Genomics, Proteomics, and Related Approaches to Physiology

CHAPTER 3

Old-time Antarctic whalers believed that some of the fish in the polar seas had no blood because when they lifted the opercular flaps of the fish to see their gills, the gills were white, and when they cut the fish, only a whitish fluid ran out. A young Norwegian named Johan Ruud, a recent graduate with an undergraduate degree in biology, was introduced to these fish by whalers during an Antarctic voyage in the late 1920s. His curiosity whetted, he remembered the unique fish throughout the middle years of his life and, 20 years later, seized an opportunity to investigate them. Convinced of their distinctive properties, he then brought the fish to the attention of biologists worldwide. That was in the 1950s, and by now the fish have become the focus of one of the most startling and instructive efforts to fuse studies of physiology and genetics.

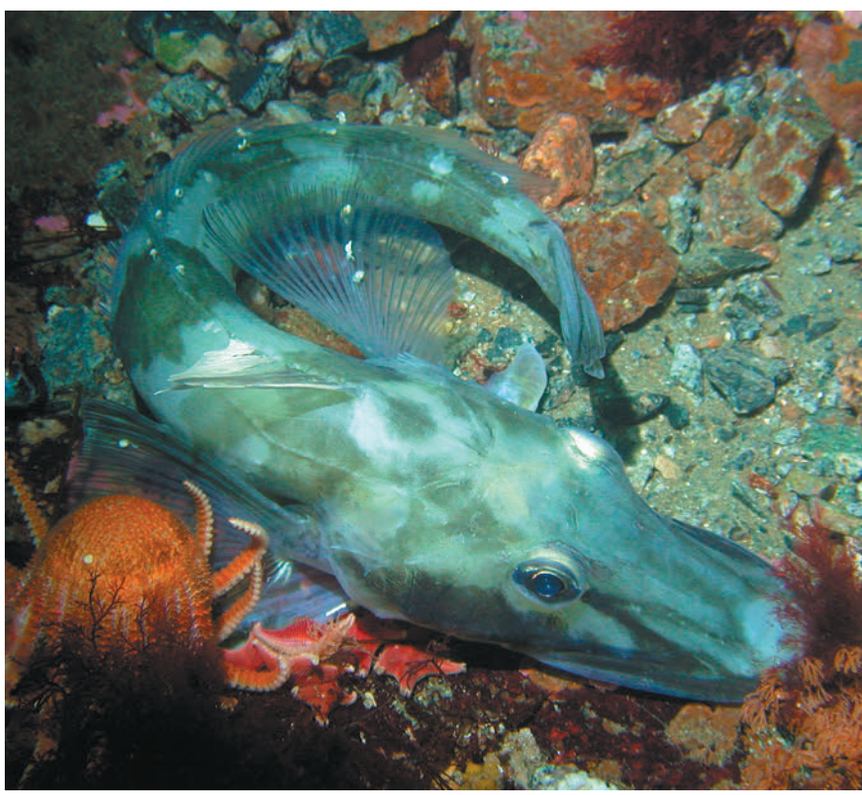
Although Johan Ruud originally referred to the fish using the whalers' term *bloodless fish*, his studies revealed that they in fact have blood. Their blood lacks hemoglobin, however, and is virtually devoid of red blood cells. Thus the blood is translucent and whitish, rather than dense red like most vertebrate bloods (**Figure 3.1**). Today the fish are usually called *icefish*, a reference to their clear blood and the icy seas they inhabit. There are about 25,000 species of fish alive today—and more than 53,000 species of vertebrates of all kinds—yet the icefish are the only vertebrates that do not have red blood as adults.

The icefish function, overall, as quite ordinary fish even though one might imagine that their lack of blood hemoglobin would be a crippling defect. They are neither rare nor small. Some species have been sufficiently common at times in the past to form commercially valuable fisheries, and several species grow to be 0.5–0.6 m long. Some are active swimmers that move between deep and shallow waters each day.

If icefish have any sort of obvious limitation, it is that they are restricted to the Antarctic seas,[1] where the waters are persistently very cold (often –1.9°C) and saturated with dissolved oxygen (O_2). The coldness tends to depress their metabolic needs for O_2, and it tends to make O_2 particularly soluble in both the seawater and their body fluids. The Antarctic seas became functionally isolated from most of the world's oceans about 30 million years ago because of dramatically altered global patterns of ocean circulation. Thereafter, the seas became much colder than they had been earlier, and the icefish evolved in that frigid context.

A question that immediately arises in considering icefish is how they came to lose their ability to synthesize blood hemoglobin. Physiologists recognize that studies of genetics can often provide

[1] One species, of undoubted polar ancestry, occurs in adjacent cold waters near the southern tip of South America.



This Antarctic fish differs from most fish in that it has no hemoglobin in its blood, giving it an almost ghostlike appearance The fish is one of 16 species known as *icefish* because their blood is clear like ice and they live in the icy polar seas around Antarctica. The icefish differ dramatically from most fish in the proteins they synthesize: They fail to synthesize blood hemoglobin, explaining why their blood is clear instead of red, but they produce antifreeze glycoproteins that are not made by the great majority of fish. (The photo is of *Chaenocephalus aceratus*; courtesy of William J. Baker.)

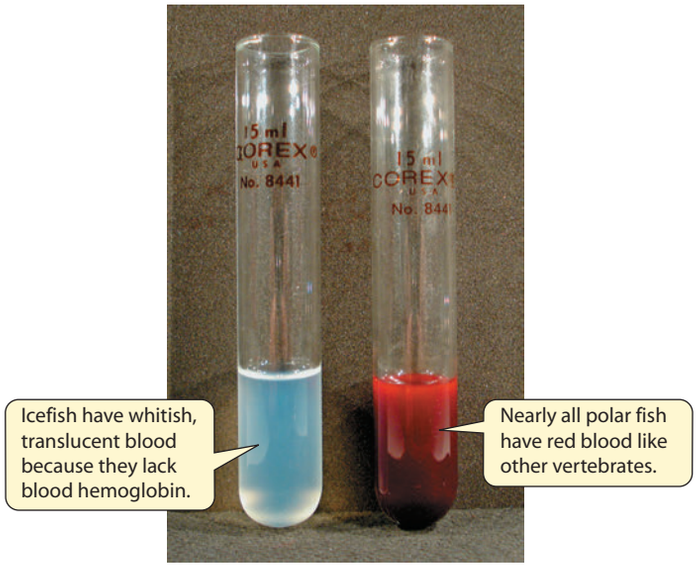


FIGURE 3.1 Freshly drawn blood from two species of Antarctic fish Both species belong to a single suborder, the Notothenioidei, the dominant group of fish in Antarctic waters. Most species in this group, such as the yellowbelly rockcod (*Notothenia coriiceps*) represented on the right, are red-blooded. The icefish, represented here by the blackfin icefish (*Chaenocephalus aceratus*) on the left, lost blood hemoglobin during their evolution. (Photo courtesy of Jody M. Beers.)

insight into such questions. Do the icefish still have the genes that code for hemoglobin and not transcribe those genes? Or have the genes become nonfunctional, or possibly entirely lost? Modern molecular genetic studies can answer these sorts of key questions.

Studies of genetics from an evolutionary perspective also may help to clarify the adaptive significance of the loss of blood hemoglobin. Fewer than 20 species of icefish exist today; most authorities say 16. If the fish in this small group turn out to lack functional genes for synthesis of blood hemoglobin, can we trace the loss of the genes back to a single common ancestor of all members of the group, meaning the genes were lost a single time? Or did certain species lose the genes independently of others during evolution? The answer, as discussed later, might help us think more confidently about whether the loss of blood hemoglobin was a disadvantageous accident or an advantageous change favored by natural selection.

The protein portion of the blood hemoglobin of vertebrates consists of alpha (α) and beta (β) globin units. Specifically, each hemoglobin molecule is composed of two α-globin units and two β-globin units (see Figure 24.1*c*). The genes that code for the α- and β-globin units are members of an evolutionarily ancient *gene family*. Biologists know the family is ancient, in part, because genes with a clear structural similarity are found in bacteria and yeasts, indicating that genes of this basic type existed before the time that animals branched off from other life forms. The ancestral genes became duplicated during evolution. Because of this duplication, each individual vertebrate animal today has multiple copies. Over the course of millions of years of animal evolution, copies of the genes in different species diversified by accumulating changes, and the multiple copies within a single species also underwent diversification. All the genes retained their family resemblance nonetheless. In modern birds and mammals, the two distinct—but structurally similar—genes that code for the α- and β-globin units are located on different chromosomes. In fish, by contrast, the two genes are found on a single chromosome, relatively close to each other.

Physiologists reasoned that if they could look at the exact DNA structure of the α- and β-globin genes in icefish—and compare those genes with ordinary fish globin genes—they might be able to determine how the DNA of icefish became modified during evolution to produce the hemoglobin-free condition. The physiologists employed knowledge of the basic DNA structure of fish globin genes to find the relevant stretches of DNA in icefish. Then they used the polymerase chain reaction (PCR) to make enough copies of the icefish DNA so that they could determine the sequences of nucleotide bases in the DNA. From research of this type carried out in just the last 15 years, the researchers found that in 15 of the 16 species of icefish, the DNA is modified in exactly the same way![2]

The relevant DNA in ordinary fish, as shown in **Figure 3.2*a***, consists of a complete α-globin gene separated by an intermediate stretch of DNA from a complete β-globin gene. In the icefish, however, as shown in **Figure 3.2*b***, the β-globin gene is completely gone, and the α-globin gene is missing parts, rendering it nonfunctional. A substantial stretch of DNA was deleted during the evolution of the icefish from their red-blooded ancestors.

To visualize *when* an event occurred during the evolution of a set of species, biologists often plot the event at the most logical position on an evolutionary tree of the species concerned. **Figure 3.3** shows the most likely evolutionary tree of the icefish and some of their closest red-blooded relatives. This tree is itself based principally on genetic information. To construct the tree, biologists determined the nucleotide base sequences of mitochondrial DNA in all the fish involved. Then they used the sequences of the

[2]The 16th species, although it does not contradict the conclusions described herein, is a special case. If you are interested in more detail, see Near et al. (2006).

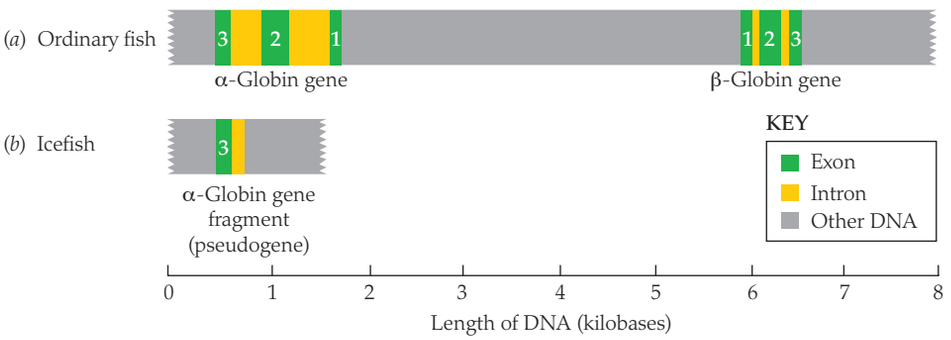


FIGURE 3.2 Genes and pseudogenes for blood hemoglobin The diagrams depict homologous stretches of DNA in (*a*) ordinary fish and (*b*) icefish. In ordinary fish (exemplified here by *Notothenia coriiceps*, an Antarctic red-blooded fish), functional genes for α- and β-globin are found near each other on a single chromosome; each globin gene consists of three exons (coded green) and two introns (coded yellow). In nearly all icefish, the entire β-globin gene, most of the α-globin gene, and the DNA between the original globin genes have been deleted. The icefish retain only a nonfunctional pseudogene, a fragment of the α-globin gene consisting of exon 3 and a part of the adjacent intron. (After Near et al. 2006.)

