in bacteria in response to toxic or nutrient-poor environments or antibiotic exposure. In bacteria, gene amplifications may be mediated by transposable-element activity (Chapter 9). Human health may be negatively impacted by amplifications occurring in pathogens. One such example concerns the virulence-correlated amplification of the cholera toxin gene (*ctx*) in epidemic strains of *Vibrio cholerae* (Mekalanos 1983).

Other cases in which selection for increased dosage has been invoked as an evolutionary explanation include the genes specifying rRNAs and tRNAs, which are required for translation (see below), and the genes for histones, which constitute the main protein component of chromosomes and therefore must be synthesized in large quantities, especially during the S phase of the cell cycle, when DNA is replicated.

One prediction of the dosage-compensation model is that following gene duplication, genes for proteins known to form multicomponent complexes or to interact in any other way with other proteins will be preferentially retained in the genome. Some evidence exists that this is indeed the case (Aury et al. 2006; Hughes et al. 2007).

We note that under the increased-dosage model of evolution, the invariant duplicated copies are created rapidly because of stringent positive selection. Under the dosage-compensation model, on the other hand, the invariant copies are actively maintained invariant by stringent selective constraints and purifying selection. Intersequence invariance may also be maintained by two processes that will be dealt with later in this chapter, concerted evolution or birth-and-death evolution

In addition to dosage, Ohno (1970) proposed a second reason for the maintenance of invariant repeats. According to this proposal, a second gene can provide **functional redundancy** in case the original gene is subsequently disabled by mutation. In other words, a second copy can increase **mutational robustness**, i.e., the extent to which a function or a phenotype will remain unchanged in the face of mutations. Many researchers have addressed this model and shown that it is not likely to have played a large role in evolution except in populations with extremely large effective population sizes and very high deleterious mutation rates (Clark 1994; Lynch et al. 2001; O'Hely 2006; Hahn 2009; Price et al. 2011).

Evolution of rRNA-specifying genes

Ribosomal RNA gene copy number varies widely among organisms. Prokopowich et al. (2003) showed that a very strong correlation exists between the number of rRNA-specifying genes and genome size (Figure 7.7). This relationship with genome size also holds for the tRNA genes. There may be three reasons for the general positive correlation between genome size and number of copies of RNA-specifying genes. First, large genomes may require large quantities of RNA and the quantity of rRNA may be correlated positively to the number of rRNA-specifying genes. Second, the number of RNA-specifying genes may simply be a passive consequence of genome enlargement by duplication. A third intriguing possibility is that the large number of almost identical copies may be maintained for reasons unrelated to rRNA production. For example, yeast has many copies of rRNA-specifying genes that are not transcribed. Interestingly, when these "extra copies" are lost or deleted, the cells become sensitive to DNA damage induced by mutagens. Apparently, a high density of heavily transcribed genes is toxic to the cells, so in yeast, a gene amplification system has evolved for maintaining large clusters of tandemly repeated copies of mostly untranscribed rRNA-specifying genes (Ide at al. 2010).

As a rule, duplicate copies of rRNA-specifying genes tend to be very similar to one another.

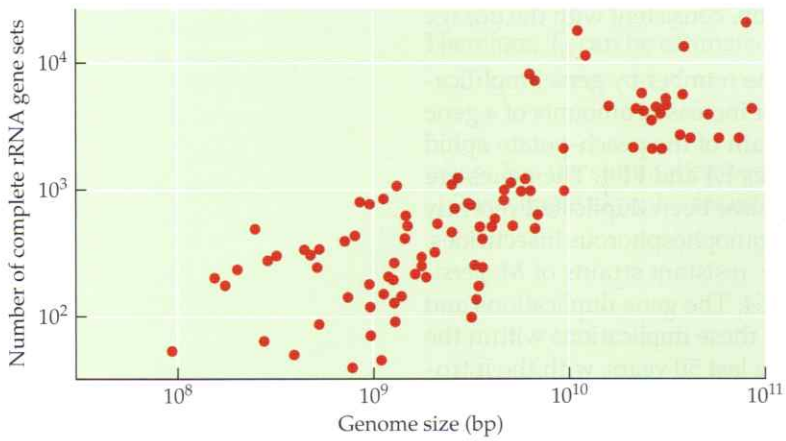


Figure 7.7 Log-log plot illustrating the relationship between genome size and number of complete rRNA gene sets in the genomes of 94 animal species. (Modified from Prokopowich et al. 2003.)

One factor responsible for the homogeneity may be purifying selection, because these genes are required to abide by very specific functional and structural requirements. However, homogeneity often extends to regions devoid of any functional or structural significance, and thus the maintenance of homogeneity requires that other mechanisms be invoked.

Neofunctionalization

Following gene duplication, one of the paralogous genes may undergo neofunctionalization, i.e., it may acquire a new function. In other words, one of the duplicates retains its original function, while the other accumulates molecular changes such that, in time, it can perform a different task (Ohno 1970). Neofunctionalization has also been called "mutation during redundancy" (Hughes 1999). We note, however, that the process of straightforward neofunctionalization, whereby a copy is freed from selective constraint and by chance acquires a different function that is then improved by natural selection, may be extremely infrequent in evolution because (1) functional changes usually require a large number of amino acid replacements, and (2) it is infinitely easier to destroy the function of a gene by mutation than to confer on it a new function by chance. Thus, neofunctionalization should be a minority mechanism as far as the retention of duplicated genes is concerned. The degree to which neofunctionalization can "win out" over nonfunctionalization depends on (1) the number of mutations required to attain the new function and (2) the possible paths within the sequence space to the new function. If the change in function can be achieved by a single amino acid replacement, such as the serine-to-proline replacement that converts the enzyme guanylate kinase into a mitotic spindle-orienting protein (Johnston et al. 2011), then the chances of the change becoming fixed in the population seem considerable. In cases where more than a single change is required, the probability of neofunctionalization decreases rapidly as the number of changes increases. An additional means by which neofunctionalization may be achieved is a single mutation that has big consequences. In the reading frame of protein-coding genes, frameshift mutations resulting from the deletion or insertion of DNA sequences that are not multiples of three nucleotides will have a huge effect on the amino acid sequence. The probability that such a mutation will have a beneficial effect and be fixed in the population is, of course, minuscule. Notwithstanding these almost insurmountable difficulties, neofunctionalization by frameshifting of reading frames does occur in nature (e.g., Okamura et al. 2006; Liu and Adams 2010).

