earlier, there is evidence that
the rise of agriculture ushered in poorer health and periodic nutritional stress, which is expected given the greater reliance upon the more precarious, narrower selection of crops in such populations.

If hunter-gatherers were not particularly prone to famine, then could we adapt Neel’s hypothesis to propose that the more widespread and severe episodes of famine that have occurred since the introduction of monocrop agriculture have been a major driver of selection for thrifty metabolism? We know from studies of lactase persistence (Box 1.1) and malaria resistance genes (Section 13.9.1) that selection over such a timescale can leave physiologically important signatures in the human genome. Indeed, it has been suggested that selection in response to major famine events caused by failure of the Indian monsoon is the cause of the tendency of individuals of South Asian origin to deposit more of their energy in metabolically labile visceral fat (Wells 2007).

Others have questioned on a number of grounds whether the selective advantage of a putative thrifty gene would be sufficient to cause it to spread widely. It is argued that the frequency and severity of famine have been insufficient, that relatively few famine events have occurred in most populations, that deaths during famine are predominantly caused by factors other than starvation, and that mortality in famines disproportionately affects the young and old rather than individuals of reproductive age (Speakman 2006). In addition, obesity is rare in well-nourished individuals from present-day hunter-gatherer and subsistence agricultural societies, who would have been expected to inherit such thrifty genes. If foraging populations do not put down excess calories as fat during times of plenty, how could the putative thrifty alleles underlying human metabolism have increased survival during periods of famine?

Additional doubt is cast upon the assumed role of famine when we consider the development of body fat. In humans, fat makes up a larger percentage of weight at birth than in any other mammal (Box 9.7, Figure 9.5). This is followed by a continued period of rapid fat deposition during the early postnatal months. In well-nourished populations, adiposity reaches peak levels during the first year of life.
are a source of nutritional stress, and indeed it is primarily through their effects on nutritional status that they compromise health and contribute to mortality during infancy and childhood. Once sick, a child loses appetite and this is often compounded by the withholding of food by caretakers. The common diarrheal diseases reduce nutrient absorption and digestion, while the fevers associated with many viral infections can increase metabolic rate and thus energy expenditure. While the specific symptoms vary by illness, the ensuing nutritional depletion has the effect of suppressing immune function, leaving the infant more prone to future infection and a compounding cycle of nutritional stress.

Human infants thus face a profound energetic dilemma: at precisely the age when they are most dependent upon caretaker provisioning to maintain the high and obligatory energy needs of their large brains, they are most likely to be cut off from that supply chain as a result of illness and the nutritional stresses of weaning. It is this confluence of factors, and the synergy between nutritional stress and compromised immunity, that accounts for much of the high infant mortality in many societies. In light of these risks, natural selection probably favored neonatal adiposity as a strategy to prepare for this difficult period. It is not difficult to imagine how infants who deposit copious quantities of energy as fat prior to weaning would be better represented among the subset who survive to adulthood to reproduce and pass on their genes. It is also important to note that these sources of energy stress have largely receded by mid-childhood: children have already acquired antibodies against the major pathogens that they are likely to confront. As a result, infections and periods of negative energy balance decline to a small fraction of their high prevalence during the post-weaning period. Because older children are far less likely to have to rely upon energy reserves for survival, it is easy to see why the human body places lower priority on maintaining sizeable body fat stores by this age. This of course would not be the case if our bodies were primarily gearing up to survive famine.

**Figure 9.5** Humans are born fat: percentage of body fat at birth in mammals. Adapted from Kuzawa (1998), with permission.
life before gradually declining in childhood, when humans reach their lowest level of body fat in the life cycle (Davies and Preece 1989; Chapter 5). If the threat of famine is what drove the human tendency to build up fat reserves, it is not obvious why children’s bodies should do so little to prepare for these difficult periods. The lower priority placed upon maintaining an energy reserve by middle childhood suggests that the background risk of starvation faced by our ancestors—“famine”—was small in comparison with the nutritional stress during the preceding developmental period of infancy. Indeed, infancy is marked by often intensive nutritional stress associated with weaning and the related problem of infectious disease, making nutritional disruption and nutrition-related mortality common at this age (Kuzawa et al. 2007; Box 9.7). Because all individuals who successfully pass on their genes to offspring must have survived this early-life nutritional bottleneck, there is likely to have been selection for building up protective fat reserves at this age in large-brained humans.

In summary, while the thrifty gene hypothesis was valuable in developing evolutionary concepts related to metabolic disease, the hypothesis itself has largely been discarded.

9.4.2 Does Evolutionary Novelty Explain Current Patterns of Metabolic Disease and Obesity?

The preceding sections provide the background for considering more recent concepts of how our evolutionary history helps explain the changing patterns of obesity and metabolic disease, in particular the recent increases in these conditions associated with the rapid nutrition and socioeconomic transition in Western countries over the last 100 years, and the transition that is now occurring in low- and middle-income countries. To recapitulate, our ancestors evolved in the absence of agriculture and a stable source of carbohydrates. Our species is characterized by being able to cope with an omnivorous diet and with a metabolism and behavior evolved on the basis of regular access to food. Selection acted to match our physiology to these characteristics and to the diet of the Paleolithic foragers from whom we are descended. Although, as discussed earlier, the concept of a typical hunter-gatherer is misleading, it is generally accepted that we were selected within and adapted to environments characterized by a relatively high protein intake and low intake of sugars and fat. Frank obesity would have been a remote possibility. It has been estimated that hunter-gatherers may expend up to 2500 kcal each day gathering food. Since the Upper Neolithic, the human diet has changed dramatically, at first in association with the development of agriculture and then, since the industrial and technological revolutions, in association with the development of various forms of highly refined foods. In parallel the physical work expended to “earn” this food is declining, again at an accelerating rate.

These two changes have occurred against a background of a marked increase in life expectancy across virtually all populations as public health measures take effect—more dramatically in high-income countries, but increasingly also in low- and middle-income countries. It is important to note that because peak reproduction occurs well before middle age, health and reproductive fitness are not identical. There will not be great selection pressure against health consequences occurring in middle age if reproductive competence at an earlier age has not been affected, although the effects on kin fitness conferred by the presence of older individuals in social groups may provide some fitness advantage for longevity. The increasing incidence of metabolic disorders among younger people refutes one possible reason for the epidemic—that it is the direct result of a longer lifespan—but does raise the question of how such early onset disease will influence the reproductive fitness of affected individuals.

The simplest evolutionary explanation of this epidemic is therefore that our species is facing a nutritional environment that is entirely novel in evolutionary history and for which our metabolism has not been selected. Instead, our metabolic repertoire of genes is based on that which would have been best adapted for the Paleolithic. For instance, a metabolism selected for a high-protein, low-fat, and low-carbohydrate diet might result in obesity and metabolic disease when confronted with a low-protein, high-fat, and high-sugar diet. This is the simplest form of evolutionary mismatch in which
The problems facing indigenous peoples living in colonized and developed economies—Native Americans, Indigenous Australians, and the Māori of New Zealand—include socioeconomic disadvantage and a disproportionately high rate of obesity and type 2 diabetes. Yet the people of the island of Nauru in the South Pacific, who a few years ago enjoyed one of the highest per-capita incomes in the world as a result of phosphate mining on their tiny island, now have the unfortunate distinction of being the world’s most metabolically dysfunctional nation, with 90% of adults being overweight and more than 40% of adults aged 55–64 having type 2 diabetes (Khammala et al. 2011). What links these two apparently disparate situations, and what might the answer tell us about the causes of obesity and diabetes worldwide?

One population of indigenous people that has been closely studied with respect to their high prevalence of diabetes is the Pima Indians of Arizona in the southwestern USA. About half of Pima Indian adults have type 2 diabetes, a rate that has increased by two- to four-fold over the past 30 years. Such a high prevalence of disease in a population of relatively homogeneous genetic background provides an ideal opportunity to identify susceptibility genes, and any such genes identified might be candidates for putative thrifty genes. Indeed, numerous genetic studies of the etiology of type 2 diabetes in the Pima have been performed—without any major success in identifying chromosomal regions with strong linkage to the trait. More tellingly, the genetically similar Pima population living in Mexico has a five-fold lower prevalence of diabetes, similar to that of neighboring non-Indian populations (Schulz et al. 2006). That observation refocuses attention on environmental and lifestyle reasons for the difference between these populations, and indeed the US Pima population are much more obese than the Mexican Pima group: the latter remain as subsistence farmers with high levels of physical activity whereas the US Pima population is characterized by a low level of physical activity in conjunction with a Westernized “supermarket” diet.

So what is common to this epidemic of diabetes among indigenous peoples living in developed economies? The pattern we see is that these populations have undergone—voluntarily or involuntarily—very rapid change in their lifestyle, from one of subsistence farming or foraging to one where highly energy-dense foods are available cheaply and easily. Such rapid change creates a mismatch between anticipated and actual nutritional environments and exacerbates the risk of metabolic disease. Moreover, such predisposition to obesity and diabetes may be transmitted across generations by exposure in utero to maternal hyperglycemia, which leads to high birthweight (and a greater number of fat cells) and subsequent increased susceptibility to adult obesity in the offspring (Osgood et al. 2011). The effect will be compounded by low socioeconomic status, which among other negative factors such as stress will increase the attractiveness of cheap foodstuffs with high energy content but low nutritional value, to create an intergenerational “metabolic ghetto” from which it is difficult to escape.

What seems clear is that the environmental setting is critical, regardless of genetic background. Pima Indians who have maintained their traditional way of life do not have high rates of type 2 diabetes, but their contemporaries consuming a different diet and engaging in reduced levels of physical activity frequently develop the disease (Box 9.8). And yet, as Figure 9.6 shows, the risk of developing diabetes can vary markedly across populations: this also requires an explanation. In some cases this can be traced to genetic differences in local populations that have been confronted with distinct nutritional histories, selecting for unique mutations which influence their disease risk. At a given BMI, populations from the Indian subcontinent are more likely to develop type 2 diabetes than are Europeans, and it has been proposed that monsoon cycles could have created famine conditions that might have selected for such a trait. In themselves these observations still fit with a simple model of mismatch caused by evolutionary novelty, but where the sensitivity to the novel environment might be genetically different across populations.

**Box 9.8 Vulnerable Populations**

The problems facing indigenous peoples living in colonized and developed economies—Native Americans, Indigenous Australians, and the Māori of New Zealand—include socioeconomic disadvantage and a disproportionately high rate of obesity and type 2 diabetes. Yet the people of the island of Nauru in the South Pacific, who a few years ago enjoyed one of the highest per-capita incomes in the world as a result of phosphate mining on their tiny island, now have the unfortunate distinction of being the world’s most metabolically dysfunctional nation, with 90% of adults being overweight and more than 40% of adults aged 55–64 having type 2 diabetes (Khammala et al. 2011). What links these two apparently disparate situations, and what might the answer tell us about the causes of obesity and diabetes worldwide?

One population of indigenous people that has been closely studied with respect to their high prevalence of diabetes is the Pima Indians of Arizona in the southwestern USA. About half of Pima Indian adults have type 2 diabetes, a rate that has increased by two- to four-fold over the past 30 years. Such a high prevalence of disease in a population of relatively homogeneous genetic background provides an ideal opportunity to identify susceptibility genes, and any such genes identified might be candidates for putative thrifty genes. Indeed, numerous genetic studies of the etiology of type 2 diabetes in the Pima have been performed—without any major success in identifying chromosomal regions with strong linkage to the trait. More tellingly, the genetically similar Pima population living in Mexico has a five-fold lower prevalence of diabetes, similar to that of neighboring non-Indian populations (Schulz et al. 2006). That observation refocuses attention on environmental and lifestyle reasons for the difference between these populations, and indeed the US Pima population are much more obese than the Mexican Pima group: the latter remain as subsistence farmers with high levels of physical activity whereas the US Pima population is characterized by a low level of physical activity in conjunction with a Westernized “supermarket” diet.

So what is common to this epidemic of diabetes among indigenous peoples living in developed economies? The pattern we see is that these populations have undergone—voluntarily or involuntarily—very rapid change in their lifestyle, from one of subsistence farming or foraging to one where highly energy-dense foods are available cheaply and easily. Such rapid change creates a mismatch between anticipated and actual nutritional environments and exacerbates the risk of metabolic disease. Moreover, such predisposition to obesity and diabetes may be transmitted across generations by exposure in utero to maternal hyperglycemia, which leads to high birthweight (and a greater number of fat cells) and subsequent increased susceptibility to adult obesity in the offspring (Osgood et al. 2011). The effect will be compounded by low socioeconomic status, which among other negative factors such as stress will increase the attractiveness of cheap foodstuffs with high energy content but low nutritional value, to create an intergenerational “metabolic ghetto” from which it is difficult to escape.

What seems clear is that the environmental setting is critical, regardless of genetic background. Pima Indians who have maintained their traditional way of life do not have high rates of type 2 diabetes, but their contemporaries consuming a different diet and engaging in reduced levels of physical activity frequently develop the disease (Box 9.8). And yet, as Figure 9.6 shows, the risk of developing diabetes can vary markedly across populations: this also requires an explanation. In some cases this can be traced to genetic differences in local populations that have been confronted with distinct nutritional histories, selecting for unique mutations which influence their disease risk. At a given BMI, populations from the Indian subcontinent are more likely to develop type 2 diabetes than are Europeans, and it has been proposed that monsoon cycles could have created famine conditions that might have selected for such a trait. In themselves these observations still fit with a simple model of mismatch caused by evolutionary novelty, but where the sensitivity to the novel environment might be genetically different across populations.

**9.4.3 Developmental Mismatch—a Contributing Factor**

It has been obvious from Hippocratic times that events in early life may have lifelong consequences. During the past 25 years, much research has emerged to show that early human development is...
responsive to metabolic signals, and that a mismatch between the signals and the actual environment in which the individual later lives could contribute to the rise of metabolic disorders such as diabetes and heart disease (Gluckman et al. 2015).

Population-based studies were the first to suggest that the developmental environment plays a role in risk of later disease. Evidence for such effects came in several forms. The most common was the finding that individuals of lower birthweight tend to have higher rates of later metabolic disorders, including hypertension, diabetes, and cardiovascular diseases (reviewed in Godfrey 2006). These studies are by their nature very long term: how else can we relate what happened to a person when they were a fetus or a growing child to their health as an adult?

One of the most important approaches was to find historical records of birthweight and early growth after birth from a period over 50 years ago, and then to trace the subsequent adults. In the early 1980s, David Barker, an epidemiologist from Southampton, and his colleagues were investigating the rates of mortality from coronary heart disease and other vascular diseases in England and Wales over the previous decade (in fact 1968–78). They noticed that the highest rates of mortality did not occur in the areas of greatest contemporary affluence, in the southeast of Britain for example. Rather, the highest rates occurred in the northwest of England and South Wales, which were areas of high unemployment and poor social conditions, particularly at the time the adults they were studying had been born (Figure 9.7). Furthermore, there was a strong geographical correlation between rates of infant mortality (a marker of poverty, and indirectly of maternal nutrition) in the early 1900s and rates of death from coronary heart disease several decades later (Figure 9.8).

What might account for such an association between early life and late-life mortality? These observations were very remote in time, and likely to be influenced by unmeasured factors, such as differences in smoking, diet, or lifestyle, which were not quantifiable. The next step was to move from gross correlations over integrated populations and across time to find out what happened to a specific population where individuals could be studied throughout their lives. What was needed was a set of records about the growth and development of children in the early part of the century that could be linked specifically to the causes of death in later life.

The largest set of records that were found related to the English county of Hertfordshire. These included weight at birth, weight at 1 year of age, and whether the baby was weaned at 1 year. The ledgers had been maintained from 1911 to 1945. Barker and his colleagues used the National Health Service Central Register at Southport to trace 16,000 men and women born in Hertfordshire between 1911

**Figure 9.6** The relationship between BMI and the risk of developing diabetes varies between populations: for example at a clinically “normal” BMI of 24, South Asians’ risk of diabetes is approximately five-fold that of White populations. Data from Chiu et al. (2011).
Heart disease is more common in socially deprived areas: rates of coronary heart disease in men in England and Wales, 1968–78 (expressed as standardized mortality ratio, SMR). The low mortality in the relatively affluent south and east (darkest blue shading) contrasts with the high mortality in the industrial areas of northwest England and South Wales (darkest grey shading). From Gardner et al. (1984), with permission from Dr. Paul Winter.
and 1930, and to determine their cause of death. The results were controversial (Barker et al. 1989; Osmond et al. 1993). What they found, for both men and women, was that risk of death from heart disease was doubled in individuals born at a weight of less than 5.5 lb (2.5 kg) compared with those born at a weight of 9–9.5 lb (4.1–4.3 kg).

While there was much controversy about these initial findings, they in fact built on many earlier observations suggesting that the conditions during pregnancy and early life had long-term health consequences. Over the next two decades, a large amount of confirmatory epidemiological, prospective clinical, and experimental studies supported the general model that an adverse start to life, which might be proxied by a low birthweight, was associated with a greater risk of later obesity, cardiovascular disease, and insulin resistance.

Other workers also showed that excess nutrition in infancy, such as associated with formula feeding, also led to altered disease risk. Given that poor fetal growth is generally associated with some rebound rapid growth in infancy these are probably different perspectives of the same phenomenon. However, further evaluation of the developmental data showed that the relationship between developmental growth patterns and disease risk is not a function of severe intrauterine growth impairment; rather, there is a continuous change in risk across the full range of developmental growth patterns (Figure 9.9). This suggests that disease risk is not a consequence of developmental disruption but is
9.4 HOW CAN CHANGE IN THE ENVIRONMENT INCREASE DISEASE RISK?

part of a developmentally plastic process that originally had an adaptive purpose. Indeed, subsequent studies have shown relationships between the conditions of pregnancy such as maternal caloric intake and later pathophysiology of the offspring (Gale et al. 2006).

The first studies were conducted in high-income countries, so of course the question arises of how relevant they are to low- to middle-income countries. There have now been studies conducted in several of these countries including India, China, and parts of South America, and they all show similar trends. Finally, evidence from modern famines (Box 9.9) further supports the role of the intrauterine environment in determining health in later life.

9.4.3.1 Maladaptive Consequences of an Adaptive Process

These epidemiological studies could not provide evidence about the biological basis of the relationship, but a consensus arose that poor fetal or early life experience altered development in such a way that growth was affected and also influenced the propensity to develop disease in later life. The term “programming,” previously and variously used to reflect the putative intergenerational effects of gestational diabetes (see later) and then the later consequences of formula feeding versus breastfeeding, was adopted to describe the phenomenon. Subsequently, Barker and Hales developed the thrifty phenotype hypothesis, in an explicit reference to Neel’s earlier proposal of a thrifty genotype, to explain programming (Hales and Barker 1992). Their concept was that the fetus adjusts its biology in response to signals from its mother of poor nutrition, allowing it to survive until birth (i.e., an immediately adaptive response) but then predisposing it to the adverse consequences of such programmed thriftiness in adulthood. They proposed that the likely mechanism was the in utero induction of insulin resistance, as insulin is known to be a key regulator of fetal growth.

However, that model failed to fit several aspects of subsequent data. It assumed that fetal growth retardation was key to the process, yet it soon became clear that most people who later developed obesity and metabolic compromise were not growth retarded at birth. Further, infants who are small at birth have insulin hypersensitivity, and insulin resistance only appears at about 3 years of age (Mericq et al. 2005). Nevertheless, the conceptual

Figure 9.9 The relationship between birthweight and death from coronary heart disease (expressed as standardized mortality ratio, SMR) is continuous across the range of birthweights. Adapted from Osmond et al. (1993), with permission.
The improvement in farming efficiency in the twentieth century arising from mechanization and use of chemical fertilizers has not ended famine. Indeed, some commentators claim that even the great famines of history were never the result of food shortages alone but always had some exacerbating sociopolitical dimension. Such claims are supported by examination of the factors precipitating some of the famines of the last century, such as the Ukrainian famine of the 1930s in the Soviet Union, associated with forced collectivization, the disastrous famine in China that followed Mao’s Great Leap Forward in the 1950s, the famine associated with Pol Pot’s regime in Cambodia, and the association of famine with political instability in the Horn of Africa that continues to this day.

Famine, whilst a regrettable situation, provides scientists with an opportunity to examine one extreme of human nutritional physiology. Such natural experiments build our knowledge of how undernutrition during pregnancy affects the growth and later health of the offspring, but the social disruption that accompanies famine usually means that medical records and the opportunity for follow-up of affected individuals are lost, and famine is usually protracted, so that separation of effects on gestation from those on infant and childhood nutrition is difficult. Few famines are temporally circumscribed events during which adequate social and medical records are kept, but some are, and several cohort studies exist of twentieth-century populations whose nutrition was affected by armed conflict. Such cohorts include those from the Spanish Civil War of 1936–9, the siege of Leningrad from 1941–4, and the Dutch Hunger Winter of 1944–5.

The rapid advance of Allied troops across western Europe after the Normandy invasion in June 1944 was halted in September by the failure of the operation to seize the “bridge too far” across the Rhine at Arnhem in the Netherlands. Reprisals on the Dutch civilian population for their cooperation with the Allies, followed by a severe winter, meant that food and fuel supplies in the western Dutch cities were extremely limited between November 1944 and the liberation of the country in May 1945. Although the population was relatively well nourished before this period of famine, and food supplies were restored quickly after liberation, energy intake during the peak months of deprivation was no more than 400–800 calories per day. Despite these difficulties, medical care for pregnant women continued and detailed records of their food intake, the course of their pregnancies, and the size of their babies at birth were maintained. The availability of these records has allowed the lifelong health of the children of the Hunger Winter, who are now about 70 years of age, to be correlated with their exposure to undernutrition in utero at various stages of gestation and compared with that of a group of controls from the same area who were born in the months before, or conceived in the months after, the famine.

As early as 1976 it was reported that these famine-exposed offspring became obese as adults (Ravelli et al. 1976). Only exposure to famine in late gestation had a marked effect on birth size, with both length and weight being reduced. Babies born small because of maternal undernutrition in the third trimester developed glucose intolerance and high blood pressure in later life. But even though exposure to famine in early or middle gestation had no effect on birth size, such babies also showed markers of ill-health as adults, including glucose intolerance, altered blood lipid profiles and blood coagulation, and increased sensitivity to stress. These abnormalities were associated with an increased risk of coronary heart disease. In addition, female fetuses exposed to famine during early gestation were more obese as adults than women who were not exposed as fetuses, and had a markedly higher risk of developing breast cancer (Roseboom et al. 2006).

Studies in other populations exposed to famines have yielded similar findings. Individuals exposed to the Chinese famine of 1959–61 are at higher risk of developing hyperglycemia and the metabolic syndrome in adulthood, especially if they were born in more severely affected areas and then consumed a Western-style diet in later life (Li et al. 2011). In utero exposure to famine during the Nigerian civil war in the late 1960s has also been linked to hypertension, hyperglycemia, and adiposity in adulthood (Hult et al. 2010).

Box 9.9 Modern Famines

The framework of the thrifty phenotype has inspired considerable research which has led to our current understandings of the developmental basis of later disease risk.

Animal studies confirmed and extended the results and helped dissect the underlying molecular mechanisms. The initial studies were performed in rats. If a rat fetus was undernourished in utero, because the pregnant dam was fed a reduced-energy diet or just an unbalanced diet (e.g., with a low protein content), it became hypertensive and obese as an adult (Langley and Jackson 1994).
These adult rats were also shown to have insulin resistance and to have shorter life expectancies than those whose mothers had been fed a balanced diet in pregnancy. The effect was magnified if the rat was placed on a high-fat diet after weaning, echoing the human situation of dietary abundance and demonstrating that the interaction between the fetal and post-natal environments determined outcome (Vickers et al. 2000).

As experimental work proceeded it became linked to other fields of biological enquiry, especially evolutionary developmental biology (Chapter 4). The realization was growing that development was an important and under-represented component in the explanation of metabolic disease.

But a key feature of the link with metabolic disease was that this was not just about those with a low birthweight. The epidemiological studies had shown a continuous relationship between birth size and later disease risk, present even in those of above average birthweight (Figure 9.9). Epidemiologists had also shown that metabolic programming could be induced by changes in the fetal environment that did not affect birthweight (Gale et al. 2006). Further epidemiological and clinical studies showed that smaller infants had relatively more visceral fat (Box 9.1) at birth, although subcutaneous fat was reduced. Such observations, together with the broad incidence of metabolic disease in the population, suggested that the developmental component did not represent the outcome of a pathological process involving disruption of fetal development. It was rather a maladaptive outcome of the generally adaptive processes of developmental plasticity.

Indeed, there is now ample experimental evidence in animals, and growing clinical evidence, that lifelong epigenetic changes in genes associated with systems such as insulin sensitivity, glucose metabolism, and the glucocorticotoid axis underpin the developmental induction of metabolic risk (Low et al. 2014). In humans, individuals exposed pre-natally to the Dutch Hunger Winter famine of 1944 showed differential methylation levels at gene loci implicated in growth and in metabolic and cardiovascular disease risk nearly six decades after exposure, showing that a transient environmental exposure can indeed have long-lasting molecular consequences (Heijmans et al. 2008). Other epigenetic marks have been found in umbilical cord tissue that relate to body composition in pre-pubertal children (Box 9.10).

9.4.3.2 Developmental Plasticity and Mismatched Signals Across the Life Course

How can these developmental components be understood in evolutionary terms? The general paradigm we have put forward is that early developmental cues have induced an adaptive, developmentally plastic response that cues the individual’s physiology in the expectation of living in a nutrient-poor or adequate environment, but then the organism ends up in a more nutritionally plentiful environment than predicted. This general model suggests that these processes are not pathological in origin, but are fundamental to biological variation operating in the normal range of ecological cues. However, as will be discussed in Section 9.4.3.3, there are other developmental pathways that reflect evolutionarily novel exposures in development for which an adaptive explanation is inappropriate, including maternal obesity, gestational diabetes, and feeding with infant formula.

In Chapter 4 we explained how the processes of developmental plasticity act to allow one genotype to give rise to a range of phenotypes and how organisms use plasticity as an alternative or additional process to adapt to environments, particularly among species faced with change during their life course. Developmentally plastic responses are induced by external cues (e.g., maternal nutrition), and depending on the fidelity of the relationship between the cue and the future environment there may be effects on fitness or health.

It was hypothesized that if the developing organism predicts a limited nutritional environment in the future, it might be appropriate to use the mechanisms of plasticity to adjust growth patterns so that later body function is optimized for a limiting nutritional environment. Conversely, if the fetus predicts a later environment with plentiful nutrition it is appropriate to have a metabolic system set up with different expectations. In the former situation of predicted nutritional threat, an appropriate response would include reduced investment in...
somatic growth (e.g., reduced muscle mass), a preference for high-fat foods, metabolic settings that favor fat deposition in times of energy excess, and altered endocrine, behavioral, and vascular controls such that the organism has reduced insulin secretion and sensitivity (Figure 9.10). Given that evolution is driven by the fitness imperative, anticipation of a threatening environment might be expected to accelerate the timing of maturation and commit more resources to reproduction, perhaps even at a cost to other traits that improve longevity (e.g., by investing less in cellular or DNA repair) (Gluckman and Hanson 2006a). Depending on the severity of the initial cue, the organism may only respond with predictive adaptation, but if the challenge in early life is more severe it will induce the immediate adaptive responses of reduced growth and/or early delivery, thus explaining the original birthweight relationships found in epidemiological studies.

The general hypothesis of predictive adaptive responses was initially tested in rats. It was shown that neonatal rats that had been born to undernourished mothers could be “tricked” into thinking they were in a high-nutrition environment by being injected with the adipokine leptin. This prevented the animals from developing obesity, insulin resistance, and the other features of metabolic compromise, and the associated epigenetic changes—even when maintained on a high-fat diet through life (Vickers et al. 2005; Gluckman et al. 2007b). But could such a hypothesis be tested in humans? Studies in a population where severe undernutrition is common showed that being born small was protective against the risk of dying from infant undernutrition, and could be best interpreted as supportive of the evolutionary hypothesis (Forrester et al. 2012; Box 7.2).

The adaptive advantage of a predictive response depends on the fidelity of the prediction. If correct predictions lead to greater chances of growth and survival to reproduce, this would be why underlying anticipatory and plastic processes have been selected through evolution. Modeling work shows that developmental forecasting does not have to be particularly accurate to confer a selective advantage (Jablonska et al. 1995). When fetal nutrition is
not a reliable cue of external conditions—as a result of maternal disease, placental inadequacy or malfunction, or simply because the environment has changed notably between birth and later life—this could produce a phenotype that is not well suited to meeting the challenges of its environment, thus placing the individual at greater risk of disease. In the metabolic domain, a phenotype of increased insulin resistance, reduced muscle mass, and increased propensity to store fat is precisely the background on which susceptibility to metabolic disease would be enhanced in a later nutritional environment of high energy availability.

Because of maternal constraint, it is reasonable to suggest that most individuals are sensitized to an obesogenic environment because the fetus will be biased towards predicting lower-nutritional environments that may have been typical of earlier epochs. Maternal constraint (Box 9.11 and Section 8.9.5) refers to a set of somewhat poorly defined mechanisms by which fetal growth is limited by maternal size (Gluckman and Hanson 2004). It probably involves limits on uterine blood flow and placental–fetal hormone interactions. The outcome is that genetic factors are less important than post-natal factors in determining fetal growth—indeed embryo transplant experiments in animals, and studies of humans born to surrogate mothers, show that the maternal phenotype rather than genotype is the primary determinant of birth size. These studies

---

**Figure 9.10** Integrated life-history responses to developmental cues. Developmental trajectories in response to predicted optimal (adequate nutrition) or threatening (deprived nutritional) environments. HPA axis, hypothalamic–pituitary–adrenal axis. Modified from Gluckman et al. (2007a), with permission.
also show that fetuses do not grow to their maximal genetic potential because of these constraining mechanisms (Hanson and Godfrey 2008), and thus are perhaps generally signaled to expect limited nutrition after weaning. The exceptions are discussed in Section 9.4.3.3.

The developmental mismatch model proposes that, in evolutionary terms, there is an advantage in using the processes of developmental plasticity to adjust the set points for metabolic homeostasis to match the predicted environment of the mature organism. This process could have had adaptive advantage in environments that were reasonably stable over decades. But the fidelity of the prediction might increasingly be lost because the mechanisms of maternal constraint limit the forecast that is possible, particularly as the post-natal environment becomes abundant in evolutionarily unprecedented ways. Additionally, modern medical care allows a greater range of fetuses to survive, many with greater degrees of maternal constraint (such as twins and those with extreme growth retardation) and perhaps therefore with a greater risk of mismatch. Maternal ill-health and placental dysfunction are other ways in which the maternal–placental transduction of environmental information can lead to a faulty prediction, and progress in obstetrics and pediatrics allows a far greater range of babies to survive to adulthood.

While the expansion of experimental and clinical research and the definition of underlying epigenetic mechanisms has established the view that...
developmental pathways are important, the relative importance of these pathways in contributing to the risk of metabolic disease remains uncertain. Clearly, disease risk would not exist were it not for the large changes in diet and lifestyle discussed in previous sections.

It is important to note that this model does not imply that developmental factors cause obesity or type 2 diabetes; rather it explains how developmental factors change the sensitivity of the individual to the obesogenic environment in which they will live as older children and as adults. Indeed, recent epidemiological studies show that there is an interaction between developmental and evolutionary mismatch, such that adult lifestyle interacts with birthweight to compound the risk of developing type 2 diabetes (Li et al. 2015).

9.4.3.3 Evolutionary Novelty in Development

The developing human may, additionally, be exposed to entirely novel exposures, such as infant formula based on cows’ milk. It is well established that feeding with infant formula predisposes to obesity, probably through mechanisms other than via adaptive developmental plasticity, and that breastfeeding has a protective effect (Cattaneo and Cogoy 2012). When infants are breastfed they regulate their own food intake. This is not the case when fed by bottle, and this may change the development of appetite control. Human breast milk has a number of components including milk oligosaccharides, neurohormones such as leptin, and other hormones such as IGF1, which may have an impact on the infant’s metabolic and satiation biology and development. Furthermore, human breast milk supports a different gut microbiota from that of formula (Penders et al. 2006); these differences persist for some time after birth and may have important long-term effects. Finally, the macronutrient content of infant formula is very different from that of breast milk, and may make obesity more likely to develop.

A second example of probable evolutionary novelty that is of rapidly emerging public health importance is gestational diabetes mellitus. Hyperglycemia in a pregnant woman leads to fetal hyperglycemia and hyperinsulinemia, thus resulting in adiposity of the offspring. Insulin is known to be adipogenic in the late-gestation fetus and infant. As a consequence, infants of diabetic mothers have an increased fat mass, in proportion to the degree of maternal hyperglycemia (Figure 9.11a). The larger number of fat cells in infancy confers a greater risk of obesity through childhood. Beyond this, there may also be epigenetic changes that also predispose to a later risk of diabetes.

It has been suggested that gestational diabetes is an evolutionarily novel exposure (Ma et al. 2013). Some degree of insulin resistance is normal in pregnancy, and is induced by placental production of growth hormone and placental lactogen that induces mild insulin resistance in the pregnant woman. This promotes transfer of maternal glucose to the fetus to promote fetal growth. But unlike other nutrients such as amino acids and fatty acids, there is no limit to placental glucose transfer. Gestational diabetes itself is more likely in women who had been born small and are consequently at greater risk in an obesogenic environment, thus reflecting earlier developmental impacts on their own lives. Left untreated, gestational diabetes results in fetal macrosomia, which in the absence of obstetric intervention leads to dystocia, posing a severe risk to both mother and child. That there is no limit on glucose transfer suggests that there has been no selective pressure for such a need, which might imply that gestational diabetes is itself a manifestation of modernity. Indeed, maternal obesity and gestational diabetes often coexist, and correlations have been shown between maternal BMI and offspring adiposity (Figure 9.11b).

Animal research is now starting to suggest that paternal obesity and insulin resistance, experimentally induced in male rodents, can also pass risk to their offspring, even when the mothers are lean and fed a balanced diet (Fullston et al. 2013, Wei et al. 2014b). This is one of the growing pieces of evidence for epigenetic inheritance (Section 3.8).

One of the most active areas of current research concerns the role of the gut microbiome in health (Box 9.3 and Section 10.3). Recent work has shown that women whose gut microbiota predominantly comprise the bacterial phylum Firmicutes had different epigenetic marks in genes functionally associated with obesity, cardiovascular diseases, and inflammation. A human infant that is delivered
Figure 9.11 The influence of maternal hyperglycemia and BMI on offspring adiposity. (a) Maternal levels of fasting glucose (upper panel) and plasma glucose 2 h after an oral glucose challenge (lower panel) are associated with neonatal adiposity in a continuous manner. Data plotted from The HAPO Study Cooperative Research Group (2009). (b) Maternal BMI in the first trimester of pregnancy is strongly associated with risk of offspring obesity at pre-school age. Data plotted from Whitaker (2004).
vaginally will acquire a microbiota similar to that of its mother’s vagina, while infants delivered by cesarean section have a very different microbiota (Funkhouser and Bordenstein 2013). There is now good evidence that the risk of later overweight/obesity is up to 25% greater in babies who were delivered by cesarean section (Darmasseelane et al. 2014). Although the causative nature of the association remains to be tested, this probably represents another example of a health risk arising from exposure to an evolutionary novelty—namely loss of exposure to the maternal vaginal microbiota.

Other evolutionarily novel exposures include environmental chemicals such as endocrine disruptors, which mimic or block hormonal action and are found in many common modern-day household products. Many of these endocrine disruptors interfere with lipid metabolism and adipogenesis, and have been implicated in the growing obesity epidemic.

Recent data in rats suggest that the adverse effects of maternal obesity on health of the offspring may be mitigated by supplementing the mother’s diet during pregnancy with methyl donors, taurine, or a cocktail of antioxidants (Vickers and Sloboda 2012). The specific mechanisms by which this occurs remain unclear, but these findings provide important proof of principle for the reversibility of developmental metabolic effects induced by evolutionarily novel exposures. In light of the modern-day obesity epidemic, the potential public health utility of these findings may be substantial.

9.5 Conclusion

This chapter has demonstrated how an evolutionary perspective can provide insights into the epidemic of metabolic diseases such as obesity and type 2 diabetes. It is the body’s limited capacity to adapt to changing environments that ultimately lies at the heart of these diseases.

Members of a species adapt to their nutritional environment through natural selection operating on genes that regulate metabolism and thus influence how the body manages its finite energy supply and expenditure. This process of genetic adaptation is slow but powerful, and through time members of a species come to have metabolic machineries that are well suited to the conditions they routinely experience.

New evidence is revealing how we also have a capacity to adjust our metabolic priorities developmentally in response to conditions and cues experienced early in life, beginning in utero. Together, these two modes of adaptation allow organisms to cope with long-term and gradual changes and with more fine-grained changes that occur across decades or generations.

Despite the best efforts of these mechanisms, rapid environmental change can outstrip their ability to accommodate it. Metabolic diseases can be understood as a symptom of the resultant mismatch between an organism’s biology and its environment. When environmental change is rapid relative to the rate of natural selection, the consequent gene–environment mismatch will drive disease.

We have abundant evidence, albeit mostly indirect, that this is an important influence on human health. The human genome was selected over millions of years for a lifestyle of foraging and continues to “expect” a diet high in protein, low in fat, and free from novelties such as refined sugars. Even if the search for specific metabolic disease-related genes has posed challenges, it seems undeniable that the bulk of the world’s ballooning waistlines reflect patterns of cultural and dietary change coming into conflict with an increasingly obsolete genotype.

But mismatch between genes and environments may only be part of the story, for there is also evidence that an individual’s nutritional experiences in early life influence how the body handles nutrition later in life, suggesting that when the pace of change is particularly rapid within a single lifetime or between generations, the body’s capacity to adjust its developmental biology in response to early nutrition can lead to a different form of mismatch: that between the biological settings established early in life and the nutritional and lifestyle conditions experienced subsequently.

This chapter has shown how both evolutionary and developmental mismatch interact to drive the modern global epidemic of metabolic diseases: gene–environment mismatch resulting from historically recent changes in diet and physical activity is largely responsible for the rise of conditions
like obesity as global phenomena. As populations experience dietary change and gain weight, the impact on health depends upon that population’s recent nutritional history. Populations with a high prevalence of individuals in the developing world who were born with a lower birthweight, for example as a result of maternal stunting, will be worse off as they gain weight. Maternal obesity and gestational diabetes exacerbate the transmission of risk to the next generation. Such models can help explain differences in disease susceptibility in populations at different points in the nutrition transition or experiencing change at different rates. In all cases, it is adaptability—or, more precisely, the limits of adaptability—that lies at the heart of this modern global scourge.

**Key Points**

- The bulk of human evolution occurred in a nutritional milieu that was very different from the modern one, and the human genotype was selected to be well adapted to that milieu.
- There is a rising incidence of obesity and its associated disorders in both the high- and low- to middle-income countries.
- There has been a rapid change in diet and behavior against the background of a genotype that cannot change rapidly, creating an evolutionary mismatch.
- Fixed genetic processes make only a small contribution to the prevalence of metabolic disease at the population level.
- Developmental processes may also contribute to the rising incidence of metabolic mismatch.
- Differences in metabolic phenotype induced in development can alter the individual’s responses to later challenges such as living in an “obesogenic” environment, resulting in greater risk of obesity and metabolic disease such as diabetes.
CHAPTER 10

Coevolution, infection, and immunity

10.1 Introduction

Rates of extrinsic mortality (the risk of death as a result of environmental hazards) have had major effects on the evolution of life-history traits such as growth rate, timing of maturation, allocation of resources to repair and maintenance, and lifetime reproductive strategy (see Chapter 5). The major causes of extrinsic mortality are biotic, and include competition for nutrient supplies with other organisms of the same or different species, predation, infection by micro-organisms, many of which rely on a parasitic relationship for their own survival, and injury, which may be accidental or the result of conspecific competition and violence. This last cause is a particular feature of human biology.

Human cultural evolution has resulted in mechanisms to ameliorate the risks of predation and accidental injury, and technological advances have reduced the impact of infection by micro-organisms in many parts of the world. Yet infectious disease still causes 25% of all deaths worldwide, with wide variations between high- and low-income countries. Evolved mechanisms to cope with the impact of infectious disease are a major feature of the human genome (Quintana-Murci et al. 2007).

Not all human relationships with micro-organisms are deleterious to the host. Our associated microflora, principally in the gut, contribute to nutrition and metabolic regulation. Beneficial micro-organisms participate in controlling the growth of pathogens, and may have a role in the evolution and individual development of the host immune system.

In this chapter we describe some fundamental aspects of coevolution and then apply them to the evolution of the relationship between humans and micro-organisms, discuss why and how some of those micro-organisms are pathogenic, and examine the benefits and occasional costs of our evolved defenses against micro-organisms through our innate and adaptive immune systems. Finally, we examine the three main types of human technological (cultural) response to the threat of infectious disease—public health measures, vaccination, and antimicrobial chemotherapy.

10.2 Coevolution

Organisms live and evolve in environments that include other organisms. Coevolution occurs when one species reciprocally influences the evolution of another. A simple example is the evolution of a predator–prey relationship where the prey evolves a mechanism to evade the predator (perhaps running faster) and the predator, faced with the selection pressure of reduced food supply, responds by evolving a countering mechanism (perhaps running faster still or developing a new attack strategy). In turn, the prey evolves a response to the countering mechanism, and so on. Such processes of adaptation and counter-adaptation have been referred to as evolutionary arms races, by analogy with the desire by the military not to be left vulnerable to a potential enemy’s developing arsenal.

Modeling studies show that coevolution is an important driving force for evolution. A common metaphor in classical evolutionary theory suggests that species evolve “uphill” towards fitness peaks in their adaptive landscapes, where they are well adapted to their physical environment (see Box 2.4). As the physical environment generally only changes slowly, the rate of evolution is predicted to slow down as the peak is approached. Inclusion...
of other species into the notion of “environment” changes this perspective, because other species, particularly microorganisms, can evolve rapidly and this rapidly changing biological environment (which thus changes the shape of the landscape) requires equally rapid counter-adaptive change from the host species; this is then matched by a further adaptive response from the interacting species. Continuous adaptations by all species in the system are required for survival, although relative fitness as determined by classical measures of reproductive success does not increase. This concept was named the Red Queen hypothesis after the remark by the Red Queen in Lewis Carroll’s Through the Looking Glass that Alice needs to undertake “all the running you can do, to keep in the same place” (van Valen 1973).

The Red Queen hypothesis as applied to host–parasite interactions has been proposed as an explanation for the evolution of sex (see also Section 8.2). The argument is that the requirement for slowly evolving multicellular eukaryotes to compete with their more rapidly evolving parasites is facilitated by the continuous generation of new genotypes containing novel combinations of parasite resistance alleles, and that can only be achieved by meiotic recombination during the gametogenesis of sexual reproduction. There is some limited experimental evidence from organisms such as roundworms that can reproduce either sexually or by self-fertilization to suggest that the sexually reproducing forms do better when exposed to parasites (e.g., Morrán et al. 2011).

Coevolution does not only occur in the context of parasite–host or predator–prey relationships, where one participant may be “running for its life, rather than just its dinner.” There are many different types of species interactions with different coevolved allocations of costs and benefits. Symbiosis refers to any long-term interactions between species: this can be subdivided into mutualism (where both participants benefit), commensalism (where one participant benefits and the other is unaffected), parasitism (where one participant benefits and the other is harmed by, e.g., reduced growth as the parasite extracts resources), and predation (where one participant benefits and the other reaches a sudden evolutionary dead end).

Pathogens can be considered as a type of parasite with an evolved strategy of causing disease in their host. Some species with complex life cycles (such as the unicellular organism that causes malaria) may be parasites in a vector species (mosquitoes) but pathogens in an intermediate host during their life cycle (in this case humans). Some commensal species, such as certain microorganisms in the human gut, may become opportunistic pathogens in their hosts when defenses such as epithelial barriers or the immune system are weakened.

10.3 Humans and Their Associated Species

Humans have coevolved with many species, ranging from large mammals to viruses. Coevolution of a mutualistic relationship of humans with domesticated animals is exemplified by cattle, where human-directed selection has resulted in multiple strains of animals selected for nutritionally and economically beneficial traits while the coevolved trait of lactase persistence in humans (see Box 1.1) has spread as milk consumption and then dairy farming became established practices.

In this chapter, however, we are mostly concerned with the relationships between humans and their associated microorganisms. Those relationships can be commensal (e.g., many skin and gut bacteria), mutualistic (many of the gut microbiota), parasitic (e.g., intestinal worms and body lice), or pathogenic (e.g., bacterial diseases such as cholera and tuberculosis and viral pathogens such as HIV and measles). Nevertheless, the distinction between a commensal and a pathogen is sometimes blurred, as for example when the normally harmless Escherichia coli acquires a virulence factor such as the bacteriophage-encoded Shiga toxin, giving rise to the pathogenic O157:H7 strain.

It is only recently that the complexity and physiological importance of the community of symbiotic microorganisms associated with humans (i.e., the microbiome) has come to be appreciated. The human body consists of about $10^{13}$ human cells, but this is outnumbered about ten-fold by the bacterial, archaeal, fungal, and protozoal cells found on all body surfaces (on the skin, in the mouth and nose, and in other orifices) and in particular


in the lower gastrointestinal tract, which contains 1 to 2 kg of bacterial cells representing at least 500 species. However, most of the mass of the gut microflora comes from 30 to 40 dominant anaerobic species, predominantly from the phyla Bacteroidetes and Firmicutes. The composition of the gut microbiome varies in type and diversity between individual humans according to factors such as age, geographical location, and diet. Families tend to share similar microbiomes, although this seems to be environmentally rather than genetically determined since monozygotic and dizygotic twins have similar levels of variability in their microbiomes and newborn infants take some months to achieve similar gut microbiomes to the rest of their family. The diversity of the microbiome, which is generally taken as an indicator of ecosystem health, is much higher in modern hunter-gatherer populations than in Westernized industrial populations.

The gut of the human neonate is generally sterile, but is rapidly colonized by maternal and environmental microflora; the composition matures during early life and then remains relatively stable, although subject to dietary and other environmental influences (Sekirov et al. 2010). Interestingly, establishment of the infant microbiome is promoted by, and the composition may be affected by, components of breast milk such as IgA and oligosaccharides, as well as by the tolerogenicity of the developing infant’s immune system. Some studies have shown correlations between the composition of the microbiome in a mother and her offspring; this can be considered a further form of non-genomic inheritance (see Box 10.5). Infants born by cesarean section as opposed to a vaginal delivery have different gut microflora from each other, as do infants fed on formula as opposed to breast milk. It may be that these evolutionary novelties in childrearing have later consequences for health; for example, it is known that breastfed infants have a lower risk of obesity and metabolic disease and greater cognitive development than do formula-fed infants (see Section 9.4.3.3).