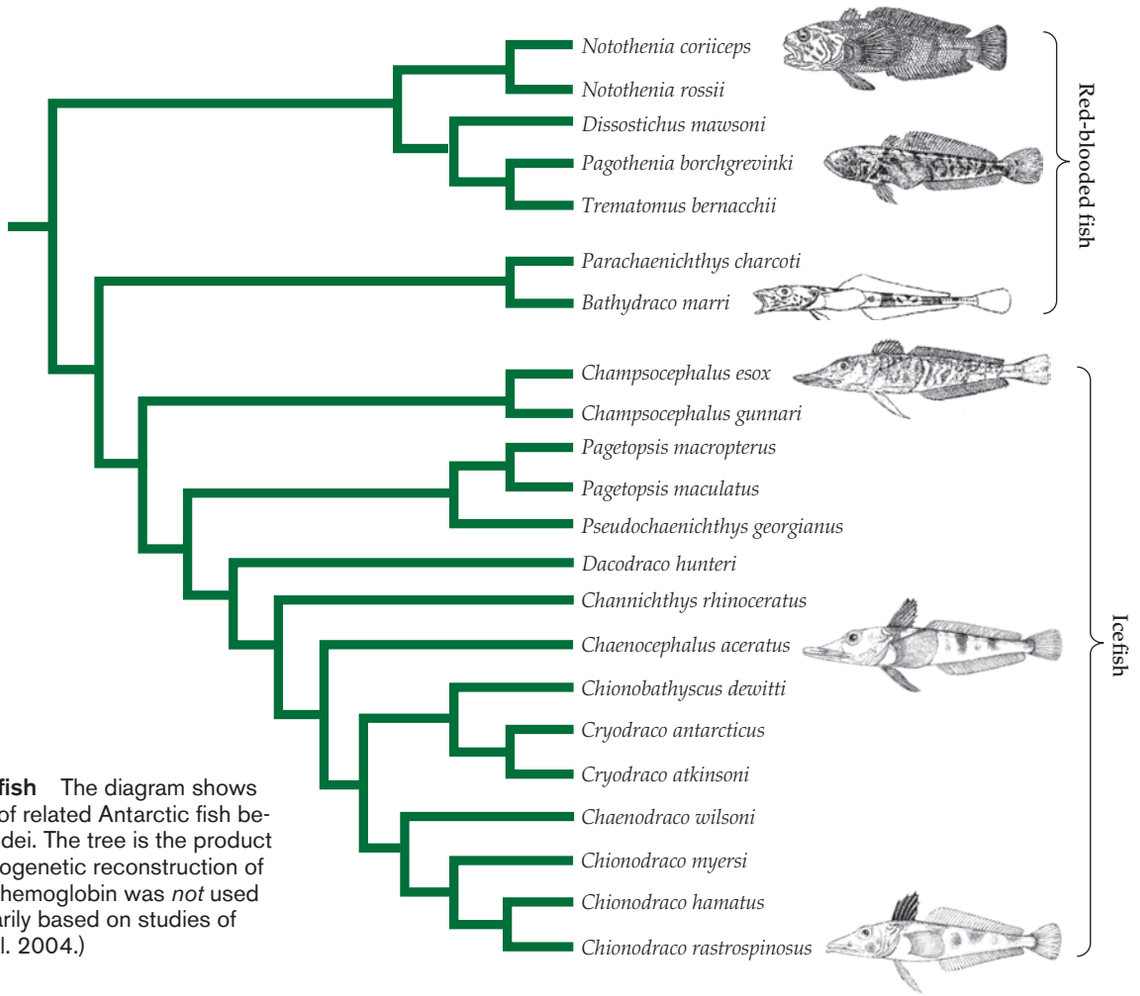


FIGURE 3.3 The evolution of icefish The diagram shows the evolutionary tree of 22 species of related Antarctic fish belonging to the suborder Notothenioidei. The tree is the product of the most recent research on phylogenetic reconstruction of this suborder. Information on blood hemoglobin was *not* used to construct the tree, which is primarily based on studies of mitochondrial DNA. (After Near et al. 2004.)

various species to identify logical relationships among the species, based on the same principles we discussed in Chapter 2 (see Figure 2.21) for interpreting amino acid sequences. No information on hemoglobin or the α- or β-globin gene was used in constructing the evolutionary tree.[3] The tree is therefore *completely independent of our knowledge of the globin genes*.

The most logical spot on the evolutionary tree to plot the loss of the globin genes is shown in **Figure 3.4**. All the lines of evolution drawn in red in Figure 3.4 end in species that have functional genes for both α- and β-globin and that synthesize blood hemoglobin. All the lines of evolution drawn in gray end in species that lack functional globin genes because of DNA deletions. Moreover, *all the species with deletions exhibit the same deletions*, those evident in Figure 3.2*b*. Therefore, the most logically coherent proposition is that the deletions occurred at the spot marked in Figure 3.4, in an ancestor of modern-day icefish. Later, as the various existing species of icefish evolved, all inherited the deletions from their common ancestor.

A deeper appreciation of these conclusions is reached by looking at another property of icefish that is similar in certain respects but dissimilar in others. In most vertebrates, the blood is not the only place in the body where hemoglobin is found (see Chapter 24). Hemoglobin of distinctive structure is found also within the cells of muscles—particularly many of the skeletal muscles and the heart muscle—where it imparts a red color to the muscle tissue. Hemoglobin within muscle cells helps increase the rate at which O_2 diffuses into the cells, and it sometimes acts as an important internal store of O_2 for the cells. Muscle hemoglobin is known as *myoglobin* (*myo-*, "muscle").

Six of the 16 species of icefish lack myoglobin in the cells of their ventricular heart muscle. The ventricle in these species is cream-colored, in contrast to the ventricle in the other 10 species of icefish, which have ventricular myoglobin.[4] **Figure 3.5*a*** shows examples of these two sorts of icefish (left and middle images). To increase understanding of the evolution of the myoglobin-free condition, physiologists took the same approach as in the study of blood hemoglobin. They examined the DNA sequences of the genes for myoglobin and asked what had happened to render the genes nonfunctional in the icefish that lack ventricular myoglobin.

The physiologists found that the myoglobin genes in some species of myoglobin-free icefish are altered in distinctly different ways from the genes in other of the myoglobin-free species. This discovery indicates that the myoglobin-free condition evolved independently more than once. In fact, based on the evidence currently available, there were four independent occasions when

[3]The globin genes are found in chromosomal DNA in the cell nucleus, not in the mitochondrial DNA used to construct the tree.

[4]Fish with red blood almost always have ventricular myoglobin.

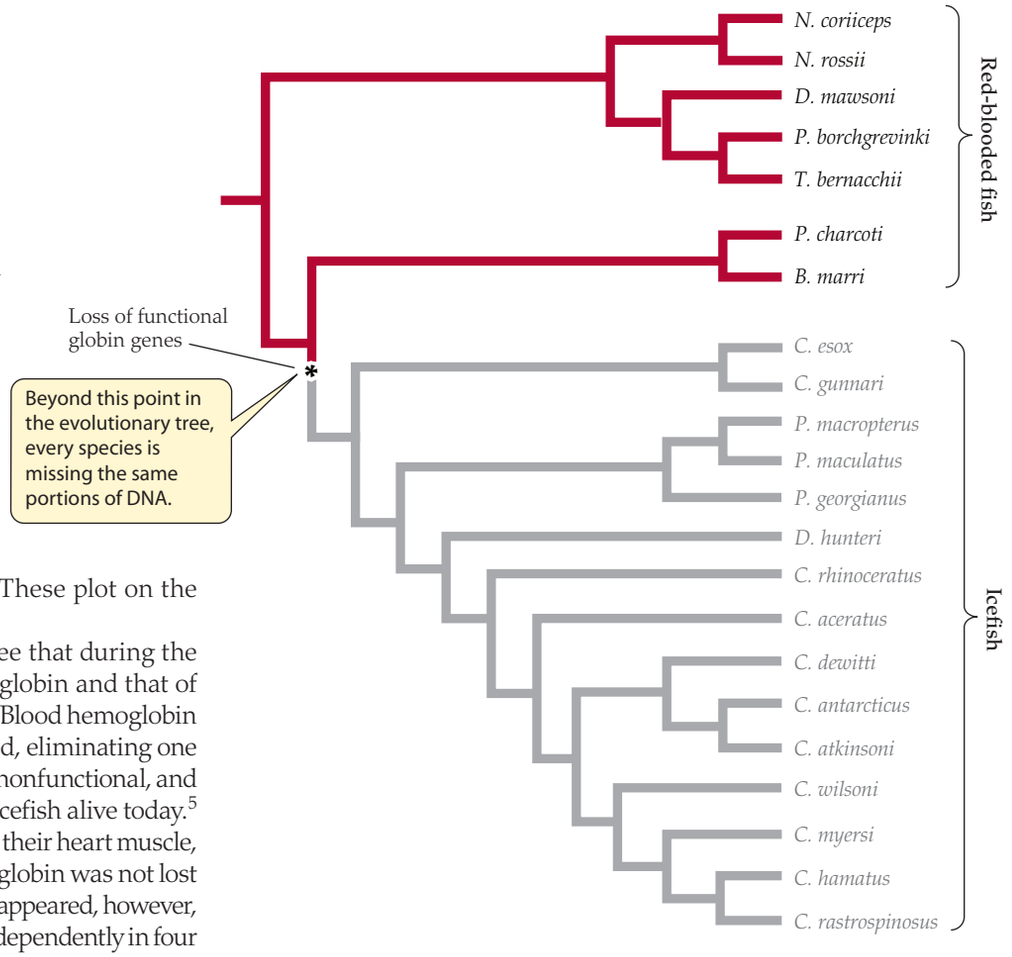


FIGURE 3.4 The most likely point at which the ability to synthesize blood hemoglobin was lost during the evolution of icefish Red and gray are used to symbolize which species and lines of evolution are characterized by the presence of functional globin genes (red) and which are not (gray). All of the 15 icefish species shown exhibit the condition diagrammed in Figure 3.2*b*: They lack the β-globin gene and possess only a nonfunctional fragment of the α-globin gene. One species of icefish (*Neopagetopsis ionah*) is omitted because, although it does not contradict the conclusions described here, it represents a special case. See Figure 3.3 for full species names.

the myoglobin genes became nonfunctional. These plot on the evolutionary tree as shown in **Figure 3.5*b***.

Comparing Figures 3.4 and 3.5*b*, you can see that during the evolution of the icefish, the loss of blood hemoglobin and that of muscle myoglobin followed very different paths. Blood hemoglobin was lost once: A large stretch of DNA was deleted, eliminating one globin gene and rendering the other irretrievably nonfunctional, and this deletion was passed on to all the species of icefish alive today.[5] By contrast, most icefish synthesize myoglobin in their heart muscle, just like most of their red-blooded relatives; myoglobin was not lost when blood hemoglobin was. After the icefish first appeared, however, mutations that eliminated myoglobin occurred independently in four of the lines of icefish evolution, and today six of the species exhibit one of those mutations and have myoglobin-free ventricular muscle.

Trying to decide if the loss of blood hemoglobin and myoglobin was advantageous or disadvantageous is a complicated matter. Surely biologists will debate some aspects long into the future. Here let's look just at blood hemoglobin.

Most researchers conclude that the initial loss of blood hemoglobin was almost surely a disadvantage for the ancestors of icefish. The genetic discovery that the loss occurred once—rather than multiple times—enhances the plausibility of this conclusion. Evolution by natural selection tends to weed out deleterious mutations. If geneticists found that blood hemoglobin had been lost multiple, independent times in icefish evolution, we would have to explain how multiple deleterious mutations were able to survive the immediate selection against them in order to argue that the loss of hemoglobin was a disadvantage. Instead, we need only argue that a single loss was able to survive despite being disadvantageous. Dramatic morphological and physiological specializations of icefish provide the most compelling evidence that the loss of blood hemoglobin was a disadvantage. Compared with related red-blooded fish of the same body size, existing icefish have very large hearts, and they circulate their blood at rates that are far higher than usual.[6] These specializations strongly suggest that the original loss of blood hemoglobin was a defect that significantly decreased the ability of the circulatory system to transport O_2, and the circulatory system thereafter became modified to make up for the defect by evolving a capacity to circulate the blood exceptionally rapidly. Looking at the evidence from both genetics and physiology, a persuasive case can be made that the loss of blood hemoglobin initially decreased the fitness of icefish and thereby favored the subsequent evolution of other attributes that compensated for the shortcoming.[7]

The *antifreeze glycoproteins* are a final set of proteins of icefish that raise intriguing genetic and evolutionary questions. All species of ocean fish with bony skeletons, including icefish, have body fluids that are more dilute in total dissolved matter than seawater is (see Chapter 28). Because of this, the body fluids of ocean fish tend to freeze at a higher temperature than seawater freezes.[8] The seawater in the frigid polar seas, in fact, is often cold enough to freeze fish even though the seawater itself remains unfrozen. Antarctic species of fish typically differ from the great majority of fish species in that

[5]The overall change from normal to altered DNA, diagrammed in Figure 3.2, probably involved several sequential mutations, not just a single large and simultaneous deletion. The change is described as occurring once because from the viewpoint of the species existing today, there was a single net alteration in the DNA, regardless of the intermediate steps involved in its occurrence.

[6]The distinctive cardiovascular physiology of icefish is discussed further in Chapter 24 (see page 638). You can see that the icefish hearts are enlarged in Figure 3.5*a*.

[7]An obvious question is how the fish that initially exhibited hemoglobin loss were able to survive. Researchers working on the subject argue that survival occurred in part because of the particular conditions existing in the Antarctic seas at the time, including relatively little overall competition and the existence of ecological refuges from competition. The arguments rest on studies of the ancient oceanography and plate tectonics of the region, as well as studies of biology.

[8]Typically, in aqueous solutions, the freezing point depends on the concentration of dissolved entities: The higher the concentration, the lower the freezing point. More specifically, doubling the concentration of dissolved entities (e.g., Na^+ ions or glucose molecules) approximately doubles the extent to which the freezing point of a solution is lowered below 0°C (see page 117). Antifreeze proteins and glycoproteins are unusual, compared with other dissolved materials, in that they exert a far greater effect on freezing point than can be accounted for by just their concentrations (see page 248).