Neofunctionalization may occur through either changes in the coding region or changes in the regulatory elements. Neofunctionalization may refer to minor changes in function, such as quantitative changes in expression pattern and timing, or it may refer to major changes in function, such as changes in substrate and ligand affinity, subunit structure, or subcellular compartmentalization. Some neofunctionalization processes may even turn an enzyme into a structural protein or vice versa.

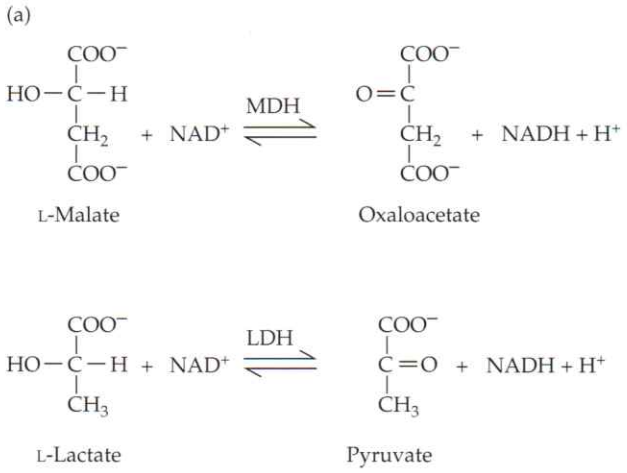
What are the telltale signs of neofunctionalization in the case of paralogous genes? First, the expectation is that the rates and patterns of evolution of the two duplicate genes will be highly asymmetrical. Second, we expect the ratio of nonsynonymous to synonymous substitution to differ between the two copies. Finally, we expect the faster-evolving paralog to have a higher ratio of nonsynonymous to synonymous substitutions than the other paralog or singleton orthologs from other species. In other words, we expect the fast-evolving paralog to exhibit evidence of an increase in positive Darwinian selection. Comparative studies indicate that neofunctionalization may exert a significant role in the evolution of about 10% of newly duplicated paralogs in primates (Han et al. 2009) as well as in the maintenance of about 6% of the retained genes following the tetraploidization event that gave rise to *Xenopus laevis* (Chain and Evans 2006). In maize, approximately 13% of all the gene pairs that have been created by genome duplication exhibit regulatory neofunctionalization, i.e., are expressed differently in leaves (Hughes et al. 2014). As far as human-specific duplicates are concerned, neofunctionalization may have

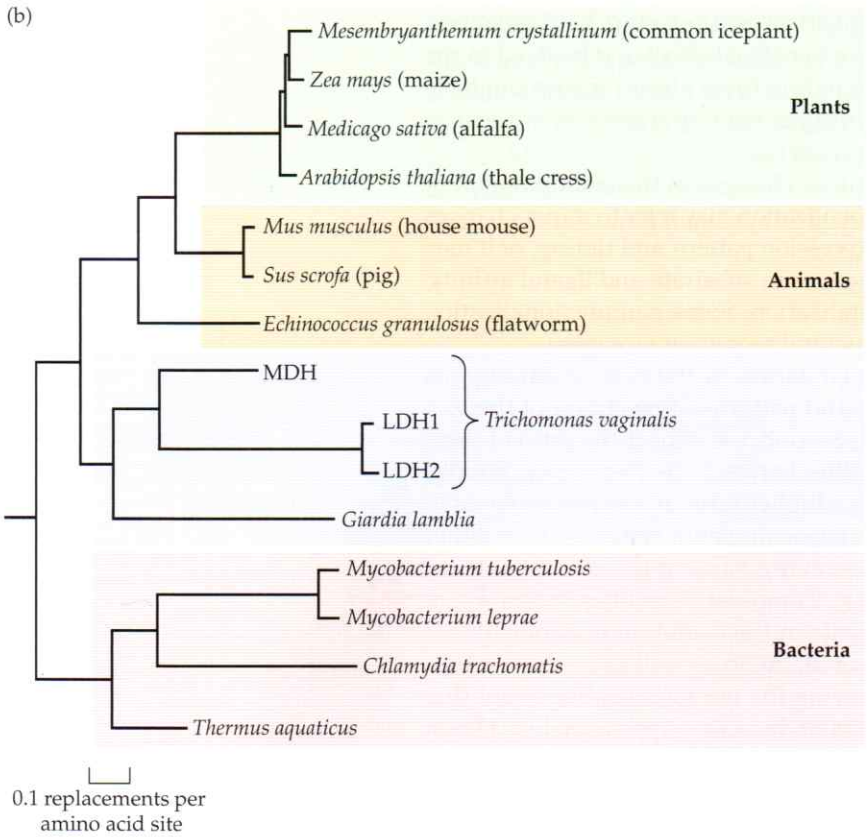
had an important role in the evolution of genes involved in neuronal and cognitive functions.

In *Drosophila*, neofunctionalization seems to account for the retention of almost two-thirds of duplicate genes (Assis and Bachtrog 2013). Because most gene duplications in *Drosophila* result in the creation of retrogenes, i.e., occur via reverse transcription of mRNA (Chapter 9), it is possible to distinguish between the **parent copy** and the **child** or **daughter copy**. (Such a distinction is not possible if the gene duplication event creates two tandemly arranged copies.) Surprisingly, novel functions nearly always originate in child copies, whereas the function of parent copies persists unaltered. Many young child copies are expressed primarily in the testes, which may be a reflection of the transcriptionally permissive environment in the testes rather than a testament to their functionality (Chapter 8). Nevertheless, the expression breadth, i.e., the number of tissues in which child copies are transcribed, increases over evolutionary time. This finding supports the so-called **out-of-testes hypothesis** (Kaessmann 2010), which posits that testes are a catalyst for the emergence of new genes that ultimately evolve functions in other tissues.