The human gut offers an attractive environment to its microbial colonizers, being warm and damp and providing regular delivery of nutrients. Nevertheless, the relationship should be seen as mutual rather than commensal, since both parties derive benefit. The gut microflora provide numerous benefits to the host, including digestion of nutrients such as dietary fiber (Box 10.1), synthesis of vitamins such as folate and biotin, and metabolism of potentially toxic food constituents. There is growing evidence for a link between the gut microflora and the brain, affecting aspects of mood and behavior (Sampson and Mazmanian 2015). Importantly for this chapter, the gut microflora also act to protect the host organism by preventing colonization of the gastrointestinal tract by pathogenic competitors (Box 10.2) and by modulating the development of the host’s adaptive immune system. We will return to this latter point in Section 10.7.2.

**Box 10.1 Seaweed on the Menu**

The gut microbiome is a highly versatile addition to its host’s metabolic capability. A striking illustration of the potential of the microbiome comes from an unexpected source.

The *nori* seaweed (actually a red alga) used to wrap sushi contains sulfated polysaccharides that humans are unable to digest, although marine bacteria have evolved enzyme systems that can break down these and other constituents of algal biomass. This set of metabolic enzymes has also been found in human gut bacteria of the genus *Bacteroides*—but only in people from Japan, who traditionally consume large amounts of seaweed (Hehemann et al. 2010). The implications are that *Bacteroides* acquired these enzymes by horizontal gene transfer after their human hosts ate seaweed contaminated with marine bacteria, and that regular consumption of algal polysaccharides creates the selective pressure to maintain this capability in the microbiome.

There are two wider points to be made from this intriguing observation. The first is how the metabolic diversity of the gut microbiome is sensitive to the environment—here, both novel food constituents and novel microbes. The second is how human symbionts can take advantage of horizontal gene transfer from transient species to acquire a metabolic trait that could be detrimental to the host if the metabolic trait happens to be some form of antibiotic resistance.
On the other hand, some parasitic species use the human intestine as a host environment. Tapeworms, trematodes (flatworms or flukes), and roundworms pass their eggs/larvae to other hosts via human feces. Some infestation is direct, by contamination of water or vegetation, and sometimes it involves an intermediate host, for example the water snail in schistosomiasis. These parasites compete for nutrients and can cause considerable malnutrition in the host, with long-term consequences such as stunting, as well as abdominal disorders such as enlargement of the liver or spleen. Gut parasitism is particularly prevalent in low-income countries.

**10.4 The Challenge of Infectious Disease**

There are over 1200 species of recognized human pathogen, although the majority of infection-related mortality and morbidity is now caused by just a few of these, particularly malaria, HIV/AIDS, and tuberculosis. This multiplicity of threats imposes a continual interplay between our defense systems and the transmissibility and virulence of these organisms.

Micro-organisms evolve in relation to their hosts and the vectors they rely on for their survival and fitness. There are great differences in population size and generation times between micro-organisms and humans, giving microbes the advantage of being able to evolve much more rapidly. Indeed, in bacteria, stress may produce an increase in their mutation rate (often termed the SOS response; Baharoglu and Mazel 2014) which makes it more likely that new variants will appear and that gene flow will be preserved. This difference in generation times is well demonstrated by the development of bacterial resistance to antibiotics, which is one of the best-characterized manifestations of evolutionary processes operating in real time (Section 10.10).

Humans with their long generation time must deal with this microbial challenge, and the biological complexity of vertebrates has allowed the evolution of a number of defensive strategies that can cope with faster-evolving organisms—in particular the innate and adaptive immune systems (Section 10.7). In historical times humans have added additional strategies in the form of technologies such as vaccination (Section 10.9) and antimicrobial chemotherapy (Section 10.10). Thus an
evolutionary arms race is played out with different weapons employed on each side—rapid mutational change on one hand, and multivalent immune and defense mechanisms on the other.

10.5 Pathogen Emergence

During human evolution the risks of infection by many pathogens changed significantly. The development of animal husbandry, settlement, and increasingly larger aggregations of population allowed infectious agents to spread more easily. Problems of waste disposal and hygiene grew with settlement. Other human endeavors also increased the risk of infection. For example, alterations made in drainage and irrigation systems in Africa led to the spread of schistosomiasis. Thus, while humans have evolved alongside their microbiotic environment, the challenges have become greater in the past 10,000 years (Cleaveland et al. 2007).

Patterns of pathogen-induced disease are not the same across the globe. Environmental factors play a part in this; for example, the conditions suitable for mosquitoes to breed are confined to tropical and subtropical pools of water. Historically, malaria was endemic in southern Europe well into the twentieth century, and climate change may mean that the disease could well become widespread there again.

However, historical considerations reveal that there may be more to explaining the patterns of disease. Why, for example, do the diseases of the tropics include chronic infections and infestations such as schistosomiasis and onchocerciasis and those of more temperate regions include infections such as smallpox and tuberculosis? One hypothesis is that the tropical pattern emerged through coevolution of the pathogens responsible in consort with ancestral hominins in Africa. Low population density and a nomadic lifestyle may have favored the evolution of chronic diseases with relatively low virulence and the need for intermediate vectors such as the mosquito, tsetse fly, and water snail. The more recent migration into temperate regions was accompanied by the development of settlements with a higher population density and of animal husbandry, favoring diseases with a zoonotic origin or with higher virulence and more direct person-to-person transmission. In addition to the adoption of a sedentary lifestyle, humans in temperate zones began to domesticate animals. Close proximity to such animals favored the transmission of their pathogens to humans, and this may be the origin of diseases such as tuberculosis. Interestingly, the overwhelming majority of human infectious diseases originated in the Old World (Africa and Eurasia) rather than in the Americas, possibly as a result of the geographical difference in the number of animals that were domesticated (Wolfe et al. 2012).

Most infections originate and are sustained by close interpersonal contact and/or by contact with animals. Many viruses that infect humans have their origin in domesticated animals. For example, the influenza virus originated in pigs and poultry, and it is possible that the human measles virus evolved from the closely related morbillivirus that causes the cattle disease rinderpest (Sharp 2002)—the close similarity between measles virus and the morbillivirus that causes canine distemper suggests a later jump in the opposite direction, from humans to dogs (Uhl et al. 2011). It is generally accepted that HIV is derived from the lentivirus simian immunodeficiency virus (SIV); this is a slow-growing subtype of retrovirus that infects Old World primates. There are many previous incidences of the transmission of such viruses between species, but the critical event for the origin of human HIV was likely to have been transmission from (depending on the strain of HIV) chimpanzees or sooty mangabeys to humans, most likely as result of blood contact during hunting and butchering of these species for bush-meat. We may never know the precise location or time of the transmission, but research suggests that the origins were south central Cameroon and possibly Guinea Bissau during the 1940s or 1950s.

What determines the transmission of an infectious disease from animals to humans? Why is it a relatively rare event? The answer seems to be that a successful shift in host from animal to human involves several distinct events. Consider the influenza A virus H5N1, a strain of avian influenza which has only infected a few humans who have been in particularly close contact with infected poultry, but about which there is much concern for the potential for human-to-human transmission. Studies of the influenza pandemics of 1918, 1957, and 1968 give some clues. Influenza viruses are
endemic in bird populations, both in the wild and in domesticated species, and in some other animals such as pigs, and they are highly mutable. Each pandemic is thought to represent the emergence of mutated forms that have escaped the immunity of previous influenza infections in humans and successfully made the transition from reproducing in an animal host to reproducing in a human host. The so-called Spanish flu virus, which produced the 1918 pandemic in which 50 million people died worldwide, may have been due to the transfer of a complete avian virus into humans, although others have proposed an intermediate mammalian vector (Taubenberger et al. 2012). In contrast, the more common influenza A virus shows relatively linear evolution. The antigenic properties of the influenza virus change from year to year as a result of mutation (antigenic drift), necessitating the production of new vaccines and annual vaccination of susceptible members of the population. But on top of this there have been several large shifts in antigenicity in the past century, such as the “Asian flu” in 1957 and “Hong Kong flu” in 1968, arising from reassortment of genes between two viral strains co-infecting an individual.

Close proximity between humans and animal hosts offers the potential for initial direct inoculation (e.g., through a cut). In the case of H5N1, the first transmission to humans was reported in 1997 and this is the stage at which H5N1 currently exists: infection from the animal host can occur, but human-to-human transmission remains unconfirmed. The critical next step will be if the H5N1 virus exchanges genomic material with a flu virus that has the potential to be transmitted between humans because it can bind to surface receptors on cells in the human respiratory tract. If this happens a pandemic could result, because there may be little resistance in the human population in any country. Given that viruses replicate with very short generation times, the chances of individual mutations are high. The cumulative effect of those mutations in inducing a virulent pathogen depends on whether the human defense mechanisms are adequate or whether technology in the form of isolation, vaccination, or medication can reduce the spread of the organism and contain the epidemic.

10.6 Pathogen Virulence and Transmission

The evolutionary aim of a parasite, like that of all other species, is to maximize its reproductive success. The parasite does this by shaping aspects of its life cycle to optimize survival and replication in its host and transmission to its next host. Two of the traits that evolution uses to optimize reproductive success are virulence and transmissibility.

By definition, a pathogen is a parasite that causes morbidity and mortality in its host. It may do this by secreting a toxin, damaging cellular function, competing for nutrients, or simply by causing mechanical damage. The extent of this damage is termed its \textit{virulence}. Virulence may appear as incidental damage to the host that does not benefit the infecting organism, or as damage that does benefit the infecting organism, such as by increasing resource extraction from the host or enhancing the transmission of the pathogen. For example, the principal morbidity caused by HIV is gradual destruction of the host’s immune system, increasing susceptibility to opportunistic infections and malignancy. This type of damage to the host does not benefit the virus by directly enhancing its sexual transmission, although the temporal pattern of the infection, with a long asymptomatic period during which numerous sexual partners can be infected, does promote transmission. Contrast this with cholera, ingested for example via water contaminated with feces containing the cholera bacterium. The organism clings to the wall of the gut and secretes a toxin which triggers the secretion of serous fluid and rapidly produces violent diarrhea. This trait is adaptive for the cholera organism in that it allows greater spread to other hosts; it may also serve to displace commensal gut microbiota, giving the pathogen a competitive advantage in the gut environment.

For a pathogen to be evolutionarily successful, it must also be transmitted to a new host. Parasite fitness is often expressed as basic reproductive ratio, $R_0$, the average number of new host individuals infected during the life cycle of the parasite. $R_0$ must be at least 1 if a parasite is to persist within its host population; high values of $R_0$ indicate that the parasite is spreading rapidly. Conversely, if $R_0$
is below 1, because of poor transmission or strong host defenses, the infection will die out.

A pathogen will evolve to an equilibrium established by the optimal trade-off of virulence against transmission to ensure that it infects the greatest possible number of new hosts. High virulence kills the host quickly, so the pathogen needs to be highly transmissible. Conversely, low virulence allows the host to survive for long enough to infect many other hosts, so the pathogen can afford to be less transmissible. Optimization of this trade-off depends on a number of factors. These include whether transmission is vertical (between mother and offspring) or horizontal (across all members of a species) as well as the method of transmission. Pathogens can be transmitted horizontally by many routes, for example by airborne droplets (exemplified by the common cold virus and influenza), by contaminated water (e.g., cholera and typhoid), by contact with infected body fluids (e.g., HIV, Ebola), by contact with dormant organisms in the environment (e.g., smallpox), by sexual intercourse (e.g., syphilis, HIV) or by an intermediate vector (e.g., malaria).

Exclusively vertical transmission of human pathogens is unknown, although many diseases, such as HIV, can be transmitted both vertically and horizontally. (Some animal parasites, such as Wolbachia, an intracellular bacterial symbiont of invertebrates, are normally transmitted only vertically.) The virulence of horizontally transmitted pathogens often depends on the method of transmission. For example, gastrointestinal pathogens such as cholera that are transmitted in contaminated water can afford to be highly virulent, as they do not depend on continued survival of the infected host to achieve high values of $R_0$. Conversely, a pathogen that is exclusively transmitted by sexual contact would be expected to have low (or, at least, delayed) virulence, as achieving $R_0 > 1$ relies on survival of the host to infect more than one partner.

The setpoint of the virulence/transmission trade-off of a particular pathogen in a particular host will depend on both pathogen and host factors and the history of the coevolutionary interaction between them (see Box 10.3). In general, we might predict that the longer the relationship between pathogen and host, the lower the virulence of the pathogen in that host. This is because the host will have had time to evolve resistance or tolerance to the pathogen and the pathogen will have evolved its virulence/transmission trade-off to a level that allows its maintenance in the host population. Such reduction in virulence can be observed experimentally (Bérénos et al. 2009) and in the field, where a reduction (although to still high levels) in the virulence of the myxoma virus was observed in the decades after its artificial introduction into Australia to control the rabbit population (Di Giallonardo and Holmes 2015).

The converse situation is relatively common, however, for organisms that normally infect other hosts but are opportunistic pathogens in humans, where their virulence can be very high. The Ebola virus is a good example. Its natural host is a species of fruit bat, in which it appears to have little or no virulence. However, when transmitted to humans, possibly via consumption of infected bush-meat, it is highly contagious and the resulting hemorrhagic fever has a very high case fatality rate. But this very high virulence limits the ability of the virus to spread to a large number of hosts, and isolation and quarantine, properly applied, are effective in further restricting the pool of potential hosts and containing the outbreak.

The virulence/transmission trade-off can be manipulated artificially, and this is the basis of the live attenuated vaccines developed in the 1950s and still used for immunization against some diseases such as poliomyelitis. The process relies on serial passage of the pathogen in the laboratory in cell or tissue culture under “ideal” conditions where transmission is assured (high $R_0$) and the host (the cell culture) cannot evolve resistance or tolerance because each passage is into naïve tissue previously unexposed to the pathogen. The result is that the pathogen evolves towards high virulence in its new host but decreased virulence in its natural (human) host, and this “attenuated” strain is used for vaccination. In Section 10.9 we discuss how vaccination programs themselves can alter pathogen virulence.

Whether or not we succumb to an infection is often not simply a question of whether we have come into contact with the pathogen, but rather the size of the infectious load to which we are exposed. Simple procedures such as hand-washing greatly
reduce the risk of cross-infection in hospitals by reducing the infectious load; the major advance in patient survival after surgery afforded by Lister’s invention in 1869 of the crude but effective carbolic acid spray was due to the reduction of the infectious load rather than achieving complete sterility, a principle that still holds true in operating theatres today. Outbreaks of, and deaths from, plague still occur, although early diagnosis and antibiotic treatment can reduce the case fatality rate to less than 10%.

*Yersinia pestis* has evolved within the past 10,000 years, probably in central Asia, from the progenitor species *Yersinia pseudotuberculosis*. This species, which is still extant, is a water-borne pathogen that causes mild gastrointestinal symptoms in mammals, and can live harmlessly in the lower intestinal tract of fleas. Some of the serial genetic alterations that permitted *Y. pseudotuberculosis* to change its transmission and virulence characteristics have been mapped (Sun et al. 2014). First, mutations in genes controlling biofilm development allowed the bacterium to colonize the upper and middle parts of the intestinal tract of the arthropod vector, aided by acquisition of a plasmid encoding a phospholipase D that enhances bacterial survival in this environment. Together, these genetic changes allowed the bacterium to be transmitted to an animal host by a flea bite. Next, and crucially for the virulence of the pathogen, was acquisition of another plasmid encoding a bacterial cell surface protease. Possession of this enzyme allows *Yersinia* to spread in its mammalian host from the site of the initial flea bite, and is sufficient for development of the pneumonic form of plague. The final change in virulence was a mutation in the plasmid protease gene that enhances enzyme activity. This mutation facilitates the rapid systemic invasion by the bacterium that characterizes the bubonic (lymph gland) form of the disease (Zimbler et al. 2015). Onward transmission as an uninfected flea bites an infected host would be facilitated by systemic dissemination of the bacterium, and it is possible that acquisition of this last mutation contributed to the emergence of pandemic outbreaks of plague. Study of the evolution of *Yersinia* demonstrates that relatively minor genetic changes can lead to the emergence of new pathogens.

## 10.7 Host Defenses

Humans have evolved in the presence of a large repertoire of associated micro-organisms. Some of these are pathogens or potential pathogens, but the human “meta-organism” (i.e., the body plus its microbiome) also contains a large number of mutualistic or commensal species, which normally vastly outnumber the pathogens. From this perspective, our evolved antimicrobial strategies should be seen not only as defenses against pathogens but also as mechanisms to ensure peaceful coexistence with our commensals. Reciprocally, not only do commensals themselves provide protection against pathogens, but exposure to commensals in early life appears to be a prerequisite for maturation of the adaptive immune system (Belkaid and Hand 2014; Section 10.7.2). In the following subsections we discuss and distinguish between these aspects of our host defense systems.
10.7.1 Innate Immunity

All multicellular animals possess an innate immune system; this protects them from infection by maintaining defensive barriers against penetration by micro-organisms and by mounting chemical and cellular defenses against any microbes that do reach the interior of the organism. Innate immunity is non-specific and, unlike the adaptive immune system, does not lead to any lasting protective immunity (or “memory”). The innate immune system is phylogenetically ancient, and many of its features are conserved across both vertebrates and invertebrates.

The physical barriers that resist invasion by micro-organisms include the skin as well as the epithelial mucosa lining the alimentary tract and parts of the respiratory and reproductive tracts. These barriers are equipped with sensory receptors that are deployed at strategic points, where they have greater sensitivity. Sensitivity is greatest on the most frequently exposed parts of our skin such as the fingers and face; mechanoreceptors can detect very small inhaled particles in the nose or larynx and, on the skin, can detect small wounds. Tight junctions between epithelial cells provide good protection against the passage of micro-organisms; the waterproofing of the skin limits passage of water-borne organisms, and the mucus-secreting glands and ciliary action of the mucous membranes help to expel ingested or inhaled organisms. Some attackers have evolved highly effective ways of penetrating these barriers, such as the proboscis of the mosquito, and other parasites such as the malaria protozoan use this as a way of gaining access to a host’s body.

Other barriers within the body have different degrees of penetrability and defensive strategies (Doran et al. 2013). The human placenta, for example, is somewhat “leaky,” prioritizing exchange and transport functions over defense, and so some micro-organisms such as the syphilis spirochete (*Treponema pallidum*), rubella virus, *Toxoplasma gondii*, *Listeria bacteroides*, and cytomegalovirus can pass from mother to fetus. At the other extreme the barrier between the blood and the brain is very tight apart from at the cribriform plate, which conducts the olfactory nerves from the nose—this provides a route for infection leading to meningitis. Amoebic meningitis, a very rare disease, can arise when contaminated water found in natural hot springs enters the nose.

The epithelial mucosae, such as those lining the respiratory tract and gut, are among the sites most exposed to the microbiota, yet they must allow the passage of oxygen and nutrients. Epithelial cells and their associated inflammatory and phagocytic cells, such as macrophages and neutrophils, secrete mucus, IgA, and other antimicrobial substances such as those of the complement system to provide physical and chemical barriers to the penetration of micro-organisms. The triggering of these defenses to invading organisms involves recognition of chemical structures, called pathogen-associated molecular patterns (PAMPs), that are common to micro-organisms but not found in more complex animals. In other words, this form of discrimination between self and non-self takes the form of a generalized ability to distinguish between multicellular and single-celled organisms. PAMPs include microbial proteins such as flagellin, cell wall molecules such as mannose and lipopolysaccharides, as well as unmethylated DNA (most eukaryotic DNA is methylated; see Section 3.7). The family of pattern recognition receptors that sense PAMPs can trigger phagocytic and inflammatory responses in mucosal tissue, leading to engulfment and destruction of invading micro-organisms (Albiger et al. 2007).

The release of cytokines such as interleukin-6 from cells of the innate immune system (e.g., macrophages) in response to PAMPs causes fever, one of the most characteristic responses to infection. In mammals, fever is mediated centrally by hypothalamic mechanisms that drive thermogenesis and promote peripheral vasoconstriction to reduce heat loss. In turn, this induced hyperthermia drives further activation of both the innate and adaptive immune systems (Evans et al. 2015). This may occur by secretion of heat-shock proteins into the circulation. Heat-shock proteins are a diverse group of proteins with multiple intracellular functions, so named because they are induced in a cell when it is exposed to elevated temperatures (Srivastava 2002). These proteins control the transcription of a range of genes, such as those involved in glucocorticoid secretion, as well as expression of cytokine
246

C oevolution, infection , and immunity

and chemokine genes. Thus a febrile response may
be part of an evolved defense mechanism that uses
the resulting expression of heat-â•‰shock proteins to
stimulate the immune system. Although microbial
replication may be less efficient in a hyperthermic
host, some pathogens such as the malaria parasite
produce their own heat shock proteins that probably serve to protect the organism during episodes
of fever (Przyborski et al. 2015).
Competition for the essential nutrient iron illustrates another form of interaction between host and
invader. Microbial growth in host tissues is limited
by the availability of free iron, and infection induces
secretion by the host of iron-â•‰binding proteins such
as transferrin, which act to sequester iron intracellularly and make it unavailable for bacterial metabolism. For their part, bacteria secrete siderophores,
which are small peptides that bind iron and transport it into their cells. In turn, the host processes
that detect PAMPs stimulate the secretion of a small
protein called lipocalin 2, which binds the bacterial
siderophores with very high affinity and deprives
the pathogens of iron. In experimental animals the
absence of this defense mechanism can result in
sepsis and death. Infection-â•‰associated anemia is a
common clinical finding, particularly with chronic
infection, leading to the question of whether iron
supplementation in such situations is deleterious
or beneficial. For example, iron deficiency appears
to be protective against severe malaria in children
(Gwamaka et al. 2012), and untargeted iron supplementation of children in an area with a high risk of
malaria has been shown to increase hospitalization
and mortality (Sazawal et al. 2006). Nevertheless,
current WHO recommendations favor iron supplementation of children in areas endemic for malaria,
in conjunction with measures to prevent, diagnose,
and treat the disease.
Protection against pathogens can also result from
direct action of the commensal microflora. The
mechanisms involved include colonization resistance, which is the exclusion of pathogens by the
commensals in competition for nutrients, and niche
manipulation, for example by altering the local pH,
as well as secretion of antimicrobial peptides and
production of metabolites such as short chain fatty
acids that can alter the virulence profile of invading
pathogens. In this way, the innate immune system

of the host has evolved barrier and defensive functions to control the growth of commensal organisms
but can also co-â•‰opt these organisms to control invasive pathogens.

10.7.2â•‡ Adaptive Immunity
The adaptive immune system is the key weapon
in the vertebrate war against microbial infection.
The adaptive immune system differs from the
innate immune system in two general ways. First,
it is able to respond to any antigen that is perceived
by the individual as non-â•‰self rather than to a non-â•‰
specific pattern signaling the presence of a pathogen. Second, although the initial response of the
adaptive immune system to a new challenge is
relatively weak, it has a “memory” that ensures
that subsequent encounters with the same antigen
cause faster and stronger responses than did the
first challenge. It is this second feature that gives the
adaptive immune system its name. The adaptive
immune system provides vertebrates with a powerful defense against microbial infection and immune
surveillance against aberrant clones of its own cells,
but its exquisite specificity has two clinically relevant consequences: susceptibility to autoimmune
and atopic disease, and rejection of transplanted
tissue.
The adaptive immune system has evolved as a
highly complex and multilayered mechanism, the
details of which are beyond the scope of this chapter. Briefly, however, it comprises two arms: cell-â•‰
mediated immunity and humoral immunity. As
its name implies, cell-â•‰mediated immunity involves
direct counter-â•‰
attack by the effector cells of the
immune system, mostly T (for “thymus”) cells and
natural killer cells. Cell-â•‰mediated immunity has its
major role in defense against intracellular pathogens
such as viruses and in immune surveillance against
host tumor cells. Humoral or antibody-â•‰mediated
immunity involves the action of circulating immunoglobulins, made by B (for “bone marrow”) cells,
which bind to their target antigens on pathogens
and inactivate them directly or mark them for
destruction by other components of the immune
system. Humoral immunity generally functions
against extracellular pathogens such as bacteria.
As well as acting as effector cells, some subtypes of


T cell also initiate and promote (T helper or T\textsubscript{h} cells) and regulate (T\textsubscript{reg} cells) immune reactions.

The ability to discriminate between self and non-self is critical to the function of the adaptive immune system, and defects in this process of immunological tolerance underlie autoimmune diseases (Romagnani 2006). The fundamental basis of recognition of self or non-self macromolecules by the adaptive immune system is the display of fragments of those molecules on the cell surface bound to proteins of the major histocompatibility complex (MHC). There are two classes of MHC molecule: nearly all cells of the body express class I molecules, which display fragments of the normal proteins of the cell, allowing the immune system to detect whether the cell’s protein synthetic machinery has been subverted by viral infection or tumorigenesis. Conversely, class II molecules are only expressed by specialized antigen-presenting cells, which phagocytose and degrade micro-organisms and display protein fragments from foreign antigens. Recognition by a T\textsubscript{h} cell of a displayed peptide as non-self triggers an immune response: in general, recognition of non-self in the context of MHC class I causes a cell-mediated response that kills the subverted cell, whereas recognition in the context of MHC class II causes both a cell-mediated and an antibody-mediated response.

The multi-subunit MHC molecules are coded for by genes on chromosome 6, each of which is extraordinarily diverse (up to several hundred alleles), much of which is expressed at the peptide-binding site of the molecule. This combination of allelic variation and subunit pairing means that the resulting MHC complexes are capable of binding a huge variety of different peptides. Such genetic diversity is best explained by balancing selection, driven by the different ability of particular MHC alleles to protect against particular pathogens. Even in a small population there will be many different combinations of MHC alleles that protect against the multiple pathogens in the environment and thwart the evolution of new pathogen epitopes (see Box 10.4).

In humans, the MHC system is also known as the human leukocyte antigen (HLA) system. The HLA system is the predominant determinant of tissue

**Box 10.4 The MHC: More Than Just Immunology**

While the MHC comprises an integral part of the body’s immune defense system, it also serves another function which, superficially, appears unrelated: it contributes to an individual’s “signature” scent, or body odor (Overath et al. 2014). Studies on animals ranging from fish to humans have revealed that the scents emitted differ from individual to individual and are highly dependent on MHC genotype. In mice, scents originate from specific combinations of volatile metabolites present in urine, and the detection of olfactory differences enables mice to recognize each other. Less is known about the nature of body scents emitted by humans, but it seems that the ability to discriminate between scents plays a role in sexual attraction, and therefore mate selection (Havlíček and Roberts 2009).

How does this link with the immunological function of the MHC? A Swiss group performed an intriguing experiment in which men were asked to wear the same T-shirt for two consecutive nights and forego scented toiletries. The sweat-stained T-shirts were collected, and women were then asked to rate the “pleasantness” of the odors. The researchers found that the women tended to prefer odors of men who were MHC-dissimilar to themselves. These results could be explained by mate preference being driven by the need to ensure that offspring have increased MHC heterozygosity, as this may be vital in disease resistance. Also, selecting a genetically dissimilar mate could be a means of reducing the chances of inbreeding (Wedekind and Penn 2000).

These notions are supported by studies on the Hutterite religious clan, a closed self-sufficient group whose members only marry within the clan. Even though the gene pool is more limited in this community, couplings nevertheless tended to be between MHC-dissimilar members, and the couples that did share MHC similarity were more likely to have higher rates of miscarriage and difficulty conceiving.

The exact evolutionary reasons for the involvement of MHC genotype in mating preferences in humans and other animals are still somewhat speculative. Nevertheless the association is consistently strong and elucidation of the basis by which the MHC modulates odor production could provide further clues on the nature of human chemical communication.
compatibility after transplantation, and the wide diversity of HLA alleles makes tissue rejection likely unless careful matching is performed. Another situation where tolerance between HLA-incompatible individuals is critical is at the maternal–fetal interface. The fetal trophoblast cells that contact maternal tissues during pregnancy do not express classical MHC class I or II molecules on their surface, and instead express a set of minor HLA genes with restricted allelic diversity and immunosuppressive properties towards maternal cells.

Two facets of the adaptive immune system use selective processes at a cellular level to ensure immunological self-tolerance and strong antibody affinity. First, recognition of self antigens is critical for immunological tolerance and the avoidance of autoimmune disease. The “education” of the immune system to distinguish self from non-self is a selective process that takes place in the thymus during early life. Immature T cells, each of which carries a T-cell receptor on its surface that binds to just one of the estimated 10 million different types of epitope recognized by the human immune system, are normally programmed to undergo apoptosis and die shortly after they are formed. In the thymus, they are exposed to the full range of self antigens, and those that fail to react are “rescued” from the apoptotic process and released into the blood. This “fail-safe” mechanism of negative selection ensures that the organism develops with a set of T cells that only respond to foreign antigens.

The second process is the clonal expansion of B cells. During their development, each B cell is programmed to produce a single one from the wide range of possible immunoglobulins, and this is displayed on its surface as part of the B-cell receptor. Once released into the circulation, if it encounters that epitope it is stimulated to proliferate, leading to the production of millions of identical B cells that secrete immunoglobulins with the correct specificity. In the course of this clonal expansion, further fine-tuning of immunoglobulin specificity occurs by a process of somatic hypermutation, which promotes erroneous DNA repair and increases the mutation rate in the variable region of the immunoglobulin gene by about a million-fold. The resulting variations in immunoglobulin amino acid sequence are random, but only those cells with the highest affinity for antigen are selected to survive and proliferate. The end results of this Darwinian process of variation and selection, which can be seen as a form of somatic evolution, are two-fold: plasma cells that secrete large amounts of immunoglobulin of high specificity, and the development of memory B cells containing the relevant DNA sequence that persist for years in lymph nodes and are activated rapidly on new exposure to the antigen. It is the existence of this immunological memory that forms the basis of vaccination (Section 10.9).

The adaptive immune system has evolved in the presence of the associated commensal microflora which, because of its extent and diversity, must represent a major antigenic challenge to the host. How is it that the adaptive immune system does not respond to this challenge, and how does it filter out signals of pathogen intrusion from the noise of the millions of commensally derived antigens? To do so, the commensal antigenic burden must either be minimized or tolerated. The first is achieved by limiting exposure of host tissue to commensals using the barrier defenses described in Section 10.7.1, while the second involves development of tolerance to the commensal microflora so that the host immune system sees it as “self” rather than “non-self” (Lee and Mazmanian 2010). Development of this tolerance occurs early in life, facilitated by the immature immune system of the neonate which is oriented towards immunoregulation—discovering “self” and establishing tolerance towards it—rather than to mounting effective immune and inflammatory responses. This early life deficit in protective immunity underlines the importance for neonatal health of the passive immunity transferred in breast milk (Box 10.5).

The importance of early exposure to the gut microflora for development of the adaptive immune system is shown by the impaired phenotype of animals reared under sterile (“germ-free”) conditions. Such animals show deficiencies in the development of their intestinal mucosal barrier and immune system tissues, both in the gut and systemically. They also have defects in immunoregulation, as indicated by the properties of their T-h and T-reg cell populations, and are more susceptible to pathogen infection. Experimentally, replacement of the microflora with a single species of gut micro-organism,
Box 10.5 Transgenerational Passage of Acquired Immune Characteristics

After birth the human neonate must defend itself against micro-organisms, but there is a daunting array of them. The lactating mother can help. Of the five classes of human antibodies, IgA dominates in breast milk. IgA antibodies are able to convey a considerable degree of passive immunity to the baby, in a system with a high degree of efficiency. The antibodies are produced only by the mother, in response to organisms that she encounters and that are likely to be highly relevant to her baby too, but they do not harm the developing microflora of the infant’s gut which are needed for digestion.

The first milk, colostrum, produced by the mother also contains interferon, which non-specifically inhibits viral growth. Oligosaccharides, which are found in human breast milk in particularly large quantities and variety, may act as metabolic substrates to encourage the growth of beneficial commensals (Zivkovic et al. 2011). For this reason, oligosaccharides such as galacto- and long-chain fructo-oligosaccharides are under trial for addition to infant formula used for bottle-feeding to impart a variety of beneficial effects, including improved immune function.

Other factors in milk include lactoferrin which interferes with bacterial iron metabolism, bifidus factor which promotes the growth of the beneficial commensal 

\textit{Lactobacillus bifidus}, and fibronectin which makes macrophages more aggressive. In addition, breast milk contains both macrophages and antibody-producing leukocytes, their activities again tuned by the exposure of the mother. They can remain active in the infant’s gut for several weeks, allowing time for the infant’s immune responses to mature.

\textit{Bacteroides fragilis}, can ameliorate these immunoregulatory defects. This effect can be mimicked by a single secretory product (polysaccharide A or PSA) from the bacterium, suggesting that factors secreted from the gut microflora can direct the maturation of the host immune system (Ivanov and Honda 2012). In Section 10.7.5 we describe how the interplay of the adaptive immune system and the evolutionary novelty of altered microbial exposure may change the pattern of autoimmune disease.

10.7.3 How Pathogens Evade Host Defenses

Pathogens have evolved mechanisms to evade and neutralize the defenses mounted by the innate and adaptive immune systems, buying themselves enough time to survive, proliferate, and be transmitted to the next host. Host barriers against infection can be overcome by the secretion of toxins that inhibit ciliary action or the production of a protective biofilm. Many bacteria shield themselves from the PAMP recognition system by secreting masking molecules on their cell surface; the intracellular pathogen 

\textit{Mycobacterium tuberculosis} uses lipids both to mask its cell surface PAMPs and to attract benign (to it) macrophages in the lower respiratory tract that engulf it and transport it into lung tissues (Cambier et al. 2014). The trick of hiding inside host cells is used by many pathogens, from the retroviruses such as HIV that hide within the genome to the malarial parasites such as \textit{Plasmodium} that live within red blood cells.

Other pathogens hide from the adaptive immune system by constantly changing their cell surface antigens, so that the host’s immune response always lags behind the pathogen’s antigenic appearance. One example concerns the contingency loci found in several genera of bacterial pathogens. These loci, usually associated with gene products that interact in some way with the host, contain short sequence repeats that are liable to mispair, resulting in frameshift mutations in the resulting cell surface proteins and therefore high phenotypic diversity that enables the pathogen to evade host responses. A somewhat similar process is used by the trypanosomes that cause sleeping sickness to generate the variant surface glycoproteins (VSGs) that act as decoys for the host immune response, although here the variants arise from an archive of pre-existing VSG genes rather than from \textit{de novo} mutations. A final example is provided by the extreme diversity in the envelope proteins and drug-target molecules of HIV, originating from mutations promoted by the low fidelity of the virus’s reverse transcriptase.

Some evasion mechanisms directly target the adaptive immune system. Viruses such as herpesvirus, which cause persistent lifelong infections,
inhibit the host antigen presentation machinery that would normally alert the immune system to their presence within cells (Zuo and Rowe 2012). Other pathogens, including trypanosomes, hepatitis C virus, and HIV, can subvert B-cell-mediated immune responses (Nothelfer et al. 2015). Some pathogens interfere with the cytokine signaling required for an effective immune response; for example, interferon is a cytokine secreted in response to viral infection, and many viruses, including papillomavirus and measles virus, are able to inhibit the production or action of interferon (Devasthanam 2014).

### 10.7.4 Over-reaction of Host Defenses Can Cause Morbidity and Mortality

The interplay between host defenses and pathogen biology represents the sum of two evolved trade-offs. The pathogen must balance how much it damages its host by its virulence and evasion mechanisms against the need for the host to be able to transmit the infection, and the host must calibrate its defensive response to the severity of the threat while avoiding collateral damage to its own tissues.

Nevertheless, exaggerated host responses can contribute to morbidity and mortality. The case fatality rate in the 1918 Spanish influenza pandemic was particularly high because the virus elicited an aberrant host immune response that caused extensive tissue damage in the lungs, which may explain its unusual pattern of age-specific mortality (Figure 10.1; Morens and Fauci 2007).

Another example of an exaggerated immune response is that to staphylococcal enterotoxins. These secretion products of *Staphylococcus aureus* cause mild gastrointestinal symptoms, but when introduced systemically act as “superantigens,” causing widespread activation of the immune system and massive release of cytokines (a “cytokine storm”). The resulting toxic shock syndrome involves vascular leakage, hypotension, and disseminated intravascular coagulation, leading to multiorgan failure (Krakauer 2013). Such a dramatic outcome can be envisaged as an unintentional consequence of the evolved virulence of the bacterium, which exists as a human commensal and opportunistic pathogen that has little evolutionary interest in the rapid death of its host.

![Figure 10.1](image-url) **Figure 10.1** Age-specific mortality rates in the influenza epidemics of 1911–17 and the so-called Spanish flu epidemic of 1918. The high virulence of the 1918 virus may have resulted from an aberrant host response. From Taubenberger and Morens (2006), with permission.
To the patient, host defenses like fever and cough can appear as debilitating over-reactions to relatively mild infections, such as with the common cold virus. In this situation, symptomatic treatment with antipyretics and antitussives makes the patient feel better and is generally unlikely to have effects on the eventual outcome of the infection. But, as we discuss in Section 13.6.1, suppressing these evolved defensive responses may not always be a good idea. For example, there is some evidence that suppressing fever in critically ill patients increases mortality (Schulman et al. 2005).

10.7.5 Autoimmune and Allergic Disorders

The autoimmune, inflammatory, and allergic disorders have their proximal cause in the dysregulation of the immune system. Diseases in which the immune system attacks the body’s own cells, as if mistaking them for non-self pathogens, are termed autoimmune diseases. Examples include type 1 diabetes and multiple sclerosis. Other conditions such as Crohn’s disease and ulcerative colitis result from inappropriate activation of the inflammatory response. Allergies such as allergic rhinitis ("hay fever") are caused by hypersensitivity of the immune system to environmental antigens that are inappropriately perceived as harmful.

In this book we have seen repeatedly that environmental change producing evolutionary novelty can pose a threat to human health. In the context of this chapter, we can contrast the fall in morbidity from infectious diseases seen in high-income countries over the past 100 years or so with the dramatic increase in the prevalence of autoimmune, inflammatory, and allergic disorders (Figure 10.2). What could explain this apparent negative association?

There are both genetic and environmental determinants of susceptibility to disorders caused by immune dysregulation. The genetic determinants are often associated with particular alleles of the HLA system. For example, HLA-DR2 is associated with systemic lupus erythematosus (more common in females) and ankylosing spondylitis is associated with HLA-B27 (more common in males). Ankylosing spondylitis develops gradually during adolescence and young adulthood, causing chronic inflammation and structural changes in joints, especially the spine. The rapid change in the prevalence of immune-related disorders (Figure 10.2) makes any genetic explanation unlikely. For example, not all HLA-B27 males develop ankylosing spondylitis, so there must be some other environmental trigger. There is evidence to suggest that the trigger in some cases is a response to Klebsiella commensals which inhabit the gut, perhaps itself triggered by an acute infection, since there are cross-reacting epitopes.
between *Klebsiella* and human cell-surface antigens (Husby et al. 1989).

There are other subtleties in the environmental associations of immunoregulatory disorders that may provide clues to the underlying mechanisms. For example, there is a lower prevalence of asthma and atopic disorders in children in rural compared with urban communities. Inflammatory bowel disease is common in highly sanitized, industrialized areas of the world, but uncommon in rural areas where living quarters are crowded and unhygienic. Breastfeeding in infants has a protective effect in reducing the later risk of asthma, and it is known that breast- and formula-fed infants are exposed to different types of bacterial flora.

Earlier in this chapter (Section 10.7.2) we described how the vertebrate adaptive immune system has coevolved with both commensal and potentially pathogenic micro-organisms, and that appropriate exposure to microbes early in life is essential for establishing immunoregulatory pathways. For instance, germ-free mice experience abnormal development of their immune system, which can be corrected by re-introduction of intestinal commensal bacteria.

A unifying mechanism to explain all these observations, termed the *hygiene hypothesis*, proposes that lack of early exposure to the full range of microbes leads to inappropriate activation of immune responses which are then manifest as allergic or autoimmune disease. In other words, such deficient exposure has changed the set-point of the trade-off between efficient pathogen defense and susceptibility to autoimmunity (Bergstrom and Antia 2006). Clinical trials involving deliberate helminth infestation of patients who have inflammatory bowel disease (the pig whipworm, *Trichuris suis*, was used since it is not pathogenic in humans) have indicated a reduction in disease activity in patients with ulcerative colitis and Crohn’s disease (Heylen et al. 2014). If such trials are successful and a clear link is established between parasite-driven immunoregulation and immune disorders, then new possibilities for treatment will become available. Conversely, efforts to clear helminth infections from low-income countries may need to take account of a possible rise in the level of allergies and inflammation (Wammes et al. 2014).

**10.8 Public Health Measures**

In addition to our immune systems, the evolved extended human phenotype now includes technological defenses against microbial infection. The three arms of these defenses are public health innovations, vaccination (Section 10.9), and antibiotics (Section 10.10).

Improved nutrition and sanitation underpin much of the fall in morbidity and mortality from infectious disease over the past 150 years. The interaction between poor nutrition and infection is particularly important in low- and middle-income countries, and especially in children. An early effect of malnutrition is the suppression of immune
function. For example, in a population of subsistence farmers in the Gambia with highly seasonal food availability, people born in the “hungry season” showed ten-fold higher infection-related morality as young adults and decreased life expectancy, suggesting that immune function may be compromised by events early in life. In this population, thymus development was sensitive to early life nutrition, with smaller thymuses observed in those born in the hungry season. Furthermore, size at birth was positively correlated with antibody responses to vaccination (Moore et al. 2006). Such observations may represent the results of trade-offs in which the body prioritizes resource use and favors immediate survival over investment in long-term defense mechanisms (see Chapter 5).

The importance of uncontaminated water supplies and efficient removal of sewage in limiting the oro-fecal transmission of disease have been known since Dr. John Snow removed the handle of a public pump in London in 1854 to curb an outbreak of cholera. Because waterborne diseases such as cholera and typhoid are easily transmitted to multiple new hosts, and retain their virulence because transmission is not dependent on continued host viability, improvement of sanitation is one of the most effective measures to combat infectious disease.

10.9 Vaccination

Vaccination takes advantage of the ability of the adaptive immune system to retain a memory of previous antigen challenges, so that future exposure to the antigen results in prompt and vigorous protective responses (immunization).

Vaccination has eradicated one human disease (smallpox), is close to eradicating another (poliomyelitis), and in high-income countries controls several acute childhood infections, such as rubella and measles (Figure 10.3). It is a characteristic of these diseases that natural infection with these agents, if not fatal, generally provides lifelong protection against a second infection since the causative viruses show little or no antigenic drift (mutation of the sites to which protective antibodies bind) in the wild; vaccine-induced immunity against these pathogens similarly results in lifelong protection. However, recent reports of a vaccine-resistant strain of poliovirus should be a reminder that antigenic stability should not be taken for granted (Drexler et al. 2014).

In contrast, pathogens that are able to evolve their antigenic properties in the wild can cause repeated infections. Influenza virus shows both antigenic drift and antigenic shift (where virus subtypes

![Figure 10.3](image-url) Estimated uptake of measles vaccination and associated decline in the global incidence of measles, 1980–2011. From World Health Organization (2012), with permission.
combine to produce a new strain), which means that the immune response to one strain is ineffective against a new strain. For such diseases, public health strategies require continual surveillance of circulating viruses and the regular administration of new vaccines formulated against their predicted antigen profile.

Where a pathogen is pathogenic by way of a toxin, immunization against the toxin (which is likely to be less variable) may address the pathogenic process while leaving the host immune system to eradicate the actual infection. This is the basis of protection against tetanus.

There are also several examples of existing human and veterinary vaccines where evolution of the targeted organism in the host has been demonstrated in response to the selection pressure of the vaccine, and in the case of one veterinary disease (Marek’s disease, a virus-induced neoplasm in poultry) this has led to large-scale failure of the vaccine. How might a pathogen respond to a vaccine? First, it might evolve by selection against the epitopes recognized by the vaccine. Such epitope shifting, potentially leading to loss of effectiveness of the vaccine, has been observed for several human viral and bacterial vaccines including those for hepatitis B, pertussis, and pneumococcal disease. For example, a resurgence of whooping cough infections in high-income countries has been linked to strains of *Bordetella pertussis* that have evolved to no longer express pertactin, a virulence factor that is an antigen included in modern acellular vaccines (Lam et al. 2014).

Secondly, the pathogen might evolve to change its virulence, a situation that particularly applies to so-called “imperfect” vaccines that are not sterilizing (i.e., preventing all infection) but are rather partially effective in preventing transmission or reduce the severity of any infection that does occur. Recall from the discussion of virulence in Section 10.6 that a pathogen will evolve to a particular level of the virulence/transmission trade-off that optimizes its reproductive success. If vaccination acts to increase the survival of an infected host, giving it more time to transmit the pathogen to the next host, then the set point of the trade-off will move towards increased pathogen virulence (Gandon et al. 2003; Mackinnon et al. 2008). Although vaccinated hosts will not be affected, unprotected hosts who do not benefit from the survival advantage conferred by vaccination will experience more severe disease. At a population level, pathogen evolution of this kind may negate the benefit of a vaccination program by increasing the cost of managing the disease in unprotected individuals. The failure of the Marek’s disease vaccine in chickens was caused by vaccine-driven increases in viral virulence.

For some infections, vaccination of an entire population is not necessarily essential to prevent outbreaks of disease. As long as a sufficiently high proportion of the population is vaccinated, any outbreak will be contained because it is likely that the potential hosts exposed to the pathogen will already be immune and unable to develop and transmit the disease. The vaccination coverage necessary to produce this “herd immunity” (or, more elegantly, “social immunity”) varies from disease to disease depending on transmissibility; it is estimated to be 90–95% for a highly transmissible disease such as measles. One advantage of herd immunity is that members of the population who cannot be vaccinated, for example frail or immunocompromised individuals, also benefit from vaccination programs. On the other hand, “freeloaders” who take advantage of the protection conferred by community vaccination without exposing themselves to the (very small) risk involved in receiving vaccines will also benefit. This, together with misplaced antivaccine attitudes, has led to vaccination coverage in some areas of high-income countries declining below that necessary for herd immunity, allowing outbreaks of previously controlled childhood diseases (Fine et al. 2011).

Current vaccine development is aimed at diseases where host immunity fails to control the pathogen and initial infection becomes chronic; such diseases include HIV and malaria. For these pathogens, the important issue is the agility of their epitope composition that presents an ever-shifting target for the host immune system and for the vaccine developer. The genetic diversity of HIV is a function of its highly error-prone reverse transcriptase (Smyth et al. 2012). The malaria parasite uses antigenic variation and polymorphism, as well as active immune evasion strategies, to slow and misdirect the immune response of the host; these properties are
likely to have coevolved with the human immune system (Pierce and Miller 2009). The failure, despite decades of effort, to develop vaccines against these diseases suggests that knowledge of the coevolutionary relationships of the human immune system might provide us with new therapeutic approaches.

10.10 Antibiotics

Antimicrobial chemotherapy kills or inhibits the growth of micro-organisms, and is used to prevent or treat infection. In this section we particularly discuss antibiotics, which are substances used against bacteria; antivirals and antifungals are also mentioned briefly.

There is a wide variety of mechanisms of action of antibiotics. Molecular targets include the bacterial cell wall (e.g., penicillin and its derivatives), DNA replication (quinolones), RNA synthesis (rifamycins), protein synthesis (macrolides and aminoglycosides), and metabolism (sulfonamides). A characteristic of these targets is that the affected pathway does not exist or is only weakly affected in eukaryotes, ensuring that the toxic effect is confined to the prokaryotic organism.