(*a*) Hearts of two species of icefish and a red-blooded Antarctic fish

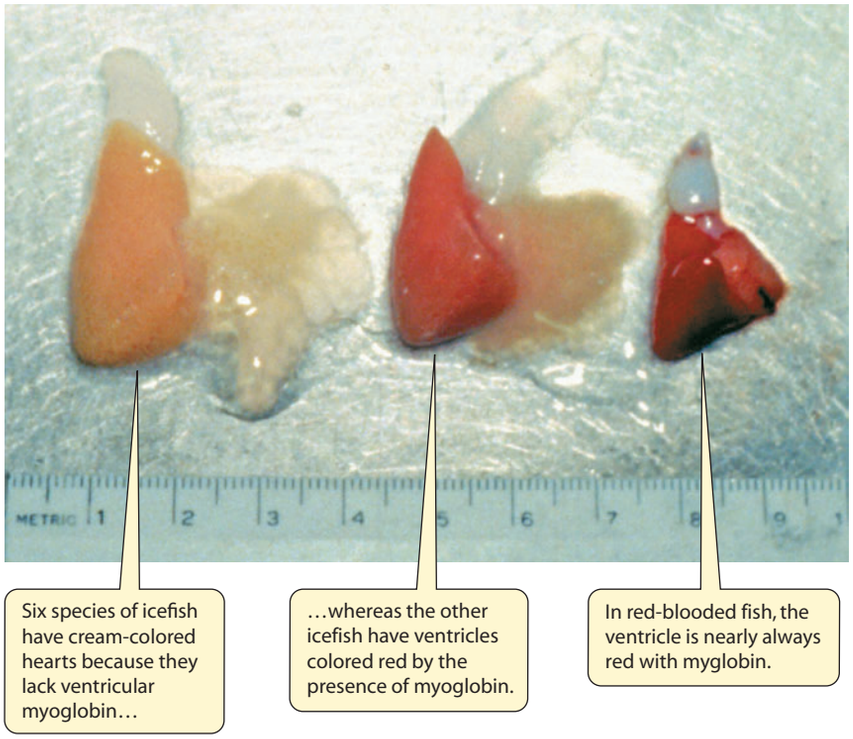
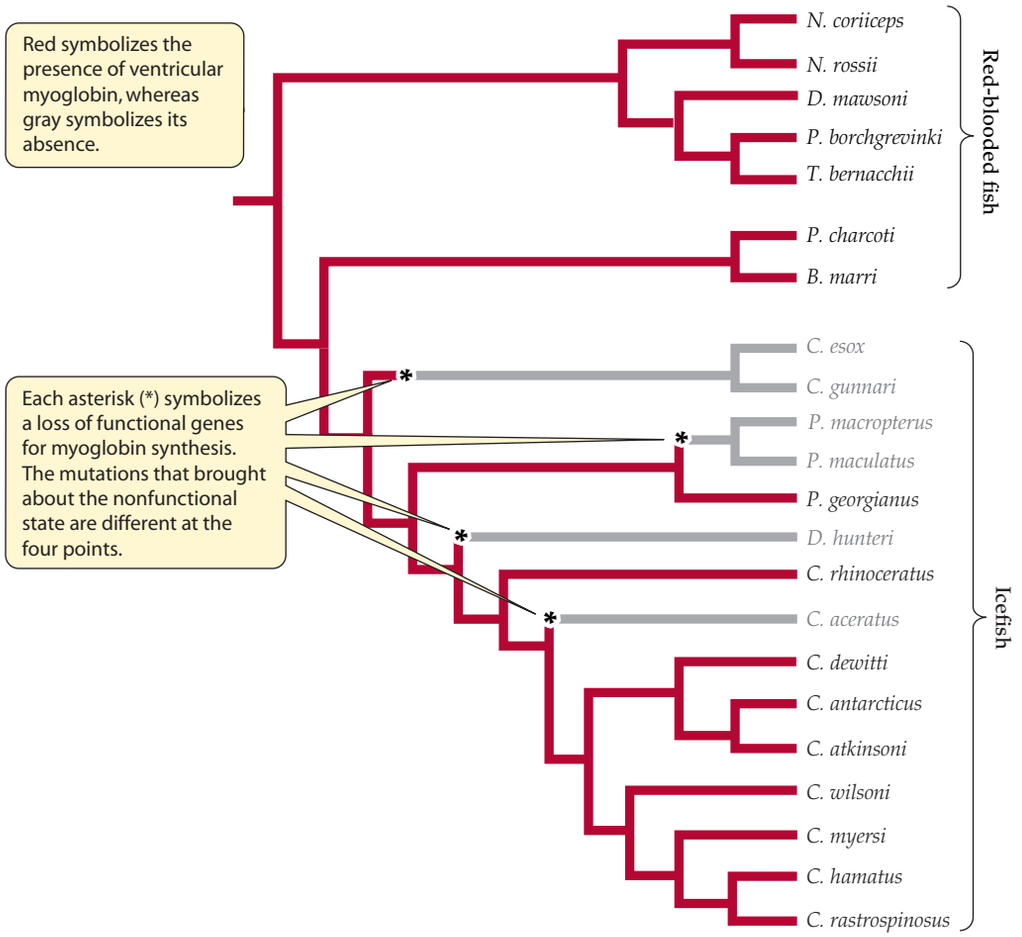


FIGURE 3.5 Presence and absence of myoglobin in the ventricular heart muscle (*a*) Three hearts—representing three species of Antarctic fish—removed from individuals of approximately the same body size. Because blood has been drained from the hearts, the color of the tissues depends on whether or not myoglobin is present in the cardiac muscle cells. The left and middle hearts are from two species of icefish; the species at the left (*Chaenocephalus aceratus*) has a cream-colored ventricle because it lacks ventricular myoglobin, whereas the species in the middle (*Chionodraco rastrospinosus*) has a reddish ventricle because it synthesizes ventricular myoglobin. The heart at the right, red with ventricular myoglobin, is from a red-blooded species of Antarctic fish (*Notothenia coriiceps*). (*b*) The points in evolution at which the genes for myoglobin synthesis in the heart ventricle became nonfunctional. The mutations at three of the four points are known to be different from each other; those at the fourth (*Dacodraco hunteri*) are presumed also to be distinct but remain to be described. See Figure 3.3 for full species' names. (*a* from Moylan and Sidell 2000; *b* after Sidell and O'Brien 2006.)

(*b*) The points at which the myoglobin genes became nonfunctional



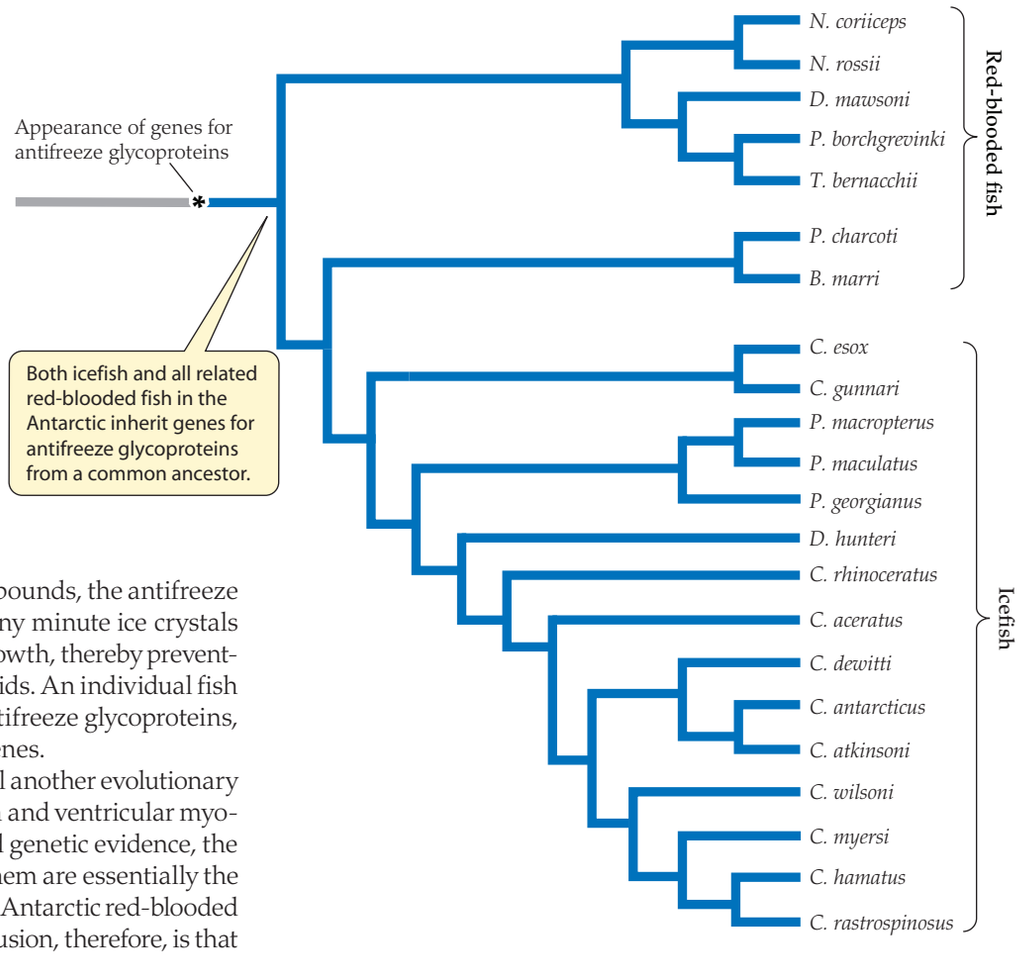


FIGURE 3.6 Evolution of antifreeze glycoproteins Blue symbolizes the species and lines of evolution characterized by antifreeze glycoproteins. Genes coding for the glycoproteins appeared prior to the evolution of icefish. See Figure 3.3 for full species' names. (After Cheng et al. 2003.)

they synthesize specialized antifreeze compounds, the antifreeze glycoproteins. These compounds bind to any minute ice crystals that appear in the body and arrest crystal growth, thereby preventing freezing of the blood and other body fluids. An individual fish synthesizes a suite of chemically similar antifreeze glycoproteins, coded by a suite of evolutionarily related genes.

The antifreeze glycoproteins present still another evolutionary scenario, compared with blood hemoglobin and ventricular myoglobin. According to available chemical and genetic evidence, the glycoproteins and the genes that code for them are essentially the same not only in all icefish but also in all the Antarctic red-blooded fish related to icefish. The most logical conclusion, therefore, is that the genes for antifreeze synthesis evolved before icefish appeared, as shown in Figure 3.6. When blood hemoglobin became deleted, giving rise to the first ancestors of icefish, those early hemoglobin-free fish already had the types of antifreeze compounds that their descendants still display today.

Modern research on icefish dramatically illustrates the fruitful use of genome science to help understand questions in animal physiology. We will return to the icefish examples several times as we now look more systemically at genomics and the disciplines related to genomics.

Genomics

Genomics is the study of the genomes of organisms. The **genome** of a species is the species' full set of genes, or—more broadly put—its full set of genetic material.

Probably the most famous aspect of genomics at present is *genome sequencing*, in which the DNA sequence of the entire genome of a species is determined. Because individuals of a species differ genetically to some degree (e.g., whereas some people have genetically coded brown eyes, others have blue eyes), the DNA sequence for a species depends a bit on the particular individual from which the DNA for sequencing was acquired. The sequence is enormously useful nonetheless, even if based on just one individual. At present, complete genome sequences have been determined for many, diverse animal species. Probably thousands of animal species' genomes will be fully sequenced within the next decade.

The study of genomics is not limited to species for which the entire genome has been sequenced. Sequencing of just a limited set of individual genes—a subpart of the genome—can sometimes set the stage for major new insights into a physiological system. An example is provided by research on the evolutionary origins of the genes that code for the antifreeze glycoproteins of Antarctic fish. Sequence comparisons of those genes with a limited but relevant set of other genes reveal that the antifreeze genes are derived from genes that in ordinary fish code for pancreatic proteins similar to trypsinogen, the precursor of the digestive enzyme trypsin. That is, copies of genes that originally coded for pancreatic digestive proteins evolved to code for the antifreezes! Like the unfolding of a good mystery story, this strange revelation from genomics has lately been joined by an equally strange revelation from the study of organ physiology to point to a dramatic, unexpected conclusion. Researchers have recently obtained evidence that the antifreeze glycoproteins—long assumed to be secreted directly into the blood by the liver—are in fact secreted by the exocrine pancreas into the intestines, only later making their way to the blood. The genomic and physiological evidence taken together suggest that when the antifreeze glycoproteins first evolved in Antarctic fish, they were pancreatic secretions homologous to the preexisting pancreatic digestive secretions.

Genomics is inextricably linked with advanced methods of information processing

Genomics, especially when entire genome sequences are studied, involves the processing of massive quantities of information. The genome of a single species, for example, may consist of a string of more than 1 billion nucleotide bases. To compare the genomes of two species, researchers often need to search for stretches of similar and dissimilar DNA in two sequences that each exceed 1 billion

bases in length. Modern genomics is defined in part by massive information processing.

Progress in genomics relies typically on the use of computer programs and robots that carry out great numbers of steps without direct human attention. The computer programming is itself sufficiently demanding that it is carried out by scientists in new specialties termed *computational biology* and *bioinformatics*. One key task for these specialties is the organization of data: The great masses of information gathered in genomic studies need to be recorded in ways that permit reliable retrieval by multiple users, many of whom were not involved in the original data collection. A second key task is to articulate the operational meaning of similarity and difference among base sequences within stretches of DNA and write algorithms that efficiently identify similarities and differences. Algorithms of this sort are used, for example, to locate apparently homologous strings of nucleotides in DNA—similar DNA sequences—in two or more genomes.

Much of genomic research is carried out by what are called **high-throughput methods**. The term refers to methods of the sort we have been discussing, in which computer programs and robots—after being designed to be as effective and error-free as possible—are "turned loose" to carry out procedures and generate results without moment-to-moment human attention or detailed human quality control. The process of adding direct human interpretation is known as **annotation**. To illustrate the interplay between high-throughput methods and annotation, consider a genome composed of 20,000 genes. When the genome is first fully sequenced, both the sequencing itself and the initial identification of individual genes will be carried out largely by high-throughput methods. Thereafter, experts on various genes in the global scientific community will directly or indirectly (e.g., by use of data catalogs) assess and add annotations to the information on genes of interest, but this time-consuming process is never complete. Possibly, therefore, the information on just 6000 of the genes will be annotated within the first few years. Knowledge of the other 14,000 genes would then consist, for the moment, only of the decisions of computer programs.

The World Wide Web is the primary vehicle by which the vast quantities of genomic information are shared among scientists worldwide. Gene and genome sequences are posted at dedicated websites (e.g., National Center for Biotechnology Information, GenBank, FlyBase, and WormBase). Web-based tools—notably BLAST programs—are available online to search for similarities among nucleotide base sequences in two or more stretches of DNA of interest.[9] Web-based tools, such as the Gene Ontology, are also available to facilitate and standardize annotation.