The process of neofunctionalization can be sped up considerably by very strong positive selection, such as that exerted by insecticides. This is particularly true if the acquisition of the novel function can be achieved by very few mutations. One such example was discovered in the sheep blowfly (*Lucilia cuprina*), where a single glycine-to-aspartic acid replacement conferred insecticide resistance by turning a carboxylesterase into an organophosphorous hydrolase (Newcomb et al. 1997).

Here we illustrate the process of neofunctionalization through an example involving a change in substrate specificity of an enzyme.





THE NEOFUNCTIONALIZATION OF MALATE DEHYDROGENASE INTO LACTATE DEHYDROGENASE FOLLOWING GENE DUPLICATION An interesting example of neofunctionalization involves the emergence of a lactate dehydrogenase enzyme following a duplication involving a gene encoding malate dehydrogenase in *Trichomonas vaginalis*, the causative agent of the most common pathogenic protozoan infection of the urogenital tract in industrialized countries (**Figure 7.8**). Malate dehydrogenase catalyzes the interconversion of oxaloacetate and malic acid with the concomitant interconversion of NADH and NAD^+. Lactate dehydrogenase, on the other hand, catalyzes a similar but distinct reaction, the inter-

Figure 7.8 (a) The reactions catalyzed by malate dehydrogenase (MDH) and lactate dehydrogenase (LDH). (b) Phylogenetic relationships among malate dehydrogenases from animals, plants, unicellular eukaryotes, and bacteria showing the nested position of the lactate dehydrogenase of *Trichomonas vaginalis* within the malate dehydrogenase family.

conversion of pyruvate and lactic acid with the same concomitant interconversion of NADH and NAD^+. A genomic analysis of *T. vaginalis* revealed the existence of two very similar copies of the lactate dehydrogenase gene, each encoding a 333-amino-acid-long enzyme. Surprisingly the two genes, which are clearly the result of a gene duplication, turned out to be closely related to the malate dehydrogenase gene family rather than to the lactate dehydrogenase family. It seems that a recent gene duplication of the original malate dehydrogenase gene created two copies of the gene, one of which was neofunctionalized into a lactate dehydrogenase gene, most probably through a nucleotide substitution that resulted in a leucine-to-arginine replacement at amino acid position 91. Subsequently, the lactate dehydrogenase gene was duplicated again, resulting in two genes, *LDH1* and *LDH2*. The neofunctionalization of malate dehydrogenase to lactate dehydrogenase seems to be quite common in nature, as it has also been recorded in other organisms, such as *Bacillus stearothermophilus* (Wilks et al. 1988).

Multifunctionality and subfunctionalization

During the pre-double helix era (i.e., from the 1910s to the early 1950s) a group of scientists put forward the idea, very controversial at the time, that genes are functionally divisible entities, i.e., that individual genes have not one but a set of functions, that each function is performed by an independently mutable region at the locus, and that each function is separable from the other functions (Emerson 1911; Dubinin 1929; Serebrovsky and Dubinin 1929; Agol 1930; Serebrovsky 1930; Muller 1932; Raffel and Muller 1940; Verderosa and Muller 1954).

Multifunctional proteins, as their name implies, perform more than one function. Multifunctional proteins were thought at first to be exotically rare, but recent evolutionary, biochemical, and genomic studies have revealed that protein multifunctionality is the norm in nature rather than the exception (Khersonsky and Tawfik 2010). **Moonlighting proteins** constitute a subset of multifunctional proteins in which the two or more functions cannot be ascribed to the fusion of protein-coding genes with distinct functions (Copley 2012). In addition, multifunctionality due to mRNA alternative splicing or proteolysis resulting in independently functional proteins is excluded from the definition of "moonlighting." (The evolution of non-moonlighting multifunctional proteins will be discussed in Chapter 8.)

The first examples of moonlighting proteins were recognized in the late 1980s. For example, neuroleukin, a protein found in human skeletal muscles, brain, and bone marrow, which prolongs the life of embryonic nerve cells, was found to function also as phosphoglucose isomerase. Crystallins in the lenses of vertebrates were shown to be identical to various metabolic enzymes, and glyceraldehyde 3-phosphate dehydrogenase in several *Streptococcus* species was found to function as a receptor for fibronectin on the bacterial surface, suggesting that it might play a role in colonization of the pharynx. Moonlighting is not only a fairly common phenomenon (e.g., Hittinger and Carroll 2007; Des Marais and Rausher 2008); it also blurs the traditional distinctions between different types of proteins, such as the difference between enzymes and structural proteins.

Subfunctionalization can most broadly be defined as the partitioning of the set of ancestral functions into different subsets among the duplicated descendants (Ohno 1970; Hahn 2009), and it is predicated on the ancestor gene producing a multifunctional product. Subfunctionalization provides an appealing explanation for the ubiquity of duplicated genes in eukaryote genomes, because the model does not require each duplication event to immediately confer a selective advantage. Because our knowledge of the functions of the ancestral gene is limited, however, it is often difficult to pinpoint different subfunctions of the ancestral gene that have been partitioned among its duplicated descendants.

Many scenarios of subfunctionalization have been described in the literature (Innan and Kondrashov 2010). Here, we will discuss several such examples: subfunctionalization of non-moonlighting multifunctional genes, specialization, gene sharing

and escape from adaptive conflict, duplication-degeneration-complementation, and segregation avoidance. We note that the different subfunctionalization scenarios are not mutually exclusive.

SUBFUNCTIONALIZATION OF NON-MOONLIGHTING MULTIFUNCTIONAL GENES Cusack and Wolfe (2007) studied a gene in which the ancestral functions being partitioned among the daughter genes are distinct and readily identifiable because they have nothing in common with each other except the fact that two products of the gene are exported to the chloroplast.

In most flowering plants, the gene for plastid ribosomal protein L32 (*RPL32*) is located in the chloroplast genome. In the lineage leading to the order Malpighiales (a large, heterogeneous plant taxon consisting of such diverse species as poplars, mangroves, cassava, poinsettia, flax, passion fruit, willows, violets, and pansies), however, the plastid gene was transferred to the nuclear genome after the divergence of the lineage from the orders Cucurbitales (which includes the gourd and begonia families) and Fabales (legumes). In these taxa, plastid ribosomal protein L32 is produced in the nucleus and later inported into the chloroplast.