Most antibiotics are derivatives of naturally occurring compounds that have evolved in micro-organisms as defenses against other micro-organisms; the classic example is the production of penicillin by the mold Penicillium. The evolution of these defensive chemicals places selective pressure on the targeted micro-organism; consequently, all microbial defense mechanisms, and their countermeasures, will have been tested by millions of years of coevolution. This implies that resistance mechanisms against most if not all naturally derived antibiotic classes will already have evolved. Samples of bacteria from a cave thought to have been isolated for over 4 million years were found to be resistant to multiple classes of modern antibiotics (Bhullar et al. 2012). The mechanism underlying resistance generally involves denying the antibiotic access to its target site. For example, this could be by the acquisition of degradative enzymes such as beta-lactamas or of efflux pumps that remove the antibiotic from the bacterial cell. Alternatively, the sensitivity of the target site to the antibiotic may be decreased, for example by point mutations in the active sites of the DNA-replicating enzymes inhibited by the quinolone antibiotics such as ciprofloxacin.

A particular feature of antibiotic resistance, with implications for the therapeutic use of these compounds, is the ease with which pathogens can acquire the trait. Resistance can evolve by de novo mutations as a result of the selective pressure of exposure to the antibiotic, because the large population size and short generation time of bacteria mean that resistance-conferring mutations will frequently occur and spread even within the human host. Additionally, resistance can be acquired by horizontal gene transfer, not only from members of the same bacterial species but also from phylogenetically distant prokaryotes. Together, these processes mean that clinically relevant levels of resistance usually appear within 2–4 years of the introduction of a new class of antibiotic (Figure 10.4; Hawkey 2008). However, the innovation pipeline for new antibiotics is narrow, with fewer new drugs being approved each year (Figure 10.5).

Hospital-associated (nosocomial) infection with resistant bacteria is a particular threat. This arises because of the high rates of antibiotic use within the hospital setting, generating strong selective pressure favoring resistant strains and clearing sensitive ones, making colonization of patients and staff by resistant strains more likely. In addition, the high turnover of patients in a hospital imposes selective pressure for a resistant strain to transmit rapidly so as to remain endemic within the hospital. In 2013, the US Centers for Disease Control and Prevention reported that each year in the USA at least 2 million people acquire serious infections with antibiotic-resistant bacteria, and at least 23,000 people die each year as a direct result (Centers for Disease Control and Prevention 2013). The financial impact of resistance may be as high as US$20 billion in excess direct healthcare costs, with additional costs to society for lost productivity of as much as US$35 billion a year. Antibiotic resistance clearly presents a considerable problem, and there is a fear that bacterial strains with multiple resistance to all known antibiotics could arise, with disastrous consequences. However, we must remember that major progress in reducing deaths from infectious disease occurred in the pre-antibiotic era as a result of advances in vaccination, nutrition, and sanitation, and so the
Antibiotic resistance is another consequence of human manipulation of the environment. Over-use through unnecessary prescribing is a major cause, but antibiotics are also added to animal feedstuffs to promote growth. Measures to prevent over-use include reduction of the use of antibiotics in animal feed, as well as education of prescribers and patients to promote “antibiotic stewardship,” using protocols that ensure antibiotics are used only when necessary and that the right antibiotics are prescribed and administered in the right way.

There has been interest in evolutionary-based treatment protocols to reduce the development of antibiotic resistance in hospitals. These include antibiotic cycling (where use of different classes of antibiotics is alternated across the whole hospital on a regular schedule) and antibiotic mixing (where individual patients receive one of several antibiotic classes used simultaneously in the hospital). The rationale behind these approaches is that if strains resistant to one antibiotic evolve, they will be susceptible to the new agent (in the case of cycling) or will not spread to nearby patients (in the case of mixing). Although there have been many theoretical studies of such protocols (e.g., Bergstrom et al. 2004), it is only recently that trials have begun to test their efficacy (van Duijn and Bonten 2014).

Another potential approach is to adjust antibiotic administration schedules to ensure a better balance between curing the infection and preventing the evolution of resistance (Kouyos et al. 2014). Traditionally, aggressive therapy with high dosages for long durations has been used to achieve maximum bacterial kill. However, clinical cure can be achieved by use of moderate dosages and durations if the treatment buys time for the patient’s immune response to kill the pathogen or develop tolerance to it (Råberg et al. 2009). In this situation, lower dosages of antibiotics could prevent the evolution of resistance, first by allowing the survival of susceptible strains to compete with resistant strains, and second by reducing the emergence of resistance in “bystander” non-pathogenic commensals and subsequent horizontal gene transfer into the pathogen.
Another approach to antimicrobial therapy targeting secreted bacterial “public goods” is described in Box 10.6.

Antivirals and antifungals also act against the survival or replication of the infecting microorganism, but the range of molecular targets of these agents is potentially limited by the characteristics of the pathogen. Viruses often use part of the host’s biochemistry for their replication, ruling out targeting those pathways. Consequently, antivirals target other virus-specific features, as exemplified by the reverse transcriptase inhibitors used to inhibit replication of retroviruses such as HIV (e.g., nevirapine). Another approach is to inhibit the release of viruses from infected cells; the HIV protease inhibitors (such as ritonavir) and the influenza virus neuraminidase inhibitors (such as zanamivir) work in this way.

Development of resistance to antiviral drugs is the result of the high mutability of viruses in the face of the selection pressures caused by the antiviral treatment. An extreme example is HIV, in which the high mutation rate conferred by an error-prone replicative mechanism leads to rapid evolution of resistance to antiretrovirals, even in individual patients. Consequently, antiretroviral therapy in patients with HIV infection is based on combination therapy to delay development of resistance, along with (in high-income countries at least) monitoring of viral load and resistance testing to guide treatment changes.

By analogy with the situation of “imperfect” vaccines described in Section 10.9, evolutionary theory predicts that intensive treatment of HIV infection in a population will increase its virulence. Transmission will be favored by rapid progression to high viral load, and this higher virulence that might lead to death of the host before transmission can be offset by treatment. Conversely, in an untreated population virulence should decrease, because lower viral load and the consequent greater host longevity favor transmission. This prediction appears to be confirmed by reports of opposing trends in HIV virulence in Europe (Pantazis et al. 2014) and Africa (Payne et al. 2014).

Fungi, like their hosts, are eukaryotes, meaning that molecular targets unique to the prokaryote domain are not available. Most antifungals target the fungal cell membrane, which differs in lipid composition from animal cell membranes; examples

---

**Figure 10.5** Timeline of antibiotic approvals. The graph shows the number of antibacterial New Drug Application (NDA) approvals in the USA over the past three decades. Intervals from 1980–2009 are 5-year intervals; 2010–12 is a 3-year interval. Drugs are limited to systemic agents. From Centers for Disease Control and Prevention (2013).
Key Points

- Many causes of extrinsic mortality are biotic, including competition for nutrient supplies with other organisms of the same or different species, predation, and infection by micro-organisms.
- Humans have evolved with their pathogens, parasites, and commensals. This process of coevolution has shaped aspects of human physiology as well as the biology of the microbiota.
- The role of the human gut microbiome in human physiology is increasingly

Box 10.6 Bacterial Resource Acquisition as a Therapeutic Target

The idea that bacteria can communicate and cooperate is relatively recent, but it is now known that they can secrete a number of molecular species to mediate such social interactions (West et al. 2007). Communication by means of quorum sensing ensures that bacteria can signal their presence to each other and coordinate their behavior. For example, production of virulence factors such as toxins and biofilms is metabolically costly, and is futile if there are only a few bacteria in the local environment. Consequently, bacteria secrete density-signaling molecules such as small peptides, and coordinated action such as expression of virulence is dependent on detection of a threshold level of signal.

Bacterial cooperation involves the production of public goods, a term taken from economics to mean items from which individuals can freely benefit and from which no-one can be excluded. In our context, public goods refer to extracellular molecules secreted by bacteria that are available for use by other bacteria. These can include components of the biofilms that allow bacteria to form colonies on biological surfaces, as well as molecules that help bacteria extract resources from the environment. A well-known example of the latter is the iron-binding siderophores (see Section 10.7.1).

As with many social interactions in humans, production of public goods by bacteria provides opportunities for strains of “cheaters” who do not themselves produce public goods but take advantage of goods produced by other strains. Remarkably, some bacteria have evolved mechanisms to “punish” cheaters by secreting toxins along with their public goods (Wang et al. 2015).

Targeting secreted public goods may represent an innovative approach to antimicrobial chemotherapy that circumvents some of the ways in which bacteria can evolve resistance to conventional antibiotics. Inhibiting an extracellular product means that bacteria cannot exclude access to the site of action, and mutations that increase the amount of secreted product will be metabolically costly for the resistant strain while also benefitting susceptible strains, leading to selection against the mutant. Such approaches have been investigated by simulation (Pepper 2012) and experimentally in a model of bacterial infection of an insect (Ross-Gillespie et al. 2014).

Humans rely on barriers, their adaptive immune system, and now their technology to deal with the threat of pathogens; pathogens rely on rapid evolution to maintain their own viability against these defenses.

Adaptive immunity is our core system for dealing with microbial evolution, and part of this involves somatic evolution at the cellular and molecular level. An equilibrium is never reached in this coevolutionary process, and we continue to have epidemics of infection. Infectious diseases remain major threats to public health in both high and low- and middle-income countries.

10.11 Conclusion

The history of the human species is one of coevolution with microbes and parasites. Humans and their symbionts have been locked in an evolutionary arms race, albeit with different sets of weapons. Even so, most members of the community of symbiotic micro-organisms that lives in association with each of us are harmless, and some are beneficial.
understood; it contributes to defense against pathogens, metabolic health, and regulation of the immune system.

- The adaptive immune system of vertebrates is a key component of protection against microbial infection. Its specificity is a consequence of Darwinian processes of variation and selection.

- Trade-offs between virulence and transmissibility influence the ways in which microbial infections progress in individuals and populations. The balance between these attributes of the pathogen can evolve and change rapidly under some circumstances, with important consequences for therapeutic strategies.

- Human technology in the form of public health measures, vaccination, and antibiotics has greatly reduced the threat from infectious disease in high-income countries. Infection is still a major cause of morbidity and mortality in low-income countries.

- The concerning development of antimicrobial resistance in pathogens can be understood in terms of an evolutionary arms race. This understanding can provide insights into how to reduce this risk by adjusting treatment regimens.
11.1 Introduction

It is a truism that humans can be distinguished from other species, at least in degree, by a large brain relative to body size. Our brain is characterized by a particularly well-developed neocortex (Chapter 6), and this feature has had a number of profound evolutionary consequences. Our capacity to communicate, to use and develop technology, and even the nature of the social structures we have evolved can be attributed to a large neocortex.

Humans are a social species. We evolved with characteristic behaviors adapted to living in groups with other members of our species. Yet our societal structure has undergone enormous changes in a few thousand years, from the small isolated clans of foraging societies to the complex organizations of cities with populations of millions. Increasingly, the social environment is changing in other ways. Family structure has changed, and communication is no longer necessarily verbal and face to face; indeed, telephone, radio, television, the internet, and social media are now all dominant forms of communication.

This chapter considers how human behavior has evolved and how its evolution has influenced behavioral morbidity, as reflected in a greater risk of some psychological and psychiatric disorders. These forms of disorder are now a large component of the current and anticipated burden of disease.

The social environment is a major part of the selective environment which has led to the evolution of our species. But at the same time, humans have evolved with a rapidly changing capacity to alter their social and societal environments. This capacity creates the potential for a mismatch between our evolved phenotype and the environment we now inhabit. This mismatch, in turn, is likely to be a potent source of psychological disorders.

11.2 Biological Determinants of Culture and Behavior

In everyday language, culture is usually understood as an amalgamation of knowledge, behavior, and tradition within a particular community or population. But creating a formal definition of “culture” has itself been contentious. There have been intense debates among social scientists over its precise definition, and some have argued that culture must be viewed as a purely human characteristic. Modern evolutionary scientists define culture as information acquired through social learning (see Box 2.6). In this sense cultural phenomena may be observed in other species, notably primates, cetaceans, and birds.

It is obvious that human culture evolves, and this cultural evolution is another form of an inheritance system with the potential for both vertical and horizontal transmission (see Section 2.3.3 and Chapter 6). Understanding the significance of the interplay between biological and cultural evolution is important. The evolution of the capacity to learn and the potential for learning to influence evolution are also important components of evolutionary science. But what has to be learned—and how it is learned—has changed dramatically and rapidly in the past 12,000 years, from experiential learning within the forager clan to the formalized and structured learning of modern developed societies.

Because of changing attitudes to what culture is and how it originates, evolutionary explanations of human behavior have had an especially contentious
history. This contention arose in part because of philosophical and political debate stemming from the ways in which various disciplines have interpreted human behavior. Some have wanted to view it as an entirely learned phenomenon, while others argue that human behavior is built on some strongly selected and therefore genetically determined components of brain function. Extreme positions have been taken, or at least interpreted as been taken, and at times the debate has been more polemical than critical. The media have not infrequently exaggerated scientific reasoning and observations into fatuous extrapolations of aspects of human behavior.  

We can reduce this discourse to the fundamental question of the extent to which human behavior is determined (or influenced) by our evolved biology. Several schools of thought (Box 11.1) have emerged to explain the evolution of human behavior, largely based on how each school conceived of culture, behavior, and learning. The debate was distorted by the strong genetic determinism of late-twentieth-century biology. The distinguished evolutionary biologist and zoologist E. O. Wilson, and subsequently many others including Richard Dawkins, put forward strongly deterministic arguments in which all aspects of behavior were essentially framed in genetic (i.e., evolved) terms (Wilson 1975; Box 11.1). Their critics, who included equally distinguished evolutionary biologists such as Stephen Jay Gould and Richard Lewontin, saw that this stance left little room for the role of active decision-making, learning, and cultural adaptation. At its most inappropriate extreme, some critics, particularly from the social sciences and humanities,
saw biological arguments being used to excuse all forms of antisocial behaviors from rape to murder, a position exploited by the media and which was a gross misuse of the actual evolutionary theories.

Social scientists take a very different view from biological scientists: they see human behavior as being produced by culture, learned from others. The basis of social science is that humans are cultural organisms, and in general social scientists perceive culture as a learned rather than innate phenomenon. They argue that humans are quantitatively and qualitatively different from other species in terms of the complexity of their culture. But culture itself is a product of evolution and, as explained earlier, is not a uniquely human characteristic. Learning can be demonstrated in many animals. For example, some aspects of foraging are clearly learned, feeding behaviors are culturally transmitted in some reptiles and birds, migratory patterns are learned in some avian species, and tool use in different groups of chimpanzees and New Caledonian crows is a learned or culturally transmitted behavior. Some cetaceans have a culturally transmitted whale song. Language provides a sophisticated capacity for communication, and we have developed varied and complex social structures. Humans are a prescient species, possessing consciousness; we use technology in sophisticated ways, and we have developed varied and complex social structures. Humans are a prescient species, possessing consciousness; we use technology in sophisticated ways, and we have developed belief systems manifesting in religion and superstition. But there is an important difference between the view that genes determine our behavior and the argument that our evolved brain is the substrate on which experience and the current environment shape our abilities and behavior.

A fair appraisal of the concepts would show that accepting that there is a genetic basis to human behavior does not mean that every aspect of human behavior is genetically determined. Rather, as we have made clear throughout this book for other systems, for neural function evolution has led to the selection of inherited neural traits that provide a substrate on which developmental and environmental influences (including the social components of the environment) can act to mold behavioral and neural phenotypes.

Evolutionary biologists have been particularly concerned with explaining how altruistic behavior and the interplay between the sexes have emerged. In this chapter we will extend the discussion to other aspects of human behavior and psychology. Each of the various schools of thought that have taken positions on the evolution of behavior can offer valuable insights, and this chapter will draw from them to describe what, in our view, is now a broadly held consensus on how evolutionary principles can, and should, be used to explain human behavior and psychology. In turn, evolutionary perspectives offer useful insights for understanding certain psychiatric disorders.

### 11.3 Evolution of the Human Brain and Behavior

In considering the evolution of human behavior it is useful to bear in mind the four questions suggested by the famous ethologist Nikolaas Tinbergen as a way of systematically understanding behavior (reviewed in Section 1.7; Tinbergen 1963). These are: what is the mechanism underpinning the behavior; what is the function of the behavior; how does it develop during the life course; and how does it evolve? Addressing each of these questions can help us to understand behavior from both proximate and ultimate perspectives.

The ratio of brain size to body size in primates is about an order of magnitude larger than that in other mammals. But even within the primate order, humans and our ancestor species have had brain sizes that are disproportionately large (Chapter 6). The investment in a large brain has major energetic considerations. The human brain consumes about 20% of the body’s total energy but makes up only about 2% of bodyweight; the proportion of energy consumed by the brain is considerably higher in infancy. So a fundamental evolutionary question is why did primates, and particularly hominins, evolve to invest so much of their energetic resources in brain growth and function? The question does not have a single answer.

A simple answer is that the sensory and processing capacities of the brain conferred adaptive advantages to the primate clade. For example, primates use senses such as color vision to a greater extent than other mammals in their search for ripe fruits. Evolutionary processes do not work on a single trait in isolation, but operate on multiple
interacting traits. Indeed there has been an interactive “ratchet” involving changing ecology, evolving social structure, better communication, and better use of technology, all of which have driven brain development.

As discussed in Chapter 6, many primate species found advantage in living in groups with a complex social structure. This social organization enabled them to protect themselves against predators and to defend food supply in larger territories. But living in a large group requires complex dynamic relationships between the individuals because of the social complexity that ensues. Neocortical volumes correlate strongly with group size in primates; in humans the neocortical volume is approximately 80% of total brain volume, whereas it is much less in some prosimians (Azevedo et al. 2009). Based on these relationships, Robin Dunbar proposed that during the course of their evolution humans lived in group sizes of about 150 individuals. Intriguingly, modern forager societies tend to be built around clans of this size. Early villages and those in modern agricultural societies (and even communes) tend to be of similar size. Sociological research shows that group sizes of about 150 are the maximum that can be sustained before complex hierarchies are required. It also suggests that while people in modern societies may have many more “acquaintances,” perhaps up to 2000 individuals, the circle of true friends, intimate acquaintances, and relatives averages about 150 people (Dunbar 2008; Table 6.2). Language appears to be the tool that allows humans to manage the complex interactions found in groups of this size. In contrast, chimpanzees and other primates live in group sizes of 50 or less, and other forms of interaction such as reciprocal grooming may play an important role in ensuring social cohesion. Likewise, cuddling remains an important element in maintaining relationships in subsets of human societies such as families.

Social complexity within a group requires the capacity of each member to interpret the intentionality of others within the group. This is not simply a matter of communication. It is a matter of interpreting intent and understanding not only the relationship between two individuals but all the potential relationships within a group. An analytical construct has been developed to describe how animals and humans interpret the mental state and intentionality of others within their species: this is termed the theory of mind (see Box 6.4). This construct provides an analytical basis for how communication and social interaction can move beyond simple alarm calls and herd behavior to the complexities of human society and culture.

It is generally accepted that most adult humans operate with about five levels of intentionality. Indeed, effective social discourse requires this level of interaction to avoid unnecessary misunderstandings and conflict. Clearly the higher the level of intentionality required, the more likely it is that errors of interaction will result. Many problems in interpersonal relationships, and even wars, have arisen as we employ these higher levels of intentionality.

Such advanced processing and engagement in social organization clearly required, and was expedited by, the development of language (Section 6.3.9). Once higher levels of intentionality had evolved, they provided the capacity for further components of human culture to develop. They also provided the basis for a high level of reciprocal altruism and detection of cheaters, which have become fundamental to the structure of human society. Higher levels of intentionality are required for the capacity to have prescient self-awareness (including that of death), to develop beliefs and superstitions (the forerunners of ritual and religion), to form political structures, and to use language to communicate via the complexities and beauty of literature.

The theory of mind develops over our life course. By the age of 4–5 years children can recognize a third order of intentionality; before that they cannot lie convincingly. The concept of a theory of the mind may be relevant to understanding autism and the related Asperger’s syndrome (Baron-Cohen et al. 1997). Those afflicted have a limited ability to interpret the intent of others, suggesting that third-order and higher levels of intentionality have not developed appropriately.

### 11.4 Evolution of Social Behavior

Humans are characterized by living in groups larger than a family and, as discussed previously, there is a compelling rationale for the view that we
evolved living in clans of about 100 to 150 individuals. Thus the social environment became a key part of our selective environment, and selection would have favored traits that promoted fitness within that environment.

However, we also evolved in parallel with our cultural repertoire. There has been a close link between our cultural evolution and our biological evolution, often referred to as gene-culture coevolution (see Sections 1.6 and 2.3.3).

The development of consistent use of tools by H. habilis is the first unequivocal evidence of culture in the hominin clade (see Section 6.3.8). Such tool-making capacity eventually evolved through learning and cultural evolution into the technological repertoire of modern society such as brain scanning, pharmaceutical development, nuclear weapons, and the internet.

There are many features of society which have undergone change, many of which are self-evident. Over the past 10,000 years, virtually all human groups have changed from being foragers living in small clans into pastoralists and city dwellers. These changes have been accompanied by rising exposure to infection and malnutrition. The causes and nature of trauma and conflict have also changed, with humans becoming their own main predators, particularly through war and religious and other persecutions. There have also been enormous changes in social structure. Organizational and thus power hierarchies became necessary, and individual skills became differentiated: a surgeon and a lawyer obviously have very different skills, and interpersonal interactions often now require higher levels of intentionality.

Culture evolves, and as it changes so does the selective environment in which an individual lives. The example in Chapter 1 of how lactase persistence coevolved with pastoralism demonstrates such coevolution. Conversely, had the gene for lactase persistence not mutated, milk could not have become a major food source for that population. It is not surprising that the outcome may be maladaptive, because culture can evolve at a different pace and in a different manner from biological function. For example, the evolution of societal structure into large aggregations of population with less structured clan support may conflict with our evolved capacity to manage best psychologically in smaller groups.

This mismatch may be the basis of some psychological disorders. In Chapter 5 we described how the young age of biological puberty is in conflict with the age at which we accept young people as adults in developed societies. Indeed, it may be that the change in societal complexity has affected the rate of neural maturation: there is evidence that some aspects of decision-making activity may not mature until after 20 years of age (Figure 11.1), and magnetic resonance imaging shows the prefrontal cortical structures that are involved in impulse control, strategizing, and judgment are not fully mature until after 25 years of age (Lebel et al. 2008; Figure 11.2). There is growing empirical evidence to show that this mismatch plays a role in teenage depression, acting-out behavior, drug abuse, and suicide. It has been found that boys who undergo earlier puberty, spending a longer period of their lives in a biologically mature but psychologically immature phase, are much more likely to be suicidal than those having a later puberty (Table 11.1).

This example highlights a challenge. Just as the introduction of an exotic species into a previously stable environment (e.g., rats, dogs, and humans in New Zealand, which was free of terrestrial mammals until the arrival of humans some 800 years ago) can drive species (flightless birds such as the moa) to extinction, or global warming can destroy frog habitats in the mountains of Costa Rica, rapid changes in our social environment can have impacts on human health.

The human neocortex largely evolved to deal with the challenges of the social environment. As biological and cultural evolution proceed in very different ways and paces, we can anticipate that the consequences of the inevitable mismatch between brain and environment will be reflected in disorders of behavior and mental health.

11.4.1 Social Competition and Altruism

One way that living with others influences our evolution is through social selection. In Chapter 2 we discussed sexual selection—that is differential fitness created either by competition between members of the same sex for the right to mate, or by choice of
mate by members of the opposite sex. Some authors have suggested that sexual selection processes are a subset of social selection (West-Eberhard 1979), the latter being a form of natural selection influenced by intraspecific social competition or choice, in sexual as well as non-sexual contexts. We have discussed the peacock’s tail as an example of a trait that evolved in the context of members of the opposite sex choosing their mate. But consider the example of American coot families, where offspring display highly ornamented and bright plumage to their (gray) parents (Lyon and Montgomerie 2012).

**Figure 11.1** Decision-making activity may not mature until about 20 years of age. Data from Cauffman and Steinberg (2000).

**Figure 11.2** Growing neurobiological evidence indicates that brain systems that play a key role in emotional and incentive-based behavior, such as the subcortical regions (top line), mature earlier than systems modulating cognitive and impulse control, such as the prefrontal cortex (bottom line). This regional discordance in structural and functional maturation is particularly marked around the period of adolescence, contributing to elevated risk of acting-out and risky behaviors during this phase of life. Adapted from Somerville et al. (2010), with permission.
Attracting the parents’ attention is crucial for provisioning food and so for offspring survival. We could say that sexual selection involves fitness advantages relative to mating (and fertilization) while non-sexual social selection influences components of fitness related to survival.

Social selection may have shaped aspects of human cooperation. Being selected as a social partner would have enhanced survival for an individual of a species that lived in a social group structure, particularly in environments where being an outcast would put survival at risk. Thus there could have been selection of prosocial traits such as empathy and altruism (Nesse 2007). One of the biggest challenges in evolutionary biology has been to explain altruism, behavior that is apparently costly to the actor and beneficial to the recipient, and which at first glance is at odds with the basic tenet that selection operates on the individual. Initially altruism was used as an argument for group selection, but problems were inherent in that concept (see Section 2.3.2.4) and it was largely abandoned. So if the unit of selection is the individual, how can behavior that apparently does not serve the reproductive interests of the individual (i.e., altruism) evolve?

Altruism is a common feature of mammalian groups. A meerkat will take up a guard position to watch for a predatory raptor even though this entails a greater risk of being spotted and becoming prey. One member, not necessarily a parent, will guard the nest while others in the clan are out hunting. Much in human social behavior appears to be for the benefit of others rather than oneself. As discussed in Chapter 2, a favored explanation is provided in part by the concepts of kin selection and inclusive fitness.

William Hamilton, the originator of the concept of kin selection, argued that an animal would behave altruistically with respect to other animals if they are likely to carry the same genes and thus assist indirectly the intergenerational flow of some of its own genetic material (Hamilton 1964a, b). This altruism would apply in the case of kin, and the closer the kinship the more likely it is that altruistic behavior would be beneficial. This is formulated in Hamilton’s rule, which states that a gene supporting altruistic behavior would be under positive selection when the benefit to the recipient of the altruistic act (in terms of reproductive fitness) is greater than the cost to the individual conducting the beneficial act. This benefit is clearly dependent on the degree of relatedness: the greater the degree of relatedness the greater the benefit in terms of potential gene flow. Indeed there is considerable empirical evidence that altruistic acts in animals are more likely when there is a high degree of relatedness. This concept of kin selection has also been used to explain the development of eusociality in insect species such as the honey bee (see Section 2.3.2.4).

But the degree of altruistic behavior shown towards kin may not just be determined by the degree of relatedness. From an evolutionary perspective, aged individuals have less impact on inclusive fitness than those of peak reproductive age. This consideration is termed reproductive value, and is a measure of an individual’s potential to contribute to the gene pool of the lineage by virtue of their age.

Generally, having a larger social network and greater number of kin has positive effects on longevity. Several studies have shown that under

<table>
<thead>
<tr>
<th>Area of risk</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females:</td>
<td></td>
</tr>
<tr>
<td>Body-image concerns</td>
<td>1.3</td>
</tr>
<tr>
<td>Smoking</td>
<td>1.4</td>
</tr>
<tr>
<td>Functional symptoms</td>
<td>1.5</td>
</tr>
<tr>
<td>Victimization</td>
<td>1.7</td>
</tr>
<tr>
<td>Sexually active</td>
<td>1.9</td>
</tr>
<tr>
<td>Males:</td>
<td></td>
</tr>
<tr>
<td>Drunkenness in previous 6 months</td>
<td>1.4</td>
</tr>
<tr>
<td>Victimization</td>
<td>1.7</td>
</tr>
<tr>
<td>Cannabis usage</td>
<td>1.8</td>
</tr>
<tr>
<td>Sexually active</td>
<td>1.8</td>
</tr>
<tr>
<td>Smoking</td>
<td>1.8</td>
</tr>
<tr>
<td>Illegal drug usage</td>
<td>2.0</td>
</tr>
<tr>
<td>Depression</td>
<td>2.1</td>
</tr>
<tr>
<td>Functional symptoms</td>
<td>2.2</td>
</tr>
<tr>
<td>Suicide attempt</td>
<td>4.9</td>
</tr>
</tbody>
</table>
extreme conditions, such as the Donner Party disaster of 1847 where US settlers caught in a snowstorm resorted to cannibalism, those with kin within the group were more likely to survive as they were more likely to receive support from their relatives (Grayson 1990). Kin selection is an important component of human behavior. There is much evidence to show, for example, that people are more willing to care for a relative’s child than for one who is completely unrelated. We loan money to relatives on an entirely different basis from non-relatives. The grandmother hypothesis, discussed in Section 8.9.7 as a favored explanation for the evolution of menopause, can only be understood in the framework of kin selection and inclusive fitness. In general, kin selection is affected by age, as more altruism is shown to younger kin members, be it in patterns of childcare or willingness to invest in medical care.

In many species, such as the lion, males will kill the offspring of another male when they are able to mate with the female who mothered those offspring, whereas they will be protective of their own offspring. This may have a parallel in humans: Daly and Wilson (1988) observed that step-fathers are more likely to abuse or murder step-children than biological fathers.

It might be argued that the practice of adoption of unrelated children, common in the West today, is inconsistent with concepts of kin selection (Volk 2011). Indeed in many traditional societies, adoption largely has a function in kin support, with relatives (most often the maternal grandmother, in line with the evolutionary predictions based on kin selection and paternal uncertainty) temporarily or permanently looking after the children of their impoverished or deceased kin. This form of kin adoption is not uncommon today, though it usually takes the form of fostering or guardianship, where ties with the biological parents are not severed. But while for the majority of adoptions in the developed world, involving entirely unrelated individuals not necessarily from the same social or ethnic group as the adopters, the explanation of kinship selection does not work, there may be other evolutionary explanations. One is reciprocal altruism (see later), with investment in non-related children generating opportunities for economic return in later life. Another form of return is social. In some traditional societies adoptions are used to forge ties between groups or rebuild damaged social relationships, similar to arranged marriages practiced, for example, historically in European dynasties or still in many Asian societies. Finally, adoption may be a by-product of powerful adaptations, for example motivations to procreate and to parent, offering an outlet for such feelings to infertile couples. It is interesting that in studies looking at the qualities in an adopted child that the prospective adoptive parents prefer, it has been found that women emphasize cues of health (presumably because of their larger investment into care of offspring) while males valued resemblance cues (presumably because paternity is far less certain than maternity).

Adoption of unrelated children (or surrogate pregnancies and egg donation without remuneration) is only one kind of example of altruistic behavior extending beyond relatives. We give blood not only for psychological reward but in the expectation that others will donate blood should we need a transfusion in the future. In forager communities, it is common for food to be shared between unrelated individuals. This can be seen as a form of insurance against potential hard times. The potlatches (festivals involving lavish gift-giving) of the Pacific Northwest peoples are an extreme example. Such non-kin-based apparent altruism can also be demonstrated in animals. For example, the vampire bat will give food (blood) to other unrelated bats in the colony that were unsuccessful in the hunt in expectation of the favor being returned at a later date.

The favored explanation for such behavior comes from another major tool of evolutionary biology, namely game theory, and in particular the concept of reciprocal altruism. The basic premise is that if A does a good turn for B in the expectation that at some later time B will reciprocate by doing a good turn for A, then both parties will benefit. However, this only works if B is not a cheat (a “freeloader”): if B is a cheat, A has lost and B has gained. Working with this simple model, game theorists such as the eminent evolutionary biologist John Maynard Smith demonstrated that evolutionarily stable strategies can emerge based on reciprocation and cooperation (Box 11.2). These strategies operate best where there is a great ability to detect and punish a cheat. Indeed, many aspects of modern society are based
Altruism can be defined as an unselfish concern for the welfare of others. Reciprocal altruism is conditional, as it is expected that the benefit provided by one individual to another will be repaid in kind at a later stage. Cheaters who exploit this system are punished by future withdrawal of aid. This behavior is comparable to the so-called “tit for tat” strategy in game theory, a branch of mathematics that analyses gains and losses for behavioral choices during strategic situations.

A classic example used to illustrate tit for tat is the prisoner’s dilemma. The scenario is that two arrested suspect accomplices are kept isolated from each other in separate cells and are both offered the same deal. They have to decide whether to betray their accomplice or remain silent: the catch is that whether they go free or are sentenced to a long imprisonment also depends on the decision their accomplice makes. Their choices can be illustrated in a payoff matrix, shown in Table 11.2.

We can see that the best outcome for both prisoners can be obtained if they both cooperate and remain silent, as their sentence will be just 1 year. Still, each prisoner could be tempted to act on a purely individualistic level, where the rational decision is to betray their accomplice and get released immediately. Yet, betrayal by both parties—the self-interested rational choice of each—leads to a worse outcome for both (a 5-year sentence) than if they were to cooperate with each other (a 1-year sentence). What would be the best strategy for playing this game, given that one’s fate is reliant on the second player?

By playing this game more than once, players have a memory of the opponent’s previous behavior and how likely they are to cooperate or betray (cheat). Doing this, it has been mathematically determined that the most robust strategy for ensuring the best outcome for oneself is to cooperate on the first iteration, then adopt the same approach as the opponent on subsequent moves; that is, tit for tat. A degree of forgiveness is also required to prevent a vicious circle of revenge and counter-revenge should a player cheat. Thus, players will eventually default to cooperating after a period of non-defection. An unavoidable consequence of this approach is that the occasional emergence of cheaters cannot be prevented.

A related idea is the hawk–dove game, as formulated by John Maynard Smith and George Price (Maynard Smith and Price 1973). In terms of evolutionary game theory, this model describes competition between two members of the same species for resources such as territory, food, or mates. Parties can adopt either a hawk or a dove strategy. As the terminology suggests, the hawk strategy involves aggressive and escalated fighting with opponents, while the dove strategy involves passive behaviors such as threat displays and pretense. Hawks have the advantage of easily gaining possession of a resource when faced with a submissive dove opponent, but the downside is the possibility of being injured or even killed by another hawk. The dove approach avoids injury, but expends time and energy in resolving the fight. Again, using a payoff matrix, we can examine how the combination of strategies adopted by each party affects resource

Table 11.2 The choices in the prisoner’s dilemma can be illustrated in a payoff matrix

<table>
<thead>
<tr>
<th>Prisoner 2</th>
<th>Remains silent</th>
<th>Betrays Prisoner 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remains silent</td>
<td>Each serves 1 year</td>
<td>Prisoner 1 serves 10 years</td>
</tr>
<tr>
<td>Betrays Prisoner 2</td>
<td>Prisoner 1 released</td>
<td>Prisoner 2 serves 10 years</td>
</tr>
</tbody>
</table>

Table 11.3 Net payoff matrix for the hawk–dove game, where two animals encounter the carcass of a prey

<table>
<thead>
<tr>
<th>Animal 2</th>
<th>Hawk</th>
<th>Dove</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawk</td>
<td>Both animals: each has a 50% chance of getting the carcass; 50% chance of sustaining injury. Net payoff is negative as the cost of injury is greater than the value of getting the carcass</td>
<td>Animal 1: gets the carcass</td>
</tr>
<tr>
<td>Dove</td>
<td>Animal 2: empty-handed</td>
<td>Animal 1: empty-handed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Animal 1</th>
<th>Hawk</th>
<th>Dove</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawk</td>
<td>Both animals: each has a 50% chance of getting the carcass; no injury sustained. Net payoff is positive as the winner gets the carcass half the time, but time is wasted</td>
<td></td>
</tr>
<tr>
<td>Dove</td>
<td>Animal 2: gets the carcass</td>
<td></td>
</tr>
</tbody>
</table>
Humans underwent positive selection for living in groups. It could be postulated that the psychological mechanisms allowing humans to live successfully in groups evolved because such eusociality allowed the clan to solve problems collectively, but this is not an argument for group selection.

In this discussion it is important to distinguish between selfish genes and selfish behavior, a distinction often forgotten by the popular press and some academics. The phrase “selfish gene” was simply a catchy shorthand introduced by Richard Dawkins to emphasize that evolutionary processes are fundamentally about preserving gene flow, and that selection may well be acting at the level of the gene rather than the whole organism (see Section 2.3.2.5). Kin selection, natural selection, social selection, and sexual selection are all manifestations of processes that attempt to preserve fitness by protecting gene flow to the next generation, either directly or indirectly. In contrast, selfish behavior is a description of an individual’s behavior, but even here its interpretation depends on the level of analysis. What may have had its origin in self-interest can
lead to non-selfish behaviors that we interpret as being altruistic. An example might be philanthropy, which originates from the desire of the donor for social recognition.

11.4.3 Emotions

Emotions are universal human attributes and may exist because they offered adaptive advantage in our evolution as a species. Emotions do not occur in isolation, but involve integrated physiological and behavioral responses to environmental stimuli, either at the time or in recollection. Darwin recognized that the physical manifestations of emotions can play a role in selective processes, and he wrote an extensive volume on emotion (Darwin 1872; Box 11.3). More recently, the evolutionary psychologists Leda Cosmides and John Tooby, who developed a relatively extreme position regarding the evolution of behavior (Box 11.1), referred to emotions as the “Darwinian algorithms of the mind,” emphasizing that these are selected traits.

Given their universality, emotions may have been shaped by their adaptive value in signaling and responding to situations that are frequently and universally encountered, such as fear, panic, and sexual desire. In general emotions are healthy phenomena. The challenge for medicine and psychology is when emotions become situationally inappropriate in nature or severity and impair the functioning of the individual.

11.4.4 Fear, Anxiety, and Response to Threat

Many aspects of our physiology defend us from external threats (see also Chapter 10), and some aspects of our emotions can be viewed as serving similar functions. Fear and anxiety serve the obvious purpose of alerting us to danger and maintaining a state of vigilance. They have evolved in response to threats from predation and violence (Box 11.4). Survival depends on the development of sensory functions and prescient capacities to detect or predict threats from other species. The development of a theory of mind allows us to understand the intentions of other humans, and we need an ability to mount so-called stress responses allowing

---

Box 11.3 Darwin on Human Emotions

In 1872, Darwin published a book on *The Expression of the Emotions in Man and Animals*, a book that complements *The Descent of Man* yet is idiosyncratic in its combination of different kinds of sources, extensive use of photographic material, and close collaboration with a physician, a leading psychiatrist of the time, the superintendent of the West Riding Pauper Lunatic Asylum Dr. James Crichton-Browne (Pearn 2010; Buklijas and Gluckman 2013; Radick 2013). The central argument of the book—that human facial expressions exhibited continuity with the expressions of animals—supported the thesis in *The Descent of Man* that modern humans evolved from “lower” organisms and thus did not have a special status in the nature.

Crichton-Browne provided Darwin with 41 photographs of patients accompanied by descriptions of diagnosis, symptoms, and behavior. Darwin understood these as offering an immediate access to human emotions “as they [the ‘insane’] are liable to the strongest passions and give uncontrolled vent to them.” In addition to the photographs and information about psychiatric patients, Darwin collected observations on infants (also seen as prone to strong and uncontrolled emotional expression), animals, people of different races, answers to questions about what emotion is being expressed in the photographs, and the study of art. But although the book provides a close comparative analysis of expressions, including a speculation on the evolutionary history of expression, Darwin, curiously, did not attempt to explain the adaptive advantage of expressing emotions. Some historians have explained this by the fact that his interest in the expression of emotions may have pre-dated his formulation of descent by natural selection. More convincing is the argument that Darwin’s main goal, in addition to showing the continuum of expression between animals and humans, was to show the sameness of emotional expressions across human races, thus providing another argument for his view that all races are descended from a single parent stock, itself human in character before the races diverged (Darwin 1872, p. 361).
There is an enormous overlap in the physiological responses to threats of danger and to social stress. Humans rely on being members of a group for their survival. Just as an isolated member of a herd species is at particular risk of predation, an isolated human would have been at a disadvantage in threat detection, in hunting and foraging, and in protecting offspring. When placed in isolation, rats and many other species have a marked endocrine stress response. Humans find social exclusion threatening and stressful, as is evidenced by studies of prisoners in solitary confinement.

It is useful to distinguish between a stressor and stress itself. A stressor has been defined as “a threat, real or implied, to the psychological or physical integrity of an individual.” Detection of a stressor by an organism elicits a coping strategy that may be an active (physical or behavioral) or passive (psychological) stress response. Responses to stressors are usually short term. However, the responses to chronic stressors may differ: the acute stress response may become either exaggerated or attenuated, making the effects of the sustained stressor more or less harmful. Either way, pathological consequences arise when the coping strategy fails.

Acute stress responses in complex organisms involve avoidance, withdrawal, or escape from the stressor, a change in central nervous system function involving greater sensory awareness and alertness, stimulation of the sympathetic nervous system,
and neuroendocrine changes leading to activation of the hypothalamic–pituitary–adrenal (HPA) axis. Other hormones, such as growth hormone, which promotes lipolysis, and vasopressin, which redistributes blood flow and affects kidney function, are also released by the hypothalamic–pituitary unit.

These stress responses are driven by the higher centers of the brain, from descending control by the frontal cortex (some people can become just as stressed by being asked a question by a tutor that they cannot answer as others are when physically attacked; the Trier test is a stress test used in behavioral and endocrine studies that is simply a test of public speaking). The limbic system coordinates the response and plays a role in mediating the potentiating or inhibiting effects of chronic stress.

The HPA axis has built-in mechanisms for feedback control mediated by glucocorticoid receptors at various levels of the axis, with the hippocampus being a major site of feedback. Hence, changes in the expression of glucocorticoid receptors during development or during chronic stress can alter the magnitude of the response. In turn, activation of the HPA axis and catecholamine release induce metabolic changes, including glycogenolysis and gluconeogenesis to provide energy resources for “fight or flight” responses and cardiovascular responses that increase heart rate and blood pressure and redistribute blood flow. Chronic stress responses include further neural changes: for example, elevated glucocorticoid levels can change mood and affect memory and induce a wide variety of secondary physiological changes, many induced by chronic hypercortisolemia. Exposure to chronic stress in early life has been associated with later risk of depression (Box 11.5).

Changes in the physical, biotic, or social environment can all act as stressors. These conditions offer different challenges, yet the response is stereotypical. The stereotypy suggests that the stress response originally evolved to deal with one set of conditions, for example threat from a predator, but then became co-opted and selected as an advantageous response to another set of conditions (i.e., it is an exaptation). An Olympic athlete poised to sprint at the start of a race shows many of the same physiological responses as a rabbit that sees a fox or a swimmer who faces a shark. The stereotypical nature of the response is fail-safe, in that it heightens awareness of a situation in which threats may arise, even if they have not yet done so. This is akin to the “smoke detector principle”: false alarms are

---

**Box 11.5 Early Stress and Later Life Risk of Mental Disease**

Exposure to an adverse early social environment, such as childhood abuse or severe neglect, has been associated with heightened HPA responses to stress and with greater risk in later life of recurrent and persistent depression, particularly when further exposed to stressful events. Such heightened sensitivity to environmental stressors may be understood as an evolved trait, as stressful environments direct development toward strategies that are adaptive in stressful conditions (Ellis et al. 2011). These strategies are likely to be harmful—in humans, childhood adversity has also been linked to poorer treatment outcomes in depression—but they may provide the optimal path to survival and reproduction under adverse conditions, similarly to the offspring of malnourished mothers. The adaptiveness of these strategies suggests that there are likely to be well-conserved mechanisms of stress response modulating a broad range of physiological functions.

A large body of work in the rat (see Section 3.8.2) has established the role that the epigenetic regulation of many genes, including the glucocorticoid receptor, plays in modulating HPA function and therefore in mediating the later life effects of developmental adversity. There is now growing evidence that similar molecular mechanisms operate in humans. For example, in samples of brain tissue from people who have committed suicide the expression of several glucocorticoid receptor non-coding exons is lower in individuals who had also experienced childhood maltreatment, and these differences were further reflected in differential DNA methylation patterns at the corresponding promoters (Anacker et al. 2014). A better mechanistic understanding of how early social adversity influences subsequent responses to treatment models for major depressive disorders may provide leads to help devise more effective therapeutic strategies.
better than failing to react, so selection may have favored the response being set on the “sensitive” side (Nesse 2001). Anxiety allows us to acknowledge current threats and anticipate and avoid potential threats. These clearly have an adaptive origin, but if expressed inappropriately can manifest in an inappropriate form as the psychiatric disorders of anxiety and phobia. More severely threatening life experiences may manifest as post-traumatic stress disorder (Box 11.6).

11.4.5 Love, Jealousy, Marriage, and Inheritance

Given the centrality of reproduction to evolutionary biology, much research in evolutionary psychology has focused on human behavior in relation to mate selection, pair bonding, sexually selected traits, familial investment strategies, the role of each gender in society, and the nature of male–female relationships. Some of these topics have been considered in previous chapters. Again it is important to recognize that whereas we will focus on evolutionary determinants here, humans will overlay other cultural behaviors on top of these evolutionarily determined fundamentals.

Strictly from the point of view of evolutionary biology, the pattern of interaction between parents of sexually reproducing species has evolved to maximize the fitness of their offspring. Where they can, females choose their mate in the expectation that the mate will be able to contribute to her fitness by supporting the nurture of their offspring. This is true for both polygynous and monogamous mating structures. Sexual selection has operated to favor females with those characteristics most likely to support successful pregnancies. Pair bonding helps reinforce this interaction, and romantic love may be a mechanism that evolved to help reinforce this bonding (Box 11.7). Equally, jealousy can be envisaged as a response to the breaking of these pair bonds.

The long nurturing period required for human offspring and the nature of human culture means that humans can continue throughout life to affect the potential fitness of their offspring. Complex social arrangements have emerged to enable them to do so. Again, from an evolutionary perspective, the concepts of property inheritance, marriage, dowry, and so forth are all mechanisms to protect the status and wealth of offspring in an attempt to promote their reproductive success. Different societies have developed complex rules of inheritance and marriage systems, and these can often be understood in terms of the ecology of a particular society. For example, in many societies male reproductive success is linked to wealth and a father’s inclusive fitness might be greater if he concentrated his wealth in the hands of fewer of his male offspring rather than benefit them evenly. This is seen in some pastoralist societies where large herds of cattle are more viable than smaller ones, and thus spreading cattle evenly might reproducitively disadvantage all of a man’s sons. In such societies reproductive success is

---

**Box 11.6 Post-traumatic Stress Disorder: Adaptive or Pathological?**

Post-traumatic stress disorder (PTSD) was first classified as a disease in the aftermath of the Vietnam War, around 1980, although delayed reactions to trauma or distress, especially following combat situations, had been recorded earlier—for example the condition widely reported in the First World War to describe the psychological trauma experienced by soldiers, “shell-shock” (Winter 2000). PTSD is defined as an anxiety disorder that can develop following a major traumatic event—assault, participation in warfare, major stress—with symptoms including intrusive memories such as nightmares and flashbacks, and negative impacts on mood and emotional reactions. Its incidence is related to the degree of exposure of an individual to a traumatic event, ranging from about 5% in New York residents who had been broadly exposed to the September 11, 2001 terrorist attack to about 15% of those directly affected, although the prevalence declines over time (Galea et al. 2003). Several of the components of PTSD could be viewed as potentially adaptive in origin; for instance, the initial hyper-responsiveness may produce avoidance of a recurring local threat; the subsequent resilience may promote advantageous fitness in a survivor, and so on. However, this is conjectural.
**Box 11.7 Genes and Social Behavior**

Some closely related species may have very different social structures. Humans and gibbons generally establish male–female pair bonds that last for considerable time, and in essentially all human societies this is the accepted cultural norm. Chimpanzees, however, do not form one-to-one male–female pair bonds, and mate with multiple partners not just over lifetime but even within a single mating cycle.