One overarching goal of genomics is to elucidate the evolution of genes and genomes

Genomics, whether based on entire or partial genome sequences, can be said to have two overarching goals:

- First, elucidate the evolution of genes and genomes.
- Second, elucidate the current functioning of genes and genomes.

[9] BLAST stands for *basic local alignment search tool*. BLAST programs are available to search for similarities among amino acid sequences in different proteins, as well as similarities among base sequences in samples of DNA. See "tools" at the National Center for Biotechnology Information website (www.ncbi.nlm.nih.gov) for useful tutorials on BLAST programs.

Genomics does not proceed in isolation in the pursuit of either of these goals. Instead, progress is most effective when genomics is integrated with physiology, biochemistry, and other disciplines.

In the study of the evolution of genes and genomes, one central topic is the elucidation of *mechanisms of gene modification*. The genome of each species is descended from the genomes of ancestral species, and genomes become modified as they evolve. Genomic studies help clarify the mechanisms by which genes and genomes become modified. One mechanism documented by genomic research is that genes sometimes become modified during evolution by the accumulation of beneficial base substitutions or other mutations favored by natural selection. This process can occur to such a significant extent that certain genes come to code for new proteins, as we have seen in the evolution of the genes for antifreeze glycoproteins in Antarctic fish from genes for pancreatic proteins such as trypsinogen. Another mechanism of modification is that genes sometimes become duplicated during evolution, and following duplication, the multiple copies within a single species often come to code for different proteins (illustrated by the evolution of the α- and β-globin genes in red-blooded vertebrates). Genes also sometimes become nonfunctional as they evolve, by partial or full deletion or by mutations that block transcription (illustrated by the icefish hemoglobin and myoglobin genes). The *coding* regions of genes are not the only parts that change by the operation of these and other mechanisms. *Regulatory* regions also sometimes change, thereby affecting the circumstances in which genes are transcribed

Besides analyzing mechanisms of gene change, another central topic in the study of the evolution of genomes is the *reconstruction of paths of evolution in the past*. In research of this sort, the evolutionary tree of species of interest is first determined as accurately as possible from genomic, morphological, and biochemical evidence. Then, as illustrated by Figures 3.4–3.6, the most likely points of occurrence of particular evolutionary developments are located on the tree. Methods exist to add a time scale, so that the amount of time between events can be approximated. Reconstructions of the sort described are useful in several ways. One is that they clarify the order of evolutionary events. For example, as physiologists reason about why certain icefish lost ventricular myoglobin, they can be virtually certain that blood hemoglobin was already gone (compare Figures 3.4 and 3.5*b*): To hope to explain accurately the myoglobin-free condition in icefish, one needs to think of the loss of myoglobin as having occurred, not in fish with ordinary vertebrate blood, but in ones with hemoglobin-free blood.

A second overarching goal of genomics is to elucidate the current functioning of genes and genomes

Because of the evolutionary continuity of life, when a new genome is sequenced, many of the genes found in the new genome are likely to be similar to genes observed already in genomes that were sequenced at earlier times. One may therefore be able to predict the function of genes in the new genome by extrapolating from preexisting knowledge of homologous genes. Suppose, to illustrate, that a particular gene of physiological importance is known to occur in the laboratory mouse (one of the most intensively studied mammals), and suppose that the mouse gene has been extensively annotated with information on how it is regulated and the physiological roles it plays. Suppose also that when researchers sequence the genome of a new mammal species, they find a gene that has a DNA sequence

closely similar to the mouse gene. It would then be reasonable for the researchers to predict that the gene in the newly sequenced species has a function like that of the gene already known. Not all such predictions prove to be accurate when tested, because genes evolve and can take on new properties. Nonetheless, homologous genes in different species often have similar functional properties.

The existence of *gene families* creates opportunities for large-scale interpretations based on the same sort of logic as just described. We have already seen an example of a gene family, namely the set of genes that code for vertebrate globin proteins. You will recall that the various globin genes within a single vertebrate species bear a family resemblance, and the globin genes in different species also do—all being related through evolutionary descent. All these genes are considered to belong to one gene family. The property that defines a **gene family** is that all the genes in a family share distinctive DNA base sequences. All the genes in a family also *tend* to code for functionally similar proteins; for example, just as the genes in the globin family code for hemoglobin proteins, the genes in another family might code for a particular type of enzyme, and those in still another family might code for a particular type of voltage-gated channel protein. With these concepts in mind, it is clear that when the genome of a species is initially sequenced, the simple process of scanning the genome for distinctive DNA sequences—the signature sequences of gene families—can be instructive. The process can provide a great deal of tentative insight into the functions of the genes present and the types of proteins likely to be synthesized.

An illustration is provided by the genome of the purple sea urchin. Figure 3.7 lists just six of the many major insights into urchin biology that arise from surveying this genome. Consider, first, observations ❶ and ❷, which state that when the urchin genome is surveyed, no genes are found that would be expected to code for gap-junction proteins or mediate the synthesis or use of melatonin and adrenaline. These observations suggest that the cellular communication system of sea urchins is unusual, compared with that of other animals, in that it lacks gap junctions and certain of the common neurotransmitter compounds. Observations ❸ and ❹ highlight that there seem to be lots of genes in the urchin genome for immune and detoxification proteins; these observations suggest that sea urchins have unusually elaborate immune and detoxification systems, possibly helping to explain why urchins are exceptionally long-lived. Observation ❺ points to a functional explanation for why the skeletons of sea urchins differ from those of vertebrates in the chemistry of their mineral composition. Sea urchins and vertebrates are thought to belong to closely related phyla (see the endpapers at the back of the book). Observation ❻ suggests that certain vertebrate genes in fact evolved in a common ancestor of sea urchins and vertebrates, rather than being exclusive to vertebrates as previously thought.

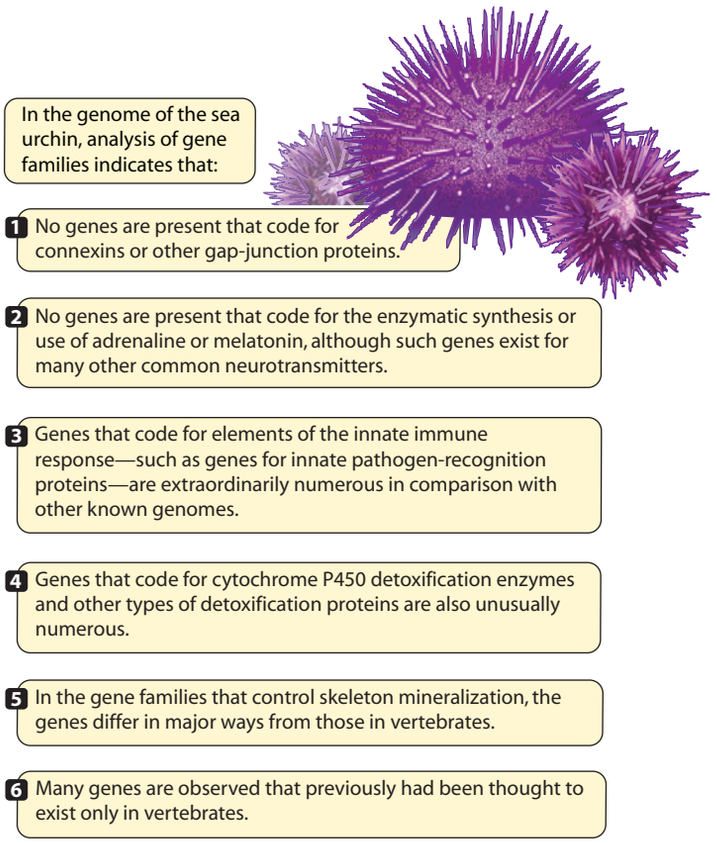


FIGURE 3.7 Characteristics of the genome of a sea urchin The six observations listed here arise from a close study of the genome of the purple sea urchin (*Strongylocentrotus purpuratus*), a common echinoderm in shallow coastal waters and tidal pools along the West Coast of the United States. Sequencing of the urchin's genome was completed in 2006. (After Sea Urchin Genome Sequencing Consortium 2006.)

Genomes must ultimately be related empirically to phenotypes

Although the genome of an animal reveals what genes are present in the animal's tissues, the *phenotype*[10] of any particular tissue at any given time is not a simple, deterministic consequence of the genes present, the tissue *genotype*. Thus efforts to predict proteins, metabolic processes, and other phenotypic traits from the genome represent just the first step in a long and essential process, namely that the genome must be related by empirical studies to the phenotype. When predictions are made from just the genome, they may ultimately prove to be wrong for several reasons. These include that (1) the actual functions of newly discovered genes may in fact not match the functions predicted by extrapolation from already-known genes and (2) even if the true function of a new gene is known, the gene may not be expressed when and where predicted. When all is said and done, predictions from the genome are *hypotheses*, and they must be tested before they can be accepted or rejected.

The process of testing genomic predictions entails, in part, the study of *which genes are transcribed and expressed* under various circumstances. It also entails the study of the *actual proteins synthesized* as a consequence of gene expression, and the *metabolites*[11] that are synthesized, used, and otherwise processed by the proteins. We will return to these sorts of studies—transcriptomics, proteomics, and metabolomics—after looking at important issues in research strategy in the next two sections of this chapter.

[10] The *phenotype* of a tissue consists of its outward characteristics—its structure, activities (such as contraction or secretion), biochemical constituents, and metabolic pathways—as opposed to its genetic material. Its *genotype* is its genetic material, its genome.

[11] A *metabolite* is an organic compound of modest to low molecular weight that is currently being processed by metabolism. An example would be glucose that is being processed by glycolysis.

The study of a species is said to enter a **postgenomic era** after the genome of the species has been sequenced. *Postgenomic* does not mean that the genome can be relegated to history. Quite the contrary, it emphasizes that in the era "after the genome is known," the study of a species' biology is forever altered. In the postgenomic era, the sequence of a species' genetic material is entirely known. The monumental task of empirically evaluating the full *significance* of this knowledge remains, however.

SUMMARY Genomics

- Genomics is the study of the genomes—the full sets of genes—of organisms. Because of the large numbers of genes, genomics depends on high-throughput methods to collect data and on advanced information processing to catalog and use data.
- One of the two major goals of genomics is to elucidate the evolution of genes and genomes. In pursuit of this goal, students of genomics seek to understand the mechanisms of evolutionary modification of genes and genomes (e.g., deletion and duplication). They also seek to reconstruct the paths followed by evolution in the past so that, for example, the order of evolutionary events is better defined.
- The second major goal of genomics is to elucidate the current functioning of genes and genomes. In pursuit of this goal, genomics uses information on already-known genes and gene families to predict the likely functions of newly identified genes and the likely ranges of action and competence of newly sequenced genomes.
- Although knowledge of an animal's genome permits many useful predictions to be made about the animal's biochemical phenotype, these predictions must ultimately be tested empirically. For example, although the suite of proteins synthesized in an animal's tissues can be predicted from the genome, the proteins must ultimately be studied directly, as by proteomic methods.

Top-down versus Bottom-up Approaches to the Study of Physiology

The traditional approach to the study of the multiple levels of organization in animal physiology can be described as *top-down*. To see this, consider the diagram of levels of organization and the chain of causation in **Figure 3.8*a***. In the traditional approach, shown at the left in **Figure 3.8*b***, the order of study of a phenomenon proceeds from the top of the diagram to the bottom. Physiologists first recognize an attribute of animal function of interest; a human example would be the *exercise training effect*, loosely defined as the improved ability of previously sedentary people to engage in exercise when they participate in a program of athletic training. After the attribute of animal function is specified, physiologists seek to identify the aspects of tissue function that are involved. In the case of the exercise training effect, a key aspect of tissue function is that skeletal muscles increase their capacity for physical work, and sometimes their size, when they are trained. Physiologists then look for the specific proteins—and the properties of the proteins—that are responsible for the tissue functions they have identified. Finally, physiologists identify the genes coding for the proteins, and they study how the expression of the genes is controlled, and the evolution of the genes. In the traditional, **top-down order of study**, investigation proceeds from animal function to tissue function, then to tissue biochemistry, and finally to genes.

Genomics sets the stage for physiologists to adopt a new *bottom-up* approach to the study of physiological phenomena, shown at the right in Figure 3.8*b*. In this approach, physiologists first sequence the genome of a species, or they approximate the genome by extrapolating from other, related species. Actually, the entire genome need not be known or approximated; physiologists need only have information on the portion of the genome that is relevant to them. After the genomic information is available, physiologists study the transcription of the genes in key tissues. For example, in research on athletic training, physiologists would study gene transcription in the skeletal muscles to determine which genes are transcribed to a greater extent after training than before (i.e., genes upregulated by training), and which genes are transcribed less than before training (downregulated). After changes in gene transcription are known, physiologists employ their knowledge of the affected genes to predict changes in tissue proteins. They also look directly at extensive suites of tissue proteins to see which are increased and decreased in amount. Then, physiologists seek to understand how the changes in proteins alter tissue function, and how the changes in tissue function are likely to affect animal function. In the new, **bottom-up order of study**, investigation proceeds from genes to gene expression, then to tissue biochemistry, and finally to tissue and animal function.

What would be the potential advantage of the bottom-up approach? The answer depends on recognizing the methods used. Researchers who employ the bottom-up approach have developed *high-throughput* methods to study all the genetic and biochemical steps. Thus not only the genes, but also the gene transcripts, proteins, and metabolites, are surveyed and monitored in great numbers simultaneously.

The bottom-up approach pursued with high-throughput methods has two distinctive assets: (1) It can be extremely thorough in searching for the genes, proteins, and metabolites that are instrumental in a physiological process; and (2) precisely because it is thorough, it can proceed without preexisting biases regarding which genes or proteins are likely be involved. Advocates of the bottom-up approach point to examples in which the traditional top-down approach missed important genes or proteins because investigators looked only at the small subsets that they imagined would be important. With the high-throughput methods employed in the bottom-up approach, investigators can, in principle, look at everything and therefore miss nothing.