As seen in **Figure 7.9**, in the Burma mangrove (*Bruguiera gymnorrhiza*), which represents the ancestral state, the one-exon *RPL32* gene was inserted in the intron separating exons 7 and 8 of the *SOD* gene. The resulting chimeric gene is alternatively spliced, producing one transcript for *SOD* and one for *RPL32*. In the lineage leading to the poplars (represented in the study by the western balsam poplar, *Populus trichocarpa*, and the silver-leaf poplar, *P. alba*), the *SOD-RPL32* chimeric gene was duplicated twice. The first duplication resulted in one copy that lost the *RPL32* exon and one copy that subsequently lost the ability to encode SOD. Thus, the first duplication resulted in the subfunctionalization of the chimeric gene, producing daughter genes that produce either RPL32 or SOD, but not both.

The RPL32-coding gene (*Poplar1*) has retained a continuous open reading frame derived from the former *SOD* gene. However, the product cannot be a functional SOD protein, as two of its exons have been affected by internal deletions and three exons have disappeared altogether (one was pseudogenized

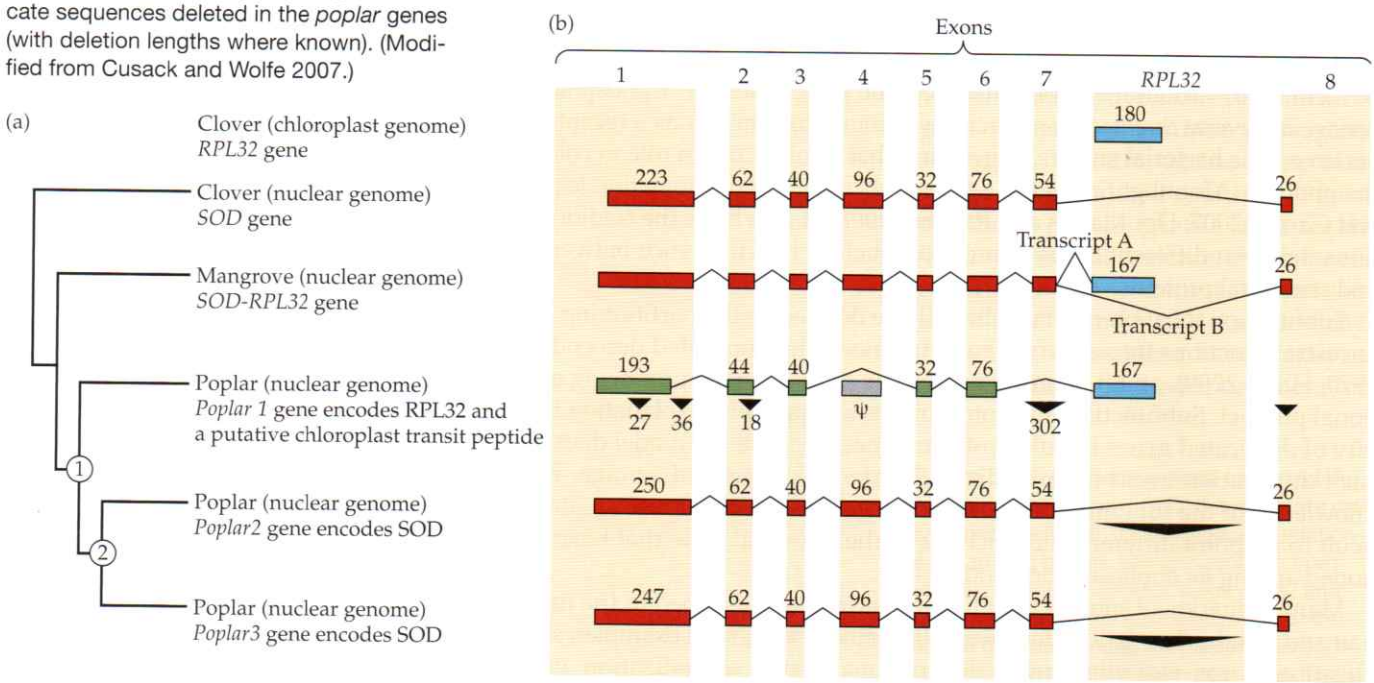


Figure 7.9 Evolution by subfunctionalization in the plant order Malpighiales. (a) The branching order of the nuclear genes is based on pairwise synonymous nucleotide substitution distances. Node 1 represents the first gene duplication in poplar, resulting in *Poplar1* and *Poplar2*. Node 2 is thought to represent a large segmental (or whole-genome) duplication in poplar, because many of the genes neighboring *Poplar2* have homologs neighboring *Poplar3*. (b) Organization of the exons of chloroplast *RPL32* gene and the nuclear *SOD* gene in the barrel clover (*Medicago truncatula*), the *SOD-RPL32* chimeric gene in the Burma mangrove (*Bruguiera gymnorrhiza*), and the *SOD*- and RPL32-coding genes in two species of poplars (*Populus trichocarpa* and *P. alba*). Boxes represent exons and a pseudoexon; the lengths (in bp) of some exons are shown. Introns are not drawn to scale. Red boxes show SOD-coding exons; blue boxes show RPL32-coding exons. Nonframeshifted remnants of five SOD-coding exons are shown in green. The pseudoexon in *Poplar1* (ψ) is shown as a gray box. Triangles indicate sequences deleted in the *poplar* genes (with deletion lengths where known). (Modified from Cusack and Wolfe 2007.)

and two have been deleted). The greatly abbreviated protein is predicted to be a chloroplast transit peptide, i.e., it has retained the ability to transport to the chloroplast while losing its enzymatic activity. Thus, the evolutionary history of *Poplar1* illustrates a second case of subfunctionalization.