Voles also exhibit diverse sexual models, with some species exhibiting monogamous pairing and fidelity and others promiscuous behavior. The underlying molecular mechanisms are increasingly being elucidated, with prominent roles being attributed to the nine amino-acid-long neuropeptides, vasopressin and oxytocin. These peptides had previously been associated in multiple animal models with social behavior: vasopressin with social cognition and territory defense, and oxytocin with social memory and mother–infant bonding (Barrett and Young 2015).

In the prairie vole, a monogamous species, pair-bond formation is promoted by central infusion of vasopressin or oxytocin, and is inhibited by antagonists of either of the neuropeptide’s receptors. The neuroanatomical distribution of the vasopressin and oxytocin receptors is very different in this monogamous species compared with that in the non-monogamous, but closely related, montane and meadow vole species. The coding sequence of the vasopressin receptor is the same in all vole species, yet the 5′ regulatory sequence differs between the monogamous and non-monogamous species, suggesting different transcriptional regulation of the receptor. Transgenic mice possessing the prairie vole receptor gene and its 5′ flanking region displayed a similar neuroanatomical pattern of receptor binding to that of the prairie vole, and unlike their wild-type counterparts displayed greater affiliative behavior in response to vasopressin administration (Young et al. 1999).

In recent years, there has been a growing body of research implicating a similar role for oxytocin and vasopressin in modulating social behaviors in humans. A large study focused on the role of three repeat polymorphisms upstream of AVPR1A, encoding the vasopressin receptor. It revealed that having a specific allele, 334, in one of the repeat polymorphisms, R53, was strongly associated with lower levels of pair bonding, as reflected in marital problems, marital status, and spousal perception of the marriage, in men (but not women) (Walum et al. 2008). This effect was dose dependent: while there was an effect of heterozygosity, individual males homozygous for the allele showed even greater impairment in pair bonding. Furthermore, having two copies of the risk allele doubled the risk of previous experience of marital crisis compared with having no risk allele. The sex differences observed accord with the animal studies, which have found a greater effect of vasopressin on social behavior in male voles.

The allele frequency of the polymorphism associated with impaired pair bonding was high (40%), and the alleles were distributed in accord with the Hardy–Weinberg equilibrium (see Box 2.8). This raises the intriguing possibility that humans evolved with two fitness-enhancing mating strategies, and that an equilibrium exists between the two: one involving strong and enduring pair-bond formation and one in which pair-bond formation is less important, perhaps reflecting the mild polygyny of past human society (Chapter 8), and that male fitness can also be achieved and continues to be so through multiple serial or parallel relationships. In contrast, at least for this allele, there appears to be no effect on the female strategy, which is optimized through sustained pair bonding.

Yet, interestingly, several human studies have reported an association between AVPR1A polymorphisms and maternal behavior. Mothers with two copies of R53 displayed a lower ability to perceive and respond to their young children’s cues, while those carrying the 334 allele provided less support and guidance during their child’s free-play sessions (Avinun et al. 2012, Bisceglia et al. 2012). Little is currently known about the neuromolecular mechanisms underlying these effects, although it has been suggested that they may be related to the roles of AVPR1A in stress and anxiety and in autism.

Human clinical studies have shown that intranasal administration of oxytocin has positive effects on multiple aspects of social behavior, including trust, perception of the mental state of others, and couple conflict resolution. Associations have now been found between one SNP in OXTR, the gene encoding the oxytocin receptor, and levels of pair bonding in women, but not men (Walum et al. 2012). In addition, the same SNP was associated with childhood social problems in girls, who later displayed weaker bonding with their partners. Once again, this sex specificity aligns with observations of the effects of oxytocin in female voles.

Although the effect size for human studies is small, these studies illustrate several points. First, fundamental components of human behavior do have genetic determinants. Second, it is apparent that human behavior has genetic and evolutionary relationships with analogous behaviors in other species. Third, evolution does lead to gender-specific behavioral effects. Such studies highlight the growing capability of molecular genetics and evolutionary biology to contribute to an understanding of the determinants of human behavior.
generally greater for older brothers than for younger ones. There are echoes of this approach seen in primogeniture (where the first-born son inherits all), which has been practiced in some European societies. Parenthetically, the economic historian Gregory Clark believes that primogeniture helps to explain why the Industrial Revolution occurred initially in England: it created a population of well-educated but impoverished younger sons who sought status through trade and innovation (Clark 2009).

11.4.6 Group Behavior and Morality

Animals living in groups have a set of behaviors that are necessary for harmony within the group. In some species, such as the hyena and wolf, this involves an obvious hierarchy within the group, with clear roles and rights for the alpha male or female. Humpback whales hunt fish together in the phenomenon known as bubble net feeding, and fish school because it reduces the risk of any individual member being eaten. Humans live within a particularly pronounced group structure. This group structure evolved because it provided a fitness advantage for its individuals, probably for cooperative food gathering and defense against predators or rival groups.

Highly social species exhibit a number of behaviors that reinforce group bonds, such as grooming in the chimpanzee and sexual stimulation in the bonobo. In Chapter 6 it was suggested that the evolution of language and gossip played a major role in generating and stabilizing bonds within early human groups. As group living requires multiple behaviors and is an integrated phenotype, all these selective pressures would have acted to determine the social and behavioral phenotype of our species. Natural selection, unlike artificial selection, does not act on any one trait in isolation. Therefore, teasing apart and arguing for greater weight for one component or another is neither practical nor sensible.

But as we have already suggested, membership of a well-bonded social group requires adherence to the rules of reciprocity. Human groups are particularly sensitive to freeloaders or cheaters. We respond to cheating behavior with exposure, ridicule, embarrassment, and punishment. Frameworks of what behaviors are acceptable or unacceptable become formalized within the group. As group size becomes greater than about 150, a formal internal structure is required for stability.

Concepts of morality may be derived in part from these context-specific frameworks necessary to control freeloaders. They may be manifest in custom, taboos, rules, and tradition. But other factors also play an important role in a particular societal view of morality. These include rules and concepts imposed in part by hierarchical organizations to protect the social structure, and in part by the formalization of belief systems.

11.4.7 Belief and Religion

Every human society is characterized by one or more belief systems that are reflected in its organization, tradition, and ritual. These belief systems generally involve some concept of the supernatural. Ritual burial implies a sense of afterlife and has existed for at least 70,000 years.

Belief in the supernatural was almost universal until modern rationalism emerged. The issue of why and how belief in the supernatural arose has been the subject of considerable evolutionary reflection since the initial musings of Freud (Boyer 2001; Norenzayan 2013). The evolutionary question is whether belief in the supernatural has an adaptive advantage or is simply an epiphenomenon related to other group behaviors. Supernatural belief is counter-factual and the adaptive advantage of suspending reality is not entirely clear. Its origin is also highly controversial. Perhaps it allowed individuals and the groups they were members of to develop an emotional stability in the face of events (such as drought) that they could not comprehend or predict. Ritual, which often accompanies superstition, helps build group cohesion. Sagas and story-telling are often part of a belief system, and these may have helped reinforce group identity and thus cohesion. In addition, belief in the supernatural as a potential external source of punishment or reward could help a group deal with the problem of freeloaders.

The 30,000-year-old wall paintings in the caves of France and Spain may well be some of the earliest representations of ritual and belief, though other explanations are possible (Box 11.8). The organization of belief into formalized religion from perhaps
11.4 EVOLUTION OF SOCIAL BEHAVIOR

5000 years ago occurred in parallel with the development of larger population groups and the associated political organization.

David Sloane Wilson and others have suggested that organized religion became a major way to control freeloaders and stabilize a large group (Wilson 2002). Some have argued that organized religion largely developed as a political tool within a hierarchical control system. Many societies have conflated political rule and concepts of deity. Until the seventeenth century, British monarchs were considered to have divine powers of healing, and today they are still nominally head of an organized religious structure.

A common hypothesis is that religion evolved as a way of confronting the problem of inevitable death, and that complying with the group’s behavior would hopefully lead to a deferred reward. But Wilson also suggested that the religious group needs signs of commitment from an individual if they are to receive the rewards of group membership. This could take the form of sacrifices, tithes, or changed behavior (e.g., not eating meat on specific days). Paying a price to be a member of a group reduces the risk of someone being a freeloader. The risk of being involved in cheating is not only exclusion from the group but also punishment by some higher authority such as deity or supernatural force.

Wilson developed this hypothesis to argue that religion evolved through a group-selection process. His position remains controversial, and this discussion, alongside related ones, plays a considerable part in the framing of multilevel selectionist arguments (see Section 2.3.2.4). Those who would focus

Box 11.8 On the Origin of Art

The existence of “art” across human societies throughout history has puzzled scholars for a long time. Stylized etchings on bone and stone dating back at least 70,000 years have been found in southern Africa. Representational art dates from at least 32,000 years ago, in the cave paintings of Western Europe, and perhaps even longer ago in the rock art of Australia. Indeed evidence of art can be found in all human societies, from the Australian Aboriginal painters to the audience attending the Metropolitan Opera House in New York or the graffiti on a subway station wall. Though some critics deny its universality, arguing that it is Western culture that has invented art as we know it, the same major forms appear everywhere in the world: music, dance, visual arts, and storytelling. Art does not appear to require formal training in the way that, for example, reading does; it is sustained despite its costs; and it provokes a strong emotional response.

Explanations of art include those that ascribe certain functions to it—for instance as means of communication or expression—but such explanations beg the question of why art would survive alongside other, less costly, means of communication and expression. The ubiquity and antiquity of art have stimulated questions about its biological origin.

Darwin thought that the “high cost, apparent uselessness, and manifest beauty” indicated the origin of a trait/behavior in sexual selection (Darwin 1871). Dance, for instance, often occurs in mating rituals. But sexual selection can only account for some aspects of art.

Drawing on a large body of evolutionary and humanities literature, the literary scholar Brian Boyd has explained the evolutionary origin of art in play (Boyd 2009). Play, which is widely found among animals and is an essential part of human early development, is understood as a way to develop, fine-tune, and practice mental, physical, and social skills within a safe context. Art is an advanced form of cognitive play that builds on several aspects that are unique or especially evolved in humans. One is the preference for patterned information. Humans search for patterns—discernible order in things, actions, and situations—in order to understand the underlying rules, to make inferences, and thus make predictions. Musical motifs, visual themes, storylines, can all be understood as patterns with which the human mind engages and plays. But art could not evolve without the shared attention and sociality characteristic of humans: think of chanting, dance, body adornment; of traditional styles of pottery and woodcarving; or indeed of folk poems and storytelling traditions. The intense emotional response provoked by art might make it more effective than other forms of communication and expression. Indeed, art fosters intense group cohesion: an example would be sports teams (and their audiences) singing national anthems before international games. These properties of art can explain why art, in various forms, was extensively used by religion: who can separate Christianity from the soaring, highly decorated forms of Gothic cathedrals or spectacular visual representations of the textual tradition, for instance Michelangelo’s paintings on the ceiling of the Sistine Chapel?
on individual-level selection and reject any concepts of group-level selection would argue that the adaptive value of reciprocal altruistic behavior and group living for an individual provides a sufficient evolutionary explanation. Within the context of the parallel processes of biological and cultural evolution, ritual and religion can be seen to have adaptive advantage for the individual and a group selection argument is not necessary.

11.4.8 Learning

Humans are born in a relatively immature neurological state compared with other primates. Nevertheless, the human infant is not born with a total inability to perceive or react to its world, and is certainly not as immature as more altricial species such as the rat. Much experimental and clinical data now show that human babies have very active sensory processes. They prefer symmetrical objects and images of an organized (rather than a scrambled) face. Their sense of smell has also developed to the point that they can identify the smell of their mother. By the age of 9 months, babies are clearly able to recognize and respond to the psychological state of others, and by 15 months they can persuade their mother to react by pointing to an object.

Learning then becomes a process of acquiring skills and changes induced through interactions with adults, which leads to new skills. Humans have evolved “goal-based” imitative learning, which allows the growing child to learn about the goal and the actions necessary to reach it. Gradually this permits children to engage in their culture, and this process is greatly accelerated by the acquisition of language. By the age of 4 years children have moved to a level of cognitive development for which at least second-level intentionality can be demonstrated.

Learning from experience is a key adaptive capacity of many species, but is particularly well developed in humans. Learning by forming associations between events, called associative learning, is present even in children and is a necessary precursor to inferential reasoning. The capacity to learn is regarded by some as simply another module of the mind in the context of orthodox evolutionary psychology (Section 11.5). Others regard it as flexible and generic, rather than domain specific. Again, we see this argument as unnecessary in the context of this book. Humans have evolved with a brain capable of assessing environmental information, storing information as memories, and thus guiding decisions and consequent behavior. Rather than trying to encode all possible responses in our genes, selection has endowed us with an advanced organic “computer” that can learn by association and experience, and make appropriate decisions. Indeed, in developmental learning, experience reinforces particular synaptic pathways and neuronal networks. Human learning is effectively a process of reinforcement by positive outcomes and avoidance following negative outcomes. Thus we learn to seek ripe fruits, but not to eat toadstools.

11.5 Evolutionary Perspectives on Psychology

We have highlighted how many aspects of human behavior can be better understood by including evolutionary as well as cultural perspectives. In doing this we have taken an integrated perspective, namely to examine to what extent a behavior can be considered to advance or protect fitness. The field of evolutionary psychology has also considered how the mind itself evolved. There are marked similarities between this discussion and considerations of how language evolved (see Section 6.3.9). Several schools of thought, based in part on different conceptual approaches, have emerged (Box 11.1). Whereas the term “evolutionary psychology” can be used narrowly to describe a single one of these schools (that founded by Leda Cosmides and John Tooby), we will use it here in its broadest and most inclusive sense. Two extreme views exist: first, that the brain is a universal tool able to respond flexibly to a variety of situations, and alternatively that the brain has evolved as a series of domain-specific modules.

The most prominent advocates of the modular view, the evolutionary psychologists Cosmides and Tooby, proposed that there were strong selection pressures for each capacity of the mind to have evolved as an independent module (Barkow et al. 1992). There were perhaps thousands of modules, each for a different behavior; for example a module
to detect freeloaders, a module to learn language, and so forth. A key concept in their thinking was that of the *environment of evolutionary adaptedness* (EEA). This was the putative environment that existed through the bulk of human existence, at least until the end of the Neolithic, during which selective pressures acted on human physiology and behavior to lead to the current portfolio of human behaviors. The modular model implies that behaviors have an adaptive origin and must largely be genetically determined. One limitation of the concept of EEA is that there was in fact no single EEA—rather a large number of different environments in which Paleolithic humans lived (Foley 1995).

In contrast, the opposing model would suggest that most behaviors are learned, but can only be learned because of the evolved neural substrate. This dichotomy is an exaggeration made by advocates of particular schools of thought in order to make specific points, and is to some extent unnecessary. What is clear is that humans have evolved with a neural infrastructure that is capable of learning, and with a series of cognitive abilities able to cope with novel situations and living within a complex social organization. But there is some stereotypy in a number of behaviors and emotions, and evidence for genetic determinants suggests that a finer grain of selection has operated. There is a renewed interest in the role of genetic assimilation (Chapter 4) as a process by which learned behaviors are converted into genetically based behaviors. Indeed, the first description of what we now term genetic assimilation was called the Baldwin effect after the psychologist James Baldwin who was one of the first theorists to describe how behavior might affect evolution (Bateson and Gluckman 2011; Section 4.7).

Despite its limitations, the modular model does emphasize an important point. The human brain evolved under very different social and macro-environmental conditions from those in which humans now commonly live. If these modules were based on appropriate psychological adaptations when they evolved, then there will now be a mismatch between those modules and the modern constructed world. There will therefore be situations where the adaptations that underlie human behaviors have lost their adaptive advantage, and may instead become manifest as maladaptive pathologies. This argument has echoes of that used in Chapter 9 to describe the evolutionary origins of metabolic disease. Key to this school of thought has been the understanding of how the original selective circumstances led to a particular module of behavior being selected. For example, a module for fear of dangerous animals such as snakes could be envisaged. Jealousy could also be conceived as a module that had an adaptive advantage, as a jealous individual was more likely to have a selective advantage over someone who took a passive view of being the victim of infidelity.

In the arguments over the origins of language, linguistic researchers view *universals*, patterns that appear in all natural languages, as having a selected origin. The debate among linguists on this issue has been extensive and somewhat vexed, and the evidence for universals is not compelling. Similarly, there are obvious universals manifest in the human emotions, in the patterns of infant development, and in many aspects of social interactions such as mate choice and avoidance of incest. These can be taken as evidence for an evolved mind. While there is considerable merit in this approach, the caveats we expressed in Section 2.4.4 are equally valid here. Certainly, not all aspects of human biology need have an adaptive origin. There is a danger of falling into the trap of “just-so stories,” since spandrels and exaptations may apply equally to neural functions as they do to other aspects of biology. For example, while jealousy has been suggested to have an adaptive origin, there are no data to suggest it has a genetic basis and it might merely be the by-product of other capacities of the brain. Obviously there are limits to the empirical proof that is possible in evolutionary psychology, although the science is no less important for this.

### 11.6 Evolutionary Psychiatry

The application of evolutionary principles to psychiatry builds on the previous discussion. The various schools of evolutionary psychology have given rise to diverse views on psychiatric states (Box 11.9). Because mental health disorders such as depression and anxiety are particularly common, affecting perhaps 25% of the Western population, it is necessary to consider why the evolved brain is vulnerable in...
Among different branches of clinical medicine, psychiatry (with clinical psychology) has one of the longest-standing and sometimes controversial links with evolutionary biology. The history of the study of the impact of early childhood upon emotional and mental health in later life presents a good example. In the 1940s, the British psychiatrist John Bowlby proposed that the tendency of young children to form close attachments to their caregivers, usually their mothers, was an innate need, a result of evolutionary pressures, because forming attachments would increase the chance of survival, through protection from predation and environmental exposure and the provision of food. These pressures would apply to all animals but especially strongly to humans, who experience an unusually long period of dependence (Chapter 5). Bowlby built on evolutionary studies of animal behavior, in particular the work of Konrad Lorenz (Burkhardt 2005), who around 1950 became famous for his concept of behavioral “imprinting”—the concept that some birds follow the first object they see upon hatching and model their behavior on it. Lorenz, who mainly worked with birds (jackdaws), argued that the child’s tie to its mother was a result of an evolved instinctual need and maternal love was essential for healthy infant development (Vicedo 2009).

In the mid-twentieth century, psychiatry was greatly influenced by the psychoanalytic school, established by Sigmund Freud in the early 1900s. The psychoanalytic school also placed great importance upon early development, though in a framework that was largely divorced from contemporary biomedical science; the link to evolutionary biology provided some scientific support. Through the 1950s and 1960s scholars used the concept of attachment to explain the cause of many psychiatric disorders. Famously, Harry Harlow raised infant rhesus monkeys with two types of doll as surrogate mothers: one covered in cloth, soft but providing no nutrition and other a hard, wire “mother” with a bottle of milk attached to it. The infants raised by the soft mother initially clung to her but then started to explore increasingly widely; while the behavior of those reared by the wire mother was likened to the “autistic behavior seen frequently among neglected children in and out of institutions.” (Harlow 1959) The idea that maternal lack of affection and warmth—popularized as the “refrigerator mother” theory—is the cause of autism and schizophrenia was very popular in the 1950s and 1960s. In the 1970s the evolutionary biologist Niko Tinbergen attempted to explain autism as an outcome of parenting and especially mothering “styles” (Silverman 2010). But with the loss of popularity of psychoanalytic psychiatry, and the rise of biological psychiatry in this period, such hypotheses found little interest or support.

Today, attachment theory and related perspectives are no longer considered to offer valid explanations for major mental disorders. However, it continues to be highly influential in the study of normal child development. The tenth edition of the International Classification of Diseases (ICD-10) as well as the latest version of the American Psychiatric Association’s Diagnostic and Statistical Manual of Mental Disorders (DSM) list “reactive attachment disorder,” a relatively rare disorder affecting children who failed to form normal attachment to their caregivers because of experiences of severe neglect, abuse, abrupt separation in the infant and toddler years, or perhaps lack of caregiver response. The disorder is characterized by inappropriate and disturbed social behavior. Furthermore, many medical experts who write parenting advice books draw on attachment theory when advising on such matters as infant sleep or the age at which children may be placed into organized childcare (see, e.g., Sears and Sears 2001).

It is interesting that advice from the perspective of attachment theory may find itself at odds with scholars writing from the evolutionary position of the parent–offspring conflict theory. The area of infant sleep is a good example. Infants tend to wake up frequently, mostly to feed but also for emotional reasons, and these frequent awakenings tend to disrupt the sleep of their caregivers, especially breastfeeding mothers. The proponents of attachment theory focus their attention on the child’s interest (the innate need to form close attachment) and encourage a prompt maternal response to infant calls. Parental sleep disruption is seen as a sacrifice for a good cause. Others, however, take a closer interest in the evolutionary interests of the mother as well as the child, arguing that night waking is an expression of parent–offspring conflict (see Box 8.4). In the evolutionary past, they argue, frequent suckling suppressed ovulation, prolonged post-partum amenorrhea, and thus delayed conception of the next child. In some developing societies short birth intervals are associated with higher mortality of the older child (Ronsmans 1996), so arguably the infant would have strong evolutionary reasons to postpone the arrival of a younger sibling. Infant nightly calls may thus not (only) be the result of an innate need to form attachment to the parent—especially the mother—but also an expression of the strategy to secure its own survival. Yet today, especially in developed societies with extensive access to artificial nutrition, the survival of the infant no longer depends on delaying sibling conception. Furthermore, under conditions of abundant maternal nutrition, frequent breastfeeding does
not extend the duration of lactational amenorrhea: energy balance appears to have a stronger influence on the resumption of ovarian function than the temporal pattern of suckling (Valeggia and Ellison 2009). So it would follow, for adherents of the “conflict” hypothesis, that infant calls are an evolutionarily obsolete strategy, at least in modern developed societies.

Yet we have argued that interpreting parent–offspring relationships from the conflict perspective may not be productive or supported by evidence (see Box 8.4). Rather, the concept of “shared phenotype,” according to which the interests of the parent and the child are closely aligned—and forming attachment would be beneficial to both parties—seems to make more sense.

the modern world. Some psychopathologies occur at high frequency and cannot be explained by single causes such as a monogenic trait. In these cases, the origin of syndromes such as affective disorders may be due to a discrepancy between our evolved biology and the current environment, resulting in maladaptive consequences of a selected (adapted) trait. The key issue is the capacity to adapt to conditions that require a particular behavior. As with metabolic physiology, there are a range of psychological responses that can be called upon to deal with a particular situation. Just as metabolic disease can develop when individuals live in environments with an energetic load beyond their selected capacity to cope (Chapter 9), psychological systems can be limited in their capacity to adapt. These limitations may then be revealed in different societal or social conditions. The limits of this plasticity are genetically and thus evolutionarily determined. It is important to note that, irrespective of the conceptual model being used, specific genes do not link to specific behaviors but rather to the functional neural networks that are involved. However, this does not mean that there are no associations between SNPs and some psychiatric disease—recent work has identified two SNPs associated with major depressive disorder in a population of Chinese women (CONVERGE Consortium 2015). The debate between different schools of evolutionary psychology is essentially over the extent to which the brain remains plastic and the extent to which it is constrained in its plasticity by genetic determinants.

11.6.1 Personality Traits and Disorders

Personality traits can be defined as particular and somewhat inflexible ways of behaving. Individuals are recognized as having quite different personalities, and indeed we can recognize distinct personalities in domestic pets and in well-studied colonies of wild primates. One view of personality traits is that they represent constrained plasticity within the behavioral system. In general, evolutionary psychiatry assumes that a number of personality traits may have originated through adaptive advantage, but have become maladaptive in the current context. For example, a paranoid or anxious tendency may have originated from an evolved and fitness-enhancing trait that helped avoid predators. Risk-taking behavior may have been a trait that was advantageous in finding both a mate and new food supplies.

When a personality trait is particularly exaggerated or constrained, it is considered pathological and is termed a personality disorder. Evidence from twin studies (despite their limitations) shows that even when reared apart, monozygotic twins exhibit concordance for a number of personality traits (Tellegen et al. 1988). This may reflect evidence of genetic determinants.

In evolutionary terms, individuals with antisocial personality disorder can be viewed as a manifestation of the cheater/freeloader. These individuals, whose personality often emerges in adolescence, are characterized by behaviors representing their willingness to take from the group without reciprocation. Game theory explains how cheaters can persist in a society made up primarily of reciprocators (Box 11.2), and this theory suggests that if they reproduce their genes will persist even if societies attempt to exclude them. A key feature of antisocial behavior is the extent to which deception is used to hide it. It is inevitable that some cheaters will persist in any population.

It is important to distinguish this type of behavior from the acting-out behaviors of adolescence. Such
behaviors are transitional and arise because physical maturation precedes complete psychosocial maturation (see Section 5.4.2.2). Thus, adolescent acting-out behaviors occur during a period in the life cycle when there may be additional value in showing exploratory and risk-taking behavior as a form of reproductive display. Indeed, males show a persistent tendency towards risk-taking behavior throughout life. Many individuals can be seen as pushy and attention-seeking, or as impulsive or aggressive. Again, such behaviors could be seen to have had adaptive value in the mating game.

There are individuals who have difficulties in maintaining interpersonal relationships and have a poor self-image. As a result they may have a tendency towards suicidal or other self-damaging behaviors, inappropriate temper, and chronic feelings of emptiness. These individuals are unable to adapt to their social circumstance, often because they have a lack of insight or a limited capacity to interpret the circumstances they are in. This manifests as a pathology known as borderline personality disorder. Affected individuals are constrained in their ability to participate in their group and their behaviors can be perceived as unsuccessful attempts to be integrated and accepted as active members of the group. The recognition that they are unsuccessful can lead to overt efforts to exit the group.

Narcissism is defined by a need for admiration, and is generally associated with a lack of empathy for others. Narcissists often have great difficulty in a social environment and in maintaining relationships, and are highly sensitive to criticism. Again their behaviors could be viewed as an exaggerated attempt to socialize despite limitations in their capacity to do so.

### 11.6.2 Disorders of Mood

Emotions are the result of the brain evolving ways of controlling physiology and behavior for advantage in particular situations. A feature of human behavior is the pursuit of individual goals which, in turn, can be mapped back to their potential fitness advantages. For example, controlling more material resources is likely to have been advantageous for a male when seeking a mate.

Humans experience a wide range of emotions in the course of pursuing their social and physical goals (Table 11.4). These emotions are part of the equipment necessary for achieving these fitness-related goals and coping with any challenges encountered in doing so. However, inappropriate or exaggerated emotions are maladaptive and can become pathological. This can occur because the substrate is abnormal [e.g., a polymorphism at the promoter region of the 5-hydroxytryptamine serotonin-transporter (5-HTT) gene has been associated with emotional disorders (Lesch et al. 1996)], or because the social environment exceeds the individual’s capacity to adapt emotionally. As we have discussed, there is a compelling argument that humans are not evolved to live in a “concrete jungle” with an enormous and broad level of interpersonal interaction. The constructs of hierarchy, family structure, and individual role are now very different from those that existed even five generations ago in Western society, and even more so from

<table>
<thead>
<tr>
<th>Goal or situation</th>
<th>Before</th>
<th>Usual progress</th>
<th>Fast progress</th>
<th>Specific obstacle</th>
<th>Slow or no progress</th>
<th>Success</th>
<th>Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical opportunity</td>
<td>Desire</td>
<td>Productive effort</td>
<td>Flow</td>
<td>Frustration</td>
<td>Resignation</td>
<td>Pleasure</td>
<td>Disappointment</td>
</tr>
<tr>
<td>Social opportunity</td>
<td>Excitement</td>
<td>Friendship and engagement</td>
<td>Gratitude</td>
<td>Anger</td>
<td>Low mood</td>
<td>Happiness</td>
<td>Disappointment</td>
</tr>
<tr>
<td>Physical threat or loss</td>
<td>Fear</td>
<td>Defensive behavior</td>
<td>Confidence</td>
<td>Despair</td>
<td>Despair</td>
<td>Relief</td>
<td>Pain</td>
</tr>
<tr>
<td>Social threat or loss</td>
<td>Anxiety</td>
<td>Defensive behavior</td>
<td>Confidence</td>
<td>Anger or helplessness</td>
<td>Helplessness</td>
<td>Relief</td>
<td>Sadness</td>
</tr>
</tbody>
</table>

Table 11.4 Expected emotions for situations that arise in the pursuit of social and physical goals. From Nesse (2004), with permission.
those that existed prior to agriculture and settlement. This dramatic change in social environment may have exceeded the capacity of some people to adjust, and the consequences are emotional disorders such as anxiety and depression.

11.6.2.1 Anxiety

There is ample empirical evidence that anxiety can have adaptive value. Within a school, timid fish are more likely than bold fish to survive a predator. Anxiety is a way of being alerted to potential danger, but, even so, no individual can sustain a state of maximum alertness all the time. Anxiety has an energetic cost resulting from physiological changes such as sweating and increased metabolic rate (the effects of catecholamine release), blood pressure, and heart rate. As discussed above, chronic stress and chronic anxiety states are very similar, so, not surprisingly, both are maladaptive and can have deleterious effects on the individual.

11.6.2.2 Phobias

Phobias may represent exaggerated forms of what would otherwise be healthy, adaptive responses that enhanced survival in their original form. Most phobias arise from situations of perceived danger from attack, predation, or trauma. Agoraphobia and claustrophobia can be interpreted as originating from fear of exposure to attack or inability to escape. Transient phobias are normal, and fear is an important part of learning what to be afraid of. Failure to be fearful would have been a selective disadvantage, and it has been pointed out that the costs of an exaggerated fear response are less than the costs of not being fearful and being eaten! A psychological disturbance arises when there is loss of the capacity to distinguish between legitimate causes of fear and responding inappropriately to innocent stimuli. In the latter case, the psychological responses become maladaptive because they inhibit the capacity of the individual to function in society. Thus phobias can be seen as having a hypersensitive “smoke detector” function and this understanding has proved valuable in developing therapeutic approaches (Nesse 2001).

11.6.2.3 Depression

Happiness and sadness are universal human emotions, and changes in mood are a normal part of the life course. Sadness is a normal response to a fitness-impairing event such as loss of a child, a spouse, material resources, or injury to the group in which the individual lives. Sadness may be a way of becoming transiently demotivated, thus helping the individual to avoid making decisions under stressed conditions that might be maladaptive in the long run. Being sad may lead the individual to stop a behavior that had caused an initial loss, and promote a period of self-reflection. It may induce a pause and a period of rethinking that could rebuild a threatened pair-bond relationship or change a hunting strategy that had led to the loss of a group member. It may stop the individual confronting a more powerful member of the group following loss of power or status, a situation likely to lead to an adverse and perhaps fatal outcome. Equally, sadness may be a way of communicating within a group the need for support from other group members.

It is argued that the emotion of sadness emerged to act as a brake on some behaviors, and happiness as a way of promoting others. Essentially, they are tools for changing the responsiveness of an individual to a particular situation and communicating with others in the group. While these emotions may have evolved in such a manner, they have been incorporated into other components of our existence. Inappropriate extremes of mood are reflected in pathological depression or hypomania.

The socioeconomic gradient in health is well recognized. A major part of our construct as a social species involves hierarchy and control. Loss of esteem, reputation, or power induces depression. This can be seen as a response to reduced status in the battle of sexual or social selection, and as a consequent form of submissiveness. It may allow the individual to survive to reproduce rather than being killed or expelled from the group. There is ample evidence that individuals who are disempowered or at the bottom of hierarchies are more stressed, and have more emotional disorders and physical illnesses (Box 11.10).
11.6.3 Psychoses

Hallucinations, paranoia, detachment from reality, and withdrawal are symptoms of psychoses, and in particular the schizophrenia syndromes, which affect about 1% of all populations. Generally, schizophrenics behave as if they are living in a different world from others in their social group. Developmental and genetic factors both appear to play a role. People with schizophrenia are characterized by having deficiencies in their ability to exhibit higher orders of intentionality, and there are data to suggest that early life factors may have disturbed the normal development of their neocortex. For example, there is an increase in the incidence of schizophrenia in offspring whose mothers experienced famine when pregnant or were exposed to viral infection during pregnancy (Susser and Lin 1992). This finding suggests that there can be developmental constraints imposed on behavior which are later exposed as psychopathology in some individuals. There are also data showing that those born small are more likely to develop depressive disorders.

However, it is well recognized that there are also genetic determinants of schizophrenia, and when a common disease with genetic determinants persists at a steady proportion in a population the question of whether there has been a selective heterozygote advantage will arise. That schizophrenia may be a result of balancing selection was first proposed as early as the 1960s by such leading theorists of evolutionary biology as Julian Huxley and Ernst Mayr (De Bont 2010). But the selective advantage of schizophrenia has remained elusive, the most popular proposals being (without any compelling evidence) enhanced creativity and novelty seeking. The role of schizophrenic and drug-induced hallucinations in the origin of belief in the supernatural and religion is an equally fertile ground for wild speculation.

Most recently the origin of schizophrenia has been explained using the “imprinted brain” hypothesis (Crespi and Badcock 2008). This hypothesis relies on the concept of genetic conflict and the phenomenon of imprinting (see Box 8.4) arguing that, because the father of a child may not father the mother’s later children, it is in his interest to maximize the child’s growth even at the expense of maternal health and future reproductive success. The mother, by contrast, would do better to limit the investment in the child to preserve some resources for future (and possibly existing) children. As a result, genomic imprinting with a maternal bias would arguably lead to a smaller child, more tractable behavior, and in its extreme form, schizophrenia. By contrast, a paternal bias would lead to a large child, willful behavior, and, in the extreme form, risk of autism spectrum disorder. This theory has possibly received some support from Danish health registry data that indicated that babies of above average size had a significantly higher risk for autism spectrum disorder and a lower risk for schizophrenia, while babies below average size had a lower risk for autism spectrum disorder and a higher risk for schizophrenia (Byars et al. 2014). The increase in risk, however, appears small, and it is also not clear if there other causal factors at play: for instance, while for schizophrenia the relative risk decreased from the smallest birth weight towards larger, in the case of autism spectrum disorder the curve appeared almost U-shaped, with small babies (under 2500 g) having a similar relative risk to babies weighing 4500 g. This remains a field full of speculation, with some association studies but no compelling conclusions possible at the moment.

11.7 Conclusion

Through evolutionary processes we have developed the capacities to be prescient, to negotiate, and to infer the motivation of others. From this perspective, the need to exhibit reciprocal altruism in order to live successfully in a group has caused us to develop concepts of morality and to create belief systems. These functions give most individuals the ability to adjust their behaviors to live successfully within the context of their own society. Thus, considering behavior as either purely genetically or culturally determined is an unsustainable position: both elements clearly interplay constantly through the life course.

Various academic disciplines have addressed the question of the evolutionary determinants of behavior. Each approach is limited by the nature of the available evidence. While this constraint has
led to a debate that has been characterized at times by extreme and sometimes misrepresented and sensationalized positions, a consensus has largely emerged: the brain is an evolved structure, and essentially all its components including the neural substrate exist because of their past or current adaptive value. This does not of course mean that all modern behaviors have to be seen as adaptive attributes. An evolutionary perspective is crucial for understanding how and why the brain operates as it does. But humans are social and prescient individuals, and our evolutionary success depends in part on our capacity for associative learning and for higher levels of intentionality. This creates a broad adaptive capacity which allows humans to exist in a wide range of social and societal environments. Because of these capacities, we have developed a cultural repertoire that is very different in nature and complexity from that of any other species. This cultural development has led to progressive changes in our social environment, and it is the interplay between our cultural and biological evolution and current circumstances and social conditions that is reflected in our behaviors.

Our evolution of a large neocortex was accompanied, and promoted, by the development of language and the invention and application of technology. It led us to develop the relatively complex group structures in which humans still live. Having this large brain determined our life history with its long period of infant dependence on adults and the unique juvenile period prior to sexual maturation, adolescence.

Human life is dependent on a complex web of interactions with other humans. We evolved with the capacity to interpret the actions of other members of our species, and to communicate and negotiate with them. The evolution of human behaviors can be understood in terms of the dynamics of mate choice, kin selection, reciprocal altruism, and social group living. Many human behaviors have evolved to address the challenge to the group posed by freeloaders. Emotions are assumed to have evolved because they have had adaptive value for a social species, but

---

**Box 11.10 The Effects of Social Stratification on Health**

The hierarchical organization of social classes within a society by wealth, power, education, ethnicity, or gender is known as social stratification. Social position essentially reflects a differential access to capital, and is positively correlated with health and life expectancy. Studies by Michael Marmot, one of the pioneering researchers on the effects of this gradient, showed that inequalities across social hierarchies exist not just between developing and affluent countries but within countries and even within societies as well (Marmot 2004).

While factors like stress, poor diet, and lack of medical care can cause poor health, social hierarchy has been dubbed the underlying “cause of the causes.” In his work on British civil servants, Marmot found that low-status workers who felt they had no control over their daily activities were much more vulnerable to stress-related disease than were executives at the top of the hierarchy, even though these are just the people whom we would normally consider to have highly stressful jobs. But importantly there was a continuous gradient of disease risk across the hierarchy, so that even second-tier people in well-paid managerial jobs had twice as much stress-related disease as did those at the top of the pyramid (Marmot et al. 1991).

Much of our knowledge about how being at the bottom of the social hierarchy results in ill-health comes from studies on baboons by Robert Sapolsky (Sapolsky 2005). These animals, like humans, live in competitive and stratified groups. Sapolsky’s field studies with baboons have shown that animals with the lowest social standing have high stress hormone levels, which in turn leads to disease.

The medical anthropologist William Dressler has studied stress in relation to cultural change, caused for example by economic development or by migration of an individual to a new community, and has introduced the concept of “lifestyle stress.” Working with populations undergoing economic change, and using blood pressure as an outcome measure of stress-related disease, Dressler showed that neither economic status nor material lifestyle alone predicted high blood pressure, whereas a discrepancy or mismatch between economic status and material lifestyle (in other words, living beyond one’s means) did (Dressler 1990).
they can become maladaptive with consequences for psychological and psychiatric well-being. This maladaptation may have arisen because of the massive change in the human social environment. Genetic or developmental variation may also affect the pathways determining behavior. Either way, the match between the individual and the environment has changed.

**Key Points**

- Human behavior is built on selected, and therefore genetically determined, components of brain function.
- The evolved brain is the substrate on which individual experience and the current environment shape abilities and behavior, giving humans the flexibility to exist in a wide range of societal environments.
- Humans are social animals characterized by living in groups larger than their immediate family. Selection has favored traits that promoted fitness within this environment, such as cooperation, reciprocal altruism, and the abilities to interpret the actions of other members of our species and to detect freeloaders.
- Emotions have adaptive value for a social species, but they can become maladaptive with consequences for psychological and psychiatric well-being.
- Such maladaptations may have arisen because of changes in the human social environment, or because of genetic/developmental factors creating functional variation in pathways determining behavior.
CHAPTER 12

Cancer

12.1 Introduction

Cancer is not a single disease entity: there are many pathways that can lead to uncontrolled replication of a cell population (carcinogenesis), the spread of a tumor into nearby tissue (invasion), and the migration of a tumor from the tissue where it originated to elsewhere in the body (metastasis). The complexities of the pathways leading to the apparently common phenomenon of uncontrollable cellular replication mean that while some specific and generally rare cancers can be traced to a singular initiating event, such as asbestos-induced mesothelioma, most cancers—even of a single type—do not have a single cause, and hence prevention and therapy are not straightforward.

Early in this book we discussed levels of selection, whether at the level of the gene, where one view suggests that an individual organism should be seen as a transient coalition of genes in which specific genes are co-opting other genes to promote their mutual survival via transmission, or at the level of the organism itself, because it is the phenotype of the organism that is the unit of functional selection and differential reproductive success (see Sections 2.3.2.4 and 2.3.2.5).

In cancer, the cell can be seen as the unit of selection, and its clonal expansion can be interpreted as a situation in which one cell lineage is proliferating at a cost to other cells in its neighborhood (and often, eventually, to their “vehicle”, the body). The cancer cell lineage is distinguished by having the capacity to override the normal processes inhibiting its growth, and its cells are effectively selected for this advantage, thus replicating and developing autonomy. Given that multicellular organisms are evolved coalitions of cells, cancer can be seen as a form of atavism where unicellular interests again become dominant, and selection operates at the level of the cell. Thus the proximate cause of cancer can be envisaged as the somatic mutations induced in the clone that give a particular lineage of cells a replicative advantage. But evolutionary processes provide the ultimate cause in that these cells win the competition for selection. And, as we will see later in this chapter, cancer cells have particular abilities that make them more able to adapt to the selective pressures of the environment.

Consider the very nature of a metazoan: its life history originates from a single pluripotent cell which ultimately differentiates into multiple cell types and forms organs with residual capacities for repair, which is in turn dependent on subpopulations of stem cells. It needs to organize and regulate its development, a process that frequently involves not only the promotion of cell growth and differentiation but also the regulation of apoptosis (programmed cell death). Mutational escape from each of these multiple layers of control is the primary proximate mechanism for the origin of malignancy. However, most cells that undergo a mutation that could develop into a malignancy do not survive the biological bottlenecks (tumor suppression mechanisms) that place limits on the ability of the cellular clone to grow and become autonomous.

Nowell (1976) realized a generation ago that selective processes of somatic evolution operate within a tumor to generate sub-clones that vary phenotypically with respect to traits such as the capacity for growth or drug resistance, and that these evolutionary processes might best be countered by designing individualized therapies for patient and tumor characteristics. Recently we have come to understand that metastases undergo further mutations, and that
each metastasis may have a different genetic profile from the primary tumor, thus leading to further therapeutic challenges and demonstrating successful (from the tumor’s perspective) evolution.

What are the selective pressures to which a cancer cell is exposed? As discussed later, these are likely to be analogous to those experienced by organisms in an ecosystem, including factors such as resource availability (distance from vasculature) and predation (chemotherapy and immune attack from the host). We will see that some of the behaviors of cancers can be explained in terms of such an ecological approach.

In this chapter, we explore the origins of cancer in metazoan evolution, consider tumor growth and metastasis as an evolutionary process, describe how evolutionary mechanisms can explain the growing incidence of certain malignancies in the human population, and explore how these perspectives can suggest new pathways for cancer therapy. We will consider the evolutionary aspects of cancer from two perspectives: first, that of an individual patient, where somatic evolution within their tumor is a critical factor in treatment; and second that of the population, where consideration of evolutionary mechanisms that affect the incidence of cancer might provide clues for disease prevention (Aktipis and Nesse 2013).

12.2 Epidemiology of Cancer

Life expectancy is increasing in high- and middle-income countries, and more recently in some low-income countries too, largely as a result of public health measures. But as the burden of deaths caused by infections decreases, and life expectancy increases, the incidence of non-communicable diseases such as cardiovascular disease, type 2 diabetes, and cancers is increasing.

The reasons for the increasing incidence of cancer are complex, and in part reflect increases in life expectancy (cancer is predominantly a disease of old age and about 75% of cancers in the UK occur in people aged over 60) and improvements in diagnosis. Still, there are wide variations in incidence by cancer type, risk factor, and geographical region. For example, in many high-income countries lung cancer rates are decreasing in men but increasing in women, reflecting changes in smoking habits. Breast cancer incidence is increasing worldwide, reflecting the changes in risk factors to be discussed in Sections 12.5.2 and 13.5.1. Rates of cervical cancer have been declining in many high- and middle-income countries, probably because of effective screening programs and the introduction of vaccination against human papillomavirus.

Malignant disease will ultimately affect more than a quarter of the population of high-income countries. In 2012, cancer caused over 8 million deaths worldwide (15% of all deaths; more than AIDS, malaria, and tuberculosis combined). In high-income countries, cancer was the second most common cause of death, after cardiovascular disease. In low- and middle-income countries, the high mortality caused by infectious and parasitic diseases meant that cancer was the third most common cause of death. However, the high burden of infection in low- and middle-income countries means that nearly a quarter of newly diagnosed cancers in these countries are attributable to infections (stomach cancer being an example) while less than a tenth of cancers in high-income countries have an infective origin (American Cancer Society 2015).

Incidence rates for particular neoplasms vary widely across the world. Rates for some cancers, for example melanoma or esophageal cancer, can vary by up to 200-fold depending on geographical location. In the gastrointestinal tract, colorectal cancer and stomach cancer show very different geographical distributions, with the former having its highest incidence in Europe and North America and the latter in Asia. Liver cancer has its highest incidence in Africa, largely explained by the high rates of infection with hepatitis B and C. The incidence of prostate cancer also varies widely (Figure 12.1), likely as a result of nutritional and life-history factors (Section 12.5.2).

Environmental rather than genetic factors play a large role in these distributions of cancer incidence, as can be demonstrated by studies of migration between areas of different risk. For example, compared with non-migrants, people from northern Europe who migrate to semi-tropical areas experience greater rates of melanoma, whereas individuals moving from high-risk to low-risk areas for stomach and liver cancer experience lower rates
of these neoplasms (Vineis and Berwick 2006). For some neoplasms, diet contributes to disease risk, possibly via an effect on the gut microbiota (O’Keefe et al. 2015; Box 9.3). In Section 12.5 we provide an evolutionary and life-history framework for these and similar observations. Nevertheless, some population groups are more frequent carriers of founder mutations affecting cancer risk, such as the BRCA1/2 mutations associated with breast cancer found in Ashkenazi Jews.

But although the increasing incidence of cancer has led some to suggest that cancer is just a disease of “modern lifestyles” this is not the case. Although most cancers occur in soft tissue and would not be well preserved in skeletal remains, evidence of cancer (bone metastases originating in soft tissue neoplasms, as well as primary tumors such as osteosarcoma) can be found in fossilized human bones, including an osteosarcoma in the “Kanam jaw” from an early Homo species dating from around 500,000 years ago. Early Egyptian writings contain descriptions of what can only be ulcerating tumors, and comment on their refractoriness to treatment. Nearly 2500 years ago, the Greek physician Hippocrates described tumors by using the Greek word for crab (karkinos), the origin of the word cancer, and his later compatriot Galen referred to tumors as oncos (swelling), providing our modern name for the discipline of oncology.

12.3 Ecology of Cancer

Cancer is a disease of multicellular organisms, and examples of cancer or cancer-like disease
Cancer can be found in many metazoan taxa, both invertebrate (nematodes and insects) and vertebrate (fish, amphibians, reptiles, birds, and mammals). Although plants do develop localized tumors, most often as a result of microbial infection, such growths are unlikely to spread to other parts of the plant because plant cells are held in place by rigid cell walls that do not allow migration.