Defenders of the top-down approach emphasize that looking at everything is not always an advantage. For example, when dozens or even hundreds of genes prove to be up- or downregulated during a physiological process—as is often the case—sorting through the implications for tissue function and animal function can be, literally, mind-boggling. Advocates of the top-down approach stress that it brings a needed focus to research because it starts with a defined phenomenon of known importance to whole organisms.

The top-down and bottom-up approaches seem certain to coexist for the foreseeable future. Each has advantages. Thus the two approaches can work together synergistically.

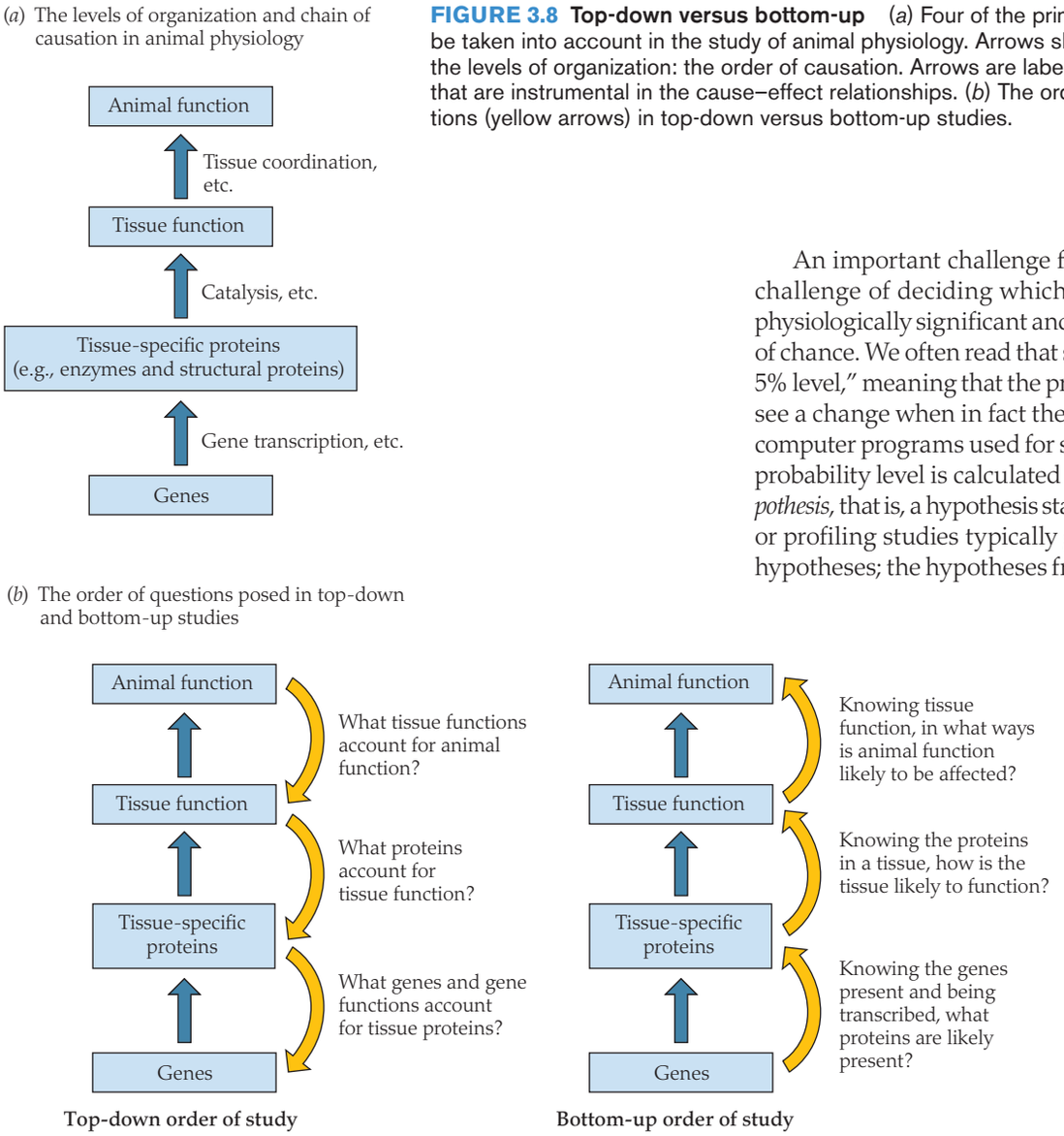


FIGURE 3.8 Top-down versus bottom-up (*a*) Four of the principal levels of organization that must be taken into account in the study of animal physiology. Arrows show cause–effect relationships among the levels of organization: the order of causation. Arrows are labeled with just a few of the processes that are instrumental in the cause–effect relationships. (*b*) The order in which investigators pursue questions (yellow arrows) in top-down versus bottom-up studies.

Screening or Profiling as a Research Strategy

At each stage of a bottom-up research program, as already suggested, the most common strategy today—termed **screening** or **profiling**—is for investigators to look as comprehensively as possible at the class of compounds of interest, whether the compounds are messenger RNAs (produced by transcribed genes), proteins, or metabolites. Screening or profiling capitalizes on the existence of high-throughput methods. This strategy also creates a demand for ever more efficient and cost-effective high-throughput methods.

In screening or profiling studies, the most common type of research design is to compare a tissue of interest before and after a change of interest. An example would be screening skeletal muscle before and after exercise training. Other examples include studies of a tissue before and after stress, or at different times of day, or in young individuals versus old. In all cases, the full suite of compounds present before a change is compared with the full suite after the change.

An important challenge for screening studies is the *statistical* challenge of deciding which observed changes are likely to be physiologically significant and which are likely to be mere artifacts of chance. We often read that statistical tests are carried out "at the 5% level," meaning that the probability of error (i.e., of thinking we see a change when in fact there is none) is 5% or less. In standard computer programs used for statistical calculations, however, this probability level is calculated on the assumption of an *a priori hypothesis*, that is, a hypothesis stated prior to data collection. Screening or profiling studies typically gather the data first, then articulate hypotheses; the hypotheses from such studies, in other words, are not a priori. Suppose a screening study examines 1000 genes and—among the 1000—identifies 100 genes that seem to exhibit increased transcription when a tissue is stressed in some particular way. Of those 100, if an investigator has simply used a standard statistical program and an error probability of 5%, there will be no doubt that 50 of the instances of increased transcription (5% of 1000) are not real and repeatable changes but chance events—false positives. Specialized statistical methods already exist and are used to reduce false positives. Moreover, research is ongoing to develop improved methods that will deal in superior ways with the problem that when great numbers of genes, proteins, or metabolites are surveyed, some of the effects observed are bound to be artifacts of chance.

The Study of Gene Transcription: Transcriptomics

The study of gene transcription—that is, the study of which genes are being transcribed to make messenger RNA (mRNA) and the rates at which they are transcribed—is known as **transcriptomics** or **transcription profiling**.[12] Because physiologists have recognized for decades that the genes that matter are those that are transcribed, physiologists have studied transcription and mRNA synthesis for a long time. *Transcriptomics* and *transcription profiling* are new terms. One of their key connotations is that they imply

[12] **Expression profiling** is another commonly used term. However, *expression* is sometimes used to mean protein synthesis. *Transcription profiling* is a more exacting term to use when *transcription* per se is the process under study. See Nikinmaa and Schlenk (2009) for a discussion.

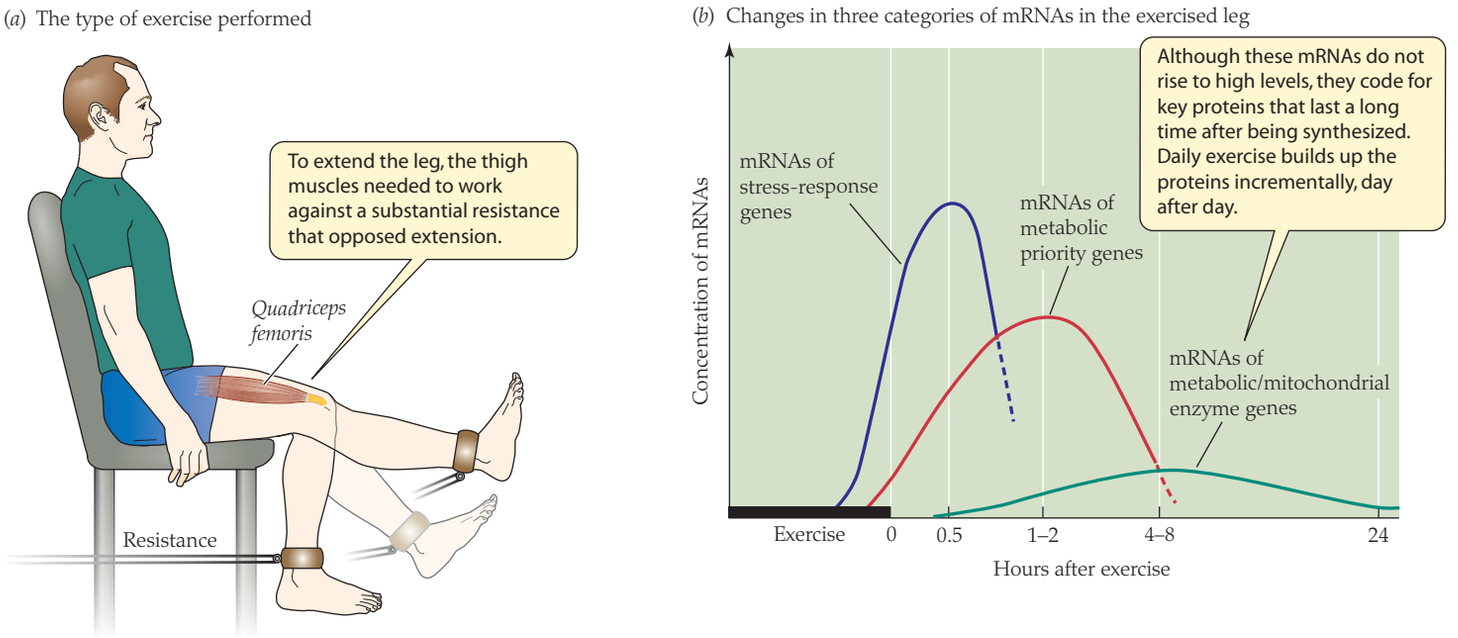


FIGURE 3.9 Exercise is followed by increased transcription of groups of genes (*a*) Subjects performed a 90-min bout of leg-extension exercise once each day for 5 days. Only one leg was exercised; the other remained at rest. On the fifth day, biopsy samples were taken repeatedly from the *quadriceps femoris* muscles of the two legs, and concentrations of many messenger RNAs were measured in the samples. (*b*) Based on the mRNA data—and using the unexercised leg as a control for the exercised one—investigators identified three groups of genes that exercise causes to be upregulated. During and after exercise of the sort performed, one group of genes—termed the *stress-response genes*—quickly exhibits highly enhanced transcription. Another group—dubbed the *metabolic priority genes*—is slower to be upregulated but also exhibits a dramatic increase in transcription. The third group—the *metabolic/mitochondrial enzyme genes*—is the slowest to be upregulated and exhibits only modest (but long-duration) increases in transcription. Note that times are not evenly spaced on the *x* axis. (*b* after Booth and Neufer 2005.)

the simultaneous study of great numbers of mRNAs, often by use of highly automated methods.

Changes in gene transcription during and after exercise nicely illustrate the transcriptomic approach. Investigators have found that numerous genes are upregulated—transcribed at an increased rate—in exercising muscles each time a person engages in an extended period of endurance exercise. In one study (**Figure 3.9*a***), subjects performed 90 minutes (min) of leg extension (kicking) exercise with one leg while the other leg remained at rest. For 24 h following the exercise, biopsy methods (see Figure 21.3) were used to remove tiny samples of muscle tissue from thigh muscles periodically, and the levels of dozens of mRNAs in the exercised leg were compared with those in the unexercised leg. Three groups of genes were identified that underwent upregulation in the muscles of the exercised leg during and following exercise (**Figure 3.9*b***).

One of the groups of upregulated genes—termed the *metabolic/mitochondrial enzyme genes*—is especially interesting. These genes code for mitochondrial proteins. As seen in Figure 3.9*b*, the metabolic/mitochondrial enzyme genes undergo just a small degree of upregulation after a single bout of exercise and thus boost mitochondrial protein synthesis just a bit. The mitochondrial proteins, however, have long half-lives; once synthesized, they last for a long time. The investigators believe they have found a key mechanistic reason for why muscles keep increasing their exercise capability, day after day, over many weeks of daily endurance training. Each day's training produces just a modest increase in transcription of the mitochondrial genes and a modest increase in synthesis of mitochondrial proteins, but these small effects cumulate when training is repeated day after day.

DNA microarrays (also called *DNA microchips* or *gene chips*) are the basis for a set of particularly important, high-throughput methods for the study of gene transcription. A single microarray can permit investigation of thousands or tens of thousands of genes at a time.

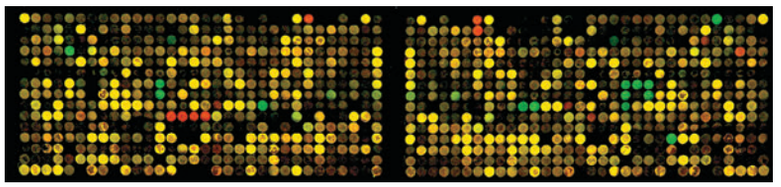
As a physical object, a microarray consists of a grid of spots of DNA placed on a glass plate or other solid substrate by a robot (**Figure 3.10*a***). Each spot might, for instance, consist of a stretch of DNA that represents a single gene or presumptive gene.[13] Given the minute physical size of each bit of DNA and the technology used to apply the spots, a grid of 10,000 spots—representing 10,000 different genes—will fit within an area of only 1 cm^2, or less.