The SOD-coding gene later became duplicated a second time, producing two virtually identical genes (*Poplar2* and *Poplar3*). Analyses of gene expression have shown that all three poplar genes are transcribed and that none of them is alternatively spliced anymore. The partitioning process of *SOD-RPL32* in *Populus* can, therefore, be unambiguously categorized as subfunctionalization, because a complementary loss of subfunctions of the ancestral chimeric gene occurred in its first two descendant duplicates.

SPECIALIZATION As a result of specialization, the products of the duplicated genes, although they continue to perform essentially the same function, acquire properties that distinguish them from each other (Ohno 1970; Otto and Yong 2002). For example, while the ancestral gene may have performed its function in all tissues, developmental stages, and environmental conditions, its descendants become specialists, dividing among themselves the different tissues, developmental stages, and environmental conditions (Markert 1964; Ferris and Whitt 1979). Thus, one descendant copy may specialize in catalyzing a certain reaction in the larval liver, whereas the second copy may specialize in catalyzing the same reaction in the adult heart. The specialization model involves positive selection, because tissue, developmental, or any other type of specialization essentially involves the improvement of function.

Good examples of variant repeats created through the process of specialization are families of genes coding for isozymes, such as lactate dehydrogenase, aldolase, creatine kinase, carbonic anhydrase, and pyruvate kinase. **Isozymes** are enzymes that catalyze the same biochemical reaction but may differ from one another in tissue specificity, developmental regulation, electrophoretic mobility, or biochemical properties. Note that isozymes are encoded at different loci, generally by duplicated genes; this is in contrast to **allozymes**, which are distinct forms of the same enzyme encoded by different alleles at a single locus. The study of multilocus isozyme systems has greatly enhanced our understanding of how cells with identical genetic endowment can differentiate into hundreds of different specialized types of cells that constitute the complex body organization of vertebrates. Although all members of an isozyme family serve essentially the same catalytic function, different members may have evolved particular adaptations to different tissues or different developmental stages, thus enhancing the physiological fine-tuning of the cell.

Mammalian lactate dehydrogenase (LDH) is a tetrameric enzyme that catalyzes the reversible interconversion of pyruvate and lactate, a key step in glycolysis and other metabolic pathways. Six paralogous *LDH* genes have been found in the human genome: *LDH-A*, *LDH-C*, and *LDH-6A* on chromosome 11; *LDH-B* and *LDH-6C* on chromosome 12; and *LDH-6B* on chromosome 15. In comparison with the *LDH-A*, *LDH-B*, *LDH-C*, and *LDH-6A* genes, which each possess seven exons, the *LDH-6B* and *LDH-6C* genes are intronless, indicating that these latter two genes may have been created through retroposition (Chapter 9).

Judging by the taxonomic distribution of the various *LDH* genes, the duplication that gave rise to the A and B isozymes occurred in an early vertebrate ancestor subsequent to the divergence of tunicates and lampreys. The duplication that gave rise to the C subunit is much more recent, as the gene for the C subunit was only found in placental and marsupial mammals but is absent from birds and egg-laying monotremes (Stock et al. 1997; Holmes and Goldberg 2009).

The A and B subunits form five tetrameric isozymes, A_4, A_3B, A_2B_2, AB_3, and B_4, all of which catalyze either the conversion of lactate into pyruvate in the presence of the oxidized coenzyme nicotinamide adenine dinucleotide (NAD^+) or the reverse reaction in the presence of the reduced coenzyme (NADH). It has been shown that B_4 and the other isozymes rich in B subunits, which have a high affinity for NAD^+,

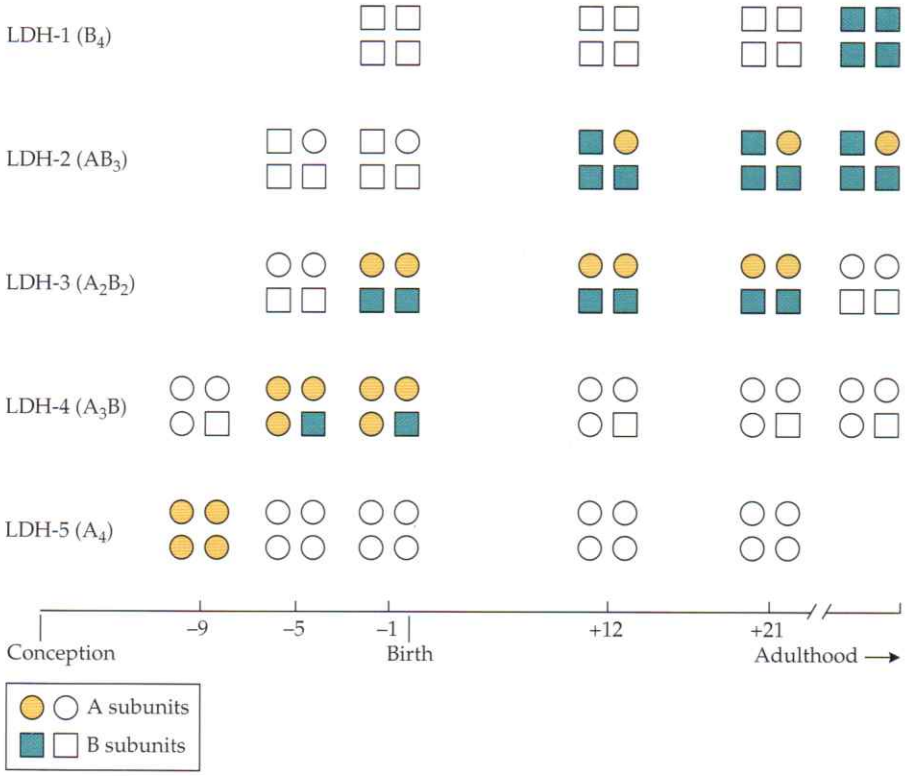


Figure 7.10 Developmental sequences of five lactate dehydrogenase (LDH) isozymes in the rat heart from conception to adulthood. Negative and positive numbers denote days before or after birth, respectively. Colored symbols indicate quantitatively predominant forms. Notice the progressive shift from a preponderance of A subunits (circles) before birth to B subunits (squares) afterward. (Data from Markert and Ursprung 1971.)

function as true lactate dehydrogenase in aerobically metabolizing tissues such as the heart, whereas A_4 and the isozymes rich in A subunits, which have a high affinity for NADH, are especially geared to serve as pyruvate reductases in anaerobically metabolizing tissues such as skeletal muscle. An interesting feature of LDH is that the two subunits can form heteromultimers, thus further increasing the physiological versatility of the enzyme.