The transition from unicellular to multicellular organisms (Box 2.9), which occurred several times around 600–1000 Mya in different lineages that gave rise to present-day animals and plants, required the evolution of mechanisms to promote cooperation among collections of cells. How this originated is unclear, but one example may be illuminating. In the multicellular alga *Volvox*, the gene controlling cell differentiation to produce germline (reproductively active) and somatic (body) cells has been co-opted from a similar gene in a unicellular ancestor that controls life-history decisions (survival versus reproduction) in response to environmental cues (Hanschen et al. 2014). As organisms became more complex, mechanisms evolved to allow temporal and spatial control of cellular division, differentiation, and migration to produce function-specific tissues and organs. These mechanisms include limitations on the ability of some cell lineages to divide indefinitely, a range of internal quality control mechanisms that identify and destroy damaged cells that might proliferate uncontrollably, epigenetic control of differentiation into particular cell lineages, and sensitivity to physical and chemical growth signals generated from the surrounding tissue or from specialized remote tissues elsewhere in the organism. In Section 12.4.1 we consider how abrogation of these mechanisms characterizes some of the “hallmarks of cancer.”

Multicellularity involves a form of cellular altruism. All cells in a multicellular organism have the same genotype (but a different epigenotype), and multicellularity requires individual cells to cooperate in a way that allows the whole organism to represent the interests of the genotype; this involves limitations on the behavior and even the existence of particular single cells. Contrast this with the “selfish” ability of a single-celled organism to behave in whatever way optimizes the propagation of its genome.

Cancer can be imagined as a re-adoption of such “selfish” behavior. The proliferating cancer cell outcompetes its neighboring normal cells and is thereby highly successful in the propagation of its particular genotype over a short timescale, but if its aberrant behavior kills the host the cancer genotype dies with it (but see Box 12.4).

12.4 The Biology of Cancer

Cancer is a disease characterized by the uncontrolled growth (malignant transformation followed by sustained proliferation) and spread (invasion and metastasis) of abnormal cells. The origin of a cancer is a single cell that acquires mutations that cause it to proliferate within its tissue at the expense of other cells, so in evolutionary terms the cancer cell has greater fitness.

There are two main types of genes implicated in the initial malignant transformation: oncogenes and tumor suppressor genes (Figure 12.2). In normal cells, proto-oncogenes (the “wild-type” precursors of oncogenes) are generally involved in growth-promoting pathways, whereas tumor suppressor genes are generally involved in repair or growth-limiting pathways. For example, the HER-2/neu gene codes for an epidermal growth factor receptor on the cell surface, and mutations that lead to overexpression of this protein, which are often found in breast cancer, cause cells to respond vigorously to the growth factor by proliferating. The drug trastuzumab (Herceptin®) blocks this pathway in individuals with this mutation. Conversely, the p53 tumor suppressor gene, which is mutated in over 50% of human cancers, normally functions to limit proliferation by monitoring cellular stress and initiating DNA damage repair and cell cycle arrest if necessary.

Cells with typical cancer-associated mutations can be detected at a surprisingly high rate in early life—much higher than the incidence of their associated cancers—suggesting that further mutational “hits” to a particular cell during later life are required to generate a proliferating clone with a selective advantage (Mori et al. 2002). Although a cancer may involve hundreds or thousands of mutations, only relatively few (perhaps five or ten) actually act to “drive” tumor formation; among
these, loss-of-function mutations in tumor suppressor genes are more common than activated oncogenes. It is these “driver” mutations that contribute to the increased fitness of a cancer cell in the tumor environment.

As the cancer cell clone expands, further mutations occur in particular cells to generate sub-clones with different genetic profiles that share only some of the “founder” mutations. Mutations in genes that enforce genetic stability, such as \( p53 \) or \( BRCA \), may in turn lead to sub-clones with a so-called “mutator phenotype,” having an increased mutation rate and an increased load of oncogenic mutations. In this way, a “successful” cancer acquires mutations that enable sub-clones to support vascularization, to break down physical barriers to allow expansion, and to evade the immune surveillance of its host organism.

Stem cells are present in small numbers in normal adult tissue and allow a degree of tissue regeneration and repair. Stem cells are able to self-renew, possibly indefinitely, by mitotic division, but are also able to differentiate into the specialized progeny cells required for tissue replenishment. This differentiation means that the progeny cells also acquire replicative senescence and a limited lifespan, itself a tumor-suppressing mechanism. Since tissue stem cells already possess the capacity for self-renewal, they are likely candidates for the founder cells of a tumor if the mechanisms that normally control their self-renewal and induce differentiation are abrogated by driver mutations. Trade-offs between tissue repair capacity and susceptibility to cancer (Section 12.5.2) may be actualized by this role for stem cells in carcinogenesis.

It remains unclear whether tumors contain particular subpopulations of cancer stem cells characterized by longevity and proliferative capacity that act as the key targets of selection, or whether all cancer cells have this property (Sprouffske et al. 2013). Certainly, primary tumors contain numerous types of non-cancer cells, including fibroblasts and epithelial cells, that together are termed the tumor stroma and that contribute to the microenvironment of the cancer cells. Because of their association with the cancer cells, these stromal cells acquire modified properties that contribute to tumor progression.

12.4.1 Hallmarks of Cancer

Although cancers can develop in most if not all tissues, and vary widely in their ability to proliferate and spread, all cancers share certain common traits at the molecular and cellular level (Hanahan and Weinberg 2011). A general feature of these hallmarks of cancer, and one that was discussed earlier in the context of the origins of cancer, is that they concern the environment of, and the relationships between, cells in the context of organized tissues.
In other words, they affect features of multicellular systems in which cells have evolved to cooperate and foster the interests of their genotype within a complex organism. These hallmarks include:

- **The ability to proliferate without the need for external growth signals.** Normal cells require a supply of growth factors for repeated cell division. Cancer cells do not, because they express their own, over-express the receptors for those growth factors (e.g., HER-2), or express mutated growth factor receptors that issue the “growth” signal without the need for an external ligand.

- **The ability to evade growth suppression.** The growth of normal cells is checked by negative signals from the extracellular matrix or from surrounding cells (“contact inhibition”). Cancer cells are insensitive to such signals.

- **Resistance to apoptosis (programmed cell death).** Normal cells carry out a form of internal quality control, checking for events such as irrepairable DNA damage or extreme hypoxia. Failing these checks causes the cell to undergo apoptosis. Cancer cells have defects in their apoptotic mechanism, causing them to become insensitive to apoptotic signals. The tumor suppressor gene p53, which undergoes loss-of-function mutation in many cancers, is part of the mechanism of programmed cell death, responsible for detecting DNA damage.

- **An unlimited ability to replicate.** Normal cells of most types cease to replicate after a certain number of divisions, a process mediated by the division-dependent shortening of DNA regions called telomeres at the end of chromosomes. In normal cells, the activity of the telomere-lengthening enzyme telomerase is very low, but many cancer cells have increased levels of telomerase, which over-rides the shortening process and allows an unlimited number of cell divisions.

- **An increased ability to induce the formation of new blood vessels (angiogenesis).** The rapid proliferation of cancer cells can be resource-limited by the supply of nutrients and oxygen. Cancer cells develop the ability to promote angiogenesis, ensuring their continued growth. The importance of resource constraints in the ecology of tumors is discussed in Section 12.4.2.

- **An uncontrolled capacity for invasion and metastasis.** The spread of cells within tissue, and their spread to other tissues, is normally tightly regulated by the control of cell–cell interactions via cell adhesion molecules, and by restricting the expression of enzymes that allow cells to move through the extracellular matrix. These control mechanisms are lost in cancer cells, allowing them to invade local tissue and to metastasize to distant sites. In particular, the increased capacity for tissue invasion appears to occur via co-option of an ancient cellular process essential for the formation of complex body patterns during embryonic development, termed the “epithelial–mesenchymal transition.”

- **Altered energy metabolism.** Although cancer cells are able to induce angiogenesis to improve their supply of nutrients, some regions of solid tumors are often inadequately perfused and are nutrient- and oxygen-poor. Under these conditions, cancer cells are often able to adapt by switching their metabolic pathways, particularly towards the use of anaerobic (non-oxygen-using) pathways such as glycolysis (the Warburg effect).

- **Evasion of immune-mediated attack.** The body’s immune system can usually identify and destroy infected or damaged cells, and such immune surveillance is part of normal tumor suppression mechanisms. However, cancer cells can evade the body’s immune system by mechanisms such as downregulating the expression of immune-mediated molecules on their cell surface or by secreting immunosuppressive cytokines.

- **Genome instability.** Progression of a cell to the state of uncontrolled proliferation and migration characteristic of cancer requires several of the above hallmarks to be present, in turn requiring sequential accumulation of enabling mutations in the expanding cancerous clone. The underlying state facilitating this high rate of mutation is called genomic instability. Instability may occur at the nucleotide level, causing changes in the DNA sequence often associated with impaired DNA repair mechanisms, or at the chromosome level, causing deletion, duplication or inversion of whole chromosomes or chromosome segments. The BRCA genes, frequently mutated in several cancers and discussed in Section 12.5.2 in
the context of life-history trade-offs in cancer, are involved in DNA repair.

- **Inflammation.** Another underlying characteristic of cancers is the role of local inflammation around the tumor site in promoting a facilitating environment for tumor growth and spread. Increased concentrations of inflammatory cells and pro-inflammatory cytokines at the tumor site may create conditions that promote tumor mutation, growth, and invasion.

### 12.4.2 Tumor Heterogeneity and its Consequences

Although cancers originate from a single cell with a discrete number of founder mutations, by the time a tumor is clinically apparent it will contain many billions of cells, each containing not only the founder mutations but also thousands more mutations acquired during rapid cellular proliferation. It is known that some bacteria are able to increase their mutation rate under stress, allowing more rapid adaptation to environmental events, but the existence of a similar “mutator phenotype” in cancer cells remains controversial. Some oncologists argue that the high prevalence in cancer cells of defects in DNA repair mechanisms, leading to high mutation rates and genomic instability, is evidence of this phenotype, whereas others argue that the mutation rates observed in normal cells together with the high proliferation rate of cancer cells are sufficient to account for the multiple mutations found in tumors (Fox et al. 2013).

Nevertheless, the observation of a high mutational load in cancers, together with demonstration of mutations conferring resistance to chemotherapy in so far untreated tumors, and of early mutations promoting metastasis in primary tumors, underlines that tumors contain high levels of genetic variation. This variation is both heritable, as tumor cells divide mitotically into daughter cells, and clinically relevant in terms of disease progression and treatment.

The existence of heritable (at a cellular level) genetic variation within a tumor implies that, if the tumor environment is not constant, some tumor cell clones will be at a selective advantage on the basis of their greater fitness within their current environment (Figure 12.3). The selective environment will vary spatially across the tumor depending on factors such as oxygen and nutrient availability, which in turn depend on proximity to the tumor vasculature. The presence of stromal and immune cells also contributes to the environment. Hence multiple sub-clones will exist within the tumor depending on their fitness in their particular microenvironment. Features of this intra-tumoral heterogeneity include different spatial patterns of mutation, convergent evolution (distinct mutations of the same tumor-suppressor genes in different parts of the tumor), and different prognostic signatures of gene expression in various tumor regions (Gerlinger and Swanton 2010). The process leading to this heterogeneity has been termed somatic evolution.

In addition to these microevolutionary events, macroevolutionary events analogous to speciation may lead to large changes in tumor phenotype, driven by genomic instability causing gross changes in the tumor genome, such as chromosomal rearrangements.

The success of chemotherapy, which applies a further selective pressure to the tumor, will depend on the relative fitness of susceptible and resistant clones. Tumors with high levels of genetic variation, and thus likely higher genetic instability that can be co-opted to evolve resistance mechanisms, would be predicted to be more resistant to chemotherapy. Moreover, if the bulk of tumor cells are sensitive to chemotherapy, killing them can provide resistant cells with a less-populated environment into which to expand. In Section 12.6.2 we discuss the implications of tumor heterogeneity for cancer treatment.

Finally, the process of metastasis can be seen as an evolutionary bottleneck analogous to a founder effect (see Section 3.4.1) in the evolution of a population. Consequently, metastasized tumor cells will, at least initially, contain only a proportion of the genetic diversity of the primary tumor. Metastases are also faced with the challenge of adapting to their new tissue environment (Section 12.6.2); this may result in the evolution of a set of mutations distinct from those in the primary tumor.

### 12.5 Cancer in the Light of Evolutionary Mechanisms

In Chapter 7 we provided an evolutionary framework for health and disease, and elucidated
pathways by which evolutionary processes can affect disease risk (Box 7.3). Table 12.1 provides examples of how these pathways apply to the origin of cancers. We expand on some of these mechanisms in the sections below.

12.5.1 A Mismatched or Novel Environment

A number of cancers appear to arise because of exposure to environments or toxins which are novel in an evolutionary sense. One of the first cancers to have a specific etiology linked to an evolutionarily novel toxin was scrotal cancer in chimney sweeps, caused by exposure to tars in soot. The relationship between cigarette smoke, which is full of evolutionarily novel compounds such as inhaled tars, and lung cancer is well recognized. The risk of cancer associated with exposure to high levels of gamma radiation may simply reflect the truism that organisms have not evolved to live in an environment with high background radiation and thus do not have DNA repair mechanisms able to cope with the level of mutational injury caused.

Different lifestyles across populations generate different risks of exposure to evolutionarily novel toxins, and these differences manifest themselves in the variable distribution of some cancers (Section 12.2). Pipe smoking is more likely to be associated with cancer of the lip and tongue, cigarette smoking with lung cancer, and tobacco chewing with oral cancer. The increase in meat consumption and reduction in dietary fiber has been linked to colon cancer (O’Keefe et al. 2015). Highly salted and pickled foods have been associated with stomach cancer, and the marked reduction in the prevalence of gastric cancer in Europeans in recent decades has been postulated to be linked to the reduction...
in consumption of such foods. Excessive alcohol intake is associated with esophageal and stomach cancer, and possibly also with breast cancer.

Another example of the carcinogenic effects of novel environments is the high rates of skin cancer, including melanoma, seen in light-skinned northern Europeans who migrate to areas with high levels of ultraviolet exposure from sunlight (particularly Australia and New Zealand, where the atmospheric ozone barrier is reduced). Here we are seeing a double effect of migration: depigmented skin evolved in human populations who had migrated outside the tropics, probably to maintain vitamin D biosynthesis (Section 13.4.1), and the reversal of that migration now results in high rates of skin cancer in those who lack the protective pigment melanin. As would be expected, migration into Australasia of darker-skinned individuals does not appear to increase their risk of skin cancer (e.g., Czarnecki 2014).

Hormone replacement therapy extends cyclic exposure to ovarian hormones beyond menopause and has been implicated in increased risk of reproductive cancers (Section 13.5.1; Chlebowski et al. 2003). Bisphenol A and a number of other plasticizer compounds as well as some pesticides can be estrogenic and/or anti-androgenic, either directly or through their metabolites, and again may increase the risk of reproductive cancers (Fenichel et al. 2013). Another factor may be the marked change in our nutrition, meaning that levels of growth-promoting hormones such as insulin-like growth factor 1 are higher from childhood and this may provide an environment in which clonal expansion of mutated cells is less restrained (Cohen and LeRoith 2012).

Finally, since most cancers are diseases of old age, it could simply be argued that just living longer allows more spontaneous somatic mutations to arise, as there has been greater cumulative exposure to environmental mutagens (see also Section 12.5.2). In recent centuries, our average life expectancy has risen dramatically from a relatively constant figure over the previous 150,000 years. This, combined with the marked changes in the environments we inhabit (that may expose us to more mutagens), generates a simple stochastic model to explain the association between cancer and ageing.

### 12.5.2 Life History and its Trade-offs

As discussed in Chapter 5, species have evolved their life histories to maximize reproductive success. Many of the traits underlying these life histories involve the relative allocation of resources among growth, reproduction, and maintenance. Thus, some species have evolved “fast” life histories that feature rapid growth to small body size, high reproductive effort, low levels of tissue repair, and short lifespan, whereas others have “slow” life histories with slow growth to a larger body size, lower (but likely prolonged) reproductive effort, investment in tissue repair, and a long lifespan (Jones et al. 2014).

Accordingly, species will have evolved tumor suppressor mechanisms appropriate to their “chosen” life history, in part explaining the similar rates of cancer in small and large species (Box 12.1). With

### Table 12.1 Pathways by which evolutionary processes can affect the risk of cancer

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>An evolutionarily mismatched or novel environment</td>
<td>Tobacco smoking; migration and skin cancer</td>
</tr>
<tr>
<td>Life-history-associated factors</td>
<td>Breast cancer</td>
</tr>
<tr>
<td>Excessive or uncontrolled defense mechanisms</td>
<td>Inflammation as a causative or contributory factor in carcinogenesis</td>
</tr>
<tr>
<td>Consequences of coevolution with microbes</td>
<td>Helicobacter pylori and gastric cancer; hepatitis B/C and liver cancer</td>
</tr>
<tr>
<td>Results of evolutionary constraints</td>
<td>No selection pressure for increased later-life tumor suppressive mechanisms</td>
</tr>
<tr>
<td>An apparently harmful allele is maintained by balancing selection</td>
<td>Possibly maintenance of BRCA mutations in the population</td>
</tr>
<tr>
<td>Sexual selection and competition and its consequences</td>
<td>High testosterone levels in males related to increased incidence of prostate cancer</td>
</tr>
<tr>
<td>The outcomes of cladal and demographic histories</td>
<td>Some evidence for variation in cancer incidence by ethnicity after controlling for economic, lifestyle, and environmental factors</td>
</tr>
<tr>
<td>Cultural evolution</td>
<td>Breast cancer and changes in reproductive schedules; hormone replacement therapy</td>
</tr>
</tbody>
</table>

Table 12.1 Pathways by which evolutionary processes can affect the risk of cancer
some exceptions (see Chapter 2), selection in general acts to maximize reproductive success rather than longevity, and this declining force of selection with age predicts that cancer suppression mechanisms will be relatively less effective in ageing organisms. This may explain in part the increased incidence of cancer as human longevity is increased by improved nutrition and more effective public health interventions.

Other more specific trade-offs may occur. For example, tissue repair processes such as wound healing require a supply of stem cells able to proliferate and migrate to or within the injured site, as well as the capacity to form new blood vessels, and inflammation occurs at the site of the wound. These capacities are reminiscent of some of the hallmarks of cancer (Section 12.4.1), and similar molecular mechanisms appear to be involved (Neves et al. 2015). It may be that species that have evolved highly effective wound healing mechanisms, possibly as a result of behaviors that risk physical injury, will be more susceptible to the development of cancer. Similarly, there may be correlations between placental invasiveness and susceptibility to tumor metastasis (Box 12.2).

There may also be trade-offs between reproduction, ageing, and susceptibility to cancer. One molecular mechanism mediating such trade-offs involves the tumor suppressor gene $p53$, which responds to cell damage by initiating cell cycle arrest or apoptosis. We have already seen evidence that larger and long-lived animals have more copies of this gene (Box 12.1). There is emerging evidence that $p53$ also has roles in ageing (probably mediated by impaired tissue repair) and reproduction (probably mediated in mammals by effects on embryonic implantation). Although $p53$ loss-of-function mutations are a major cause of human cancers, the

### Box 12.1 Peto’s Paradox

“The mere existence of whales suggests that it is possible to suppress cancer many-fold better than in humans” (Caulin and Maley 2011). If cancer is initiated by one or a series of mutations in a single cell, and the likelihood of such mutations occurring increases over time, then large long-lived animals (e.g., whales) would be predicted to have higher rates of cancer than small short-lived animals (e.g., mice). If the blue whale developed colorectal cancer at the rate that is typical of humans, then half of all blue whales would have developed this malignancy by the age of 50 years. But they don’t, and this species lives to well over 100 years. This lack of a correlation across species between size and cancer rates is called “Peto’s paradox,” after the Oxford epidemiologist who first drew attention to it (Peto et al. 1975).

A number of mechanisms have been suggested to explain this apparent paradox, although rather little experimental evidence is available—whales and elephants aren’t ideal laboratory animals! First, larger animals, which tend to be long-lived and reproduce slowly, may have lower rates of metabolism, generating fewer of the mutagenic by-products known as reactive oxygen species; their cells may also divide less frequently, reducing the opportunity for copying errors during mitosis. Second, tumor suppression mechanisms may have evolved to be more effective in longer-living animals as a consequence of greater investment in repair mechanisms—it is known that elephants have multiple copies of the key tumor suppressor gene $p53$ whereas humans have just one (Caulin and Maley 2011). Since there seems to be a negative correlation (trade-off) between oncogene activation and tumor suppression on the one hand, and fertility on the other hand (Section 12.5.2), the lower cancer rates in larger slowly reproducing animals may represent one extreme value of the trade-off.

There are some exceptions to the paradox. The mole rats, a rather unattractive underground-living species of small rodents, live for 20–30 years in contrast to the 4–5-year lifespan of similarly sized domestic rats and mice, and are reported to be “almost cancer-proof.” Mole rats appear to have evolved rather unique antioxidant and tumor suppressive mechanisms that probably contribute to both their longevity and their cancer resistance (Gorbunova et al. 2014).

Finally, the lack of correlation between body size and cancer rates does not appear to operate within species. Larger dog breeds have shorter lives and are more susceptible to cancer than their smaller conspecifics. Taller humans have been shown to have a higher risk of developing cancer, with a relative risk of 1.1 for every 10 cm increase in height (Green et al. 2011). This may reflect higher circulating concentrations of growth factors in both larger dog species and taller humans.
naturally occurring \( p53 \) polymorphisms, which seem to be the targets of selection in some human populations (Shi et al. 2009), appear to have more subtle effects; there are differences between alleles at position 72 in their relative capacities for tumor suppression (R72 allele > P72 allele) and ageing-related functions (P72 > R72). The maternal P72 allele also appears to increase the risk of twin births in humans and, intriguingly, cancer incidence was also higher among the first-degree relatives of these mothers (Tagliani-Ribeiro et al. 2012).

Trade-offs between fertility and tumor suppression are examples of antagonistic pleiotropy (Section 5.2.4). A similar trade-off involving the tumor suppressor gene \( BRCA1/2 \) is described in Box 12.3 and Figure 12.4. Another demonstration of a relevant life-history trade-off in humans comes from the multigenerational Framingham study introduced in Box 5.2, where the negative correlation between female reproductive success and ageing appeared to be mediated by genes linked to cancer susceptibility (Wang et al. 2013).

Some neoplasms are directly linked to periods of particular developmental plasticity, and this can explain the age distribution of certain cancers. For example, tumors of tissues that primarily proliferate in early life generally present as childhood tumors, such as Wilms’s tumor of the kidney or rhabdomyosarcoma of muscle. Osteosarcoma, a tumor of bone, generally appears during periods of rapid peri-adolescent bone growth. In contrast, the risk of most epithelial bone growth is increased as an individual ages because these tissues replicate through life and the risk of multiple mutations increases with age.

Breast cancer, even in the absence of \( BRCA \) mutation, represents an example where cultural change affecting life-history phasing in the form of reproductive behavior has clearly played a role in altering the risk of disease. It has been known for centuries that celibate women are more likely to develop breast cancer (but have less cervical cancer, which is induced by infection with the human papillomavirus via intercourse). Celibate women are characterized by an uninterrupted cycle of exposure to estrogen and progesterone from menarche to menopause and by an absence of lactation. The more children a woman has, the lower her risk of developing breast cancer (Ewertz et al. 1990). Whereas a hunter-gatherer woman might only have 100–150 menstrual
Box 12.3 BRCA1/2 as an Example of Antagonistic Pleiotropy?

The human BRCA1/2 genes are involved in DNA damage repair, and loss-of-function mutations significantly increase the lifetime risk of breast and ovarian cancer. Some germline BRCA mutations show evidence of founder effects, in which the same mutation can be detected in a particular group at a rate higher than in the general population. Although the persistence of these mutations can be at least partly explained by weak selection associated with the generally post-reproductive onset of their associated cancers, a significant proportion of cancers do occur before menopause, raising the question of whether BRCA carriers have some fitness advantage that maintains the alleles in the population through balancing selection.

Indeed, under natural fertility conditions women who were retrospectively identified as carriers of BRCA mutations had on average nearly two more children than did non-carriers (Figure 12.4; Smith et al. 2012). The difference was attenuated in later cohorts of women with access to effective contraception. Although the analysis is complicated by the mild protective effect against breast cancer of high fecundity in BRCA1 carriers, an effect also seen in the general population (see Section 12.5.2), the results do indicate some fertility advantage for BRCA mutation carriers. The molecular explanation for this putative antagonistic pleiotropic effect of BRCA may involve telomere lengthening (Section 12.4.1) after loss of function of BRCA; longer telomeres are associated with increased lifespan but also increased cancer risk, and also enhanced reproduction via an effect on ovarian ageing.

![Figure 12.4 BRCA1/2 mutation carriers have increased fecundity. Under natural fertility conditions (assumed to be mothers born before 1930 who would not have had access to modern contraception), BRCA1/2 mutation carriers in a US population have higher fecundity than non-carriers. The difference is attenuated in mothers born later, who will have had some access to contraceptive methods. From Smith et al. (2012), with permission.](image)

Cycles during her life because of the interruptions of pregnancy and lactation, a modern Western woman may be exposed to 500 cycles. Thus the nature of the hormonal exposures is very different. The result is that the modern woman has constant proliferative stress on the ductal epithelium driven by estrogen and progesterone without any period of intervening lactation. Lactation itself leads to a loss and renewal of ductal epithelial cells, thus removing many cells with somatic mutations. In one Chinese population where women tend to feed only from the right breast, cancer is more common in the left breast, showing that exudation of epithelial cells in milk is indeed protective (Ing et al. 1977).
In men, the incidence of prostate cancer increases strongly with age and shows wide geographical variation (Figure 12.1). The proliferation of this tumor is frequently testosterone dependent, and therapy involves anti-androgenic medication. There is evidence that the risk of prostate cancer is related to lifetime exposure to testosterone, and in turn testosterone levels are influenced by age, by energy intake, and (as well demonstrated in animal studies) by behavioral factors related to aggressiveness. This suggests that human populations characterized by good nutrition and/or a high level of social interaction requiring aggressive behavior should have high testosterone levels in young men and a high incidence of prostate cancer in older men. Such a relationship has been demonstrated by a study showing that testosterone levels of young men are significantly associated with population disparities in the incidence of prostate cancer in older men, providing evidence for a life-history trade-off mediated by testosterone between early reproductive effort and later health (Alvarado 2013).

12.5.3 Coevolution with Microbes

Some 15 to 20% of human cancers are thought to be the result of infections. Examples of these causative organisms include viruses (e.g., herpesvirus-related Kaposi’s sarcoma, hepatitis C virus and liver cancer, papillomavirus and cervical cancer), bacteria (Helicobacter pylori and gastric cancer), and parasites (schistosomiasis and bladder cancer). The causal nature of these organisms is clearly demonstrated by the rapid drop in incidence of their associated cancers after the introduction of vaccination or eradication programs.

Why do infections lead to cancer? One explanation is that the inflammation associated with infection drives oncogenesis, for example by increasing levels of mutagens (such as reactive oxygen species produced by neutrophils) or proliferative signals (growth factors) or by increasing tissue permeability allowing tumor invasion. It is likely that cancers associated with bacterial (e.g. Helicobacter) and some viral (hepatitis C) infections arise in this way. But some tumor viruses act more directly on the host genome by inserting oncogenic genes that dysregulate cellular signaling systems, leading to loss of endogenous control over cell replication and proliferation. For example, human papillomavirus can integrate into the host genome, and the resulting oncoprotein products inactivate host tumor suppressor proteins (Nguyen et al. 2014).

Why has evolution failed to develop mechanisms to avoid infection-related cancers? Since only a small proportion of infections with causative organisms lead to oncogenesis, we can conclude that infection is necessary but not sufficient for oncogenesis and that other host or environmental factors are required. Additionally some tumor-causing viruses, such as adenovirus, are oncogenic in some species but cause only mild infections in others. Oncogenesis mediated through the inflammatory pathway is arguably one of the costs of this defensive mechanism. Therefore, infection-related oncogenesis can be considered as an aspect of virulence or infection-related morbidity, which was discussed in Chapter 10 in the context of our coevolution with microbes.

12.6 Implications of an Evolutionary Approach for the Prevention and Treatment of Cancer

12.6.1 Prevention

How can evolutionary perspectives on cancer inform our approaches to prevention and treatment? First, appreciation of the role of environmental and life-history factors can help to prevent the development of the disease: examples include avoidance of causative environmental factors (e.g., smoking) and modification of exposure to life-history factors (e.g., use of estrogen receptor antagonists such as tamoxifen to modulate exposure to estrogen in high-risk women). An appreciation of the role of inflammation in carcinogenesis and metastasis helps us to understand the apparent protective effect of long-term use of anti-inflammatory medications such as aspirin and non-steroidal anti-inflammatory drugs (NSAIDs) against the development of colorectal cancer and possibly also breast cancer. Finally, knowledge that the efficacy of tumor suppressor mechanisms differs across species according to their
evolved life history can prompt research into new modalities to prevent malignant transformation or tumor metastasis in humans.

12.6.2 Treatment

The primary treatment strategy for cancers in humans is complete or partial surgical removal of the tumor, where possible. This may be preceded and/or followed by systemic chemotherapy with cytotoxic drugs, or by localized radiotherapy, with the aims of reducing tumor bulk before surgery and/or killing or depleting any tumor cells remaining after surgery. Sometimes, particular aspects of the tumor cell phenotype, such as surface receptor expression, are targeted with specific agents (e.g., tamoxifen for estrogen-receptor-positive breast cancer). All these approaches are usually effective initially, although at the cost of considerable toxicity to the patient. However, in most cases the
tumor recurs in a form that is resistant to the initial chemotherapy.

As discussed earlier, most tumors and their metastases are formed of multiple subpopulations of cells with different suites of mutations, and consequently different phenotypes, evolving according to the spatially and temporally differing microenvironment across the tumor. The initially predominant sub-clones can be viewed as those having the highest fitness in their environment; since resistance to chemotherapy has costs, such as diversion of metabolic effort to drug detoxification, then it is likely that these fittest sub-clones will be sensitive to therapy. However, exposure to cytotoxic chemotherapy, while killing sensitive cells, will select for a population of resistant cells that, although they may have been less dominant (i.e., had lower fitness) in the original tumor environment, will expand to fill the ecological niche left by the death of the original population of sensitive cells.

There are several implications of this evolutionary perspective. First, tumors containing high sub-clone diversity will contain more genetic variation from which resistance to therapy can be selected. This implies that, before treatment, multiple biopsies are required to assess the phenotypes of the predominant sub-clones within a tumor in order to optimize targeted chemotherapy, since a single biopsy may only sample a fraction of the sub-clones. Indeed, the extent of genetic diversity (mutational load) within a tumor can itself be used as a prognostic indicator for the development of resistance and disease progression.

Secondly, the genotypes and phenotypes of the predominant tumor clones emerging after relapse from chemotherapy will be very different from those before treatment. Understanding this can help guide subsequent rounds of treatment.

Thirdly, although the goal of most current cancer treatment is to eradicate all tumor cells by using chemotherapy at the maximum dose intensity that can be tolerated, it has been suggested that this may not be the best approach to balance the course of the disease with the patient’s quality of life. Modeling and experimental studies show that “adaptive therapy,” which uses continuously modulated dose...

---

**Figure 12.5** Adaptive therapy stabilizes tumor burden in mice carrying an ovarian cancer cell line and treated with carboplatin. Mice were inoculated with an ovarian cancer tumor cell line and subsequently treated with carboplatin administered according to a standard high-dose regimen or an adaptive therapy regimen in which the dose was adjusted to maintain a stable tumor volume. The tumor burden 120 and 180 days after inoculation is shown. Plotted from data in Gatenby et al. (2009).
schedules of low-dose chemotherapy aiming to keep tumor size constant, results in stable disease and prolonged survival with minimal therapy-related toxicity (Gatenby et al. 2009; Figure 12.5). Adaptive therapy aims to allow the survival of a significant number of chemosensitive cells, and these “fitter” cells will suppress the growth of the population of resistant cells by avoiding rapid proliferation of a treatment-resistant clone.

The fourth point concerns the treatment approach to metastatic cancer. Metastases will contain only a proportion of the genetic diversity of the primary tumor, and metastasized tumor clones must adapt to the host tissue environment. There is evidence that metastases from different primary tumors adapt convergently to the environment of the new host tissue. For example, liver metastases from different primary cancer sites appear to evolve towards a common liver-adapted phenotype. These observations suggest that treatment of metastatic disease should take account of the features of the host tissue as well as of the tissue at the site of the primary tumor (Cunningham et al. 2015).

12.7 Conclusion

We have discussed the utility of an evolutionary perspective for the prevention and treatment of malignant disease. We increasingly understand clonal evolution in tumors to be at the heart of cancer biology, but incorporation of this knowledge into treatment protocols is still at an early stage. Our increasing knowledge of cellular and developmental biology has revealed that cancer can be viewed as representing a reversion to a unicellular phenotype, in which control mechanisms that evolved to allow cells to cooperate to produce a multicellular organism are no longer functioning appropriately. Although individual cancers evolve as they progress, often with a lethal outcome for the patient, the “successful” clone(s) dies with the host and cannot evolve further to be the basis of treatment resistance in another patient. An evolutionarily informed appreciation of the origin and biology of malignancy will open up new perspectives in cancer, as we use our expanding knowledge of the cellular control mechanisms absent from cancer cells to design therapies that attack these deficiencies.
13.1 Introduction: Understanding Health and Disease from an Evolutionary Perspective

13.1.1 Multiple Medical Histories

In this book we have focused on how evolutionary history and processes have potential consequences both for our risk of disease as individuals and for population health. We have described multiple factors that need to be considered and have presented a simple categorical framework for how evolutionary pathways can affect patterns of health and disease (Chapter 7; Williams and Nesse 1991; Nesse and Stearns 2008).

When a patient presents with a complaint, medical assessment traditionally starts with the taking of a medical history—an analysis of the immediate symptoms that the patient recognizes, followed by an enquiry into the patient’s prior medical history, and, on occasion, the medical history of close relatives (family history). A more holistic view of the medical history would also include the patient’s developmental history, from the point of conception (although such information is usually not available).

But, as we have suggested, while not part of the formal medical history there is also explanatory value in the health professional considering the patient’s evolutionary history at several levels: the totality of biological evolution over nearly 4 billion years; human biological and cultural evolution since our species evolved about 160,000 years ago; our history since our ancestors initiated settlement and invented agriculture from about 12,000 up to 200 years ago (depending on the society); and the more recent history of the patient’s lineage (demographic, ethnic, and cultural history). A more holistic view of the medical history, beyond its more formal components, would thus assist both ultimate and proximate explanations.

13.1.2 An Evolutionary Evaluation of Clinical Signs and Symptoms

This broader evolutionary perspective can help understand the origin and significance of a patient’s symptoms and signs, many of which represent an appropriate and evolved response to the challenge that the patient confronts, and this may have implications in deciding whether to offer symptomatic relief or not.

Many consequences of infection, such as cough associated with a respiratory tract infection, mucus in the common cold, or diarrhea and vomiting associated with gastroenteritis, can be interpreted as attempts to expel the invading organism. For example, there is evidence that under some circumstances suppressing a cough can prolong an illness of the respiratory tract. But when these evolved mechanisms themselves become injurious and thus maladaptive—such as dehydration associated with severe diarrhea—then clearly medical treatment becomes essential.

Malaise is frequently associated with infection and may have evolved as an adaptive response to reduce workload, energy expenditure, or social contact during illness when energy consumption
is diverted to supporting immune function. There are experimental and some anecdotal human data to suggest that vigorous exercise during infectious illness can be detrimental to the host.

In Chapter 11 we discussed how emotional states such as anxiety and depression may be evolved responses of an individual to fearful situations, loss, situations of strife within the social group, or a change in group dynamics. Evidence of distress may thus elicit assisting responses from the group. Both acute and chronic illness can similarly induce both anxiety and depression as an inherent part of the symptom complex. Inappropriate exaggeration or prolongation of these emotional states themselves can lead to psychiatric illness or be maladaptive, slowing the processes of recovery and rehabilitation.

13.1.3 Implications for Prevention and Therapy

Without appreciating the patient in the context of his or her proximate and ultimate medical histories, comprehension of the complaint and its potential remedies will be limited. Clearly most attention must focus on approaches that address proximate causation, but an understanding of the condition and the choice of intervention can be influenced by evolutionary considerations (see Section 12.6.2). Beyond the individual patient, an evolutionary perspective has particular implications for the design and evaluation of public health strategies; several examples illustrating this statement are given in this section.

Many people with an infection have relative iron-deficiency anemia, arising in part from cytokine-driven extracellular and intracellular sequestration of iron. The resulting concentration of available free iron is several orders of magnitude lower than that required to support microbial growth, thus providing a bacteriostatic extracellular environment. There is evidence that treatment of some individuals, particularly young children, with iron supplements can therefore aggravate infection or infestation. Iron supplementation in children may worsen infectious outcomes such as malaria and diarrheal disease in low and middle-income countries (see Section 10.7.1). The relative anemia often seen in the presence of chronic infection may therefore be protective. Thus the decision to give iron must be given in the full context of understanding the patient; it may be that overriding the evolved protection conferred by anemia could be harmful.

The rise in use of assisted reproductive technologies is the inevitable result of both technological development and social change, enabling the choice to delay conception to achieve other life goals. But assisted reproduction is not without costs—it is intrusive, often fraught with emotion for the prospective parents, and expensive. The outcome includes more twin and multiple pregnancies with their complications, and is occasionally associated with a higher incidence of imprinting disorders in the offspring. An evolutionary perspective would argue that age-related reproductive decline is inevitable, as is the psychological burden of infertility; women (and men) need to be aware of these issues. Evolutionary biology would suggest that a more appropriate response would be to address the ultimate cause—the mismatch between cultural and biological evolution—by developing societal structures so that women can have their children at a biologically more advantageous age while still allowing them to meet their own goals. From a social point of view, this evolutionary argument may of course not be acceptable.

The obesity epidemic is a global concern. The results of simple measures to encourage healthy eating and more exercise have been disappointing. There are evolutionarily inevitable reasons for this. We evolved to appreciate certain foods with sweet and salty tastes, and modern foods have been designed to meet those preferences and to create flavors which are generally associated with fat. Infants start to develop food preferences and taste in utero and during lactation from exposure to constituents of the maternal diet. These preferences are further reinforced by food experiences and dietary practices in the period after weaning (Beauchamp and Mennella 2009). In evolutionary terms, this development can be seen as a form of intergenerational learning, to assist the independent infant to choose foods that are safe and provide essential nutrients. Further, there are pathways induced in early life (reviewed in Chapters 4 and 9) that make it likely that people
who are nutritionally “mismatched” have a tendency to poor appetite control, reduced muscle mass, and reduced energy expenditure as well as a propensity for deposition of visceral fat. These are evolved adaptive processes, and adult lifestyle intervention alone may be rather ineffective because the underlying epigenetic changes were induced in early life and may not be reversible. An obesity prevention program, therefore, should not just focus on the obese adult or adolescent but should also consider interventions aimed at optimizing fetal and infant development, and then encourage a closer match between the predicted and actual environments or lifestyles.

In Chapter 10 we discussed the development of bacterial resistance to antibiotics. The widespread use of antibiotics in medicine and agriculture favors the development of such resistance. The emergence of methicillin-resistant *Staphylococcus aureus* and multidrug-resistant tuberculosis have now become major public health concerns. The inappropriate use of antibiotics by doctors and farmers is a major reason for bacteria evolving as the winners in this arms race. Regulatory and educational approaches are needed to address this misuse.

But for the patient, one of the most valuable insights from evolutionary thinking is perhaps that it provides an explanation of “why” something has occurred, whereas most medical explanations tend to be very mechanistic.

### 13.2 Testing Evolutionary Hypotheses in Medicine

In Chapter 7 we discussed the challenges of developing and testing evolutionary hypotheses in medicine. When we evaluate clinical problems from an evolutionary perspective the various approaches discussed there must be considered. At the outset, we must emphasize again the importance of avoiding teleological arguments when testing evolutionary hypotheses.

Genomics has become a core tool in reconstructing human evolutionary history. A common approach has involved identifying a phenotypic trait under study (e.g., lactose intolerance, adaptation to altitude, or dark skin); identifying the gene loci that may underlie the phenotypic trait; and comparing the frequency of alleles at specific loci between populations that do and do not exhibit that phenotypic trait. Finding a genetic difference between phenotypically different populations is, of course, only the beginning of the story as it requires an explanation—which may need not only a geneticist’s tools but also those of an archeologist or a historian. For instance, it might be desirable to estimate the age of the mutation, or how long it took to spread to its current frequency, to find if there are signatures of positive selection or if the trait was indeed under selection rather than changing frequency via other processes such as drift or bottleneck. An explanation of the likely selective pressure may require a rather detailed understanding of the historical, biological, social, and physical circumstances under which the mutation spread. Thus, in many cases, creating evolutionary hypotheses to explain our genetic makeup will require a good knowledge of evolutionary biology, medicine, and social history.

### 13.3 Clinical Examples

Building on the taxonomy developed in Chapter 7, the following sections provide some further examples that illustrate the ultimate perspective afforded by the evolutionary approach. It is important to note that multiple pathways may operate together in providing an evolutionary interpretation of disease risk. Indeed, given the nature of evolutionary trade-offs, this is to be expected. For example the discussion of scurvy in Box 6.3 highlights our evolutionary lineage from frugivore ancestor species, the role of neutral mutation (Pathway 8 in Box 7.3), and the role of changing and novel environments (Pathway 1 in Box 7.3) in exposing the phenotypic effects of a previously neutral mutation. The operation of several processes in parallel is also true of many of the examples given in this chapter.

### 13.4 An Evolutionarily Mismatched or Novel Environment

Natural selection has operated to adapt organisms to the environments in which they evolved.
Migrations and environmental changes, generally a result of human cultural evolution, have always been part of human history and drivers of evolution. But in recent decades the pace of environmental change has been rapid. Technological innovations and social developments—from wind-powered vessels to jet planes; the slave trade; imperialism; conflicts displacing populations—have resulted in unprecedented movement of people across the globe. Cultural and technological innovations have influenced where and how we live, what we eat, how we dress, communicate, and work, and how we live in social groups.

The map in Figure 13.1 shows examples of polymorphisms that have been associated with human adaptation: in environments that differ from those in which they confer selective advantage they may increase the risk of disease. All except LCT (Box 1.1 and Section 3.3.3) are discussed in this chapter.

### 13.4.1 Sun, Skin, and Disease

Ultraviolet radiation (UVR) from the sun has been a major selective pressure through human evolutionary history. On the one hand, UVR is necessary for the synthesis of vitamin D, a critically important nutrient. On the other, UVR has a destructive effect on the skin. Melanin synthesized by melanocytes in the epidermis offers protection against UVR, and its production must be finely balanced to allow sufficient vitamin D to be synthesized (see Section 6.3.6 for a discussion of the evolution of skin color). The initial light color of the early hominins that were still covered with hair evolved into the adaptive dark skin in hairless Homo species in tropical regions; depigmented phenotypes evolved again on several independent occasions as early humans spread worldwide into latitudes with lower sun exposure.

Today, in high latitudes (including the UK and large parts of North America) there is insufficient UVR during much of the year to allow...
vitamin D synthesis, even in light-skinned people. Consequently, there is a high prevalence of vitamin D insufficiency (serum levels of 25-hydroxyvitamin D less than 40–50 nmol/l) in people in these areas. People with dark skin require up to 10 times more exposure to sunlight than those with light skin to produce the same amount of vitamin D. For that reason, the prevalence of insufficiency is highest in dark-skinned migrants to high latitudes: 52–76% for African Americans, 18–50% in Hispanics, and 8–31% for people of European origin. Perhaps surprisingly, vitamin D levels appear to be positively associated with latitude in Europe, with the highest levels in Scandinavia, which is probably a result of cultural factors such as preference for sun exposure rather than sun avoidance, a diet traditionally based on vitamin D-rich foods (fish, animal liver), and the use of supplements. Culture is probably the explanation for high levels of vitamin D insufficiency in dark-skinned women who cover their bodies entirely and live in sunny regions such as the Middle East or Australia (Munns et al. 2012).

The reason for the evolutionary importance of vitamin D is its role in the regulation of calcium and phosphorus metabolism, particularly in the development and maintenance of bone. A deficiency of vitamin D during the growth period, before the closure of the epiphyseal plates, causes rickets. This is a disease in which, due to the malabsorption of calcium and phosphate, cartilage fails to mineralize adequately, with consequences such as bowing of long bones and a deformed pelvis and ribcage. In adults, vitamin D deficiency produces osteomalacia, literally “bone softening,” similar to conditions caused by inadequate mineralization of the adult bone. These are certainly serious conditions—but how do they cause such a reduction in reproductive fitness to merit vitamin D becoming a central selective pressure in a trade-off with the major and ubiquitous pressure of sunshine?

A pelvis deformed by childhood rickets can cause problems in pregnancy, such as obstructed labor, and this would reduce fitness. But another factor that might affect fitness is vitamin D supply/synthesis during pregnancy and breastfeeding. At these times, the demand for calcium by the maternal–fetal dyad is high because of the need to build the fetal skeleton. During lactation in particular, maternal bones may go through a period of significant demineralization to meet the infant’s demand (Kalkwarf 2006). There are a number of physiological adaptations to obtain additional calcium during pregnancy and lactation: in the third trimester of gestation, the time of intense fetal skeletal growth, there is an increase in intestinal calcium absorption regulated through an increase in vitamin D-binding protein. And while there is absorption of maternal bone towards the end of the pregnancy, and especially during breastfeeding, it appears that this bone loss is transient and that bone mass and density quickly increase after the cessation of lactation. The need for calcium during these reproducively crucial periods probably played a role in making the adequate availability of vitamin D a major selective pressure—but perhaps not only in bone metabolism.

Recently, the broader role of vitamin D in organ systems other than bone has been explored in immune, cardiovascular, muscular, reproductive, and skin tissues (Jablonski and Chaplin 2012). Indeed, vitamin D is now considered to be a steroid hormone system rather than just a vitamin. It has been suggested that this system has a role in cell proliferation and differentiation and modulation of the immune response. Defects in the vitamin D endocrine system have been associated with cancers, cardiovascular disease, and autoimmune disease as well as other immune abnormalities. A study in South Africa reported that vitamin D supplementation increased the number of circulating lymphocytes and slowed down the replication of HIV-1 in freshly isolated peripheral blood mononuclear cells (Coussens et al. 2015). Clinical and experimental studies have shown an association between vitamin D deficiency and male as well as female infertility. The exact mechanisms remain elusive, though some form of disruption of cellular calcium signaling has been suggested. These various non-mutually exclusive explanations might provide the basis for why selection to protect vitamin D homeostasis has occurred.