In one common type of experiment using a DNA microarray, a single array is used to carry out a direct comparison of the mRNAs produced by a tissue under two different conditions.[14] For illustration, let's assume that we are doing a microarray study of muscle before and after exercise. The mixes of mRNA molecules present under the two conditions are extracted (❶ in **Figure 3.10*b***), and all the mRNA molecules in each extract are then labeled with a distinctive fluor,

[13]The nature of the DNA spots depends on the method used to prepare them, and several methods are in common use. One method is to reverse-transcribe mRNA molecules to produce DNA sequences (cDNA) that code for the mRNA molecules; this method, you will note, does not require a sequenced genome. Another method is to start with the entire genome and essentially cut it up into many, often overlapping, pieces of DNA (producing a *tiling* array). Still another method is to synthesize DNA sequences from scratch (essentially from raw nucleotides) using knowledge of the genome to select the sequences made.

[14]This experimental design is said to employ a *two-color spotted microarray*.

(*a*) A fluorescent image of a DNA microarray, greatly magnified



(*b*) A study to compare gene transcription under two different conditions

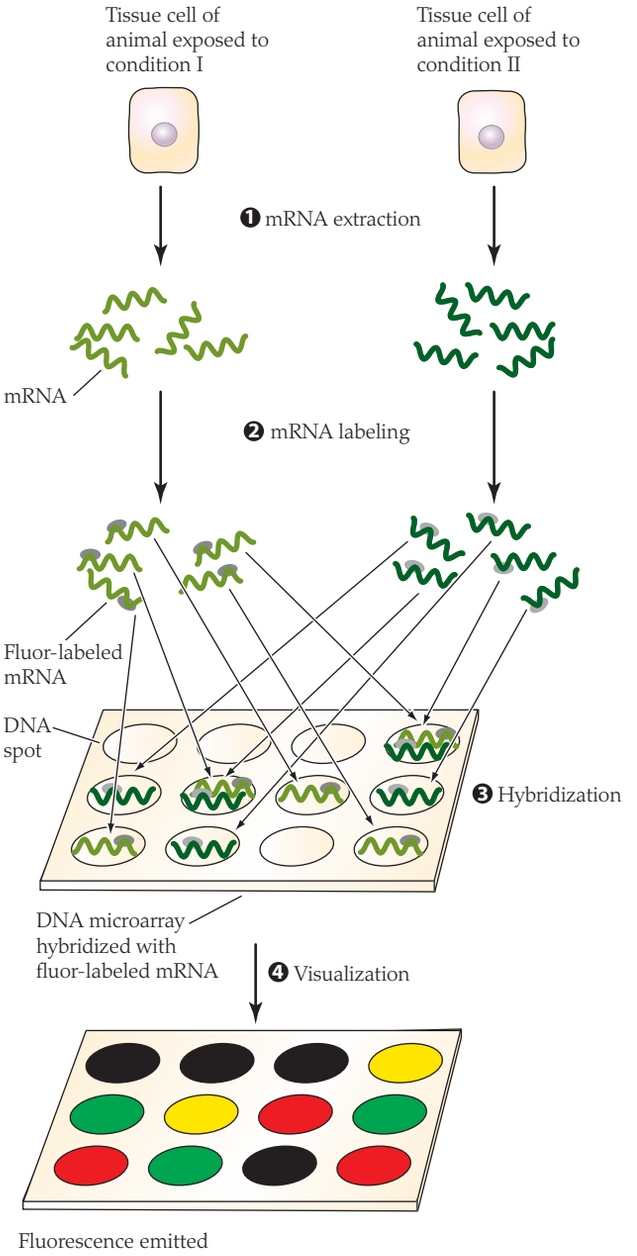


FIGURE 3.10 With DNA microarrays, the transcription of thousands of genes can be studied at once (*a*) A DNA microarray. Each spot represents one gene. The color of a spot after processing indicates whether the gene was transcribed under either or both of two conditions being investigated. (*b*) An outline of the procedure followed to compare directly the mRNAs produced by a tissue under two different conditions, I and II. The mRNAs extracted from the tissues of animals exposed to the two conditions are labeled with two different fluors (green and red) prior to hybridization with spots of DNA on the microarray. Each mRNA hybridizes with the specific DNA that represents the specific gene that codes for the mRNA. When the fluors are visualized, spots emit either green or red fluorescence if they represent genes that were transcribed under just one of the conditions. Spots that represent genes that were transcribed about equally under both conditions emit yellow (the combination of green and red). Spots emit nothing and appear black if they represent genes that were transcribed under neither condition.

a compound that has the potential to emit light by fluorescence.[15] A common approach (❷ in the figure) is to label one extract (e.g., that from muscle before exercise) with a fluor that will emit green light after it is fully processed, and the other extract (e.g., that from muscle after exercise) with a different fluor that will emit red light. The two fluor-labeled mRNA preparations are then permitted to hybridize with the microarray. Each mRNA hybridizes with the DNA spot representing its specific gene (❸ in the figure). Thus, at this step, each spot in the microarray potentially becomes labeled. A spot is labeled with mRNA from both preparations if the gene it represents was being transcribed under both conditions. A spot is labeled with mRNA from only one preparation if the gene it represents was being transcribed under only one of the two conditions. A spot that represents a gene that was not transcribed under either condition is not labeled at all. Finally, the fluors bonded to the microarray are visualized by laser scanning (❹ in the figure) so they emit their distinctive fluorescent wavelengths: green or red. If the gene represented by a spot on the microarray was not being transcribed under either condition, the spot emits nothing and appears black. If the gene was being transcribed approximately equally under both conditions, the spot emits both green and red and thus appears yellow. The most interesting spots are those that emit only green or only red, because those represent the genes that were being transcribed under only one of the two conditions studied.

Transcription profiling often identifies large numbers of genes that exhibit altered transcription in response to environmental or other conditions

Table 3.1 illustrates that animals routinely modify the transcription of hundreds of genes—by up- or downregulating them—in response to environmental changes of many sorts, or in response to other changes such as exercise or hormone exposure. Physiologists never imagined until recently that such large numbers of genes would be involved. Determining the *significance* of gene transcription changes numbering in the hundreds will obviously be a challenging and revelatory phase in the advance of physiological knowledge. One of the "At Work" chapters in this book, Chapter 21, provides considerably greater detail on exercise responses.

Transcription profiling reveals that many genes routinely undergo daily cycles of transcription

Periodic cycles of gene transcription—*transcription rhythms*—are common, based on recent research. Most known cycles in gene transcription are daily cycles in which the pattern of falling and rising transcription repeats approximately every 24 h. For example, in

[15] For conceptual simplicity, we here describe the procedure as if the raw mRNA itself were used, although in reality, technical steps must be taken to stabilize the mRNA.

TABLE 3.1 Numbers of genes studied and discovered to be up- or downregulated in transcriptomic research on seven phenomena using DNA microarrays

Phenomenon and tissue studied (reference in parentheses)	Number of genes studied	Number of up- or downregulated genes[a]	Some of the functions controlled by affected genes
Endurance exercise in humans: Thigh muscle 3 h following lengthy (~75 min), high-intensity bicycling compared with before bicycling (Mahoney et al. 2005)	8432	126	Mitochondrial biogenesis, tolerance of oxidative stress, membrane ion transport, nuclear receptor function (all categories mostly upregulated)
Tissue freezing in freeze-tolerant wood frogs: Heart muscle after frogs experienced extracellular freezing compared with before (Storey 2004)	>19,000	>200	Glucose metabolism, antioxidant defense, membrane ion transport, ischemia-related signaling (all categories mostly upregulated)
High water temperature in killifish: Liver after fish were exposed to high temperature compared with ordinary temperature (Podrabsky and Somero 2004)	4992	540	Heat-shock protein synthesis, cell membrane synthesis, nitrogen metabolism, protein biosynthesis
Continuous swimming for 20 days in rainbow trout: Ovaries in fish that swam compared with resting fish (Palstra et al. 2010)	1818	235	Protein biosynthesis, energy provision, ribosome functionality, anion transport (all categories mostly downregulated)
Transfer from freshwater to seawater in eels: Several tissues in eels after transfer to seawater compared with before (Kalujnaia et al. 2007)	6144	229	Transport across membranes, cell protection, signal transduction, synthesis of structural proteins
Exposure to juvenile hormone in developing honeybees: Animals treated with a juvenile hormone analog compared with ones not treated (Whitfield et al. 2006)	5559	894	Foraging behaviors, RNA processing, protein metabolism, morphogenesis
Ocean acidification in sea urchin larvae: Larvae exposed to high CO_2 and low pH compared with ones not exposed (Todgham and Hofmann 2009)	1057	178	Biomineralization, energy metabolism, cellular defense responses, apoptosis

[a]The numbers of affected genes are quoted from the references cited; different criteria might have been used in different studies.

a recent study of gene transcription in the lung tissue of laboratory rats, more than 600 genes exhibited clear 24-h cycles in their transcription. These genes participate in just about every aspect of lung function. Some, for example, play roles in tissue maintenance, others in defense against airborne foreign materials, and still others in the genesis of asthma. Because the genes are transcribed more rapidly at some times in the 24-h daily cycle than at others, lung repair and defense are likely to occur more effectively at some times of day than others, and treatments for asthma may be more effective at certain times. All the genes are not synchronous; some reach their peaks and troughs of transcription at different times than others. Some of these transcription cycles are undoubtedly under control of biological clocks (see page 410). Other cycles, however, are probably direct responses to day–night changes in the outside environment.

Additional studies in a wide variety of animals indicate, similarly, that 2–40% of genes exhibit cycles in their transcription on approximately 24-h periodicities. For example, hundreds of genes in the malaria mosquito *Anopheles gambiae* exhibit daily cycles (**Figure 3.11**). Inasmuch as some of these genes play roles in defending the mosquitoes against toxic environmental agents, the mosquitoes may be inherently more vulnerable at some times of day than others to agents (e.g., insecticides) used to combat them (see Figure 3.11).

Manipulations of protein synthesis can be used to clarify gene function

We are concerned in this chapter mostly with the unmanipulated chain of events by which genes are naturally transcribed and translated, leading to changes in proteins and other aspects of the biochemical phenotype. A significant aspect of gene expression, nonetheless, is that it can be manipulated experimentally as a way of gaining insight into gene function.

One strategy of this sort is **gene deletion** or **gene knockout**, in which a gene is manipulated so that experimental animals lack functional copies of the gene. The animals, therefore, cannot synthesize the mRNA ordinarily associated with transcription of the gene, and consequently they do not synthesize the protein (or proteins) coded by the gene. In principle, such animals will be deficient or inferior in one or more phenotypic traits, and their deficiencies will reveal the function of the missing gene. A converse strategy is **forced overexpression**, in which tissues are subjected to

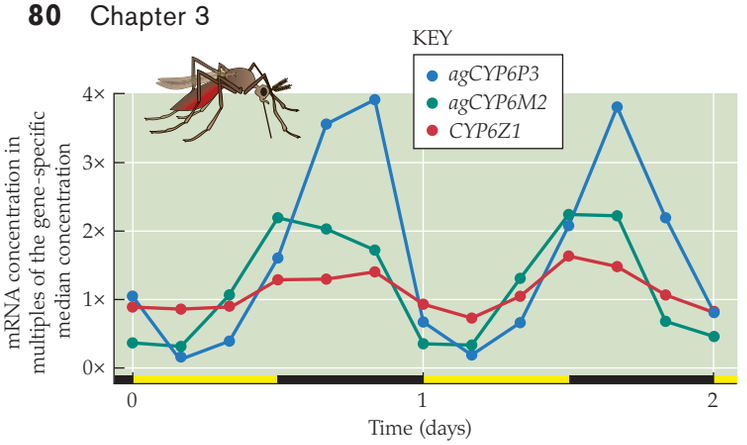


FIGURE 3.11 Daily cycles in three mRNAs coding for detoxification enzymes in the malaria mosquito *Anopheles gambiae* The mRNAs produced by transcription of three genes—*agCYP6P3*, *agCYP6M2*, and *CYP6Z1*—were measured over 2 day–night cycles (see scale at bottom: yellow = day, black = night). The genes and mRNAs code for three detoxification enzymes in the P450 system (see page 48). Because the enzymes help the mosquitoes resist pyrethrin-based insecticides, the cycles in gene transcription suggest that vulnerability to such insecticides is likely cyclic. mRNA levels are expressed on the *y* axis as multiples of the gene-specific median concentration (e.g., 2× for a particular gene indicates a concentration twice as high as that gene's median concentration over the course of the study). Because the mosquitoes studied here could not actually see day and night, biological clocks probably controlled the transcription cycles shown. (After Rund et al. 2011.)

experimentally increased synthesis of the mRNA associated with a gene of interest.[16]

Several considerations can cloud the interpretation of gene knockout or overexpression studies. *Compensation* is one of the most important: When animals artificially lack a certain protein, for example, they often exhibit other phenotypic alterations that tend to make up for the loss of function they would otherwise exhibit. Mice engineered to lack functional genes for myoglobin,[17] for example, have more blood capillaries in their cardiac muscle than ordinary mice, and they circulate blood faster through the muscle. These compensations, and others, prove that the lack of myoglobin is physiologically significant. Because of the compensations, however, the lack of myoglobin does not have a simple deterministic effect on the overall phenotype of the animals, and the mice are actually quite normal in their overall vigor and appearance.