Figure 7.10 shows the developmental sequence of LDH production in the heart. We see that the more anaerobic the heart is (specifically, in the early stages of gestation), the higher the proportion of LDH isozymes rich in A subunits will be. Thus, the two duplicate genes have become specialized by different tissues and to different developmental stages by subfunctionalization of expression.

GENE SHARING AND ESCAPE FROM ADAPTIVE CONFLICT A single-copy gene may acquire a new function in addition to its original function. If the two functions become essential for the organism, the gene will be subjected to two or more independent sets of evolutionary pressure. In some cases, the distinct functions may be incapable of further improvement through selection because of detrimental pleiotropic effects of one function on the other function. That is, the multiple functions of the protein cannot be simultaneously optimized by natural selection. This situation has been dubbed **adaptive conflict**. **Gene sharing** (Piatigorsky et al. 1988) refers to the case in which a single gene possesses two or more distinct functions that are not independently selectable.

Gene sharing was first discovered in crystallins, the major water-soluble proteins in the eye lens whose function is to maintain lens transparency and proper light diffraction (Wistow et al.1987; Wistow and Piatigorsky 1987). Crystallins perform their lenticular function by forming a uniform concentration gradient, with the highest protein concentration at the center of the lens. Since the structure and function of eye lenses are almost identical among all vertebrates, it was assumed that all vertebrate crystallins would be correspondingly similar. Comparative studies, however, have revealed an unexpected sequence diversity and taxon specificity among lens crystallins.

Many crystallins turn out to be moonlighting proteins (**Table 7.3**). For example, α-crystallins, which are present in all vertebrate lenses, also act as heat shock proteins and molecular chaperones protecting proteins against physiological stress. Crystallin ε from ducks and crocodiles turns out to be identical in its amino acid sequence to lactate dehydrogenase B and to possess identical enzymatic activity (Wistow et al. 1987). Subsequent work has shown that these "two" proteins are in fact one and the same, encoded by the same single-copy gene (Hendriks et al. 1988). Consistent with the idea that many lens crystallins have additional functions, they are also expressed outside of the lens.

Which function came first? Since the eye is a relatively recent evolutionary invention, whereas metabolic enzymes are ancient proteins, it is presumed that the enzymatic functions came first and the lenticular function later. We also note that enzymes are strictly constrained and fine-tuned by function (e.g., susbstrate recognition and proper catalysis), while crystallins need merely be hydrophobic and maintain a proper

Table 7.3

Examples of gene-sharing and paralogous relationships in taxon-specific eye lens crystallins

Crystallin type	Taxonomic distribution	Gene-sharing enzyme	Paralog
ε	Crocodiles, ducks	Lactate dehydrogenase B	—
ζ	Guinea pigs, degus, rock cavies, camels, llamas	NADPH:quinone oxidoreductase	—
η	Elephant shrews	Aldehyde dehydrogenase I	—
δ	Ducks, chickens	Argininosuccinate lyase	Argininosuccinate lyase
λ	Rabbits, hares	—	Hydroxyacyl CoA dehydrogenase
μ	Kangaroos, quolls	—	Ornithine cyclodeaminase
ρ	Frogs	—	NAPDH-dependent reductase

Source: Modified from Wistow (1993).

size. Thus, crystallins are "borrowed" or "recruited" proteins. As opposed to the case of the mammalian *LDH* genes discussed above, the recruitment of preexisting enzymes as crystallins illustrates a model of molecular evolution in which the evolution of new functions and new patterns of gene expression occur before (rather than after) gene duplication.

Adaptive conflict can be resolved through two main pathways (**Figure 7.11**). First, the conflict can be resolved by reversal, i.e., the loss of one function, usually the new function. This situation is illustrated by the loss of the δ-crystallin function in a bird, the chimney swift (*Chaetura pelagica*). Second, the conflict can be resolved by gene duplication and separation of function by subfunctionalization. This situation is illustrated by the gene duplication and subsequent separation of function between argininosuccinate lyase and δ-crystallin in chicken. Of the two tandemly arranged chicken δ-crystallin genes, δ*1* became specialized for lens expression and produces more than 95% of the lens δ-crystallin mRNA; by contrast, the δ*2* gene, which encodes the enzymatically active argininosuccinate lyase, produces most of the mRNA in nonlens tissues. Interestingly, an intermediate stage has also been found. In duck there are two almost identical δ-crystallin/argininosuccinate lyase genes, which have most probably been derived from a very recent gene duplication event (Piatigorsky 1998a,b). Thus following the emergence of gene sharing, a gene duplication has occurred, but it has not yet been accompanied by functional divergence.