But while the need for vitamin D drove the evolution of light skin as we moved away from the equator, the opposing cline has increased the amount of protective pigment in the skin. There is lively debate about the main reason for the evolution of skin pigmentation. Skin cancers have been
suggested as the main reason, and the exceptionally high rates of skin cancer in (not uncommon) African albinos, people who due to a mutation completely lack eumelanin in their skin, favor that hypothesis (Greaves 2014). Others have argued that albinos are not a suitable model for ancestral pale skin prior to the evolution of pigmentation, because modern apes (and therefore probably ancestral hominins) do have active melanocytes in their skin that are capable of producing eumelanin upon exposure to strong sunlight (Jablonski and Chaplin 2014). Even so, it is questionable whether skin cancer would be a strong selective force. Pre-cancerous cellular changes take years to become clinically apparent, and given the probable short reproductive careers of early Homo it is likely that reproduction would have been completed before a skin cancer could become disabling.

An interesting but uncorroborated hypothesis has linked the evolution of skin pigmentation with protection against the destructive effects of UVR on folate biology (Jablonski 2013; Luccock et al. 2014). Folate has an important role in DNA synthesis and as a methyl-group donor in the establishment and maintenance of epigenetically mediated gene expression patterns. Periconceptional folate deficiency has long been known to cause congenital malformations, in particular neural tube defects. The adverse effects of UVR on folate stability in humans was first suggested nearly 40 years ago in a study that compared serum folate concentrations between patients receiving UVR treatment for skin conditions and healthy controls (Branda and Eaton 1978). The potential significance of this link has been highlighted as interest in folate biology has boomed with the recent rise of epigenetics and recognition of the central role of one-carbon metabolism in DNA methylation. Thus it has been proposed that increased skin pigmentation may protect the early embryo, and perhaps also have an activity in men too by protecting spermatogenesis in situations of high UVR (see Section 13.10); periconceptional solar exposure may also select for embryos with corresponding nuclear folate pathway polymorphisms (Lucock et al. 2014).

This example illustrates the complexities of evaluating evolutionary explanations (including those of interactions between different selective environments), the need to think about the plausibility of explanations in terms of the time in the life history when they may be operating, and also the interactions between our biocultural adaptations and novel, or changed, environments.

13.4.2 Cultural Change and the Emergence (and Disappearance) of Disease

How do cultural changes influence the emergence of disease? In the 1950s, anthropologists working in the Highlands of New Guinea with the members of the Fore linguistic group observed unusual “spirit possession” characterized by involuntary muscle twitches, lack of control over the limbs, tremors, and pupil dilatation (Anderson 1992). Several years later, these observations were given clinical meaning when it was recognized that the condition, locally known as “kuru,” was an unknown form of encephalitis. The symptoms were initially observed among women and then in children—the first case is believed to have occurred relatively recently, perhaps in the early twentieth century. The women in this tribal group also tended to practice cannibalism, which in the Fore group was associated with kinship; the bodies of relatives who died from injury, accident, or from some diseases were, following a mourning ritual, cooked and eaten (Glasse 1967). Children ate whatever food they were given by their mothers, but men ate human flesh only very rarely as it was believed that cannibalism robbed a man of his vitality. And even if they did practice cannibalism, they never ate women, as that was believed to be extremely harmful to men.

By 1963, building on the laboratory evidence of transmission of the disease from cell-free suspensions of human brains to chimpanzees, it was suggested that an unknown slow virus was the etiological agent, but such a virus was never found. Some years later, the etiological agent of kuru was shown to be an infectious protein particle, named a prion. Prion diseases arise from an interaction between the normal prion protein (PrP\textsubscript{c}) found on cell membranes and its infectious isoform (PrP\textsubscript{sc}), which is able to convert PrP\textsubscript{c} into the diseased form through conformational change and aggregation. The disease-associated prion could then be passed,
by cannibalism, to another individual where the diseased protein could again transform the host’s protein, and so on. These plaques of aggregated protein produce the neurological symptoms. By then, thanks to the cessation of cannibalism in the Fore, kuru had disappeared. The disappearance highlights the relationship between culture and disease.

Yet just a decade later, another closely related disease emerged in a high-income country. Bovine spongiform encephalopathy (BSE), commonly known as “mad cow disease,” is a fatal neurodegenerative disease of the central nervous system in cattle. The epizootic agent was first recognized in 1986 and was later identified as a prion, transmitted to cattle by the practices of intensive farming in which the herbivorous animals were fed meat and bone meal derived from infected carcasses (Woods 2011). By 1996, a link had been established between the epizootic and the emergence in humans of a variant of Creutzfeld–Jakob disease (CJD). CJD is another fatal neurodegenerative disease causing spongiform transformation of the central nervous system. It may occur spontaneously or by the injection of pituitary growth hormone extracted from cadavers which, prior to the development of recombinant growth hormone made in bacteria, was the only form of growth hormone available for therapy. Many individuals in Europe and the USA who had been treated with growth hormone as children were infected by contaminated batches of pituitary growth hormone, which contained at least one pituitary gland from an individual with an undiagnosed case of CJD.

It appears that humans contract variant CJD by eating beef infected with the causative agent of BSE, or occasionally by inhaling the pathogen when working with fertilizers containing bonemeal. Over 200 deaths have so far been reported, mostly in the UK in the 1990s and early 2000s. Although improvements in agricultural practice have curbed transmission of the disease, occasional cases still occur because of the long incubation period. The epizootic and its human consequences created havoc in the British agricultural industry in the 1990s, leading to the slaughter of millions of animals amid widespread public concern. By this time, BSE had come to be associated with a whole range of spongiform neuropathies increasingly understood to be caused by a prion: CJD, scrapie (a sheep disease first recorded in the 1750s and shown in the 1930s to be transmissible), and kuru.

A footnote to the prion story provides an example of the strong selective pressure exerted by disease. Human predisposition to prion diseases is genetically determined, such that polymorphism at codon 129 of the normal prion protein gene (PRNP) is a strong susceptibility factor (Mead et al. 2009). Codon 129 heterozygosity (found in about 50% of people of European origin) appears to be protective; only one case of variant CJD has been reported in a heterozygote (Saba and Booth 2013). Interestingly, a novel adjacent polymorphism, G127V, is found exclusively in the region in which kuru was prevalent. It was found in half of the otherwise susceptible women in the highest-exposed region, and genealogical analysis revealed that the incidence of kuru in lineages harboring this polymorphism was significantly lower than in control families from the same area, with 1 of 36 parents from G127V genealogies in control families from the same area, with 1 of 36 parents from G127V genealogies having died from kuru compared with 33 of 128 in the control genealogies. The polymorphism is highly restricted geographically, suggesting a very recent common ancestor, probably about 10 generations ago. These findings suggest a strong selective pressure exerted by the high mortality rates of kuru and its distribution within kinship groups because of the nature of funeral rites and kin-related cannibalism.

13.4.3 Sleep

Sleep is essential for cognitive function and health in humans, yet we know little about why sleep evolved and the nature of sleep patterns in our ancestors. Even the definition of a “normal night’s sleep” in modern humans is a subject of debate. While asleep, animals appear to do nothing that increases their reproductive fitness: they do not look for food or for mates, and they are more vulnerable to predation (McNamara et al. 2010). Yet sleep is found across the phylogenetic tree. Even insects exhibit behavioral phenomena that are consistent with sleep, with periodic reductions in activity as well as increased rest (sleep) duration following a period of sleep deprivation.
Sleep is part of the diurnal rhythmic patterns that affect our physiology, including neuroendocrine function, thermoregulation, and the autonomic nervous system. Reversal of that pattern, for example in night workers, is associated with metabolic compromise as the coordination of biological rhythms is disturbed (van Drongelen et al. 2011). A reduction in sleep, even in young children, is associated with a higher risk of obesity. In that context sleep can be viewed as a regular period of fasting, and during sleep there are changes in metabolic control such as slow-wave sleep-related release of growth hormone to promote lipolysis. Multiple functions have been proposed for sleep, and it seems likely that it first evolved to serve one of these functions and was exapted over evolutionary time for others.

Most research interest in sleep has focused on its possible restorative functions and on memory processing (Lesku et al. 2009). Sleep deprivation impairs the working memory, which is essential for making information available for higher-level cortical processing. There is some evidence that both rapid-eye movement (REM) and non-REM sleep play a role in memory processing. Thus, it may be that while sleep may have evolved for metabolic and restorative functions in more distant taxa such as insects, it took on additional functions in later-evolving species that relied on associative learning and strategizing. Human neonates and fetuses have more REM sleep than older children and adults, and it has been suggested that sleep may have an important role in brain development. Humans have the highest proportion of REM sleep among primates, raising the question of whether there is an association with brain evolution. It remains speculative whether dreaming has an adaptive function: despite the enthusiasm of psychoanalytic psychiatry, dreaming may be a by-product of the random firing of brain circuits during REM sleep which appears important in memory reinforcement, or an adaptive process in which memories become consolidated and organized.

Sleep also has close links to immune function. Sleep architecture appears to change in response to infection, with the time spent in non-REM sleep increasing with rising fever, and time in REM sleep decreasing. In the clinical context, sleep deprivation impairs wound healing and immune function. Increases in sleep duration in mammals are associated with an improvement in immune defenses. A very speculative argument has thus been put forward that combating infection was a selective pressure in the evolution of sleep (Preston et al. 2009).

So what is the normative pattern for human sleep? The amounts and patterns of sleep vary enormously even within the mammalian order, and the environment in which animals live can exert significant pressures on sleep architecture. For example, there are major differences between aquatic and terrestrial animals. Dolphins and whales appear to lack the REM phase and never completely sleep, although they change neural activity in one hemisphere at a time to produce a kind of unilateral sleep. A major selection pressure on the sleep pattern appears to be the relative position within the food chain and the risk of predation (Capellini et al. 2010). Sleep is shorter in prey (herbivores) than predators (carnivores) and also among those animals sleeping in the open compared with those sleeping in fully enclosed spaces such as dens and tree holes. But interestingly, it appears that animals that sleep in groups sleep for shorter periods than animals that tend to sleep alone: is there a trade-off between socializing and sleeping?

Humans have the shortest sleep of all the primates, so could it be that our sociality (Chapters 6 and 11) affected the duration of our sleep? Of all the primates, great apes are the only ones to build sleeping platforms or dens and they also appear to have deeper, more efficient sleep than monkeys. Experiments in which great apes were provided with material from which they could fashion a “bed” showed that bedding increased sleep quality and cognitive performance the next day. This example of niche construction probably applied to the evolution of sleep in hominins (Samson 2013).

But some environmental factors that affect dominant sleep patterns in Western society have only been acting for a few hundred years. We tend to think that we need 7–9 hours of largely uninterrupted sleep, yet a historical study supported by persuasive evidence has argued that this pattern is no more than 200 years old and is a result of the
introduction of artificial light (Ekirch 2005). Before then, Europeans at least appear to have slept in two stages separated by a period of wakefulness in the middle of the night that was filled by reading, eating, or even visiting the neighbors. In conclusion, an evolutionary perspective shows that our sleep may be more plastic than we usually assume, with the range of normality, in particular when it comes to sleeping for long uninterrupted periods, being much wider than we tend to assume.

**13.5 Life-history-associated Factors**

Chapters 5, 8, 9, and 12 discuss multiple examples of life-history-related processes that can affect morbidity. There are many ways in which deviations from the evolved normative life history of humans can affect health. Indeed, there are health consequences to our normative life history, in particular those associated with old age. One important life-history mechanism discussed earlier is antagonistic pleiotropy, where an allele may confer advantage early in the life course but have deleterious consequences at a later, often post-reproductive, stage (see Sections 5.2.4 and 12.5.2). For example, high levels of IGF-1 in childhood, which have an important genetic determinant, may promote growth and thus, indirectly, fitness (Rogers et al. 2006), but high levels of IGF-1 in adulthood are associated with higher risks of prostate and breast cancer later in life (Renehan et al. 2004).

**13.5.1 Menopause and Health**

If we compare female and male mortality rates (Section 8.8), it appears that the female advantage, which is particularly evident in the adolescent and reproductive years, is much reduced after the menopause, when female mortality increases due to a rise in cardiovascular events and certain malignancies. Osteoporosis, a progressive loss of bone density, is especially common among post-menopausal women and may lead to fractures of the vertebrae and hips, and thus to higher mortality rates due to complications. The perimenopausal (“climacteric”) syndrome is the most common condition affecting women entering menopause. It is a diverse collection of symptoms affecting various organ systems, from erratic thermoregulation leading to hot flushes and sweating, loss of libido, and vaginal dryness, to mood changes and insomnia. These symptoms do not, by and large, cause excess mortality, but they can last for years and significantly reduce quality of life.

This increased risk of mortality and morbidity is generally explained by the dramatic drop in estrogen levels that occurs at menopause. Human females produce all of their oocytes in utero; by puberty, 80% of the 2 million oocytes present at birth are lost and this loss continues through life and accelerates towards menopause (Sievert 2011). At menopause the endocrine activity of ovarian follicles declines drastically and circulating estrogen levels fall.

But, if menopause is to be understood as an adaptive trait, as a life-history stage essentially unique to humans and evolved for its benefit to inclusive fitness (see Section 8.9.7), then how should we explain the potential health and fitness costs of the post-menopausal period? Furthermore, the perimenopausal syndrome, so common in the modern developed world, appears rare or even absent in traditional small-scale societies. Why, then, does menopause appear to be an unhealthy state in developed populations?

One answer to this question would be that all adaptations come with a potential cost, and all the traits are based on trade-offs of some kind (Peccei 2001). So it may be that while menopause does increase a woman’s mortality risk, the overall benefit in terms of (inclusive) fitness may still be sufficiently high. Another, popular, argument is that the morbidity and mortality patterns that we see in women in high-income countries today are very different from the corresponding demographic indicators in women in the past, because lifetime estrogen exposures were much lower in earlier generations and circulating estrogen levels were also likely to be lower because of lower body fat (Eaton et al. 1994, Aktipis et al. 2015). As we discussed in Chapters 8 and 12, life-history patterns in the past were different: women entered puberty later, had more pregnancies and breastfed for longer yet had less access to food and so achieved a positive energy balance later—all of which resulted in longer lactational...
amenorrhea—and they generally entered menopause earlier. Indeed, the final lactational amenorrhea commonly transitioned into menopause. For all of these reasons, their lifetime exposure to estrogen was relatively lower than today (Jasienska et al. 2006; Ziomkiewicz et al. 2008). In short, the fall in estrogen levels in the past, or indeed in those populations whose life histories are still largely following the patterns of the past, would have been far less dramatic.

It is a common physiological principle that the magnitude of change in a variable is often of greater significance than its absolute level. Thus the general solution to menopausal health problems has been to reduce the difference in the estrogen levels before and after menopause. Until recently, the medical solution was to offer hormone-replacement therapy to relieve perimenopausal syndrome and counter increased health risks—a strategy that, by continuing exposure to sex steroids in females beyond the menopause, introduced an evolutionarily novel state. Evidence then emerged showing that hormone replacement therapy is associated with a greater risk of breast cancer; this is not unexpected from an evolutionary perspective (Rossouw et al. 2002). Thus the use of hormone replacement therapy after the menopause is now generally restricted to those women in whom there are major psychological or other health reasons for such an approach. An approach that has a stronger evolutionary backing would be to try and lower lifetime estrogen and estrogen exposure before menopause. While the cultural changes in reproductive practices make such an approach impractical, reducing estrogen exposure—by increasing physical activity, changing nutrition, and reducing weight if necessary—would produce positive health effects overall.

### 13.5.2 Risks of Being Born Early

Pre-term birth, defined as birth before 37 completed weeks of gestation, increases a child’s risk of mortality, morbidity, and developmental impairments and delays. Pre-term birth affects about 11% of all live births globally and is a substantial health, social, and economic burden. Indeed, the proportion of pre-term births has been increasing over the past two decades, although there is substantial variation among countries (Zeitlin et al. 2013). This increase has been driven by a variety of cultural factors: the rise in multiple pregnancy rates associated with the use of IVF, increased maternal age, and higher BMI, as well as medical advances that allow survival of those neonates for whom mortality rates would have been very high just few decades ago.

About half of all cases of pre-term birth are idiopathic, and the molecular mechanisms triggering early parturition remain unclear. One of the known risk factors is intrauterine infection, which has been suggested to account for up to 40% of pre-term births. The most common route for uterine infection involves the ascent of bacteria up the cervix, continuing to the chorion and amnion, then entry to the amniotic fluid and fetus. This induces an inflammatory response in the uterine muscle, with the release of prostaglandin stimulating contractions and initiating labor. There is also animal and, increasingly, human evidence that maternal periconceputal nutrition can modulate gestational length, with maternal undernutrition at the time of conception being associated with mildly pre-term birth (Gluckman et al. 2005a).

In Chapter 4, we discussed the range of responses that may be made by the developing organism upon exposure to cues that range from the normative to those of a more severe nature. If a fetus is exposed to an infected uterine environment—a relatively severe cue—it may make immediately adaptive responses including elevation in cortisol levels to accelerate organ maturation, in preparation for an early birth. The latter may be caused by the infection increasing the release of prostaglandins from the gestational membranes, which then have a paracrine and stimulatory effect on uterine muscle contractions. This early delivery would maximize the fetus’s chance of survival in comparison with continued gestation that is more likely to lead to intrauterine death. Likewise, severe undernutrition appears to prompt the early embryo to modulate its endocrine maturation and so pursue a similar life-history strategy of pre-term birth. These effects would only have a survival effect for mild prematurity (35–37 weeks), but globally such mild prematurity makes up the bulk of prematurity observed: the more extreme and fatal (prior to modern neonatology) forms are relatively uncommon in comparison.
Pre-term birth, which necessarily involves incomplete intrauterine growth and maturation, inherently involves trade-offs. In general, in the short term, the infant will face substantially higher risks of mortality and major morbidities including complications of the respiratory, gastrointestinal, immune, hematological, cardiovascular, and central nervous systems (Behrman and Butler 2007). The impact of such morbidities extends to the long term, with later life impairments in neurological development and hearing being much more likely. Additionally, metabolic health in later life may also be compromised. Lower insulin sensitivity has been observed in pre-pubertal children and in adults born prematurely (Hofman et al. 2004; Mathai et al. 2012), and there is also increasing evidence for predisposition to greater adiposity in mid-adulthood (Mathai et al. 2013). This suggests that developmental priming has occurred, although the nature of the inducing cue(s) is not clear: it may be due to the higher lipid profile of the post-natal diet or to the altered hormonal milieu resulting from premature loss of exposure to placental hormones, or to an altered pattern of cortisol exposure in the perinatal period.

The multiple lifelong health consequences of pre-term birth are thus a good example of trade-offs incurred through an altered life-history strategy. Being born prematurely almost invariably has deleterious effects on health; nevertheless there is still an adaptive advantage to adopting such a strategy, as reproductive fitness—even if potentially impaired—is still preserved, compared with the alternative of likely fetal death.

13.6 Excessive and Uncontrolled Defense Mechanisms

As discussed in Chapter 7 and earlier in this chapter, many symptoms of disease are physiological responses that have evolved to be protective against an environmental challenge, be it infective, physical, or social. Some further examples are given here.

13.6.1 Antipyretics: Beneficial or Harmful?

Fever is a fundamental defense mechanism against infection that has been conserved in vertebrates for at least 600 million years (Evans et al. 2015). In Section 10.7.1 we discussed the mechanisms through which increased core temperature mobilizes components of the immune system. Historically it has long been understood that heating may produce positive health effects, either applied generally (in the form of saunas or in the hot mineral baths used across societies worldwide) or topically (in musculoskeletal disorders). In the pre-antibiotic era, artificially induced fever was applied to treat neurosyphilis, with a remission rate of about 50% (Whitrow 1990). The therapy was abandoned when more effective antimicrobial drugs came into use. But because fever is such an integral part of response to infection, methods of lowering fever have historically been an essential part of infection management. The first synthetic antipyretic drugs, starting with aspirin (acetylsalicylic acid), came on the market in the late nineteenth century but the fever-lowering properties of the bark of the willow tree (which contains naturally occurring acetylsalicylic acid) have been known for millennia.

Today antipyretic treatment is administered widely to people suffering from high or even just mild or moderate fever. It is given routinely to critically ill patients with sepsis as well as to children and adults suffering from moderate fever accompanying viral infections such influenza, which is common and occasionally fatal (Eyers et al. 2011). Yet the evidence has been mounting that a moderate increase of body temperature is associated with improved clinical outcome. Pre-clinical studies in rabbits affected with rinderpest virus also found an increase in mortality when fever was inhibited using aspirin, with 70% of animals in the aspirin-treated group dying compared with only 16% of animals in the control group (Kurosawa et al. 1987). Retrospective studies show a positive correlation between the presence of fever and survival, and between failure to mount fever and increased mortality (Ryan and Levy 2003).

The first prospective studies of the role of fever were conducted in 2013 (Kushimoto et al. 2013). A large prospective study followed patients admitted to an intensive care unit with severe sepsis. Twenty-eight-day mortality was the lowest for the group with a fever of between 38.6 and 39.6°C, followed by those with slightly lower but still
increased temperature, while the group with the highest mortality was the one with the lowest temperatures. Data collected from over 500,000 patients in 300 intensive care units across multiple countries show that the lowest risk for sepsis patients in the first 24 hours in an intensive care unit is for those having temperatures between 38 and 39.4°C (Young and Bellomo 2014).

However, the beneficial effect of fever needs to be separated from both the beneficial and harmful effects of antipyretic treatment itself. This is complicated by the fact that many antipyretics have multiple activities, combining analgesic (e.g., acetaminophen, also known as paracetamol), anti-inflammatory (e.g., ibuprofen), or anticoagulant (e.g., aspirin) effects. A further issue is that severe fever can cause seizures, particularly in children, but such febrile seizures generally have no long-term significance.

### 13.6.2 Is Depression Adaptive?

This question is, of course, misleading: a disease cannot be adaptive, although an affective change within the normative range can be. But the persistently high prevalence of clinically significant depression—projected to affect 10% of the population in high-income countries by 2030 (Figure 7.1)—does beg a question of whether there may be an evolutionary explanation for the widespread susceptibility to this condition.

Yet before venturing an explanation, we need to define what depression is. While in the colloquial language depression stands for low mood, diagnostic manuals today define it as a set of symptoms that, in addition to the obligatory low mood, also include a combination of signs as varied as fatigue, loss of libido, loss of appetite, psychomotor agitation (or retardation), loss of ability to enjoy once pleasurable activities (anhedonia), feelings of worthlessness and guilt, or suicidal thoughts. It is not clear that potentially very different combinations of these symptoms (although meeting the threshold score for the diagnosis of depression) should be regarded as a single disorder: the quest for pathognomonic biological markers of depression has not been productive, and genome-wide association studies have found few significant associations with polymorphisms (but see Section 11.6) (Fried and Nesse 2015). While newer antidepressant therapies are of value, they do not have an equally strong effect on all symptoms, and indeed may have no impact on some. The clinical landscape is further complicated by the existence of post-partum depression and seasonal affective disorder that arise in response to particular contexts and are likely have specific underlying proximate mechanisms.

Thus the argument has been made that, to make progress, we need to go back to the individual symptoms of depression, which should then be studied from both ultimate and proximate perspectives (Berrios 1996; Fried and Nesse 2015). Understanding the evolutionary origin of a particular symptom, potentially originating in an adaptive mechanism, might be more helpful in making progress towards understanding the syndrome of depression, and mental illness more broadly (Nesse and Stein 2012).

Evolutionary explanations of depression have largely focused on the lead symptom of “low mood,” a trait that may have a selective advantage. In Section 11.4.3 we discussed the idea that emotions and moods (moods defined as longer in duration and less tightly tied to specific cues than emotions) exist because they offered some sort of adaptive advantage in our evolution as a social species. This argument is supported by the universality of moods across cultures.

It has been argued that “depressive” mood is elicited by the inability to make progress towards a goal, rather than by stresses or losses: if the progression towards a goal (a valued resource) is fast, mood is high; if the progression is slow, mood is low (Nesse 2009). Mood tends to return towards normality if an individual gives up on an unachievable goal. The usefulness of low mood is revealed by those cases where high mood prevails regardless of the external situation: in manic episodes, for instance, too many projects are started, and impulsive and costly decisions undertaken. So in this framing a low mood can be understood as an evolved mechanism to help the individual recognize an unachievable goal that may be diverting resources from more productive efforts. Another evolutionary argument is that low mood signals an individual’s state to other members of the social group (and thus recruits assistance).
The value of an evolutionary perspective on moods is primarily in the distinction between a normative behavioral response and a pathological state. Low mood may be elicited by the situation in which it is an appropriate reaction, while pathological depression is an exaggerated response, in some cases induced independently of its normative triggers and resistant to endogenous recovery. Distinguishing between the two may be crucial for therapeutic decisions, because treating an appropriate evolved response might be harmful.

If the low mood is an evolved trait, could other symptoms that appear in depressive disorders be explained in ultimate terms too (Keller and Nesse 2006)? Emotional pain or sadness, like other types of pain, draws the attention to the source of pain and leads to avoidance of actions that might result in repeating whatever action led to the loss. Crying elicits empathy from others and helps attract support from the social network. Self-reproach also motivates avoidance of actions that might lead to repetition of loss. It also involves self-analysis, and may signal culpability to the social group so as to avoid expulsion. Fatigue assists energy conservation and may prevent repetition of goal-pursuit actions that cannot succeed. Anxiety promotes hypervigilance, particularly towards threats.

Recent studies have shown there are distinct patterns of symptoms that are linked to different causes of low mood and depression, thus reinforcing these adaptive ideas and providing some support for this hypothesis. For example, crying is more likely to be associated with social loss, whereas fatigue and pessimism are higher in response to failure to reach a goal. It may be that such analyses, which give greater granularity to the classification of depression, might lead to better therapeutic approaches. This ultimate perspective might have a particular value in framing therapeutic interventions between psychiatrists and patients.

13.7 Consequences of Coevolution with Microbes

In Chapter 10 we discussed the constant coevolutionary interaction between humans and viruses, bacteria, and parasites. But while microbes, with their short generation times, might seem to have the upper hand in this relationship, there is a dynamic and changing response to pathogens that creates a complex coevolutionary relationship between us and microbial threats. The association between humans and many pathogens changed when human populations started to form resident communities and domesticate animals. This increased the rates of transmission of infectious agents between individuals, and exposed humans to many potential new pathogens of zoonotic origin. Our evolved responses to these pathogens may result in collateral damage to other aspects of our physiology, as illustrated by the following examples.

13.7.1 Gout among Polynesians: a Signature of Ancient Pathogen Exposure?

The evolutionary history of Polynesians and Micronesians has long been of interest. On their long journey across the Pacific Ocean, probably leaving the area of Taiwan about 7000 years ago, these populations have passed through multiple bottlenecks created by small groups migrating by boat. This, together with other demographic events such as disease, hunger, accidents, and warfare, has reduced their genetic variation. Novel environments, together with traditions that have culturally evolved from those brought from ancestral lands, fostered different kinds of social organization, which in turn influenced the history and evolution of these populations. The phenotypes of Polynesians today are shaped both by these long-standing evolutionary pressures and short-term, more recent, influences.

In terms of disease prevalence, while there are considerable differences among island populations, certain non-communicable diseases are found at unusually high rates in Polynesians and these patterns of prevalence have been explained by selective pressures in the past. Various candidates for “thrifty genes” (Section 9.4.1)—selected to promote energy storage at times of deprivation during Pacific voyages, but causing metabolic disease today—have been proposed, although the evidence remains indirect and very speculative. For example, a variant of PPARGC1A, a gene involved in insulin signaling, mitochondrial regulation, and thermogenesis,
is found at high frequency among Tongans and has been suggested to be a factor implicated in the very high rate of obesity and type 2 diabetes in the Tongan population (Myles et al. 2011).

Gout is one of the chronic conditions found at particularly high prevalence among Polynesians and Micronesians. Gout is characterized by sudden and severe bouts of joint pain, most frequently at the base of the big toe. Gout is caused by the excess accumulation of urate salts, in turn produced by high concentrations of uric acid, a product of purine metabolism. In most mammals urates are further degraded by the enzyme uricase and easily excreted in urine, but as a result of multiple mutational silencing humans and great apes do not have a functioning uricase. They can thus develop high plasma concentrations of uric acid (hyperuricemia), especially if excretion is ineffective (usually due to kidney disease or with certain medications) or purine levels are high because of diet and associated obesity.

The epidemiological observation of a high frequency of hyperuricemia in Polynesians and Micronesians, but not among Melanesians, indicates that the condition may have originated in Southeast Asia. It has been proposed that hyperuricemia could be explained through the action of natural selection in response to the selective pressure of malaria (Gosling et al. 2014). Malaria has been an important force in shaping the human genome (Section 13.9.1), and while it is not prevalent in the Pacific, it was common in Southeast Asia and some of the areas through which Polynesians passed on their voyages. Urate has multiple roles in cell metabolism and has been found to stimulate host inflammatory responses in malaria, so the hypothesis is plausible, though highly speculative. For instance, malaria has been present in the coastal regions of Papua New Guinea for a long time; indeed the early habitation of the malaria-free highlands may have been driven by the presence of Plasmodium in the low-lying regions. Yet inhabitants of Papua New Guinea who do not have Southeast Asian ancestry have low urate concentrations. In Section 13.9.1 we discuss the diversity of adaptive responses to the pressure of malaria: could it be that Melanesian Papua New Guineans responded to the pressure with a different adaptive mechanism, for example the lack of Duffy antigen on their erythrocytes?

### 13.7.2 Sleeping Sickness and Renal Disease

African trypanosomiasis, also known as “sleeping sickness,” is a severe insect-borne illness that has exerted some influence on the human genome. It is caused by the parasite Trypanosoma brucei, transmitted by the bite of the tsetse fly and found in sub-Saharan Africa with its epicenter in the Democratic Republic of Congo. The most significant symptoms of infestation with T. brucei are related to the central nervous system, and include disruption of the sleep cycle (daytime sleepiness and episodes of nighttime wakefulness), confusion, tremor, behavioral changes, and paralysis. If left untreated, the disease is fatal. While trypanosomiasis has been endemic to the region for a long time, historical studies indicate that intense epidemics can be traced back to environmental and social disruptions caused by colonialism just over 100 years ago.

Genetic studies have suggested that trypanosomiasis has created a positive selection pressure for variants of the human APOL1 gene (called G1 and G2), located in a region of the genome bearing a strong signal of selection. The product of APOL1 is ApoL1, a serum parasite-lysing factor that confers resistance to some subspecies of T. brucei, although not to the subspecies T. b. rhodesiense and T. b. gambiense responsible for sleeping sickness; carriers of either the G1 or G2 variant are resistant to T. b. rhodesiense. The G1 allele is found at low frequency across Africa, but at high frequency in the Yoruba of western Africa; the G2 allele is found at moderate frequency across Africa. Neither allele is present in individuals of European, Japanese, or Chinese ancestry. There is also genetic evidence for a third variant, G3, in the Fulani population of Central Africa that may confer resistance to T. b. gambiense (Ko et al. 2013). Patterns of linkage disequilibrium indicate that all these genotypes have been positively selected.

Carriage of a single allele of these APOL1 variants confers protection against trypanosomiasis. However, carriage of two of the alleles, G1 and G2 at least, is a risk factor for chronic kidney disease, and in particular HIV-associated nephropathy (Genovese et al. 2010), although the molecular mechanism is unknown. The predominantly West
African origin of African-Americans has led to a high frequency of APOL1 risk variants in that population and a corresponding high prevalence of kidney disease. This association between beneficial and deleterious effects of the same allele, depending on gene dosage, provides a further example of the balancing selection we will encounter again in relation to the malaria-related hemoglobinopathies (Section 13.9).

13.8 Results of Evolutionary Constraints

The inherent design constraints of our evolutionary past lead to many compromises that can have health consequences. There are several well-known examples (see also Section 7.4.5). The looping path of the long recurrent laryngeal nerve due to the origin of the gill arches in lower taxa can cause problems during neck surgery. The frequent complaint of lower back pain is a direct consequence of the human spine evolving from one suited to quadrupedal locomotion. The high incidence of detached retina is a result of how the mammalian eye evolved with the vascular layer external to the nerve layer (see Section 2.3.1.3). Appendicitis is the result of having a persistent remnant of a large herbivore cecum. But the examples that follow demonstrate more subtle interactions between evolutionary constraints and our contemporary environments.

13.8.1 Can Shoes be Doing More Harm than Good?

Although most of us would not think twice about wearing shoes, cladding our feet in cushioned footwear is essentially an evolutionary novelty. The oldest direct archeological evidence of footwear, some woven sandals, dates to somewhere between 8325 and 7675 BP in North America, but anatomical inference suggests that some form of protective footwear may have become common in a number of populations by around 30,000 BP (Kuttruff et al. 1998; Trinkaus and Shang 2008). It is known that those who usually go barefoot or wear minimal footwear have stiffer and stronger arches than those who are habitually shod. Does modern footwear alter the mechanics of how we run, and could this be related to the high frequency of injuries observed among those engaging in high-impact activities such as endurance running?

This hypothesis was tested by studying the running patterns of barefoot versus shod runners—more specifically, whether the runner tended to land heel first or forefoot first, as the former pattern generates a higher collision force upon contact with the ground (compare parts a and b of Figure 13.2 with part c) (Lieberman et al. 2010).

This study found that habitually shod runners mostly landed on their heels, a pattern aided by the elevated and cushioned heels of modern running shoes. In contrast, habitually barefoot runners mostly landed fore-foot (or sometimes mid-foot) first, generating about one-third the impact. Notably, when those who normally run with shoes were studied whilst running barefoot, the impact load on the heel was seven times higher than that for habitual barefoot runners. While the foot strike gait of early hominins is not known, it was posited that the evolution of a strong longitudinal arch in modern humans, which is absent in early bipedal hominins such as Australopithecus afarensis, confers advantages of protection and efficiency during forefoot landing. This suggests that shoe-wearing interferes with our evolved propensity not to land on our heels, and therefore that repetitive stress injuries in knees and ankles linked to prolonged running may represent a situation of mismatch created by our technological development of shoes.

13.8.2 Evolutionary Pressures Modeling our Faces

It has been known for decades that over the last 10,000 years, but especially the last 400 years, human skulls have undergone a general facial reduction, with a decrease in the size of the jaws when compared with the skulls of our ancestors. Investigations of skulls from cultures around the world have pinpointed a change from foraging and nomadic lifestyles to farming and settlements as the key event (Unger and Teaford 2002). It was suggested that cooking food in ceramic pots softened its consistency so intense chewing was no longer necessary, reducing the forces that needed to be exerted by the jaw through the action of the
Figure 13.2  Force generated by different patterns of foot strike. Rear foot strike without (a) or with (b) shoes generates an abrupt collision force (indicated by blue shading) of up to three times bodyweight; this so-called impact transient is absent in a forefoot strike running pattern (c). Adapted from Lieberman et al. (2010), with permission.
masseter muscles. The teeth have decreased in size too, but this reduction has not quite kept pace with the decrease in the size of jaws, and the ensuing discrepancy has resulted in dental crowding and malocclusion, i.e., incorrect alignment of the two dental arches as the jaws close. Dental malocclusion not only causes aesthetic problems but may also increase pressure on the temporomandibular joint, causing pain and difficulty eating, and increasing the risk of tooth decay (Lieberman 2011, p. 275).

More recent cultural changes have had a further effect and emphasize the role of developmental influences. While prolonged breastfeeding and weaning onto relatively coarse foods was the evolutionary norm, feeding behaviors and practices during infancy and early childhood have been increasingly replaced, from the start of the Industrial Revolution but especially through the twentieth century, by bottle feeding through an artificial nipple followed by weaning onto very soft foods. These appear to be the key pressures modeling the facial portion of the skull in the early, plastic period. In breastfeeding, the infant’s mouth is opened wide to receive not just the nipple but also the surrounding tissue—the breast molds itself to the mouth, the tongue stays in its neutral position in the mouth, and the movement is executed from the temporomandibular joint. Bottle feeding, by contrast, can push the tongue back and the soft palate upwards. The result may not only lead to the change in the jaw morphology but also in the palate shape. Breastfed children have wider and shallower palates than their bottle-fed peers, a difference corresponding to that found in hunter-gatherer versus modern populations. The shape and position of the hard palate influences the position of the soft palate, which then may lead to an increased risk of pediatric sleep disorders—the consequences of which, in turn, range from minor disturbances to significant emotional, behavioral, and cognitive dysfunction (Boyd and Sheldon 2014).

While the association between changed feeding practices and a change in the shape of the face seems well supported, it is not easy to separate out influences operating through the genome from those operating developmentally. It seems unlikely that the change in facial shape had a direct effect on reproductive fitness; and even if there was an effect through some unknown mechanisms, was it sufficiently strong to produce such obvious changes in 10,000 years—or a much shorter period, as many studies have shown?

Studies of bottle versus breastfeeding in recent years, as well as animal studies of environmental influences on the developing mandible, show that there is considerable developmental plasticity in oral development (Anderson et al. 2014). This plasticity gives rise to a further possibility when discussing the reduction in the size of the jaws. Cooking led to a reduction in the size of the muscle, as the force needed to chew was much reduced. Across the life course, muscle growth and thus force on its attachments decreased. Perhaps at first this was simply a developmentally plastic response, which was later fixed by genetic assimilation. This comparative historical research has inspired clinicians to propose starting the treatment of malocclusion and dental crowding in the early years, and to focus on stimulating the growth of the jaw (Rose and Roblee 2009).

An alternative possibility is that of sexual—or indeed social—selection, which can operate more quickly. A well-known experiment at a silver fox farm in Siberia has demonstrated that selecting for tame behavior also leads to changes in physical form, with docility being associated with infantile (“neotenic”) features including a high craniofacial ratio. The two effects are associated through changes in the HPA axis (Belyaev 1979). It has been suggested that in humans, too, selection for cooperative behavior has resulted in changes in the facial features, and that facial beauty—the standards of which are remarkably similar across societies—signals greater level of friendliness (Elia 2013). More supporting research is needed to strengthen this interesting hypothesis.

13.9 An Apparently Harmful Allele is Maintained by Balancing Selection

Balancing selection is frequently suggested as a mechanism to explain the maintenance of a harmful allele, but proving a hypothesis based on this mechanism is difficult. It requires one to show that the allele of interest is in Hardy–Weinberg equilibrium (Box 2.8) and to demonstrate the fitness advantage
of the heterozygote and the selective pressure leading to that advantage. These are difficult criteria to demonstrate, and with the exception of the protective influence of some alleles in several genes associated with red blood cell function against malaria other convincing examples are rare (but see Section 13.7.2). Other explanations that need to be considered when a potentially harmful allele persists in a population include antagonistic pleiotropy—where there is a fitness advantage of the allele earlier in life (Sections 5.2.4 and 12.5.2).

13.9.1 How Malaria Shaped the Human Genome

Malaria caused by the mosquito-borne parasite Plasmodium—and in particular its severe form caused by *P. falciparum*—is the strongest force known to have shaped the human genome (Kwiatkowski 2005). A particular example of the impact of malaria is sickle cell disease, the best known case of balancing selection or heterozygote advantage (Section 3.5.2). Sickle cell disease in homozygotes (and sickle cell trait in heterozygotes) is the result of a mutation in the gene coding for the structure of the alpha-globin chain of hemoglobin A. Similar mutations causing structural variations of globin chains are responsible for hemoglobin E and hemoglobin C that confer some level of protection against malaria.

It is not clear how this protection is achieved (Gong et al. 2013).

Downregulation of the production of either alpha- or beta-globin chains causes alpha and beta thalassemia, with the effect being severe disease in homozygotes but only mild anemia, combined with protection from malaria, in heterozygotes. While sickle cell disease is prevalent in sub-Saharan Africa and parts of the Middle East, the thalassemias have emerged in the Mediterranean and Asia. These mutations exemplify convergent evolution—that is, two distinct evolutionary pathways to the same phenotype (Section 2.4.2).

But while mutations affecting the production of globin chains are the best known results of evolutionary pressure exerted by malaria, natural selection has worked across the entire erythrocyte to make it inhospitable to the parasite (Figure 13.3). Starting from the cell surface, most populations around the world, except for sub-Saharan Africans and some Papua New Guineans, carry on the erythrocyte membranes the so-called Duffy antigens. *Plasmodium vivax* binds to the Duffy antigens and invades the erythrocyte. The lack of this antigen in sub-Saharan Africans explains the remarkable absence of *P. vivax* in parts of Africa. Deletion in the gene encoding glycophorin C, an erythrocyte membrane protein that serves as a receptor for *P. falciparum*, results in reduced invasion of this parasite.

![Figure 13.3](image)

**Figure 13.3** How malaria shaped the erythrocyte. This scheme shows different parts of the erythrocyte that, through polymorphisms propagated under the selective pressure of malaria, have been shaped by this disease (*Hb* = hemoglobin).
This deletion is found in populations of coastal Papua New Guinea. Also, deletion in the gene coding the membrane protein called band 3 results in a form of ovalocytosis common in parts of Southeast Asia and shown to be protective against certain forms of malaria.

Furthermore, once inside the cell, malarial parasites break down hemoglobin, which then releases by-products such as iron that may cause oxidative stress in many types of cells. A form of defense against oxidative stress is production of the electron donor NADPH by G6PD, encoded by G6PD on the X chromosome. There are many variants of G6PD, and those that significantly compromise G6PD activity result in hemolytic anemia; their geographical distribution is consistent with evolutionary selection by malaria and they seem to confer protection. G6PD deficiency has further clinical significance because certain drugs, including the anti-malarial primaquine, may trigger acute hemolytic anemia in people with G6PD mutations (Luzzatto and Seneca 2014).

While malaria has shaped the erythrocyte, its impact has not stopped there. Following infestation of the cell with parasites, the next critical event is its sequestration in small blood vessels, which on the host side involves various adhesion molecules on the vascular endothelium, platelets, macrophages, and other erythrocytes. It seems that polymorphisms of genes encoding antigens used by P. falciparum to bind to host cells (e.g., CD36 encoding CD36, a receptor for a range of molecules including long-chain fatty acids and expressed by platelets, dendritic cells, and endothelium) are associated with protection against some types of malaria. Furthermore, a number of genes involved in the immune system appear to be associated with different malaria phenotypes: these include cytokine receptors, complement activators, and some cytokines. The case of gout described in Section 13.7.1 indicates other ways in which malaria may have had an effect on our bodies.

While many of these possible interactions between human genomes and malaria beyond erythrocyte function require more evidence, it is helpful to keep in mind the possibility of an evolved trait when encountering unusual polymorphisms and phenotypes in individuals from a population that were exposed to malaria in the past.

13.10 The Consequences of Sexual Selection

Sexual selection is an important evolutionary mechanism (see Section 2.3.2.3) and allows for the particularly rapid evolution of a trait. Yet demonstrating that the selection of a particular trait in humans has been driven by sexual, rather than natural, selection or other mechanisms such as drift is difficult. It requires evidence that the trait in question makes the individual more competitive in the mating process: by being more attractive to the opposite sex or more able to dispose of competitors or retain the mate. Yet for all of the traits that have been studied from this perspective, a selective advantage has been extraordinarily hard to prove. One reason is that there is a strong influence of cultural practices in human mating behavior, and these are both varied and often rapidly changing (Laland et al. 2010). A further reason is that providing a plausible explanation for why a particular trait should be targeted by sexual selection is by no means straightforward and often wildly speculative.

Take the example of hair and eye color, phenotypic traits that are frequently associated with sexual selection. The diversity of hair and eye color in northern and eastern Europe compared with the rest of the world has been explained by “rare color advantage,” the tentative mechanism according to which the rarity of (in this case) blonde or red hair, or green or blue eyes, caught the attention of prospective mates (Frost 2006). This selection, it is argued, only took place in certain parts of Europe because in the harsh environment of the ice-age tundra men procured most of the food yet could not support multiple wives and had high mortality. Therefore, male scarcity was solved not through a polygynous mating structure but through intense female competition for men, leading to the diversification of hair and eye color (and, possibly, accentuation of other visible sexual dimorphisms). It will be apparent that there are many questionable assumptions built into this argument, making it nothing other than a “just-so story.”

The prime targets of explanations invoking sexual selection are genes associated with externally visible phenotypes. Many of them, including melano-cortin 1 receptor (MC1R), associated with skin color,
and ectodysplasin (EDAR), associated with hair texture, show strong signatures of positive selection (Laland et al. 2010). In both cases, sexual selection has been raised as an explanation. For example, the pressure of UVR provides the best explanation for the overall gradient of skin color observable among humans (Section 13.4.1). Intriguingly, the darkest skin color (determined by the greatest number of active, pigment-producing melanocytes) coincides with the human reproductive career: while children generally have a paler skin color, melanocytes begin to activate towards puberty (Jablonski 2013, p. 71). Females achieve their darkest pigmentation around menarche, while males continue to grow darker into their late teens. Past 35 years of age, pigment production slows down and old people tend to be paler than younger adults. This pattern of skin coloration as well as the fact that women are consistently paler than males in any indigenous population has raised the possibility of sexual selection.

Yet why should pale skin be more attractive? Some scholars have proposed that the attractiveness for males of females with light skins lies in the association between light skin tone and infancy, eliciting care-taking responses in adults and inhibiting aggression (Frost 1988). Alternatively, perhaps the female skin tone is the baseline and it was female choice that pushed the male skin to darker tones, as males with darker skin had better folate levels and therefore better sperm production, as cell division requires folate (Section 13.4.1) (Aoki 2002). As a further explanation, a natural selection hypothesis was proposed according to which women tend to be slightly paler to ensure greater vitamin D synthesis and thus sufficient reserves of calcium in their body, to support pregnancy and breastfeeding (see the example in Section 13.4.1) (Jablonski 2013, p. 89). Light skin would then be a sign of greater potential fitness.

Similarly, the selection of a mutation in the EDAR gene that is linked to coarser hair in East Asian and Native American populations has been tentatively explained as a result of sexual selection—a preference for thicker hair (Laland et al. 2010). Yet, as discussed in Section 6.5.2, this gene has effects on a range of other structures and organs, such as teeth and sweat glands, so the change in hair morphology may simply be a spandrel.

The caution with which we approach sexual selection as an evolutionary explanation for human traits including behavior (and, by extension, human disease), should not, however, be interpreted as reluctance to consider it. In the area of human disease in particular sex differences are substantial (Ober et al. 2008). While some of them may at least partially be explained by the direct action of endocrine systems, for example the effect of estrogen on female health described in Section 13.5.1, or immune changes associated with reproduction, no satisfactory explanation has been found for others. Some differences will be due to sex-specific differences in gene regulation, which may be driven by natural selection but have sex-limited effects, or by sexual selection.

In evolutionary psychology, and particularly in popular accounts, sexual selection is often used to explain gender-specific difference in behavior. Most of these claims are poorly founded, but it seems likely that sexual selection played some part in determining the rather different behaviors of males and females in child rearing and food gathering in our Neolithic past.