RNA interference (**RNAi**) is a recently discovered cellular process that, among other things, has great promise for being manipulated to gain insight into gene function. When selected double-stranded RNA molecules are introduced into cells and processed by the RNAi pathway, specific mRNA molecules that are naturally produced by the cells are destroyed. In effect, the genes that produce the targeted mRNAs are silenced because the mRNAs produced by transcription of the genes are rendered inoperative before they lead to protein synthesis. The consequence is in many ways similar to that of gene deletion: Certain proteins are not synthesized, providing an opportunity to learn what the proteins normally do. Unlike the case when gene deletion is used, however, animals with *normal genomes* can be manipulated with RNAi because gene action is blocked following gene transcription.

[16]This might be achieved, for example, by genetically engineering a tissue to have an unusually large number of copies of the gene.

[17]Myoglobin aids intracellular O_2 transport and storage, as discussed earlier in this chapter.

SUMMARY The Study of Gene Transcription: Transcriptomics

- Transcriptomics or transcription profiling (also sometimes called expression profiling) is the study of which genes are transcribed in a tissue and the degrees to which they are transcribed. Transcription is evaluated by measuring the messenger RNAs (mRNAs) produced in the tissue.
- DNA microarrays are a major tool in modern transcriptomic research. A microarray consists of a grid of thousands of DNA spots, each representing a particular gene. Each mRNA produced by a tissue binds to the DNA that corresponds to the gene that produced the mRNA. The DNA spots that thus become labeled with mRNA when exposed to the mix of mRNAs produced by a tissue collectively mirror the genes being transcribed in the tissue.
- DNA microarrays are often used to carry out a direct comparison of the mRNAs produced by a tissue under two different conditions. The mRNAs produced under the two conditions are labeled with green and red fluors. Gene-specific DNA spots in the microarray then glow with green light, red light, or yellow light (the result of green + red), depending on whether the corresponding genes were transcribed under just one of the two conditions or both.
- Studies using microarrays often identify hundreds of genes that undergo changes of transcription when an animal is exposed to a change in conditions. Many genes also routinely exhibit daily cycles of transcription.
- One way to study gene function is to manipulate gene transcription or translation and observe the consequences. Gene knockout, gene overexpression, and RNA interference are three of the major methods used. All modify the capacities of cells to produce specific proteins.

Proteomics

Proteomics is the study of the proteins being synthesized by cells and tissues. The term implies the simultaneous study of large numbers of proteins, even to the extent of screening all proteins that can be detected. One reason for the study of proteomics as a separate discipline is that the proteins coded by many genes are unknown, and therefore biologists cannot *predict* the full list of proteins in a cell from knowledge of gene transcription alone. A second reason for studying proteomics as a separate discipline is that, even if the proteins coded by mRNAs are known and the mRNAs have been quantified, protein concentrations cannot necessarily be predicted because protein synthesis is sometimes only loosely correlated with mRNA synthesis. Cell proteins constitute a part of the cell phenotype. Proteomics is thus a branch of the study of the *biochemical phenotype*.

The set of proteins assayed in a proteomic study can be narrowed down by a variety of methods. For example, a particular proteomic study might examine just the set of proteins that bind ATP—such as ATPase enzymes—or just the set of phosphorylated protein kinases (types of regulatory proteins; see Chapter 2).

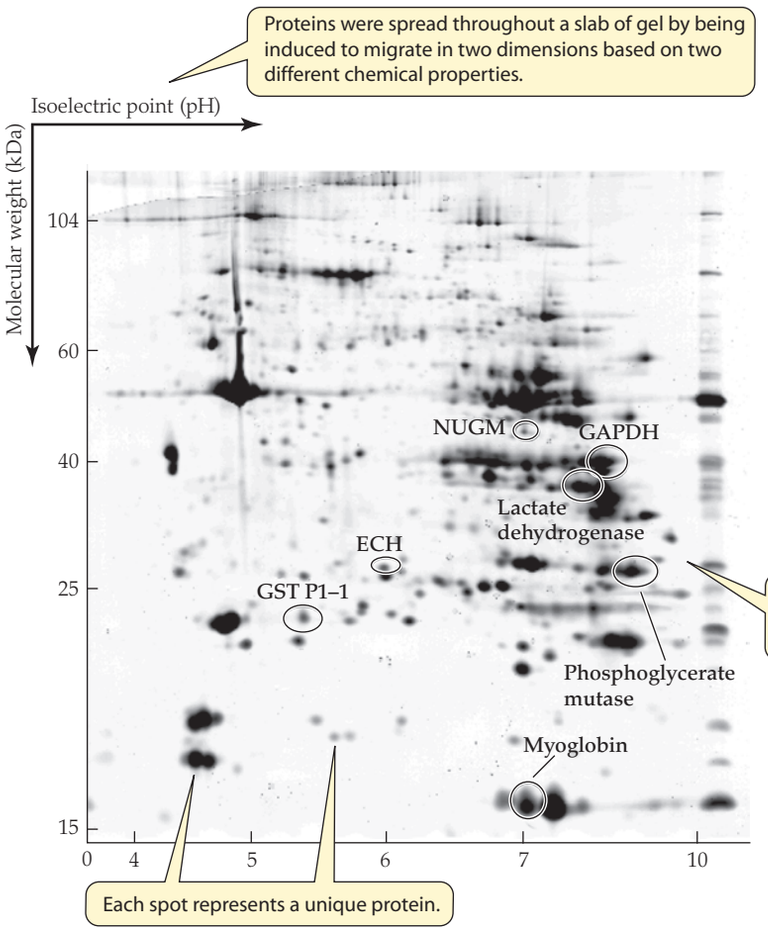
Two-dimensional (2D) gel electrophoresis is a particularly common proteomics method. In this method, the proteins in a

tissue sample, after being extracted, are first forced to migrate linearly through a gel material that separates them according to their isoelectric points, and thereafter the proteins are forced to migrate at a right angle to the first migration through a second gel material that separates them according to their molecular weights. In this way, the proteins in the original mixture are spread out in two dimensions—based on two different chemical properties—as seen in **Figure 3.12**. The proteins are then chemically identified by, first, excising spots of interest from the gel and, second, analyzing the spots by mass spectrometry or another analytical method.

Figure 3.12*a* is a gel from a proteomic study carried out to understand how Tibetan Sherpas are able to climb with great endurance—often carrying heavy loads—at altitudes above 5500 m (18,000 ft) in the Himalaya Mountains. Tissue was taken by biopsy from a thigh muscle of six Tibetans who had spent their entire lives at altitudes of at least 3500-4500 m and from six Nepal natives who had lived at 1300 m. Proteins in the muscle samples were then subjected to 2D electrophoresis. Seven of the proteins—those circled in Figure 3.12*a*—proved to be different in concentration in the high- and low-altitude people. One of these, myoglobin, was hardly a surprise; the high-altitude people had relatively high levels of myoglobin, which as we earlier discussed, helps O_2 diffuse into muscle cells. The discovery of very high levels of another protein, the enzyme glutathione-S-transferase (GST P1-1), in the high-altitude people was a surprise, however. Investigators now must learn what role this protein plays at altitude; possibly it helps defend against oxidative damage that can cause muscle deterioration at high elevations. When biologists first started to study high-altitude physiology, they focused on breathing and blood circulation. Later, they realized that tissue biochemistry is also of great importance, as discussed in Box 23.1. Proteomics now promises to expand rapidly our understanding of which tissue proteins are important.

A proteomic study of host-behavior modification by a parasite provides a second example. A hairworm that parasitizes certain

(*a*) A gel employed for protein identification



(*b*) A modern-day Sherpa carrying a heavy load at high altitude

FIGURE 3.12 Proteomics: A study of tissue protein response to life at high altitude (*a*) For a proteomic study of muscle proteins in high-altitude human populations, tiny (15-mg) samples of muscle tissue were cut by biopsy from the *vastus lateralis* muscles in the thighs of volunteers and subjected to two-dimensional gel electrophoresis. This gel resulted from the processing of one sample. The proteins in the sample were induced to migrate both from left to right and from top to bottom. When the proteins migrated from left to right, they were separated by isoelectric point (measured in pH units). When they migrated from top to bottom, they were separated by molecular weight (measured in kilodaltons, kDa). The proteins were in solution and invisible during migration. Afterward, however, they were visualized in the gel by staining, as seen here. The seven circled proteins were found to be present in significantly different concentrations in high- and low-altitude people. (*b*) An individual typical of the two human populations studied. ECH = Δ^2-enoyl-CoA-hydratase; GAPDH = glyceraldehyde-3-phosphate dehydrogenase; GST P1-1 = glutathione-S-transferase; NUGM = NADH-ubiquinone oxidoreductase. (*a* after Gelfi et al. 2004.)

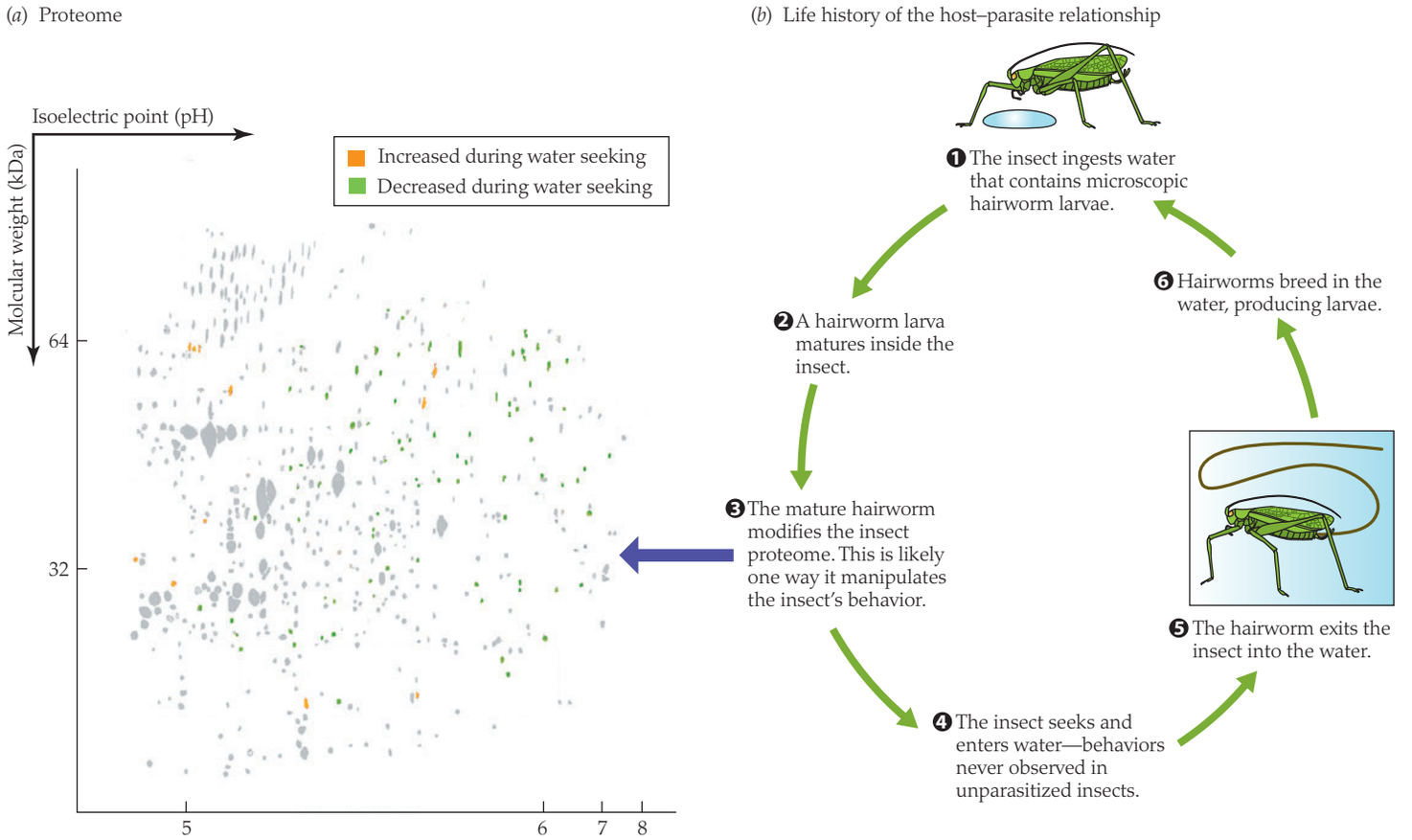


FIGURE 3.13 Proteomics: Parasites sometimes alter the protein profile of the central nervous system in their hosts (*a*) Proteins in the brains of katydids (*Meconema thalassinum*) were separated in two dimensions by the methods described in the caption of Figure 3.12*a*. Unlike in Figure 3.12*a*, however, the image seen here is not just that of a single gel. Instead, this image is a computer-generated synthesis of evidence from multiple gels, some from control katydids and some from katydids that were in the act of seeking water. Orange-colored proteins were at elevated concentrations during water-seeking behavior. Green-colored proteins were at reduced concentrations. Water-seeking occurs only when katydids are parasitized by hairworms (*Spinochordodes tellinii*), which manipulate the katydids' behavior in ways advantageous to the parasites. Although the insects are called katydids in the United States, they are more often called grasshoppers in Europe, where this study was done. (*b*) The life history of the parasite-host relationship. (*a* after Biron et al. 2005.)

katydids is known to manipulate the behavior of the insects, so that—entirely contrary to their normal behavior—the katydids jump into bodies of water at night. This strange behavior is induced when the hairworm within a katydid has grown to full size, and the behavior enables the hairworm to exit its host into water. Investigators reasoned that the parasite might exert its effect on the behavior of its host, in part, by increasing or decreasing the levels of brain proteins. They also reasoned that the host might upregulate defensive proteins. The proteomic study confirmed that many proteins are increased or decreased in the brain of a parasitized katydid at the time it bizarrely seeks water and jumps in. All the proteins colored orange or green in **Figure 3.13** are significantly altered. The next step will be to determine the roles the protein changes are playing.