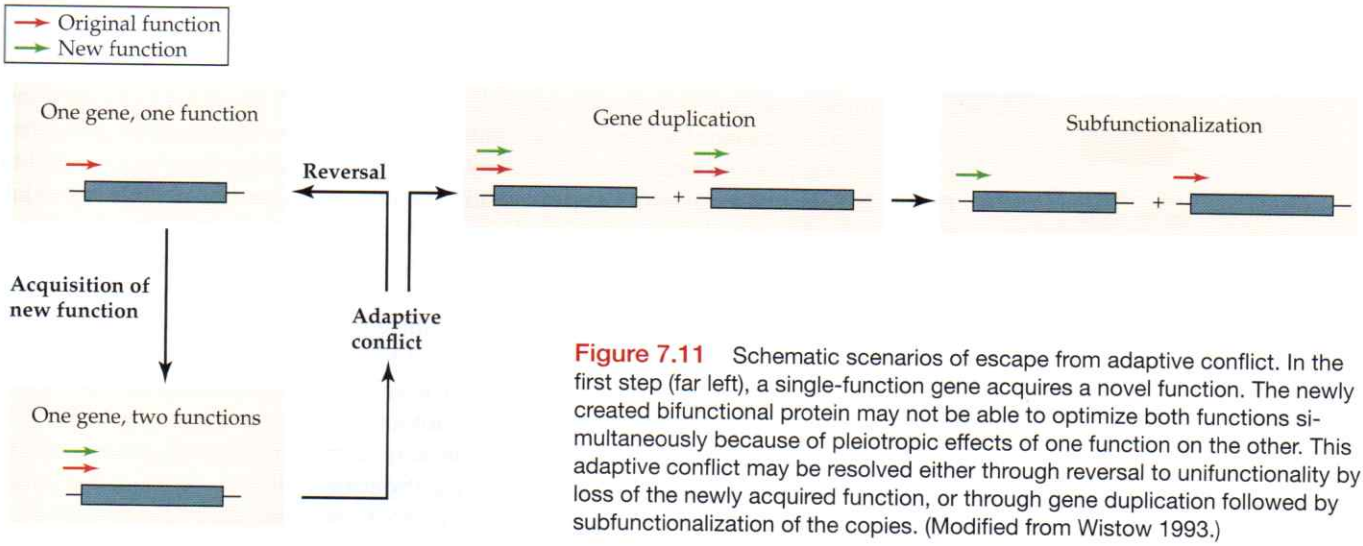


Figure 7.11 Schematic scenarios of escape from adaptive conflict. In the first step (far left), a single-function gene acquires a novel function. The newly created bifunctional protein may not be able to optimize both functions simultaneously because of pleiotropic effects of one function on the other. This adaptive conflict may be resolved either through reversal to unifunctionality by loss of the newly acquired function, or through gene duplication followed by subfunctionalization of the copies. (Modified from Wistow 1993.)

In recent years, the "escape from adaptive conflict" scenario has been validated empirically through the reconstruction and synthesis of ancestral enzymes. In one such case, the ancestral gene encoding a large family of fungal glucosidase genes has been reconstructed in the laboratory (Voordeckers et al. 2012). The original enzyme was primarly active on the disaccharide maltose, which consists of two glucose molecules linked by an $\alpha(1\rightarrow4)$ bond. The enzyme also exhibited some lesser activity on a related disaccharide, isomaltose, in which the two glucose molecules are linked by the slightly different $\alpha(1\rightarrow6)$ bond (Robinson 2012). Structural analyses and activity measurements on resurrected and present-day enzymes suggested that the two activities are incompatible, i.e., they cannot be both fully optimized concomitantly. The "escape from adaptive conflict" was possible through gene duplication followed by the accumulation of substitutions that in some copies optimized the affinity of the enzyme for maltose, while in other copies it optimized the affinity of the enzyme for isomaltose.

Additional evidence for this scenario comes from the crystallins. In humans, the single-gene encoded αB-crystallin performs a chaperone activity in addition to its lenticular function and is expressed in many tissues. In the zebrafish, on the other hand, αB-crystallin is expressed mostly in the lens and its chaperone-like activity is much reduced compared with that of its human ortholog. Interestingly, zebrafish has two nonidentical copies of genes encoding αB-crystallin-like proteins (Smith et al. 2006). The two copies have evolved through gene duplication, after which one paralog maintained its widespread chaperone role while the other adopted a more restricted, nonchaperone role in the lens.

It has been claimed that evolution of most new functions requires a period of gene sharing in which a single protein must serve both its original function and a new function that has become advantageous to the organism. Subsequent gene duplication and subfunctionalization allow one copy to maintain the original function, while the other diverges to optimize the new function. McLoughlin and Copley (2008) have provided some experimental evidence for this hypothesis.

There seems to be a pleasant historical subtext to the story of gene sharing and crystallin evolution, for it has contributed to the partial unraveling of a puzzle that has distressed students of evolution from Darwin onward. It concerns the evolution of the eye and the possibility of evolving a highly complex, totally novel character by natural means. In *The Origin of Species*, in a chapter entitled "Organs of Extreme Perfection and Complication," Charles Darwin wrote, "To suppose that the eye, with all its inimitable contrivances … could have been formed by natural selection, seems, I freely confess, absurd to the highest possible degree." At least as far as the lens and the cornea of the eye are concerned, it seems that their "inimitability" is not all that inimitable and that some of their constituent proteins can be traced back in evolutionary history to ubiquitous housekeeping enzymes that are found throughout all life-forms on Earth.

DUPLICATION-DEGENERATION-COMPLEMENTATION The duplication-degeneration-complementation process (Force et al. 1999; Lynch et al. 2001) refers to the post-duplication complementary loss of different subfunctions from duplicated genes, after which both copies are needed to maintain the original function. This is illustrated in **Figure 7.12**. In one version of the duplication-degeneration-complementation model, the two duplicate paralogs perform the same function but are expressed at lower levels than that of the original gene, so after duplication both copies are required to produce a sufficient amount of the gene product. Duplication-degeneration-complementation is a selectively neutral process, because the degeneration process merely reflects the accumulation of mutations that would have been deleterious in the original single-copy gene but are neutral because the subfunction destroyed in one copy is complemented by the corresponding intact subfuction in the second copy.

Here, we illustrate the duplication-degeneration-complementation process by using one example of gene duplication in yeast (Sommerhalter et al. 2004; van Hoof 2005). Ribonucleotide reductase is an enzyme that catalyzes the reduction of ribo-

nucleotides to the corresponding deoxyribonucleotides. Ribonucleotide reductase is responsible for maintaining proper levels of DNA precursors for replication and repair and, therefore, plays a crucial and conserved role in all organisms. In eukaryotes the enzyme is a tetramer consisting of two subunits, R1 and R2, with R2 using a diiron (Fe_2) as a prosthetic group. Each of the two subunits is a dimer of the proteins RNR1 and RNR2, respectively.