13.11 The Outcomes of Cladal and Demographic History

13.11.1 Our Traveling Companion: Leprosy and Human Migrations

Leprosy is a disease that has played an important role in our social and cultural histories. Although leprosy is a chronic and only mildly contagious disease, and so not nearly as dramatic or as dangerous as plague, in medieval Europe and the Middle East the appearance of the sick and the stigma of the disease provoked so much fear that unprecedented social strategies were invented to exclude people with leprosy from communities. In the West, leprosy began to disappear by the 1700s, possibly because of the plague’s selective effect against the immunologically compromised. Yet long after Hansen’s description of Mycobacterium leprae in 1873, and despite the availability of effective treatments, this obligate parasite, which is adapted almost exclusively to humans, remains a problem in some parts of the world.
Molecular and phylogeographic analysis of extant and now extinct *M. leprae* from archeological specimens has built a picture of the spread and evolution of its strains worldwide, and has also examined how this microbe became a human pathogen (Monot et al. 2009). Strains from Tamil Nadu, Brazil, Thailand, and the USA were sequenced and compared, selecting 84 informative SNPs. Then, the informative SNPs were tested in isolates from 28 countries, supplemented with ancient *M. leprae* DNA obtained from medieval European and Middle Eastern bone samples. The study showed that the progenitor strain of leprosy, SNP type 2, left East Africa together with early humans, then gave rise to SNP type 1 that spread with human migrations eastward into Asia and SNP type 3 that spread westward into Europe and the Middle East, before giving rise to SNP type 4 found in West Africa and in the countries linked to that region by the slave trade (Figure 13.4). Leprosy in Asia appears to have arrived by two different routes: a southern route, with new *Homo sapiens* settlers (compare with Figure 6.4) and a northerly route, through Central Asia to China, Korea, and Japan. The latter corresponds with the trading route called the Silk Road; it is thought that the worst plague epidemic in human history, the Black Death of 1347–51, had reached Europe from China that way (see Box 10.3). Also, it appears that leprosy was brought into the Americas not by early human migrations across the Bering Straits but rather much later by European immigrants.

A puzzle that remains is the discrepancy between the low genetic diversity of *M. leprae* (99.995% sequence identity among the four strains sequenced) and the exceptionally large number of pseudogenes present in these strains. Furthermore, the period at which pseudogene formation is thought to have occurred—an indicator of a radical change in the lifestyle/environment of the organism, such as a change from free-living to parasitic—does not fit with human evolutionary history. If an ancestor of *M. lepra*e infected an early hominin or primate and was then

![Figure 13.4](image-url)
transmitted within species, one would expect more genetic diversity. This has led to the hypothesis that while the genomic decay of M. leprae as indicated by the large number of pseudogenes is indeed ancient, human parasitism as a lifestyle for this microbe is not, and that an ancestral form of M. leprae infected an invertebrate host (perhaps an insect) that acted as an intermediate vector. Only then did it infect humans. Selection on this variant, which favors the human host, keeps genetic diversity low. The question remains open, but it is an intriguing hypothesis.

13.11.2 Altitudes and Ancestors: Tibetan Adaptation to Low Oxygen Pressure

Humans have colonized a range of environments across the planet, including the rather inhospitable high-altitude plateaus of Central Asia and South America. The Tibetan plateau has probably been inhabited for more than 20,000 years and the Andes for around 11,000 years; permanent settlements in both regions are found up to 3500–4500 m above sea level.

But most humans are maladapted to the low oxygen pressures found at these altitudes. New arrivals to high altitudes, especially those who have rapidly ascended, develop a set of symptoms known as altitude sickness including headache, dizziness, sleep disturbance, acute lung and neurological conditions, and occasionally death. In contrast, indigenous populations who live at high altitudes exhibit circulatory, respiratory, and hematological adaptations to low oxygen pressure. Yet Andeans and Tibetans have adapted very differently. Compared with Andeans, Tibetans have a suite of physiological adaptations that fine-tune their respiratory physiology, including a lower arterial oxygen content, increased resting ventilation, lack of pulmonary vasoconstriction, lower incidence of low birthweight (with pregnancy presenting a particular hemodynamic challenge at high altitudes), and hemoglobin concentrations similar to that of populations at sea level (Petousi and Robbins 2014). In particular, the (relatively) low hemoglobin concentration of Tibetans has attracted attention.

For most people, compensatory increases in hemoglobin concentrations in response to tissue hypoxia are the key adaptation to low inspired O₂ pressures. But as the hemoglobin level rises, so does blood viscosity—which can cause cardiovascular problems. In contrast to Andeans, who have high hemoglobin levels, Tibetans maintain normal hemoglobin concentrations because of their blunted stress response to altitude. These changes take place through a sequence variation in EPAS1 (endothelial PAS domain protein 1, also known as HIF2α) encoding the oxygen-sensitive subunit of HIF-2 (hypoxia-inducible factor) (Beall et al. 2010). SNPs in and around EPAS1, EGLN1 (a regulator of HIF), and PPARA (a transcriptional target of HIF) are strongly associated with lower concentrations of hemoglobin relative to concentrations found in lowland residents moving to high altitudes, or in Andeans. These gene variants in the Tibetans seem to have been under strong positive selection, resulting in near fixation over several thousand years.

It has been suggested that this highly unusual, yet highly prevalent (within the Tibetan population), haplotype resulted from interbreeding with Denisovans, a now extinct hominin species (see Box 6.1). The relevant haplotype pattern in Tibetans is identical to that of Denisovan DNA sequenced from one individual, and is completely different from their neighbors, the Han Chinese. While, to our present knowledge, Denisovans resided in the Altai mountain range that is not as high as the Himalayas, the altitude may have been sufficient to create selective pressure (Huerta-Sanchez et al. 2014).

In addition to illustrating demographic influences, namely interbreeding and introgression from another hominin species, this pathway also illustrates how selective pressures of a particular physical environment can act upon human phenotypes. Andeans exemplify the response possible within the constraints of the ancestral norm of reaction (see Box 4.1). Tibetans show what happens when a novel mutation appears and is positively selected, providing the basis for new modes of adaptation.

13.12 Value and Limits of an Evolutionary Medicine Perspective

In this book we have used humans for our illustrative examples wherever possible. This approach
contrasts with other evolutionary biology texts that use observations collected across multiple taxa, and thus describe systems that may best illustrate a particular process but are not necessarily relevant to humans. The primary purpose of this book has been to show that just as evolutionary biology provides substantial explanatory power in understanding other components of the biotic world, the same principles and processes provide a critical perspective on the human condition, whether viewed from the standpoint of an individual or as part of a broader ecological system.

We have attempted to show that an evolutionary perspective adds considerably to the traditional proximal focus of much of clinical medicine and public health. It demonstrates the value of a broader ecological perspective, offers causal explanations at a different level from those of proximal mechanisms, and provides important new insights for therapeutic or preventative approaches. Using a range of examples, we have illustrated how evolutionary explanations can suggest new directions that inform basic, clinical, and public health research.

Evolutionary explanations tend to focus on “why” a clinical problem has emerged rather than the more mechanistic “how” question. Because of this, they offer the clinician a valuable and additional approach to explaining symptoms and disease that many patients will find both informative and intelligible. In some cases, such as the management of phobias or perimenopausal symptoms, evolutionary explanations can themselves be part of therapy or can inform the therapeutic choices that the patient may make.

In this chapter we have added to the cases described in the second half of the book, both to expand the range of illustrative material and to make some additional points. While the taxonomy of evolutionary explanations provided in Chapter 7 and in this chapter is a useful heuristic, the essential nature of evolutionary processes means that multiple pathways are often involved and a case can be viewed from several angles. This does not diminish the value of a mechanistic taxonomy: rather, it highlights the point that evolutionary processes are not unitary but interlinked. This broader aspect must be maintained. For example, influenza can be, and generally is, viewed simply as a respiratory viral infection of varying severity and morbidity. But the example of influenza highlights cultural evolution leading to greater interaction with animals, the rapid evolution of a virus and its life-history strategy (virulence versus transmissibility), and the technological arms race of maintaining effective vaccines. Each of these dimensions has direct relevance to public health strategies. Attention becomes focused on animal hosts for public health surveillance and, in the case of an outbreak, on increasing the barrier between the animal host and humans; the genomic sequence of the influenza virus is constantly being screened to study its evolution and to inform vaccine development and prophylactic public health strategies. Consciously or not, all of these actions are drawing on evolutionary principles.

As highlighted in Chapter 7, testing evolutionary hypotheses is complex and often more indirect than testing hypotheses in traditional biomedical sciences. This has unfortunately led to much teleological speculation and “just-so stories,” which can be considerably exaggerated by their protagonists and the media. We think it is critical that all health professionals have a sufficient understanding of evolutionary biology to be able to see through such arguments and to help themselves, their patients, and the population to more rigorous application of evolutionary understanding. Equally, we have tried to demonstrate the large gaps in our knowledge and the consequent enormous research agenda for the field, as well as the limits of evolutionary explanations given these knowledge gaps.

The fact remains that evolutionary science is the key integrating principle of all biology. Humans and their health and diseases cannot escape this principle, and the more we understand the multiple histories of patients from their immediate illness to their macroevolutionary past, the more we will be able to assist in promoting a better quality of life and longevity for everyone.
Key Points

• A holistic approach to medical history includes, in addition to personal and family history, the history of the patient’s lineage, human biological and cultural evolution, and broader biological evolution.
• An evolutionary perspective on clinical signs and symptoms considers their possible adaptive function and may influence therapeutic decisions.
• Explaining “why” a symptom or disease has occurred may be of particular value to the patient.
• Constructing and testing evolutionary hypotheses requires a good understanding of evolutionary biology, medicine, and social history.
• Examples in this chapter illustrate these points and also show how multiple pathways interact to influence disease risk.
14.1 Introduction

The primary goal of this book has been to demonstrate how knowledge about evolutionary principles can add to a broader appreciation of health and disease. Indeed, evolutionary biology integrates our understanding of the human species from the level of the molecule to the level of society. It provides the basis for conceptualizing how ultimate biological processes lead to greater understanding in proximate biology, the core concern of most medicine. Yet while evolutionary biology is an established scientific field based solidly on empirical data and with a coherent conceptual basis, it is also a discipline that has been the subject of an unusual amount of controversy and opposition. If evolutionary medicine is to be useful, those who employ it must understand these issues, some of which are addressed in this final chapter.

There are several reasons for such strong reactions. First, evolutionary biology removes much of the mystery about who we are, and it takes a biological and materialistic approach to understanding human origins, human nature and behavior, and the possible future of humankind. Many find this approach challenging: does this mean that humans are biological automatons or that human culture has neither a learned nor a moral component? Furthermore, concepts of descent with modification and of natural selection contest the myths of human origins and the place of humans in the world: these are at the core of most religions as well as cultural traditions.

Evolutionary biology is not the only field of scientific discovery that has challenged religious dogma—consider the response of the Church to the arguments of Copernicus and Galileo that displaced the Earth from the center of the universe. But evolutionary biology’s concern with the living world, including humans, has been seen as particularly provocative.

Further controversy arises from the history of Darwin’s theory of evolution: its interpretations, use, and misuse in areas ranging from national politics and social welfare to clinical medicine and biomedical research. Some critics of evolutionary science draw a direct line from Darwin to the horrific genocide committed by the Nazis (Weikart 2004). Others extend their criticism to the present day, suggesting that Darwinian arguments have promoted what they regard as unacceptable practices such as voluntary abortion and the modern use of pre-natal diagnosis. Others hold antagonistic views because they have seen Darwinian arguments misused to provide grossly flawed and strictly hereditarian explanations of differences in academic achievement and socioeconomic status among different ethnic groups.

To appreciate such criticisms and concerns fully, so as to be able to respond to them, we must first place Darwin’s evolutionary thought and its relationship to medicine and human society within the historical context in which it originated and developed. Then we need to take an overview of the history of the relationship between evolutionary science and medicine, and highlight certain social, cultural, and legal challenges that evolutionary medicine has faced and is facing, as well as those that may emerge in future.

14.2 Origins of Darwin’s Theory

Until the eighteenth century, the question of human origins was the remit of religion. In the
West, Christianity combined the story of origins contained in the Jewish Old Testament with Greek philosophy. Thus the account of a young Earth, swift creation of the world through divine will, and the eventual arrival of a single pair of humans who would give rise to the world’s population was brought together with the belief that the purposeful complexity of nature could not have emerged through a blind force (Ruse 2013). Other cultures developed their own creation myths, but given the origins of evolutionary biology in Western science it is the responses within the Western tradition that are the focus of this discussion.

It is no coincidence that the first theories of evolution appeared in the period of rapid social, demographic, economic, and political change that followed the upheaval of the Reformation, the Industrial Revolution, and the emergence of global travel. Imperialism and colonial expansion brought the West into close contact with other religions, human cultures, and different stories of origin. The new wave of philosophical ideas in Europe, and the rise of natural science from the seventeenth century on, had much to do with these experiences (Drayton 1998). Scholars ordered the extraordinary natural objects collected worldwide into classification systems, of which the best known is the binomial nomenclature developed by the Swedish systematizer Carl Linnaeus (1707–78). In Linnaeus’s divinely inspired and directed natural world, each species was the subject of separate design and creation (see Box 1.3).

There were several different attempts to order the diversity of organic life and to explain its history, including that of humans. Many scholars transferred the idea of progress that underpinned so much of contemporary social change—the belief that it is possible to improve one’s lot through effort—to the living world, ranking the organisms on a ladder of increasing perfection (Bynum 1975). Georges-Louis Leclerc, Comte de Buffon (1707–88), was among the first to suggest that humans had emerged from an earlier, more primitive state. Buffon also suggested a notion of organic evolution, alluding to the possible common ancestry of humans and the apes. Also writing in pre-Revolutionary France, the philosopher Denis Diderot (1713–84) suggested “Just as in the animal and vegetable kingdoms, an individual begins, so to speak, grows, subsists, decays and passes away, could it not be the same with the whole species?,” foreshadowing the concept of emergence and extinction of species (Ruse 2013). Later in his career, Linnaeus too was to suggest that many species had arisen following creation through hybridization from primal species. Charles Darwin’s grandfather Erasmus Darwin (1731–1802), a country physician and natural philosopher, proposed that organisms could develop over generations towards a higher order through accumulation of experience (Porter 1989). He wrote verses on organic change, which, for him, was tied to the idea of cultural progress. In his 1794 book *Zoonomia*, he espoused a view of life as a continuous chain from the most primitive forms to humans.

These early “transmutationist” ideas were by and large theoretical, lacking an understanding of the fossil record, geographical distributions of species, or comparative anatomy. In the French Revolution (1789–99), the traditional hierarchical social order, with the aristocracy at the top enjoying social privileges and peasants at the bottom having none, was replaced by one in which everyone—at least in theory—had the same rights. Post-Revolutionary France established the first institutions to focus primarily on the study of the history of nature. The emerging discipline of geology provided evidence for the history of the Earth and of the life on it; comparative anatomy examined extant organisms and fossils to construct the history of species; while individual embryonic development came to be seen to repeat the important stages through which the ancestors passed during the evolution of the species (Temkin 1950; Cunningham and Jardine 1990; Hopwood 2009).

Arguably the most sophisticated and best known early model of evolution was proposed by the French naturalist Jean-Baptiste Lamarck (1744–1819) (Burkhardt 1995). Lamarck was a prolific biological theorist, a professional botanist turned zoologist, from minor but impoverished nobility. He managed to survive the French Revolution to achieve a professorship in the Museum of Natural History in Paris, but was to end his career in an impoverished and ignominious state. Lamarck maintained that the simplest species emerged by spontaneous generation and new species then appeared by progress
up a ladder of development. He proposed two laws. The first was that use enhances, and disuse atrophies, an organ. The second was that the effects of use or disuse were passed through reproduction to the next generation, a concept frequently referred to as the “inheritance of acquired characteristics” and popularly illustrated by the example of the giraffe’s long neck resulting from stretching in each successive generation to reach ever higher tree branches.

Lamarck’s *Philosophie Zoologique* (1809) received a critical and negative reception from his influential younger Parisian rival Georges Cuvier (1769–1832), a great comparative anatomist and influential paleontologist. Until that time it was generally believed that species did not become extinct, and that the large fossil animal bones found in Europe belonged to species still extant in Africa. Cuvier showed unequivocally that this was not the case, and argued that most species that had ever lived were now extinct, destroyed in various catastrophic events: this was the school of “catastrophism.” He was highly skeptical that species could change. Rather, he maintained, they became extinct, and his ideas were highly influential in the opposition to early evolutionary concepts, even after his death.

Cuvier’s main opponent was the third member of the Paris natural history professoriat, Étienne Geoffroy Saint-Hilaire (1771–1844), who had developed a concept of species transmutability somewhat similar to that of Lamarck (Appel 1987). Using comparative anatomical methods he demonstrated homologies between organs such as fins and limbs across phyla. He even noted the homology between the middle ear bones of mammals and the jaw bones of fishes, which we now recognize as an important example of exaptation.

In the twentieth century, after the “Modern Synthesis” interpretation of Darwinian evolution through the lens of genetics, Lamarck was often pitched in opposition to Darwin (Mayr and Provine 1980). But as we will show a little later, inheritance of acquired characteristics was a widely accepted and non-controversial phenomenon well into the second half of the nineteenth century. Some have suggested that modern concepts such as transgenerational passage of epigenetic marks (Section 3.8.3) have echoes of Lamarck. Yet between Lamarck’s concept of use and disuse and the effect of the environment upon parental physiology and epigenetically mediated gene expression, there are 200 years of research into the mechanisms of heredity, normal and pathological development, and the effects of the environment on development. Superficial similarity should not mask important essential differences. Modern concepts of non-genomic inheritance do not refer at all to the inheritance of acquired characteristics, but rather to environmental memory recorded by the multiple mechanisms of non-genetic inheritance (see Section 3.8). The use of the term “neo-Lamarckian” may have biased many scientists against considering the issues of non-genomic inheritance seriously, so it should be avoided.

While the ideas of transmutation contradicted the religious stories of creation, by the late eighteenth century many no longer took the creation myth literally. One could easily imagine a God who had created both the universe and the laws by which the universe works, and then let it run without further involvement. A version of this approach was later adopted by some Christian churches (Section 14.6). However, this did not mean that the idea of evolution—or rather, transmutation, as it was known then—was accepted wholeheartedly. Many prominent philosophers and scientists of that period, including Immanuel Kant (1724–1804) and Georges Cuvier, were arguing that the purposefulness of organisms is incompatible with a blind force operating to produce such outcomes (Ruse 2013).

Opposition to this emergent evolutionary thinking in the early nineteenth century was bound up with a general hostility towards the idea of progress, which grew stronger in the aftermath of the French Revolution as conservatives attempted to rebuild the old social order. In the tumultuous Britain of the 1830s, support for the arguments of Lamarck and Geoffroy Saint-Hilaire was associated with radical politics and support for Cuvier with conservatism (Desmond 1989). Of particular significance was the publication in 1802 of *Natural Theology or Evidences of the Existence and Attributes of Deity* by the English clergyman and philosopher William Paley (1743–1805). Paley espoused a teleological view of nature according to which God had designed each organism with a purpose and with characteristics to match its ecological niche (Brooke 2009). He famously likened God to a watchmaker: because
organisms were as complex as a watch, there had to be an active designer. This argument is echoed by those who today advocate “intelligent design,” and Paley’s analogy gave Richard Dawkins the title for his 1986 book *The Blind Watchmaker*, one of several books in which he demonstrates the power of evolutionary biology to create complex designs without an active designer (Dawkins 1986). Paley’s book was very popular, and was read by Darwin, influencing his thinking about how and when to present his arguments for natural selection.

This was the world into which Charles Darwin, a descendant of a wealthy family, was born in 1809 (Desmond and Moore 1992). Like his father and grandfather, he intended to become a doctor, but while studying at Edinburgh he found medicine not to his taste. One of his teachers at Edinburgh was the radical anatomist Robert Grant (1793–1874), a committed Lamarckian and a close friend of Geoffroy Saint-Hilaire. Darwin became an enthusiastic biologist under the mentorship of Grant, studying sponges and other marine organisms that would become the subjects of his first scientific presentations. Not wanting to pursue medicine, he moved to Cambridge where he took the “ordinary” degree (a separate theology degree was introduced decades later), intending to go into the church. But he became immersed in the excitement of the emerging natural sciences of geology, botany, and zoology. In particular, advances in geology captured Darwin’s interest. A new world view, namely that the world was old and had been molded by the gradual actions of forces such as water and ice, emerged to replace the belief in “catastrophism,” particularly that of a biblical flood. The most influential of the new wave of geologists was Charles Lyell, whose three volumes of the *Principles of Geology* (1830–3) and friendship heavily influenced Darwin. Darwin’s first significant scientific writings were on the emergence of coral reefs and atolls, and in these he developed concepts which are largely still valid today. Lyell, although a deist, was more than any other responsible for the new view that the natural world was old and that change had occurred slowly. He argued that the study of processes that were still observable was a key to understanding the past, and thus geological change was slow: this “uniformitarian” approach was in marked distinction to the “catastrophic” school.

Upon graduation, Darwin accepted the post of a naturalist and companion to Captain Robert Fitzroy, under whose command the ship HMS *Beagle* was to take a 5-year surveying journey around the world. He started out on that voyage in 1831 as a firm believer in natural theology as exemplified in William Paley’s writings. He returned in 1836 with ideas about speciation and evolution that he was starting to consolidate, married his cousin Emma, and, supported by family wealth, spent the rest of his days at Down House, Kent, as a gentleman scientist: experimenting, writing, and corresponding.

Much has been written on the various influences that contributed to Darwin’s theory of natural selection. Beyond his family and education, experiences from the voyage of HMS *Beagle* provided the cornerstone of his thinking, although in many ways he realized their significance only years later (Box 2.2). His observations on the Tierra del Fuegians on board the *Beagle*, whom Fitzroy was returning to their homeland, influenced his thinking about human evolution. He worked with breeders of pigeons and domestic animals to understand what he was to term “artificial selection;” breeders and physicians provided him with examples of hereditary traits (Bynum 1983). The key theoretical inspirations arguably came soon after the *Beagle* trip, from reading the work of the highly influential British social theorist Thomas Malthus (1766–1834). In his 1798 work *Essay on the Principle of Population* Malthus argued that a population always tended to outstrip its food supply, which inevitably led to poverty, starvation, and death. Poverty and famine were thus natural checks that kept the population size at a manageable level. Malthus inspired Darwin to propose that those individuals best adapted to their environment will survive to reproductive age and breed, while those who are poorly adapted will die. In the next generation, the better-adapted lineage would be more numerous, thus changing the course of evolution. Note that Darwin used the word “descent” instead of “evolution;” from the mid-seventeenth century, the latter term (coming from the Latin *evolutio* or unfolding) was used to describe embryological development. The earliest use of “evolution” with a meaning close to the modern sense of the word was by Charles Lyell in the 1830s, but its popularity may probably be attributed
to Herbert Spencer, who in the 1860s began to use “evolution” to describe transmutation. Darwin used the term “evolution” in the introduction to *The Descent of Man* (1871), but overall made little use of it in the bulk of his writing (Bowler 1975).

It took Darwin more than 20 years to publish his ideas. Indeed, were it not for a letter he received from Alfred Russell Wallace, Darwin’s book might have waited longer. Wallace was a self-made man who traveled extensively seeking natural history specimens for wealthy private collectors. While in Southeast Asia, and also reading Lyell and Malthus, he came to an idea similar to that of Darwin’s natural selection. To secure Darwin’s precedence, his supporters Charles Lyell and Joseph Hooker arranged to read an excerpt from his unpublished essay *On the Tendency of Species to Form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection* alongside Wallace’s presentation to the meeting of the Linnaean Society on July 1, 1858. The relative contributions of Darwin and Wallace to the discovery of natural selection and the debate over the event around the joint announcement have been the subject of many books: Wallace himself acknowledged Darwin’s primacy (van Wyhe 2013).

In 1859, Darwin published what would become the first of the six editions of *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (Darwin 1859). *The Origin* was meant as an abstract of a much longer and never-completed book. Darwin’s major contributions are described throughout that volume and include those of natural and sexual selection, the latter being the focus of his second great book *The Descent of Man and Selection in Relation to Sex* (Darwin 1871); the importance of variation and heritability; and the importance of cumulative small change in the transmutation of species. There was wide public interest in Darwin’s writings; his early travelogue about his voyage on HMS *Beagle* had already been highly popular. The publication of *The Origin* led to satire, parody, and debate, both scientific and popular, but the issues it raised were much deeper. The challenge of evolutionary thought was that it replaced concepts of divine order with a focus on biological materialism and determinism. The best known confrontation was at the meeting of the British Association for the Advancement of Science in Oxford in 1860 where Samuel Wilberforce, Bishop of Oxford, and Thomas Henry Huxley, Darwin’s most strident supporter, had an exchange. While accounts of the meeting differ, the most widely told, but not validated, anecdote is that Wilberforce asked Huxley whether he was descended from monkeys on his grandfather’s or his grandmother’s side. Huxley muttered: “The Lord has delivered him into my hands” and replied that he “would rather be descended from an ape than from a cultivated man who used his gifts of culture and eloquence in the service of prejudice and falsehood.”

For all the controversy that it created, Darwin’s reputation in Britain and overseas only grew. He attracted dedicated followers and popularizers. Some, such as the German Ernst Haeckel, had such an impact themselves that in their countries their work would come to stand for Darwinism (Richards 2008b). Despite the opposition by Samuel Wilberforce, Darwin had cordial relationships with many prominent Anglican clergymen and he was never subject to any formal denunciation. On the contrary, when he died in 1882, he was buried with full honors in Westminster Abbey (Pallen and Pearn 2013).

### 14.3 From Darwin to “Social Darwinism”

Darwin and his theory were certainly children of their time. The German political philosopher and intellectual father of modern communism Karl Marx (1818–83) noted the parallel between natural selection and unregulated capitalism, the type of economic system that first developed in the oldest industrial power, the British Empire. While the term “survival of the fittest” is commonly associated with Darwin, it was coined by the political and social philosopher Herbert Spencer, a firm supporter of unrestrained free enterprise who argued that the state should have no control over citizens. It is Spencer who is seen as the father of “social Darwinism,” a set of various theories, ideologies, and practical programs that emerged in the late nineteenth century and that applied an arguably Darwinian evolutionary approach to thinking about human society (Paul 2003). Yet the usefulness of this term is dubious if
we keep in mind that none of the supposed “social Darwinists” called themselves thus. Rather, “social Darwinism” was an umbrella term invented and applied by critics who opposed the idea of unrestrained capitalism. For example, although Spencer accepted the role of natural selection, he was actually a Lamarckian who believed in the inheritance of acquired characteristics. He argued that competition stimulated (heritable) self-improvement.

Even if we do use the term “social Darwinism,” we have to acknowledge that Darwin had no input on how the generation that followed him would interpret and apply his ideas which, by the end of nineteenth century, had a wide currency. For example, the German philosopher Friedrich Nietzsche (1844–1900) extended the biological metaphor of the survival of the fittest to argue that selfish behavior and even violence between individuals was logical and would lead to the evolution of an appropriate human meritocracy with a superman elite on top. At the same time, the anarchist movement, under the guidance of the Russian Pyotr (Peter) Kropotkin (1842–1921), proposed an entirely different interpretation of evolution. For the anarchists, evolution had a progressive direction; the human was the most advanced species of all, the highest form of evolution being based on cooperation and altruism rather than conflict, and so in turn cooperation was an evolutionary force. Kropotkin was convinced that selection acted only because of conflict against the physical world, and rejected the idea that it involved competition with other species or members of the same species. Thus the ultimate manifestation of anarchism would be a society based on what he termed “mutual aid.”

14.4 Eugenics

Probably the best known area often associated with “social Darwinism” (and one in which medicine was heavily involved) was the movement for social reform through the control of reproduction, known as eugenics in the Anglo-American context or social and/or racial hygiene in Central Europe (Kevles 1985; Paul 1995; Levine and Bashford 2010). The term “eugenics” was coined by Darwin’s cousin, the polymath, scholar, and explorer Francis Galton (1822–1911), whose personal interest in the hereditary basis of human abilities led him to argue that human character depended on heredity only, and that environment and free will had no impact on human action. Like Darwin, he was a descendant of an eminent family. Worried by the falling birthrate among his own, educated class, and appalled by the sight of urban slums, Galton argued that the poor reproduced much more quickly than other sections of Victorian society, and that this trend should be discouraged in order to enhance the quality of society.

Although Galton is considered the pioneer of eugenics, he was by no means the only person to worry over the supposedly diminishing quality of the human “stock.” Even Darwin was concerned about it, although he never went so far as to propose a (eugenic) solution to this problem. Before Darwin and Galton, there had been widespread concern about the perceived deterioration of humankind being exacerbated by civilization. In 1857, the French asylum psychiatrist Benedict August Morel (1809–73) published his *Traité des Dégénérescences Physiques, Intellectuelles et Morales* in which he introduced the concept of “degeneration” that quickly gained popularity (Pick 1989). Degeneration was seen as the product of both organic and social factors and was heritable and cumulative over generations—an evolutionary reversal of a sort. It started with simple “nervous ailments” caused by the stress of modern life, and then descended, via alcohol addiction and criminality, to insanity and mental retardation. Once a family was on the downhill slope, there was no return; prevention of their reproduction offered the only solution for society. With the popularization of Darwin’s theory, this concept was placed in an evolutionary framework. While in the past, the argument went, natural selection took care of removing the “unfit” from the population, modern medicine and welfare negated this salutary effect of natural forces.

The growing popularity of theories of degeneration and eugenics was an indicator of the shifting views on heredity. The very word “heredity” was first used in the biological context in the 1830s (López-Beltrán 2007). The term borrowed from the older, legal, context, as until then there was no notion of straightforward transmission of traits across generations within biology. Before heredity,
every organism was thought to have been created in the act known as “generation,” out of maternal and paternal contributions and under a strong influence of the environment before and after conception (Müller-Wille and Rheinberger 2012). Well into the nineteenth century, most people believed in an essentially Lamarckian idea of inheritance where heredity could be modified by environmental influences. Indeed, Darwin’s own concept of inheritance, “pangenesis,” was essentially Lamarckian: he believed that every part of an organism can throw off minute granules that are dispersed through the body, develop into units, and then are collected from all parts of the body to constitute sexual elements. If the organ changed, so did these “gemmules” (Endersby 2009).

But advances in cell biology illuminated aspects of heredity. These began with the idea that the cell is the basic unit of the organism and that cells multiply through cell division (rather than, e.g., budding from an amorphous matrix), continuing into the understanding of the nucleus as the residence of the hereditary material, and the explanation of mitosis and then of meiosis, which accounted for how the amount of hereditary material is not augmented in the subsequent divisions.

In the early 1880s, the German cell biologist August Weismann proposed his theory of germplasm, which would become the cornerstone of “hard heredity” (Churchill 2015). Figure 14.1 illustrates Weismann’s distinction between germ cells (gametes) and all the other (somatic) cells, separated by an impermeable barrier. In his view—and in contrast to Darwin’s “pangenes”—germ cells were the only active agents of heredity. Changes in somatic cells (triggered by environmental influences) could not cause changes in the hereditary determinants that constitute “germplasm” or the substance that “has remained in perpetual continuity from the first origin of life.” These hereditary determinants—later named “genes” or, collectively the “genotype” by the Danish scientist Wilhelm Johannsen—and the mechanisms of their inheritance would soon become the focus of a new discipline, genetics.

In some countries such as France, where Lamarckian ideas persisted into the twentieth century, much of eugenic philosophy focused on the improvement of the developmental environment through better pre-natal care, improved housing, and temperance movements. “Positive eugenics” in France received a boost following the Franco-Prussian War in 1871 and then again after the First World War when the loss of young men created an urgent need to replace them with fit individuals.

But elsewhere, and especially in Germany and the United States, there was a decisive shift by the 1880s towards so-called “hard heredity.” Once the idea of such hard heredity was accepted, the doors were open for what came to be known as “negative eugenics.” This term encompassed all medical, public health, and social measures aimed at reducing or stopping the reproduction of those deemed unworthy (“feeble minded,” “racially inferior,” criminals, mentally and otherwise ill, or just destitute), ranging from marriage laws, segregation, and institutionalization to, in some countries, sterilization.

Today we associate such severe interference with individual bodily and reproductive freedoms with extremist totalitarian regimes. Yet it was the United States that built eugenic concerns into its 1924

---

**Figure 14.1** A scheme illustrating Weismann’s theory of heredity. From Wilson (1925, p.13).
Immigration Restriction Act, primarily targeted at restricting immigration from southern and eastern Europe and barring immigration from parts of Asia, “to preserve American homogeneity.” The United States was also the pioneer in the sterilization of the institutionalized “feeble minded” and criminals. The state of Indiana was the first to introduce a eugenic sterilization statute in the United States, in 1907 (Lantzer 2010). In California, today famous for its liberal attitudes, between the enactment of a eugenic sterilization law in 1909 and its final repeal in 1974, more than 20,000 sterilizations were performed on young people deemed to be of low intelligence (Stern 2010). While sterilizations did abate in the mid-1950s, many were performed after the Second World War. In Germany, a law permitting sterilization was first drafted in 1932, under a democratic government and before the Nazi rise to power. How can we account for such extreme transgressions into personal rights and, arguably, the misuse of evolutionary concepts, in supposedly democratic societies? To understand this, we need to know that the idea that individual reproductive rights should take precedence over reproductive responsibilities towards family, society, state, and, ultimately, humankind only developed widespread acceptance after the Second World War. In the early twentieth century, eugenics was seen as a modern, technological, scientifically well-grounded solution to a social problem, akin to artificial selection of animal or plant breeds. If modern science was giving society tools to improve nearly every aspect of human life, the argument went, then why should reproduction not be included in such “progress.” Knowing this helps explain why eugenics, in any of its many guises, enjoyed so much intellectual support, including that of many famous progressive individuals such as H. G. Wells and notable scientists such as Julian Huxley.

Yet by the 1930s the support for eugenics began to wane. One reason was that genetics, using carefully designed experiments on selected animal models, brought an increasingly sophisticated understanding of mechanisms of heredity (Kevles 1985). From the geneticists’ vantage point, a eugenics approach to the study of human heredity, characterized by the reduction of complex conditions to simple traits, and the use of imprecise pedigree charts, was seen as shoddy science. The founder of modern genetics, Thomas Hunt Morgan, criticized the Immigration Restriction Act, arguing that a genetic basis of behavior is difficult (if possible at all) to reveal in any given human group, and comparison between groups with different levels of educational attainment even more so. Studies such as Lionel Penrose’s Colchester Survey of 1938 into the causes of mental retardation in children revealed little evidence for heritability of this condition. Furthermore, as the economic crisis of the Great Depression deepened in the 1930s, no one could argue that all of the poverty was caused by intrinsic and hereditary traits. Scientists, many of them left-leaning in their politics, argued that only in an equal society where everyone enjoyed the same opportunities would we be able to truly distinguish hereditary influences. This is one of the points of the famous paper “Social biology and population improvement,” known as the “Geneticists’ Manifesto,” rejecting eugenics on the eve of the Second World War, co-signed by leading geneticists of the era including Theodosius Dobzhansky, J. B. S. Haldane, Julian Huxley, and Herman Muller (Crew et al. 1939).

So while there is a popular belief that the end of eugenics came from the discovery of atrocities committed by the Nazis and their allies to “improve the race,” followed by the foundation of the United Nations Convention on Human Rights in the immediate aftermath of the Second World War, the truth is a little more complicated. The year 1945 was not a significant threshold (Kevles 1985; Hanson 2013). As shown earlier, by the 1930s the original concept of eugenics had been in decline, but the term persisted in the names of journals and societies well into the 1960s. Sterilization laws, such as the one in California, were repealed even later. And to make matters even more complicated, those who retained and indeed fought for what they saw as the true meaning of eugenics were often anti-racist—such as the left-wing geneticist Julian Huxley—and those who pursued racially motivated science avoided the name of “eugenics.”

This broader history should help us understand why drawing a line from Darwin to negative eugenics and racism (including anti-Semitism) of the early to mid-twentieth century makes little sense. Darwin had considerable humanity, as evident in
his descriptions of his time on the Beagle, although he did assume the superiority of the White race over others—in the same way that he assumed the superiority of his own social rank over the working classes of Britain. But these were opinions of his time, universally shared among his contemporaries. At the same time, he came from an abolitionist family and had strong opinions about slavery; he firmly believed in a single origin of humankind at the time when some of his other colleagues advocated a polygenic view of human origins (Desmond and Moore 2009). He never stood for any form of eugenics, which would go against his moral views based on his Christian heritage as well as his “soft” ideas of heredity. While some of the language of extreme exclusivist ideologies was indeed borrowed from Darwin, their origins are to be found elsewhere: in long-standing beliefs about other races and ethnic groups, in particular Jews, going back much before Darwin; in the economic, social, and political forces operating in Europe between the two world wars; and in traditions and ideas regarding the reproductive responsibility of the population towards the nation and the state (Roberts 2013).

The legacies of eugenics are still evident in various aspects of interventions in reproduction, from birth control, pioneered by fervent eugenicists, to sperm banks, first advocated and indeed launched by left-wing eugenicists such as the Nobel prize-winner Hermann Muller, who proposed separating reproduction from love and sex (Richards 2008a).

14.5 The “Modern Synthesis,” Human Evolution, and Medicine

For all the supposed importance of the theory of natural selection as a way of thinking about human heredity and breeding, it was precisely in the heyday of eugenics and the early days of genetics in the late nineteenth and early twentieth centuries that Darwinism went through its darkest period (Bowler 1983). It was not that people did not believe in evolution—quite the contrary—but many doubted that natural selection (which, as we mentioned before, was not based on a convincing mechanism of heredity) was an important mechanism. Evolution could proceed by many other means. Some supported a Lamarckian idea of change through inheritance of acquired characteristics. Others, leading geneticists among them, proposed that variation occurred through mutation but did not see the need for selection. In short, it was not at all obvious how new ideas about heredity could be married with the “old” idea of natural selection, or whether this integration was necessary at all.

Yet from around the 1920s, interest in natural selection was revived. Geneticists expanded their methods and scope beyond experimental science and model organisms: they now began to use mathematical models to build hypotheses and then test them on natural populations. Evolutionary biology, which had not changed much since the days of Darwin and still mostly involved collecting and comparing specimens and describing lineages, was refreshed by contact with genetics. The “Modern Synthesis” redefined evolution in genetic terms as a “change in allele distribution in a population” (see also Section 2.3.4). New mechanisms in addition to selection were proposed and new work in evolutionary biology began to flourish (Mayr and Provine 1980; Smocovitis 1996).

In this period, human genetics took over from eugenics as the leading discipline of human heredity. Human geneticists looked for traits that could be associated with a gene and that could, on one hand, be informative about human evolutionary history and, on the other, be of use to public health and clinical medicine. However, not many traits were easily linked to an underlying gene. In addition to blood groups, only those rare genes transmitted in clear Mendelian fashion and which produced easily identifiable pathological traits, such as albinism or hemophilia, were identified, and these were of limited value in an evolutionary context.

The concept of balancing selection was first proposed in the late 1940s, based on studies of hemoglobinopathies resulting from mutations inherited in a clear Mendelian fashion that proved adaptive under the pressure of malaria (see Sections 7.4.6 and 13.9). From the 1940s onwards, geneticists teamed up with anthropologists and traveled around the globe to study “natural” human populations that lived in an “untouched” environment (and thus were seen as a proxy for our Paleolithic ancestors), were reproducitively isolated, or were exposed to novel selective pressures. Some went to South American jungles to
study remote tribes; others, such as the “father of medical genetics” Victor McKusick, recorded many heritable conditions while working with the endogamous Pennsylvania religious sect, the Amish (Lindee 2005). In 1966, McKusick published the first edition of what would become the ever-expanding catalogue of heritable traits in humans, *Mendelian Inheritance in Man* (McKusick 1966). The young human geneticist James Neel worked with the Atomic Bomb Casualty Commission to uncover the genetic effects of high doses of radiation received by the populations of Hiroshima and Nagasaki in 1945 (Lindee 1994). The possible long-term impact of radiation upon human evolution captured the public imagination in the “atomic era,” as the many depictions of “mutants” in popular comic books, novels, and films show (Boyer 1985).

Geneticists started to consider ways in which novel selective pressures could have changed the human genome. Neel proposed that the unprecedented access to food in modern developed societies and the growing incidence of diabetes mellitus exposed the ancient selection pressures of feast and famine cycles (Section 9.4.1), while others attempted to explain the persistence of the highly deleterious yet common condition of schizophrenia by speculating about the benefits it might have conferred (Section 11.6.3) (Buklijas and Gluckman 2013). These suggestions later attracted criticism, but they had the value of inspiring further research. The newly described mechanisms of population bottlenecks and founder effects were employed to explain the distribution of certain rare Mendelian disorders (Section 7.4.8).

Geneticists also took the lead in redefining the idea of race, as the four statements on this topic produced by UNESCO between 1950 and 1967 show (Stepan 1982). They argued that while there may be hereditary, and indeed important, differences between human populations, traits visible with the naked eye—such as the skin color and hair texture that had been used to classify races through human history—were irrelevant. More recently, modern genetics has largely rejected the concept of race (Section 6.7).

Under Haeckel’s influence, evolution was once taught at German medical schools, but later curricular reforms of medical education worldwide largely left it out (Buklijas and Gluckman 2013). In the mid- to late twentieth century, the study of evolution had relatively little impact on human medicine and related disciplines. In Chapter 11 we discussed the exchange between evolutionary biology and psychiatry in the 1950s (Box 11.9), but such examples were rare. Around 1970, the new school of sociobiology (Box 11.1) explained animal behavior using tenets of evolutionary biology, and then applied this knowledge to an understanding of human behavior. It soon found itself the target of strong criticism. Much of the criticism should be understood as a reaction to the power of eugenics and its association with evolutionary biology in the preceding decades (Section 14.4) as well as the profound contemporaneous social changes at the end of the 1960s and through the 1970s. These changes could be summarized as a challenge to authority and a rise in the perception of the importance of individual and minority rights. It was in this period that the notion of reproductive responsibility towards the nation, society, or species was replaced by concepts of reproductive autonomy that refused governmental intrusion into such matters. This is the world in which we live today.

Interestingly, the level of manipulation of our reproduction seen today far exceeds anything that our eugenicist predecessors could have imagined. This includes the timing of reproduction through easily available and dependable contraceptives; assisted conception, through IVF and other reproductive technologies; screening of the early embryo through pre-implantation diagnostics, and of the fetus by ultrasound, genetic, and other blood tests; fetal surgery; and legalized abortion. Whether these interventions in reproductive processes constitute a changed selective environment that will affect the direction of human evolution, for instance by effectively delaying the age of reproductive suppression and menopause, remains to be seen (Sections 5.5.3 and 13.5.1).

The 1980s brought an increasing rapprochement between evolutionary biology and medicine. Historians have associated this rise in interest with larger concerns about the future of humans in a changing environment, or the “ecological approach” (Anderson 2004). From around 1940 to the late 1970s it was widely believed that infectious
diseases would soon be relegated to history, but the appearance of AIDS in the West and the increase in antibiotic resistance put an end to this optimism. Indeed much of the early work in evolutionary medicine focused on the evolutionary “arms race” between humans and pathogens (Ewald 1980). But the real boost to the study of the evolutionary origins of human disease came from a young psychiatrist with an interest in evolution, Randolph Nesse, today the most prominent advocate of evolutionary medicine, and an older, well-established evolutionary biologist, George C. Williams, who was most noted for debunking standard models of group selection (Williams and Nesse 1991). Starting from infections but broadening their outlook to include other pathological human conditions and symptoms—from genetic disorders to mechanical injuries and fevers—in a series of theoretical articles and then a 1994 book entitled Why We Get Sick: The New Science of Darwinian Medicine, they asked the question “Why has evolution left our bodies vulnerable to disease?” and launched a new field of enquiry (Nesse and Williams 1994). This book follows in that tradition.

14.6 Evolution, Society, and Religion

The basis of objections to the application of evolutionary theory to understanding the human condition is the fact that it affects the way we think about the relationship between humans and the natural world, and about the question “what is human?” If humans are part of a continuum of evolution, does this mean that they hold no special place in the world and that there is nothing distinct about human nature, human thought, and human culture? Is human morality anything more than the equivalent of the social behavior and instincts of other animals, morphed for adaptive advantage by the processes of evolution? If behaviors are described in evolutionary terms, does this deny the existence and importance of free will?

Some would argue that the highest function of human thought is philosophical, the ability to reflect on our lives, what we are, and our inevitable death. This argument is closely linked to questions of human consciousness and is often extended to the view that whereas other animals may have a degree of consciousness, we do not have any substantive evidence that they can reflect on their condition or possess self-awareness. We do not know if they sense that they will perish, even if they see other members of their species doing so. Philosophers have argued for centuries over whether humans can actually be objective about their own existence. This raises the question of whether our reflective thinking is adaptive and has evolved through selection in a similar way to other human characteristics. Our possession of a large brain and use of tools and language are all evolved aspects of our species (Chapter 6). Are human achievements such as art, poetry, and music evolved characteristics too (Box 11.8), or are they something qualitatively different that we have been able to develop beyond the constraints of natural selection?

One way of dealing with the difficult questions about human thought and human nature is to deny evolution altogether. As we discussed in Section 14.2, before the emergence of concepts of evolution in the late 1700s, the problems of human origins and our place in nature were firmly in the realm of religion. The rise of biology as a science, including early concepts of evolution dating to before Darwin, provoked responses by theologians, of which the most famous is the one described earlier by William Paley (Section 14.2). But it was the publication of Darwin’s The Origin in 1859 and the wide-reaching response it received that created a challenge that Western established religions could not afford to ignore. Responses were diverse and dynamic; negative reactions of some religious institutions in the nineteenth century by no means foreshadowed continuing animosity in the twentieth. At the same time, some societies that were generally receptive towards evolutionary biology in Darwin’s times developed strongly critical positions later on. In this section we focus on the positions of two most numerous religious communities in the West, Protestantism and the Roman Catholic Church. For a variety of historical reasons, understanding the reception of evolution in Islam is limited (Riexinger 2013). Responses of Judaism have been extremely diverse, yet rarely outright hostile, and arguments based on the literal reading of Torah (akin to creationism, see below) are rare (Swetlitz 2013).
Possibly the easiest to trace, because of its hierarchical structure and largely unified position, is the history of the standpoint of the Roman Catholic Church on Darwinism (Haught 2013). In the early days, official Catholicism expressed considerable hostility towards Darwinism. This attitude arose mostly from the suspicion that Darwinism is inseparable from materialism and related philosophical positions, such as socialism, which were hostile to religion and were gaining popularity in the last third of the nineteenth century. Yet over the decades the Catholic Church considerably warmed towards evolution, and in 1950 Pope Pius XII’s encyclical *Humani Generis* officially acknowledged evolution (though insisting that the human soul is directly created by God). From then onwards, Catholicism has increasingly embraced evolution, and in an official statement in 1996, Pope John Paul II maintained that the evidence for evolution is strong and that it is “more than a hypothesis.” Protestant churches have generally had positive or neutral attitudes towards evolution: numerous Protestant theologians had respect for Darwin and his evolutionary ideas, others contested them, and yet others completely rejected any relevance of science to religion (Finnegan 2013).