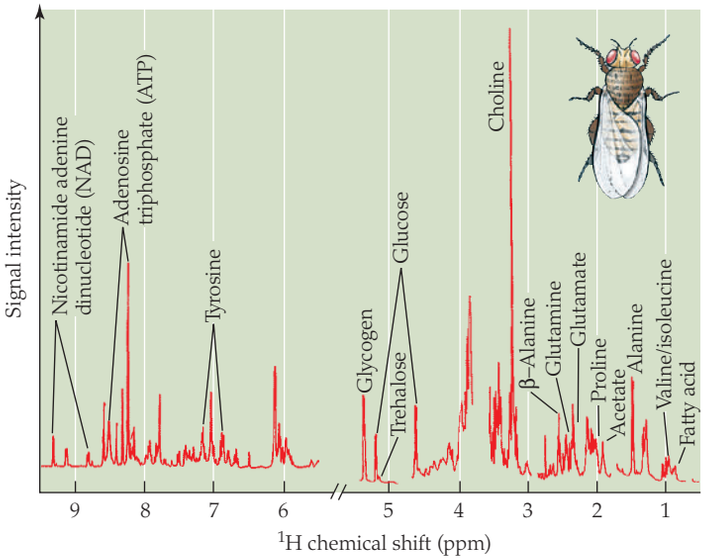
Metabolomics

Metabolomics parallels and complements proteomics. Like proteomics, metabolomics seeks to describe the biochemical phenotype. Unlike proteomics, metabolomics does not focus on gene products.

Metabolomics, in a few words, is the study of all the organic compounds in cells and tissues other than macromolecules coded by the genome. The compounds encompassed by metabolomics are generally of relatively low molecular weight (roughly <1500 daltons). They include many types of small molecules found in cells and tissues, such as sugars, amino acids, and fatty acids. Most of these molecules are metabolites, that is, compounds currently being processed by metabolism. A central goal of metabolomics is to clarify the metabolic pathways operative in cells and the ways the pathways are modulated. A shortcoming of metabolomics is that when tissues are extracted for a study, compounds from various cell compartments typically become mixed, meaning that additional research is required to clarify whether the compounds measured were together, in certain cell compartments, in the native cell.

To illustrate metabolomics, let's consider a study of fruit flies subjected to heat stress (**Figure 3.14**). Some flies were exposed for 1 h to a heat-stress temperature of 38°C. Their tissues and the tissues of control flies were then extracted with solvents and analyzed by nuclear magnetic resonance (NMR) spectroscopy, a method capable of detecting and quantifying a great diversity of compounds. Nu-

(*a*) The NMR spectrum obtained from analysis of a single fruit fly



(*b*) Average concentrations of 10 metabolites that were compared in control and heat-stressed flies.

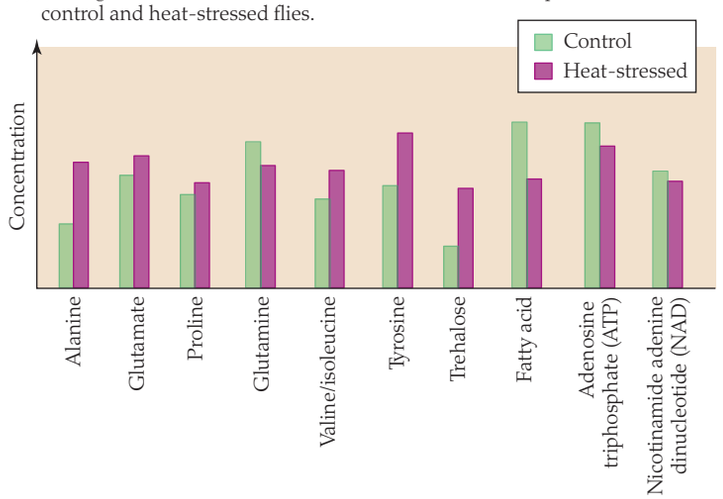


FIGURE 3.14 Metabolomics of heat stress Fruit flies (*Drosophila melanogaster*) were exposed to 38°C—a stressfully high temperature—for 1 h and compared with control flies that lived always at 25°C. Whole flies from both groups were analyzed immediately after the heat-stress period. (*a*) The NMR spectrum obtained from a single fly. The signatures of multiple metabolites are readily seen, and the amplitudes of the signatures can be used to calculate metabolite concentrations. The portion of the spectrum at 5.5 and higher was recorded at greater amplification than the portion at 5.4 and lower. (*b*) The average concentrations in control and heat-stressed flies of 10 metabolites that exhibited statistically significant changes in heat-stressed flies as compared with control flies, based on the NMR data. The concentrations are in arbitrary units and should be compared only *within* metabolites because different metabolites are scaled differently. The metabolite labeled "fatty acid" was identified only in general terms, as being a fatty acid–like compound. (After Malmendal et al. 2006.)

merous metabolites could be observed—because of their distinctive signatures—in the NMR spectrum obtained from each fly (see Figure 3.14*a*). Researchers therefore could paint a very broad picture of metabolite changes in the cells of the flies (see Figure 3.14*b*). The results obtained have been used to generate new hypotheses that will need to be tested in future research. For instance, the rise in alanine (see Figure 3.14*b*) in the heat-stressed flies suggests that flies might depend to an increased extent on anaerobic catabolism when they are under heat stress, because alanine is a product of the anaerobic pathways in flies. From the rise in tyrosine, one might hypothesize that synthesis of several hormones is accelerated during heat stress, because tyrosine is involved in the synthetic pathways.

The metabolomic approach differs from traditional metabolite studies in that traditionally just one, two, or three compounds would typically have been studied and the others disregarded. In a metabolomics study, however, as many compounds as possible are measured, so effects are studied without preconceived limits. A metabolomics study often leads to hypotheses for future research rather than to definitive conclusions. In part, this is true precisely because metabolomics measures any and all effects that are detectable, including ones for which there may be no existing background knowledge. All the areas of *-omics* biology discussed in this chapter are characterized by similar assets and limits.

Study Questions

1. A case can be made that the sedentary way of life characteristic of many people today is a very recent development in human history, dating from at most a century or so ago. According to this line of argument, physical exertion was a regular part of daily life for most people during most of human evolution. From this point of view, in what way do recent gene-transcription data suggest that physical exertion is in fact probably essential for full health? (For reading on this question, see the article by Booth and Neufer in the References.)
2. Geneticists can synthesize any gene—any stretch of DNA—desired. Suppose you use a tree of genome evolution to predict the structure of a now nonexistent, ancient gene. What insights might you obtain by synthesizing the ancient gene and inserting it into a living animal?
3. The solubility of O_2 in aqueous solutions—such as seawater or blood plasma—increases as temperature decreases. Moreover, the rates at which tissues metabolically consume O_2 tend to decrease as temperature decreases. With these facts in mind, explain why—in the world's oceans—polar seas are the most likely place where fish could survive a mutation that deletes their ability to synthesize blood hemoglobin.
4. Suppose you carry out genomic, transcriptomic, and proteomic studies on a single tissue. What type of information would you obtain from each sort of study? For understanding physiology, what are the uses and shortcomings of each type of information?
5. Compare and contrast the top-down and bottom-up approaches to the study of physiology.
6. One of the most famous quotes in the history of inquiry is Francis Bacon's statement: "The truth emerges more readily from error than confusion." What did he mean? High-throughput methods are sometimes accused of producing such an information overload that they generate confusion. Carefully consider the pros and cons of high-throughput methods as viewed from Bacon's perspective.
7. We tend to think that each expressed gene has a particular effect on an animal's phenotype. Why, therefore, if a gene is knocked out by genetic engineering methods, is the consequent change in an animal's phenotype not necessarily a direct reflection of the gene's effect? For example, how is it possible to knock out a gene that in fact has a phenotypic effect and yet not, in some cases, see

any measurable change in an animal's overall ability to function? How might your answer to Study Question 2 be affected by the considerations raised here?

8. The icefish alive today that lack ventricular myoglobin are often cited as examples of "natural knockout animals" because they lack functional genes for myoglobin synthesis. Suppose geneticists engineer some ordinary, modern-day fish to lack functional genes for myoglobin synthesis. Compare the icefish—the "natural knockout animals"—with the first generation of engineered knockout animals. Explain how these two groups of animals would be likely to differ in the mechanisms available for compensating for their lack of myoglobin.

9. Suppose you want to determine if excessive lipid ingestion alters gene transcription. Describe and explain all the steps you would carry out to do a microarray-based study of the question. Include your choice of subjects and tissues to be investigated. (The statement of the problem to be studied—effects of "excessive lipid ingestion"—is deliberately vague, and you should refine it by specifying and explaining an operational procedure.)

10. Suppose you want to determine if dehydration changes the types or amounts of enzymes. Describe and explain all the steps you would carry out to do a two-dimensional gel study of the question. As in Study Question 9, refine the experimental objectives, and include your choice of subjects and tissues.

Go to **sites.sinauer.com/animalphys3e** for box extensions, quizzes, flashcards, and other resources.

References

Booth, F. W., and P. D. Neufer. 2005. Exercise controls gene expression. *Am. Sci.* 93: 28–35.

Cheng, C.-H. C., and H. W. Detrich III. 2007. Molecular ecophysiology of Antarctic notothenioid fishes. *Philos. Trans. R. Soc. London, B* 362: 2215–2232.

Gibson, G., and S. V. Muse. 2009. *A Primer of Genome Science*, 3rd ed. Sinauer Associates, Sunderland, MA. A widely applauded introduction to not just genomics but also related topics such as proteomics and bioinformatics. This book emphasizes practical techniques (computer-based and lab-based) that are used in these fields as well as the theoretical perspectives that drive research in the fields. The book is a good place to seek information on methods.

Hardison, R. 1999. The evolution of hemoglobin. *Am. Sci.* 87: 126–137. An accessible treatment of research on the evolution of a protein, emphasizing a genomic perspective.

Hollywood, K., D. R. Brison, and R. Goodacre. 2006. Metabolomics: Current technologies and future trends. *Proteomics* 6: 4716–4723.

Jarvis, E. D., S. Ribeiro, M. L. da Silva, D. Ventura, J. Vielliard, and C. V. Mello. 2000. Behaviourally driven gene expression reveals song nuclei in hummingbird brain. *Nature* 406: 628–632.

Lockhart, D. J., and E. A. Winzeler. 2000. Genomics, gene expression and DNA arrays. *Nature* 405: 827–836.

Miller, K. M., and 14 additional authors. 2011. Genomic signatures predict migration and spawning failure in wild Canadian salmon. *Science* 331: 214–216. A truly intriguing effort to use genomic methods to find out why a salmon population is failing.

Near, T. J., S. A. Parker, and H. W. Detrich, III. (2006) A genomic fossil reveals key steps in hemoglobin loss by the antarctic icefishes. *Mol. Biol. Evol.* 23: 2008–2016.

Nikinmaa, M. and D. Schlenk. 2009. Editorial. *Aquat. Toxicol.* 92: 113.

Palstra, A. P., D. Crespo, G. E. E. J. M. van den Thillart, and J. V. Planas. 2010. Saving energy to fuel exercise: swimming suppresses oocyte development and downregulates ovarian transcriptomic response of rainbow trout *Oncorhynchus mykiss*. *Am. J. Physiol.* 299: R486–R499. A wide-ranging study of the potential conflict between sexual maturation and long-distance swimming in migratory fish.

Peck, L. S., and 11 additional authors. 2005. Genomics: Applications to Antarctic ecosystems. *Polar Biol.* 28: 351–365. A wide-ranging discussion of genomic concepts and applications, which gains interest from its focus on a single natural system, the Antarctic.

Rund, S. S. C., T. Y. Hou, S. M. Ward, F. H. Collins, and G. E. Duffield. 2011. Genome-wide profiling of diel and circadian gene expression in the malaria vector *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U.S.A.* 108: E421–E430.

Schlieper, G., J.-H. Kim, A. Molojavyi, C. Jacoby, T. Laussmann, U. Flögel, A. Gödecke, and J. Schrader. 2004. Adaptation of the myoglobin knockout mouse to hypoxic stress. *Am. J. Physiol.* 286: R786–R792.

Sidell, B. D., and K. M. O'Brien. 2006. When bad things happen to good fish: The loss of hemoglobin and myoglobin expression in Antarctic icefishes. *J. Exp. Biol.* 209: 1791–1802.

Storey, K. B. 2006. Genomic and proteomic approaches in comparative biochemistry and physiology. *Physiol. Biochem. Zool.* 79: 324–332. A thought-provoking—sometimes opinionated—overview of what can be expected from the brave new world of gene/protein studies and how the new agenda will interrelate with (and sometimes commandeer) earlier agendas in the study of animal physiology. Although the paper seems awkward at times because it is focused on a particular symposium, it has the virtue that the writing is readily accessible.

Sukumaran, S., W. J. Jusko, D. C. DuBois, and R. R. Almon. 2011. Light-dark oscillations in the lung transcriptome: implications for lung homeostasis, repair, metabolism, disease, and drug action. *J. Appl. Physiol.* 110: 1732–1747.

Thornton, J. W. 2004. Resurrecting ancient genes: Experimental analysis of extinct molecules. *Nat. Rev. Genet*. 5: 366–375. A discussion of the startling but realistic proposition that genome science can be used to reconstruct the DNA base sequences of extinct genes, permitting the genes to be resurrected.

Todgham, A. E., and G. E. Hofmann. 2009. Transcriptomic response of sea urchin larvae *Strongylocentrotus purpuratus* to CO_2-driven seawater acidification. *J. Exp. Biol.* 212: 2579–2594.

See also **Additional References** *and Figure and Table Citations.*