The central battlefield of evolution versus religion, the United States, went through several distinct periods (Numbers 1985, 2006). In the nineteenth century, the opposition was largely focused on refuting Lyell’s uniformitarianism. By proposing a slow and continuing change, the uniformitarian principle argued against the Biblical flood and the story of re-creation of the world through those species saved on Noah’s Ark. At the time, the opposition to the evolutionary account was known as “flood geology.” By the 1920s the anti-evolution movement had centered on Darwin, and in particular on the teaching of evolution in state schools. The movement soon petered out, but it returned in later decades of the twentieth century in various forms such as “young Earth creationism,” “scientific creationism,” and “intelligent design” (Scott 2005). The reasons for this expansion of creationism are tied up with the rise of American Christian religious fundamentalism, based in a tradition that places the close study of the Bible and the literal interpretation of Genesis at its heart (Emerson and Hartman 2006).

Also, opposition to evolution is generally associated with social conservatism and a broader skepticism towards science (Woodberry and Smith 1998). While creationism has been by and large an American phenomenon, there are now pockets of this movement worldwide.

Are religion and evolutionary biology truly irreconcilable? Many religious scholars and many scientists have taken this position. We discussed the standpoint of religions above; from a scientist’s viewpoint, one argument is that because both are competing to find fundamental truths about universe, they cannot be simultaneously true (Coyne 2015). An alternative position is to seek a compromise between two apparently very different perspectives, most often by suggesting that a deity created the natural laws through which evolution acts, or by suggesting that religion and belief in the supernatural on one hand, and science and its demonstration of evolutionary biology on the other, are different domains that do not have the capacity to inform each other. These are both positions that distinguished evolutionary and molecular scientists such as Francis Collins (one of the founders of the Human Genome Project) and the late Stephen Jay Gould (the influential evolutionary biologist and science writer) have taken. From the side of religion, the Catholic priest, philosopher, and respected paleontologist Pierre Theilhard de Chardin (1881–1973), called both a visionary and a charlatan, attempted to interpret the Fall and original sin as references to the imperfect, continually changing and evolving world, with imperfection always having a dark side (Haught 2013).

The third position is to consider that supernatural belief and religion themselves are the consequences or side effects of the evolution of human consciousness and social complexity, and thus have no fundamental validity as explanations of human origins (Wilson 2002). This view has been a major plank in the rise of rationalism as a philosophical approach.

It is not necessary for understanding and applying evolutionary biology to adopt a personal position regarding these competing views, except for rejecting the approach which denies anything but a purely deistic control of all components of a species’ design and creation: this attitude would deny the importance of science to medicine.
14.7 Evolutionary Thought and the Human Condition

Evolutionary biology can explain who we are, where we come from, and how we are related to other organisms. It can explain, without the need for a deity, how we have come to have our less than perfect design and biology, which can affect our risks of ill-health. While there are many gaps in our knowledge, it can lead to a better understanding of much of our behavior and of our interaction with our physical, biotic, and social environments.

But evolutionary biology is a largely value-free epistemology and thus has critical heuristic applications even if one has a commitment to other value-rich epistemologies, including religious belief. The questions that drive a desire by many for an explanation of human life which goes beyond a purely scientific, mechanistic approach are complex and have multiple layers. They pose a challenge to the unqualified acceptance of evolutionary thought in describing human nature by some who are otherwise willing to accept the role of evolution in speciation, including the evolution of Homo sapiens. Such individuals feel that, when we come to consider the intellectual, moral, and creative aspects of our lives, evolutionary processes play little part. But it is fundamental that our abilities to generate culture, altruistic behavior, and group rules are indeed important attributes of our evolved nature.

Applying evolutionary thought to improving the human condition is ultimately what an understanding of evolutionary medicine provides, whether in addressing a particular clinical problem or a wider public health concern. This is true, irrespective of the personal beliefs of either practitioner or patient. To paraphrase Dobzhansky, “evolution is the only integrating principle of medical science”: hence the importance of understanding evolutionary biology to all those concerned with medicine and public health.

**Key Points**

- Evolutionary biology inspires strong responses because it contests myths of human creation and origins, and takes a biological view of human nature.
- Darwin’s theory of evolution by natural selection built on the early nineteenth-century sciences studying organic change, contemporary ideas about social progress, and Darwin’s own observations to provide a persuasive account that has stood the test of time.
- Eugenics, the early twentieth-century movement for social reform through the control of human reproduction, had much more to do with the acceptance of “hard heredity,” which excludes any inherited effect of the environment, than with Darwinian evolutionary theory.
- The “Modern Synthesis” of evolution and genetics has underpinned much of biological science since its beginnings in the 1930s, but has had limited impact on clinical medicine.
- The origins of evolutionary or Darwinian medicine lie in the ecological approaches of the 1980s, but have since expanded from an initial preoccupation with infectious disease to encompass the entirety of medicine.
- Religion and evolutionary medicine need not be in opposition.
References


Baharoglu Z and Mazel D 2014. SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiology Reviews* 38: 1126–1145.


suggest treatment strategies based on specific meta-


REFERENCES


Przyborski JM, Diehl M, and Blatch GL. 2015. Plasmoidal Hsp70s are functionally adapted to the malaria parasite life cycle. *Frontiers in Molecular Biosciences* 2: 34.


Scharff C and Petri J 2011. Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366: 2124–2140.


REFERENCES


Wellcome Trust Case Control Consortium 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.


Index

A
acetylsalicylic acid (aspirin) 299, 313
acquired characteristics (somatically) 96
inheritance of (Lamarck’s) 329, 332, 335
acquired immunity see adaptive immunity
acting-out behavior 281–2
adaptation (to new environment and selection pressure) 8, 39–48, 46
debated areas and limitations of adaptationist argument 39–48
definition/meaning 20, 21
in development 86–90
predictive 88–90, 165, 230
environment of evolutionary adaptedness 279
eye evolution and adaptive landscape 28
local see local adaptation
maladaptive consequences 227–9
menopause and 201
metabolic 205–36
microevolutionary (=contemporary evolution) 19, 46, 153–7
nutritional see nutrition
polygenic 64
polymorphisms associated with 306
radiation see radiation, adaptive rapid 46
adaptive (acquired) immunity 246–9, 252
pathogen evasion 249
vaccination and 253
adaptive therapy (with chemotherapeutic drugs) 301–2
adipose tissue (body fat stores) 207, 208, 210
adolescence 115–16
acting-out behavior 281–2
duration and events 114
adoption 268
puberty and 121
adrenarche 114, 117
adulthood/later life:
developmental environment and disease in 223–35
duration and events 114
affective (mood) disorders 281, 282–3, 314–15
affluence and life expectancy 128
Africa:
migration out of 60, 132, 134, 135, 148–51, 173
origins of humans in 132, 144, 157
tool-making 144
Alzheimer’s disease 113, 151
Alzheimer’s disease 113, 151
Alzheimer’s disease 113, 151
Alzheimer’s disease 113, 151
Allen’s rule 113, 151
amylase, salivary 153
analysis of variance 38
Andeans and altitude 324
anemia:
hemolytic see hemolytic anemia of iron deficiency 304
sickle cell 35, 66–7, 156, 172–3, 320
aneuploidy 26
Angelman syndrome 91
angiogenesis in cancer 292
animal to human transmission (zoonoses) 241, 315
anisogamy 179
ankylosing spondylitis 251
antagonistic pleiotropy 107, 166, 169, 297, 298, 311, 320
anthropogenic changes in human environment 9
antibiotics 255–8
gastrointestinal infection and 240
resistance 171, 255–7, 305
antibodies 246
breast milk 249
antifungal drugs 257, 257–8
antigen(s) 246–54
human leukocyte see human leukocyte antigens
influenza, antigenic drift and shift 253–4
self 248
see also vaccination
antimicrobial agents in breast milk 200
antimicrobial drugs 255–8
see also antibiotics; antifungal drugs
antipyretics 170, 251, 313, 314
antiretrovirals 257
antisocial behavior 263, 281
antiviral drugs 257
anxiety 271–4, 283
apolipoprotein (APO):
apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1 (APOBEC1) 148
apolipoprotein E gene (APOE) 70–1
apolipoprotein L1 gene (APOL1) 316–17
apoptosis and cancer 292
appendicitis 171
archaea, evolution 22, 23
Ardipithecus 132, 135
bipedalism 135, 136
arginine vasopressin receptor 1A gene (AVPR1A) 275
Aristotle 81
arms race, evolutionary (between predator and prey) 16, 237
pathogens 171, 181, 241, 258, 305, 337
art 277
hominin 143–4
cave paintings 145, 276, 277
artificial selection 10, 11, 29, 330
asexual reproduction 23, 24, 178, 179, 180, 181
Asia, migration from Africa to 150
aspirin 299, 313
assisted reproduction 304
association studies 69
genome-wide see genome-wide association studies
assortative mating 60, 154, 174
asthma 252
attachment theory 280
attention-seeking 282
Australia:
migration from Africa to 150
rabbits 151
Australopithecines 132, 133, 137, 138, 139, 140, 143
bipedalism 135, 136
autism 264, 280, 284
autoimmune disease 248, 251–2
avian influenza virus H5N1 241, 242
AVPRA 275
B
B cells 246, 248, 250
clonal expansion 248
babies see infants; neonates
back pain 182
bacteria:
evolution 22, 23
infections, drugs used in see antibiotics
resource acquisition, therapeutic targeting 258
see also pathogens
Bacteroides 239, 249
Bacteroidetes 208
balancing selection 30, 31, 38, 66, 67–8, 172–3, 284, 317, 319–21, 335
Baldwin effect 92–3
band 3 erythrocyte membrane protein 321
basic reproductive ratio, parasites 242–3
Beagle (HMS) 20, 330, 331, 335
Beckwith–Wiedemann syndrome 91, 192
behavior 17, 261–86
biological determinants of 261–3
genes 33–4, 263, 270–1, 275–6
imprinting 280
see also specific types
Bergmann’s rule 112–13, 151
bioinformatics 64, 65
biology (and biological determinants):
behavior 261–3
cancer 290–3
culture see culture
see also molecular biology
bipedalism (upright posture) 43, 135–6
metabolic disease and 232
birds, sex chromosomes 181
see also avian influenza
virus H5N1
birth:
trifitial vs precocial strategy 15, 142–3, 194
preterm 312–13
see also childbirth
birthweight 224
later life disease and 224, 227, 229, 230
neonatal mortality and 88
blood vessel supply to cancer tissue 292
body:
composition in puberty 116
mass, and metabolic rate 98–100
plans and patterning 29, 45, 93, 94, 292
shape see shape
size see size
vulnerability to disease 165–6
Bogin, Barry 125
bone in pregnancy and vitamin D and calcium 307
borderline personality disorder 282
bottle feeding (formula milk) 233, 249
breastfeeding vs 227, 233, 239, 319
bovine spongiform encephalopathy (BSE) 309
Bowlby, John 280
boys, height-velocity curves 115
brain 139–43, 263
energy need 113–14, 142, 263
epigenetics and 148
male 173
metabolic rate 113–14, 220
microbes and 245
schizophrenia and imprinting of 284
size/mass (encephalization) 139–40, 263
DUF1220 and 148
human ancestors 132–3, 139–43, 263
over life history 110, 111–12, 126
social groupings and 141, 145
breast 172
cancer 172, 288, 295, 297–8
BRCA1/2 and 65, 107, 166, 292–3, 298
breast milk (human maternal milk) and breastfeeding 5, 6, 114, 199, 199–200, 233, 239, 307, 319
bottled (formula) milk vs 227, 233, 239, 319
breast cancer and 298
calcium 307
immunity and 200, 249
see also lactation
BSE (bovine spongiform encephalopathy) 309
Buffon, Comte de 328
cytochrome P450 2D6 56
cytokines in immunity 245
cytotoxic drugs in cancer see chemotherapy

cytokine storm 250
darwin, charles 9–11, 24–5, 29, 330–1, 338

D
Darwin, Charles 9–11, 24–5, 29, 330–1, 338
Darwin, Erasmus 329
Dawkins, Richard 33, 34, 35, 262, 270, 330
death see mortality
defenses (host) to pathogens and foreign material 170, 244–50
over-reactive/
uncontrolled 250–1, 313–15
sex and 180–1
see also immunity
defenses against see defense; immunity
degeneration (concept of) 332
demes 32, 174
demographic transitions and contemporary evolution 155
denises 134, 135, 149, 152, 324
dentition:
hominin 137–8
malocclusion 168
depression 283, 314–15
design 8, 146
development 79–96, 223–35
adaptation in see adaptation buffers 85
critical windows 13, 86
emotional 280–1
environment in see environment evolutionary novelty in 233–5
historical ideas 81–2
intrauterine see embryology; fetus mismatches 223–35
plasticity 12, 74, 84–6, 96, 163, 229–33
neoplasms and 297
stunting and 14
theoretical concepts 80–1
trade-offs in life histories and 13, 87, 97–8, 99, 100–2, 168–70, 295–9, 311, 312
variation in 110–13
diabetes mellitus 207–10
gestational 168, 227, 233
paternal 233
type 2 21, 68, 70, 168, 207–10, 219, 223, 233
vulnerable populations 223
worldwide prevalence 207
diapause 88, 194
diet see nutrition and diet
directionality in evolution 39–40
disease and illness 161–75, 303–26
clinical signs and symptoms 303–4
complex, searching for genetic causes 69
cultural change and emergence and disappearance of 308–9
definitions 5–8
genes and see gene; genetic diseases
hominin 147
infectious see infections
in later life/adulthood, developmental environment and 223–35
multifactorial 49, 69
proximate causes 17, 18, 163
risk see risk of disease
see also specific (types of) disease
divergence 41, 43
diversity, human 131–58
social and medical implications 155–7
DNA:
methylation see methylation
mitochondrial see mitochondrial DNA
recombination 26, 56, 57–8, 178–9
repair, cancer and 292–3, 293, 294
dominant alleles 26, 65–6
disorders 66
Dobzhansky, Theodosius 161
Down syndrome
(trisomy 21) 86, 137
Drosophila crossedineless 85
drug-metabolizing enzyme
polymorphisms 156–7
see also specific (types of) drugs
DUF1220 148
Duffy antigens 316, 320
dobzhansky, theodosius 161
down syndrome
(trisomy 21) 86, 137
drosophila crossedineless 85
drug-metabolizing enzyme
polymorphisms 156–7
see also specific (types of) drugs
DUF1220 148
Duffy antigens 316, 320
dobzhansky, theodosius 161
down syndrome
(trisomy 21) 86, 137
drosophila crossedineless 85
drug-metabolizing enzyme
polymorphisms 156–7
see also specific (types of) drugs
DUF1220 148
Duffy antigens 316, 320

economic dimensions see affluence;
socioeconomic dimensions
ectodyplasmin 322
EDAR (ectodyplasmin gene) 322
egg/oocyte/ovum 192, 193
fertilization 192
menopause and 127
ovulation 200, 201
sex determination and 181
embryology (and embryo development/embryogenesis):
body patterning 93
diapause 88, 194
historical perspectives 82
implantation of embryo 192–3
emotions 271, 285–6, 304
development 280–1
encephalization see brain
ENCODE project 52, 73
endocrine control of puberty 116–17
energy:
balance and imbalance 68, 208–9
cancer cells and energy metabolism 292
expenditure 99–100, 119, 205, 208
in life history 100–2
need/intake/
consumption 205, 208
brain 113–14, 142, 263
refined foods dense in 218
storage strategies 206–10
environment(s):
anthropogenic changes in human environment 9
cancer and 294–5
prevention of 299
definition 4
developmental 86–90
fetal/intrauterine see intrauterine environment
later life disease and 223–35
disease risk and 89, 167–8, 218–36
of evolutionary adaptedness 279
life history and 98
matched (with organism) 5
mismatched or novel 5, 166, 167–8, 305–11
adaptation to see adaptation
cancer and 294–5
developmental plasticity across life course and 12–13, 229–32
human health and 6, 294–5, 305–11
niche construction 4, 34, 45, 76, 310
separating genetic effects from effects of 39
sleep and 310–11
variation and changes in 8
epigenesis 81, 83
epigenetics 71–3, 75, 79
  brain and 148
cancer 91
development and 79, 82, 83, 90–1
epigenetic marks/markings
  (epimutations) 20, 72, 74, 75
twins 90, 92
lifelong epigenetic changes 229
variation and 71–3
epiphenomenal view of
menopause 201
epistasis 15, 30, 68
epizootic agent 241
equations of evolution 38–9
erthrocytes (red blood cells) and
malaria 172, 173, 320
estrogens 188, 192
  cancer and 299
  menopause and 311, 312
  puberty 118
see also
erone replacement
therapy
Ethiopian Jews, migration to
Israel 216
ethnicity and race 156, 157
eugenics 332–5
eukaryote:
  definition 20
evolution 22, 23
see also multicellularity
Europe, migration from
Africa to 150
evolution, definition 3–4
exaptations 15, 42, 136
eye 27, 28, 171
color 321
fastness (speed) see rapidity
fat:
  body
    babies 220–1
    stores 206, 207, 210
dietary sources 208, 210, 214–15,
  216, 217, 220
father see paternal effects;
patrilinearity
fatigue in depression 315
see also malaise
fatty acids, omega-3 and
  omega-6 217
fear 17, 271–4, 283
see also fight or flight response;
phobias
fecundity 31
trade-offs 102, 177
see also fertility
females:
  ancestry/lineage, and
mitochondrial DNA 56, 149
fertility
  decline in 127, 177, 200–2
  puberty and 126–7
fitness, demographic
  transitions 155
life expectancy 105, 127
menarche 116, 118–23, 126, 182
pregnancy see pregnancy
puberty 116, 118, 119, 120
  and fertility 126–7
reproductive cost 182–3
trade-off between current vs future
reproduction 100
see also breast; breast milk; girls;
  sexual dimorphism
fertility 31, 177
decline (reproductive decline) 127,
  177, 200–2, 304
puberty and 126–7
tumor suppression and 297
see also assisted reproduction;
  sterilization law
fertilization 192
fetus 196–8
  adaptive responses (in
development) 87–8
  environment see intrauterine
  environment
growth see growth
  head size 110, 198
hemolytic anemia 195
maternal interactions with 194–6
maternal–fetal conflict 190, 191,
  192, 195, 195–6
nutrient reduction 87
see also birth
fever 170, 251, 313–14
fight or flight response 17, 273
finches, Darwin’s 20
Finnish populations 63
Firmicutes 208, 233, 239
Fisher, Ronald 32, 37, 38, 184
fitness (biological) 3–4
  adaptation and 21
  advantage 60–1
  heterozygote 66–8, 172–3,
  319–20
menopause and 202
sex and 179, 185
direct 24
human demographic transitions in,
females 155
inclusive 24, 33, 202, 267, 311
indirect 24
landscape 28
relative 23, 39
traits reducing 46
folate and folic acid 13, 308, 322
deficiency 94, 139, 308
food:
  chewing/mastication 138, 317–19
  cooking 126, 138, 317, 319
see also foraging; hunting;
nutrition; nutrition and diet
footwear 317
foraging 125, 136, 146, 172, 210, 215,
  218, 222, 261
hunting and (hunting and
gathering) 100, 106, 125, 210,
  211, 212, 219–20, 222, 272
modern-day 211–12, 213, 217–18
formula feeds see bottle feeding
founder effects 41, 61–2, 63, 149,
  150, 173–4
freeloading 32, 33, 167, 254, 268, 276,
  277, 279, 281, 285
Freud, Sigmund 280
FTO 68–70, 209
fungal infections, antifungal
drugs 257, 257–8
G
Galton, Frances 332
game theory 33, 268–70, 281
gametes 26, 34, 57–8, 179, 192–3
see also egg; sperm
gastrointestinal tract (gut):
hominins 138
infection, treatment 240
microbiota and microbiome 43,
  171, 208–9, 239
energy balance 208–9
health and 208–9, 233
gender (social construct) 181, 205
see also sex
genes(s):
  behavior and 33–4, 263,
  270–1, 275–6
coevolution of culture and 16, 35,
  153, 154, 265
definitions 50, 52, 79
diseases and 49–50
  common 68–71
dosage 56, 65
  sex determination and 181

gene(s): (Contd.)
flow 42, 44, 45
horizontal transfer 23, 35, 45
obesity see obesity
promoter methylation 91
selfish 33–4, 270–1
sex and trading of 179
speech and grammar and 144
thifty (and thifty genotype) 70, 219–22, 227, 315
as units of selection 33–4
see also Mendelian inheritance; mutations
genetic accommodation 85
genetic assimilation 81, 84, 85, 92, 279
genetic causes of complex diseases, searching for 69
genetic diseases:
 Alu-mediated mutations and 55
rare 174–5
Finnish populations 63
single gene see Mendelian inheritance, disorders
genetic drift (random drift) 11, 27, 41, 59, 60, 61, 174
genetic effects and environmental effects, separating 39
genetic inheritance see Mendelian inheritance
genetic isolation see reproductive isolation
genetic landscape 83
genetic variation see variation
genome(s):
 conflict within 34, 191
human-specific sequences 147–8
imprinting 91, 92, 192, 284
instability in cancer cells 292–3
interindividual different 69
genome-wide association studies (GWAS) 69
obesity 223
genotype 57
definition 79
frequency across populations 37
phenotype links with 15
thifty (and thifty genes) 70, 219–22, 227, 315
geography, local adaptations to 152
geological and human timescales 4
Germany and eugenics 333, 334
gestational diabetes mellitus 168, 227, 233
girls, height-velocity curves 115
globin chains (and genes) 172–3, 320
glucose 209–10
see also hyperglycemia
happiness 283
hard heredity 333
Hardy–Weinberg 37–8
Harlow, Harry 280
hawk–dove game 269–70
head size (over life history) 109, 110
fetal 110, 198
health (human) 161–75, 303–26
hominin 146–7
life history and 168–70, 311–13
migration and see migration
novel environments and 6, 294–5, 305–11
public 18, 252–3, 303–26, 325
social stratification 285
health behavior and socioeconomic position 129
heat-shock proteins (Hsps) 85, 193, 245
height see stature
helminth worms 240, 242
hemoglobin breakdown in malaria 321
see also hemolytic anemia
hemoglobinopathies 153, 156, 172–3, 335–6
malaria and see malaria
hemolytic anemia 321
fetal 195
hemo philia 174
HER-2 ( neu ) 290, 292
herd immunity 254
heredity 332–5
see also inheritance
heterogeneity, intratumoral 293
heterozygote advantage 66–8, 172–3, 319–20
high altitude 324
histocompatibility complexes 180
major (MHC) 67, 185, 247–8, 300
see also human leukocyte antigens
histones, post-translational modifications 72
historical perspectives:
eugenics 332–5
time of evolution 327–33, 337–8
history-taking (medical) 303
HIV (human immunodeficiency virus) 137, 139, 244, 249, 254, 257
HLA see human leukocyte antigens
holistic views and approaches 18, 161, 303, 325
homeobox genes (Hox) 82, 90, 94, 95, 181
homeostasis:
  environmental variation and feedback systems 13
hominids, definition 131
hominins 10, 132–8
  art see art
definition 131
diet 210–12
evolution 132–8
hominoids:
clad 131–2
definition 131
Homo erectus 105, 131, 133–4, 135, 137, 140, 143–4, 149, 150, 210
Homo ergaster 133
Homo floresiensis 137
Homo habilis 133, 140, 143, 210, 265
Homo heidelbergensis 134, 135, 149
Homo naledi 133
horizontal transfer of genes 23, 35, 45
hormonal (endocrine) control of puberty 116–17
hormone replacement therapy 312
cancer and 295
hospital-associated infections and antibiotic resistance 255
host defenses see defenses
Hox (homeobox genes) 82, 90, 94, 95, 181
Hsps (heat-shock proteins) 85, 193, 245
human(s):
  condition of being 339
diversity 131–58
evolution/origins 131–58
coevolution see coevolution
see also hominins
individual see individuals
migration see migration
timescales in geology and in 4
human immunodeficiency virus (HIV) 17, 153, 171, 241, 243, 249, 254, 257
human leukocyte antigens (HLA) 247
HLA-B27 251
HLA-DR2 251
human-accelerated regions (genome) 148
humoral immunity 246–7
hunting 145, 211, 212, 213
  and gathering/foraging 100, 106, 125, 210, 211, 212, 219–20, 222, 227
Huxley, Julian 284, 334
Huxley, Thomas Henry 331
hygiene hypothesis 252
hyperglycemia:
famine and risk in later life of 242
maternal 223, 233
hypermorhosis 124
hypertension, sexual differences 188
hyperuricemia 316
hypocalactasia, congenital 5
hypothalamic–pituitary–adrenal (HPA) axis 89, 319
stress and 273
hypothalamus:
  energy and 208
  puberty and 117, 118
hypothesis formulation and testing 174–5, 305, 325
IGF see insulin-like growth factor
illness see disease and illness
immediately adaptive responses in development 87–8
immunity 245–52
  evasion by cancer cells 292
  by pathogens 249–50
  morbidity and mortality due to over-reaction 250–1
  neonatal and infant 200, 249
  tolerance 247
immunization (active) 253–5
  innate immunity 245–6
  insulin 209
  resistance 17, 21, 68, 80, 87, 170, 207, 233
  fetal environment and 227–9, 313
  paternal 233
  syndrome of (=metabolic syndrome) 207, 209
insulin-like growth factor (IGF):
  IGF1 107, 108, 311
  IGF2, and its receptor (IGFR2) 192
  insulin receptor substrate (IRS) 2 108
  “intelligent design” 330, 338
  intentional 141
intrauterine (fetal; prenatal)
  environment:
  growth in see growth
  mismatches 224, 226, 227, 228, 231–2, 233
  stress (and stressful factors) 87, 120, 169, 193, 227–9
  infection 312
  insulin sensitivity/resistance and 227–9, 313
  placental insufficiency and dysfunction 169, 232
  twins and 39
  see also fetus
introduction 45–6
invasion:
cancer cells 292, 297
placental 297
iodine deficiency, Sherpas 8
iron:
  deficiency and supplementation 304
  host–microbe interactions and 246
Israel, migration of Ethiopian Jews to 216

J
  jaw 137–8, 317–19
  hominin 137–8
  mastication/chewing 138, 317–19
  myosin gene and muscles of 148
Jews from Ethiopia migrating to Israel 216
  see also Judaism
Judaism 337
  see also Jews
  “just-so stories” 43, 175, 279, 321
  juveniles 114–15
  duration and events 114

K
  Khoisan people (!Kung san) 149, 177, 212
  kin 267–8
  selection 32–3, 200, 267, 268
  Klebsiella commensals 251–2
  Kleiber’s law 98–100
  Kropotkin, Pyotr 332
  !Kung san (Khoisan people) 149, 177, 212
  kuru 308
  kwashirkor 65

L
  labor, obstructed (dystocia) 172, 196–7, 233
  lactase and its persistence 5–6, 7, 15, 16, 35, 53, 61, 153, 265
  lactation 199–200
  breast cancer and 298
  calcium and 307
  lactose 5
  intolerance 6, 35
  Lamark, Jean-Baptiste (and Larmarkian ideas of inheritance) 96, 328–9, 330, 332, 333, 335
  language 144–6, 279
  learning 278
  Leclerc, Georges-Louis (Comte de Buffon) 328
  leprosy 322–4
  leptin 208–9, 230
  life, origins of 22–3
  life course 12–13
  developmental plasticity and mismatched signals across 12–13, 229–32
  epigenetic changes over 229
  life expectancy 3, 104–5, 106
  affluence and 128
  cancer and 288
  female 105, 127
  sexual differences 188
  see also longevity
  life histories (theory) 13, 97–129, 168–70, 311–13
  contemporary 128–9
  general overview 98–107
  living fast 113, 128–9, 295
  risk (to health/of disease) 168–70, 311–13
  trade-offs in 13, 87, 97–8, 99, 100–2, 168–70, 295–9, 311, 312
  lifespan 104–5
  caloric restriction and 108
  cancer and 289, 294–5
  fecundity vs 102
  increases in modern times 169
  lifestyle: changes and transitions in 218
  variation caused by in 153
  stress 285
  linkage 56–9
  disequilibrium 26, 38, 58–9, 64, 316
  mapping 69
  Linnaeus, Carl 12, 25, 328, 331
  “living fast” life history strategy 113, 128–9, 295
  local adaptation (to selection pressures) 148–53
  nutritional 152, 211, 233
  longevity 105
  cancer and 295
  fecundity vs 102
  see also life expectancy
  louse, human 143
  love 274–6
  maternal 280–1
  lymphocytes:
    B cells see B cells
    T cells 246, 247, 248
  Lysenko, Trofim 83
  macromutation (speciation; evolution of new species) 3, 9–10, 43–6
  development and 93
  phenotypic novelty and 93
  tumors and 293
  major histocompatibility complexes (MHC) 67, 185, 247–8, 300
  immunity and 247–8
  see also human leukocyte antigens
  malaise in infection 303–4
  see also fatigue
  malaria 152, 241, 246, 254, 316, 320–1
  glucose-6-phosphate dehydrogenase and deficiency 173, 321
  hemoglobinopathies and 153, 320–1
  sickle cell anemia/disease 35, 66–7, 156, 172–3, 320
  thalassemia 153, 172, 320
  males:
    brain 173
    competition with other males 31, 127, 272
    fertility and puberty 126–7
    prostate cancer 107, 288, 289, 299
    reproductive cost 182–3
    risky behaviors 173, 185
    see also boys; sexual dimorphism; Y chromosome
  malignancy see cancer
  malnutrition (poor nutrition):
    immunity and 252–3
    infant 165
    modern times 215–18, 228
  malocclusion of teeth 168
  mammals:
    milk see milk
    placenta see placenta
    reproductive cost 182–4
  marsupials 100, 193, 194
  mass see brain; weight and mass
  mastication (chewing) 138, 317–19
  malocclusion of teeth 168
  maternal . . . see mother
  mathematical equations of evolution 38–9
  mating:
    assortative 60, 154, 174
    choice of mate 183–6
    systems/structures 183, 186, 190
    monogamous 185, 275
    polygynous 184, 185, 188, 272, 274, 321
  matrilinearity see mother
  MC1R 138–9, 321
  MC4R 68, 209
  McKusick, Victor 336
  meat-eating 212, 215–17, 294
  vegetarian diet vs 212–13
  see also cannibalism
  Medawar, Sir Peter 106
  medical history 303
meiosis 26, 56
meiotic drive 34
melanocortin-1 receptor gene (MC1R) 138–9, 321
melanocortin-4 receptor gene (MC4R) 68, 209, 306
memes 35
menarche 116, 118–23, 126, 182
Mendel, Gregor 11, 37, 65
Mendelian (monogenic; single gene) inheritance 10
disorders/diseases 65–8
selection and 66–8
meningitis 245
menopause 127, 200–2, 311–12
menstruation 190–2
mental (psychological/psychiatric) disorders 279–84
early life and risk later in life of 273
single nucleotide polymorphisms and 275, 281
messenger RNA (mRNA) 50, 73
metabolic adaptations 205–36
diseases/disorders/dysfunction 170, 235
evolutionary novelty and 222–3
increasing rates and epidemic of 205, 218, 222, 224
migration and 216
predicting risk of 230–1
upright posture and 232
metabolic programming (in development) 227, 229
metabolic rate:
body mass and 98–100
brain 113–14, 220
metabolic syndrome 207, 209
metastasis 290, 292, 293, 297, 307
metazoan:
complexity 94
multicellularity see multicellularity
methylation:
DNA 71–2, 75, 90, 91
twins 92
mRNA 73
microbes (and microbiome and microbiota):
coevolution of humans with 16, 171, 238–40, 315–17
commensal see commensal microbes
gut see gastrointestinal tract
pathogenic microbes see pathogens; toxins
vaginal, delivery and 233–5
microevolution 3, 9–10, 43, 46
cancer and 293
Micronesian, gout 315–16
microRNA 73, 75
migration 148–52
from Africa 60, 132, 134, 135, 148–51, 173
health and 216
cancer 288–9
leprosy 322–4
mapping 149
stature/height and 123
variation caused by 151–2
milk (mammalian) 5–6, 8, 41, 199–200, 233
cow’s see cow’s milk
mature see breast milk
mind, theory of 140–1, 264
missing heritability 69
mitochondrial DNA/ genome 50–1, 56
female ancestry/lineage and 56, 149
hominins and 149–50
human migration and 149
mapping 135
Mitochondrial Eve 149
Modern Synthesis 17, 38, 82, 335–7
modern times see contemporary times
molecular biology 43, 49–77
hominoid/hominin clade 131–2, 157
inheritance 11–12, 49–77
variation 50–9, 71–3
molecular clock and divergence 41
monogamous mating structures 185, 275
monogenic inheritance see Mendelian inheritance
mood (affective) disorders 281, 282–3, 314–15
morality 276
morbidity:
defense over-reaction causing 250–1
life history and 168–70, 311–13
sexual differences 188
Morel, Benedict August 332
morning sickness 197–8
morphological evolution 93
morphology see shape
mortality (death):
causes (US) 4
defense over-reaction causing 250–1, 313–14
extrinsic and intrinsic 102–4
infant and/or neonatal 224
birthweight and 88
menopause and 311
sexual differences 188
mother (maternal...):
breast milk see breast milk
constraints 14, 74, 100, 196–8, 231, 232
fetal interactions with see fetus
inheritance (maternal line and maternal effects) 73, 74
mitochondrial DNA and 56, 149
obesity 233, 235, 236
schizophrenia (offspring) and 284
see also breast milk; intrauterine environment; pregnancy
Muller’s ratchet 180
muticellularity 94
cancer and 290
multifactorial diseases 49, 69
multilevel selection 33
mutations 25–6, 51
Alu-mediated 55
cancer-associated 290–1
cis-regulatory 93
estimating rates of 53–4
harmful, apparently, maintenance 172–3, 319–21
neutral 11, 41, 305
non-synonymous 53, 64
random 93
recurrent 66
somatic see somatic mutations
synonymous 52–3, 64
see also transmutationism
mutualism 238
Mycobacterium leprae 322–4
Mycobacterium tuberculosis (and tuberculosis):
cystic fibrosis and 67
immune evasion 249
MYH16 and jaw muscles 148
myopia 76, 168
myosin gene (MYH16), jaw muscles 148
N
narcissism 282
natural immunity (innate immunity) 245–6
natural selection 11, 25, 29–30, 40, 305–6, 331, 332
1920s revival of interest in 335
forms 31
in modern times (humans) 155
Nauruans, metabolic dysfunction 223
nausea see vomiting and nausea
Index

Nazi Germany, eugenics 334
Neanderthals 64, 134–5, 137, 140, 143, 144, 147, 149, 150, 152, 153, 157
nearsightedness (myopia) 76, 168
Neel, James 219–20, 336
negative selection 29, 63, 195, 248
Neolithic Revolution 143, 212–14
neonates (newborns):
   immunity in infants and 200, 249
   mortality see birthweight
neoplasms, malignant see cancer
neoteny 124
Netherlands (end of WWII), famine 87, 228
neu (HER-2) 290, 292
neutral mutations 11, 41, 305
newborns see neonates
niche:
   construction 4, 34, 45, 76, 310
   re-creation 73, 74–5
Nietzsche, Friedrich 332
non-genetic inheritance 10, 73–7, 96, 239, 329
   cultural inheritance 35, 75–6
   non-synonymous mutations/polymorphisms 53, 64
nori seaweed 239
nosocomial infections and antibiotic resistance 255
novelty:
   environmental see environments
   evolutionary 85, 93–4, 205
   autoimmune and allergic disease and 251
   in development 233–5
   metabolic disease and obesity and 222–3
NPC1L1 68
nutrition and diet 205–36
   abnormal see famine; malnutrition; undernutrition
   adaptations to 205–36
   local 152, 211, 233
   cancer and 294–5, 295
   evolution history of diet 210–12
   female puberty and menarche and 118–19, 120
   modern transitions in 214–15
see also food

O
obesity 205
   epidemic 205, 235, 304
   genes 209, 219
      FTO 68, 70, 209
      MC4R 68, 209
   maternal 233, 235, 236
   paternal 233
obstructive sleep apnea 172
Oceania and Pacific Islands:
   gout 315–16
   migration from Africa to 150–1
Oct-4 90
offspring:
   conflict between parents and 191–2, 200
   maternal obesity and risk to 235
   number per pregnancy 177
   of trade-off between quality 100
   paternal obesity and insulin resistance and risk to 233
old age:
   defining 106
   duration and events 114
   omega-3 and -6 fatty acids 217
   oncogenes 290, 291, 296, 299
   oncogenesis (carcinogenesis) 291, 297, 299
   oocyte see egg
   organ development 84, 86
   Origin of Species 10, 29, 30, 41, 46, 331
   origins of life 22–3
   osteosarcoma 289, 297
   ovarian cancer 298, 301
   ovulation 118, 126, 190, 193
   covert 187–8
   menopause and 200, 201
   oxygen, reactive species of 139
   oxygen pressure at altitude 324
   oxytocin 275
P
p53 290, 291, 296, 297
Paciﬁc Islands see Oceania and
   Pacific Islands
pair bonding 184, 185, 274, 275
   Paleolithic humans 105
   puberty 119, 121
   Paley, William 329–30, 337
   PAMPs (pathogen-associated molecular patterns) 245
   pangensis 37, 333
   parasites 12
   basic reproductive ratio 242–3
   definition of parasitism 238
   gastrointestinal 240
   reproductive success of 242
   sex and defense against 180–1
see also pathogens
parental effects 74
   maternal effects 73, 74
   parental–offspring conﬂict 191–2, 200
   parthenogenesis 178, 179, 181
   paternal effects (paternal lineage) 56, 74, 149
obesity and insulin resistance 233
schizophrenia 284
see also Y chromosome
pathogen(s):
   arms race between host/prey and 171, 181, 241, 258, 305, 337
   drugs used against 255–8
   emergence 241–5
   gout and 315–16
   menstruation and shedding of 190
   toxins see toxins
   transmission 242–3
   virulence see virulence see also infections
pathogen-associated molecular patterns (PAMPs) 245
patrilinearity see parental effects
Pax6 94
Pediculus humanus 143
penetrance 65, 66, 70
peroxisome proliferator-activated receptor gamma gene (PPARG) 70, 315
personality traits and disorders 281–2
Peto’s paradox 296
phenotype:
   definition 79
   extended 34–5, 75–6
   genotype links with 15
   novelty, and macroevolution 93–4
   robustness 27, 65, 69
   shared 281
   thrifty 227–8
   philosophy 337
   phobias 283
see also fear
phylogeny 42, 43
physical barriers to microbes 245
pigmentation see color; skin
Pima Indians 223
Pinker, Steven 146, 272
Pitx 93
placenta 193–4
   insufficiency and dysfunction 169, 232–3
   invasion 297
   microbes and 245
   plague 244
   plasticity, developmental see development
pleiotropy 15, 30, 106–7
   antagonistic 107, 166, 169, 297, 298, 311, 320
   ploidy abnormalities 26
   polygenic adaptation 64
   polygynous mating systems 184, 185, 188, 272, 274, 321
polymorphisms:
  adaptation-associated 306
drug-metabolizing enzymes 156–7
single nucleotide see single
  nucleotide polymorphisms
social behavior and 275
structural 54–6
Polynesians, gout 315–16
polyphenism 79, 80
populations:
  allele frequency across 37, 74–5
  bottlenecks 41, 61–3
effective size 59–63, 150, 154
  genotype frequency across 37
positive selection 31, 42, 63, 65, 153,
  157, 267, 270, 316, 322, 324
  post-translational modifications of
  histones 72
  post-traumatic stress disorder 274
PPARG 70, 315
Prader–Willi syndrome 91
precocial (vs altricial) strategy 15,
  142–4, 194
predation (and predator–prey
  relationships) 16, 46, 97, 237,
  238, 283
  arms race see arms race
predictive adaptive responses in
  development 88–90, 165, 230
pre-eclampsia 195
preformationism 82
pregnancy:
  calcium in 307
  failure rate in humans 193
gestational diabetes mellitus 168,
  227, 233
  offspring number per 177
  sickness in 197–8
  trade-offs in 100–2
premature (preterm) birth 312–13
prenatal environment see intrauterine
  environment
present-day see contemporary times
preterm birth 312–13
prevention (prophylaxis) 304–5
  cancer 299–300
primates:
  brain size 111, 263
  definition 131
primogeniture 276
prion diseases 308–9
prisoner’s dilemma 269
probiotics and Clostridium difficile 240
progenitor cells 84
programmed cell death and
cancer 292
programming (in development),
  metabolic 227, 229
prokaryote, definition 20
proliferation (of cells) in cancer 292
  promoter methylation 91
prophylaxis see prevention
prostate cancer 107, 288, 289, 299
Protestantism 337–8
proximate causes 17, 18, 161, 163
  psychiatric states 279–81
  see also mental disorders
psychoses 284
puberty 115, 116–23
  duration 114
  events 114
  growth spurt 115, 116, 117, 118,
  126–7, 186
timing/age 118–23, 169, 186, 265
pubic hair 116, 117, 118, 138
public-goods production by
  bacteria 258
public health 18, 252–3, 303–26
  cancer 297, 298
  in treatment of cancer
  (radiotherapy) 300
random drift (genetic drift) 11, 27, 41,
  59, 60, 61, 174
random mutation 93
randomness in selection, absence 40
rapidity (fastness; speed): of evolution 46
  “fast” life history strategy 113,
  128–9, 295
rapid adaptation 46
reaction norm 79, 80, 81
reactive oxygen species 139
recessive alleles 26, 65–6
  disorders 65, 66
reciprocal altruism 37, 141, 264, 268,
  269, 284
recombination 26, 56, 57–8, 178–9
red blood cells (erythrocytes) and
  malaria 172, 173, 320
Red Queen effect 16, 181, 238
relationships (interpersonal/social):
  difficulties maintaining 282
  sexual 264–6
relative fitness 23, 39
religion (and spirituality and the
  supernatural) 276–7
  theory of evolution perspectives from
  327–8, 337–8
creationism 19, 21, 328, 329, 338
REM sleep 310
replication ability of cancer cells,
  unlimited 292
reproduction, asexual 23, 24,
  177–204, 178, 179, 180, 181
reproduction, sexual 177–204
  contemporary manipulation of 336
  life cycle (humans) 190–202
  strategies 182–3
  trade-offs in 100–2, 177, 182, 184
  cancer and 296
  see also assisted reproduction; basic
  reproductive ratio; fertility;
  mating; sterilization law
reproductive decline (decline in
  fertility) 127, 177, 200–2, 304
reproductive isolation (genetic
  isolation) 43–4, 45, 154
  definition 20
  modern times and 153
  reproductive selection 31
  reproductive success 11, 21, 24–5,
  40, 274–5
  parasites 242
  retina 171
  retrotransposons 55, 91
  retroviruses 171, 241
  drugs acting against 257
  endogenous 55
  Rhesus D antigen 195
  rickets 307
risk of disease 163, 165–6, 167–74
  environment and 89, 167–8, 218–36
  evolutionary processes affecting 163–4, 166
  polymorphisms 306
  prediction, in metabolic
  disease 230
race and 157
sexual dimorphism 188
social environment see society
  ultimate mechanisms affecting 167–74
evolutionary change within see microevolution of hominins 132–8
variation within (=individual variation) 11, 25, 77, 157
speed see rapidity
sperm 75, 183, 192
fertilization by 192
initiation/onset of spermatogenesis 116, 118
menstruation and 190
spirituality see religion
SRY 182
stabilizing selection 31, 40, 80–1
standing variation 63–4
stature (height): final 123, 169
height-velocity curves for boys and girls 115
short see short stature; stunting
stem cells 84, 291
adult/persisting 169, 291
cancer 291
embryonic 72
sterilization law 334
stress:
intraterine see intraterine environment
lifestyle 285
response
acute 272–3
chronic 293
embryo 193
see also post-traumatic stress disorder
structural variation 54–6
stunting 123, 170
development plasticity and 14
sun (UV) exposure 14, 139, 151, 306–8
supernatural see religion
surface antigens in immune evasion 249
symbiosis 238
sympatric speciation 44
symptoms and signs 303–4
syncytin genes 55
synonymous mutations/
polymorphisms 52–3, 64
systemic lupus erythematosus 251

T
T cells 246, 247, 248
Tasmanian devil facial tumor disease 300
taxonomy 12, 325
technological innovation:
contemporary 163, 306
hominin evolution
contemporary 143–4
migration to Oceania and 150–1
teeth see dentition
telemetry 8–9, 40, 43, 305, 325, 329
temperate vs tropical region
diseases 241
temporal dimensions see time
thalamus and puberty 118
thalassemia 153, 172, 320
tumor suppressor genes 290, 291, 292, 296, 297
see also cancer
twin(s):
epigenetic marking 92
studies 39

U
ultimate causes 17, 18, 51, 175, 304
ultraviolet (UV) exposure 14, 139, 151, 306–8
undernutrition (reduced nutrition):
infant/childhood 108, 165
lifespan and 108
maternal (and prenatal/fetal) 80, 87, 108, 228, 230, 312
see also famine
United States, eugenics 334–5
universals:
emotions 271, 283
language 279
life-history traits 98
senescence 104
upright posture see bipedalism
uric acid and urate 139
gout and 316
USA, eugenics 334–5
uterus:
environment within see intraterine environment
menstruation and shedding of pathogens from 190
UV exposure 14, 139, 151, 306–8

V
vaccination 253–5
vaginal microbes and vaginal delivery 233–5
variable number of tandem repeats (VNTR) 54–5
variance, analysis of 38
variation 11–12, 24, 25–9, 49–77, 151–4
constraints on 26–9
copy number 26, 54, 56, 148
factors affecting or causing and mechanisms of 59–65
epigenetic 71–3
lifestyle 153
migration 151–2
in growth and development 110–13
variation (Contd.)
  interindividual 11, 25, 77, 157
  intratumoral 293
  molecular basis 50–9, 71–3
  novelty and 93
  origin of 93
  standing 63–4
  structural 54–6
    see also diversity
  vasopressin 275
  vegetarian diet, meat-eating
    vs 212–13
  vertebrate eye 27, 28, 171
  *Vibrio cholerae* and cholera 67, 242, 243
  violence in society 272
  viral infections, drugs
    acting in 257
    see also pathogens
  virulence of pathogens 242–3

changes in 254
  vitamin C 139, 167
  vitamin D 14, 24, 306–8
    deficiency/insufficiency 14, 113, 307
  vomiting and nausea/sickness 303
    pregnancy 197–8

W
  W chromosome 181
  Waddington, Conrad 81, 83, 85, 92
  Wallace, Alfred Russell 10, 11, 12, 24, 29, 331
  water supplies, infection acquired
    from 253
  weight and mass (body):
    birth see birthweight
    metabolic rate and 98–100
    see also obesity
  Weismann, August 82, 106, 333

Westermarck effect 185
  Wilberforce (Bishop of Oxford) 331
  Wilson, David Sloan 277–8
  Wilson, EO 262
  worms, helminth 240, 242

X
  X chromosome 181

Y
  Y chromosome 56, 150–1, 180
    human migration and 149
    male ancestry and 56
    recombination absent or low in 56
  *Yersinia pestis* 244

Z
  Z chromosome 181
  zoonoses 241, 315