In the yeast species *Saccharomyces kluyveri*, RNR2 is encoded by a single gene, so the R2 subunit is, by necessity, a homodimer composed of two identical polypeptides. In *S. cerevisiae*, the gene encoding RNR2 has undergone duplication, most probably together with the entire genome. As a result, *S. cerevisiae* has two paralogs, RNR2 and RNR4. Thus, *S. cerevisiae* can produce three R2 subunits, two homodimers (of either RNR2 or RNR4) and a heterodimer composed of one copy of RNR2 and one of RNR4. Interestingly, only the heterodimeric R2 is functional. The reason is that RNR4 lacks the proper residues to accommodate the diiron group, while in RNR2 an invariant phenylalanine has been replaced by threonine, causing structural problems in one of the iron-binding structures. The defects are evidently complementary, since the replacement of *RNR2* gene by *RNR4* or the replacement of *RNR4* by *RNR2* in *S. cerevisiae* result in various degrees of inviability, while the *RNR2* from *S. kluyveri* can replace either and both genes in *S. cerevisiae*.

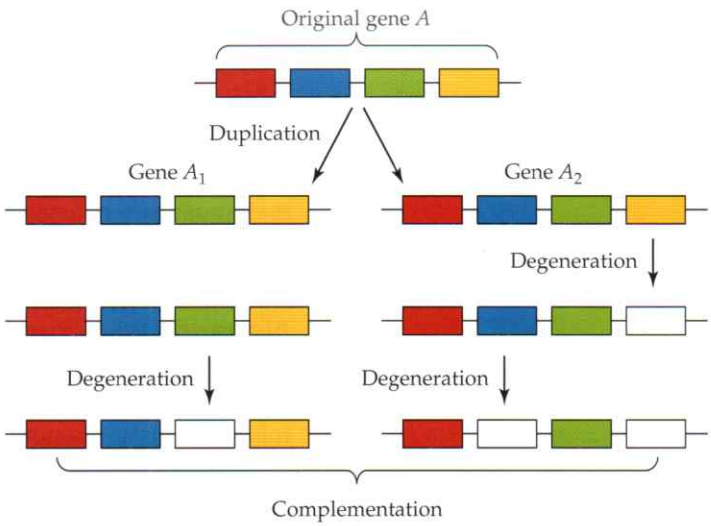


Figure 7.12 Schematic representation of the duplication-degeneration-complementation process. Each box represents a subfunction, with colored boxes representing intact subfunctions and empty boxes representing subfunctions that have been obliterated. The partially functioning duplicated genes A_1 and A_2 are both maintained in the population because the complete set of functions of the original, unduplicated gene can only be met by the two "degenerate" genes together.

SEGREGATION AVOIDANCE The premise of the segregation avoidance model (Spofford 1969) is simple. If overdominant selection occurs at a locus via heterozygous advantage, then lower-fitness homozygotes will be produced each generation regardless of the strength of selection (Chapter 2). This is because it is impossible to maintain populations composed of heterozygotes only. The mean fitness of the population will, therefore, be lower than the highest possible fitness—a phenomenon called **segregational load**. But if an unequal crossing over occurs in a heterozygote that creates a locus duplication containing the two alleles, then individuals can attain a situation akin to permanent heterozygosity, and the population will avoid segregational load (**Figure 7.13**). Hahn (2009) made the astute observation that the segregation avoidance model is not easily classifiable. Is it an instance of retention of ancestral function (gene conservation), because no change to the ancestral sequences was required? Or, is it an instance of subfunctionalization, because the multiple functions of a single locus are subsequently carried out by two loci?

An example of segregation avoidance occurs in the *ace-1* locus of the house mosquito, *Culex pipiens* (Labbé et al. 2007). This locus encodes acetylcholinesterase, the target of organophosphate pesticides. Two alleles are known to segregate at this locus: a susceptible wild-type allele, $ace\text{-}1^S$, and a resistant allele, $ace\text{-}1^R$. There is only one amino acid difference between the protein produced by the resistant allele and the protein produced by the wild-type allele, a glycine versus serine at amino acid position 119 of $ace\text{-}1^R$ and $ace\text{-}1^S$, respectively. Carriers of $ace\text{-}1^R$ have a greatly reduced susceptibility to organophosphate insecticides, thereby increasing their fitness when exposured to the pesticide. Concomitantly, the $ace\text{-}1^R$ allele drastically

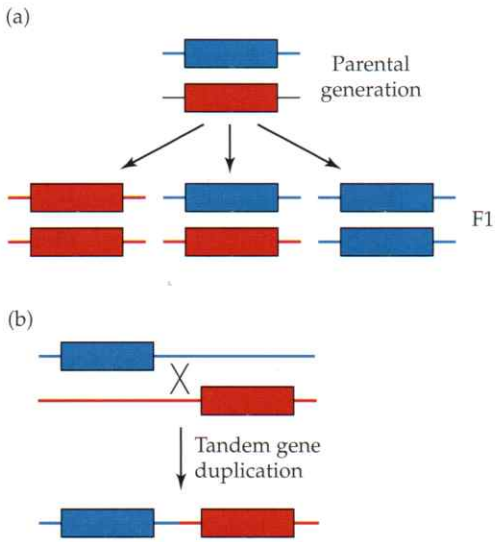


Figure 7.13 A heterozygote at a locus (blue and red rectangles) produces both homozygotes and heterozygotes. (a) If the heterozygote has a higher fitness than either homozygote, a segregational load will be incurred. (b) If, on the other hand, a gene duplication occurs through unequal crossing over, a situation akin to permanent heterozygosity is attained.

reduces the fitness of its carriers in the absence of organophosphate insecticides. Interestingly, a duplication combining resistant and susceptible alleles of the *ace-1* locus (*ace-1*^{D}) has recently appeared in the population. The duplication combines the advantages of the resistant allele at times of pesticide exposure with the advantages of the susceptible allele at all other times, thereby diminishing segregational load.

Segregation avoidance may also explain the evolution of resistance to the acutely toxic organophosphate insecticide monocrotophos in the two-spotted spider mite, *Tetranychus urticae*, through the gene duplication of the acetylcholinesterase gene (Kwon et al. 2009).

Neosubfunctionalization

The subfunctionalization mechanisms described above are not mutually exclusive, although there are other examples that clearly are. Moreover, one can envision hybrid models, such as **neosubfunctionalization**, whereby a new function evolves as a consequence of the subfunctionalization of one of the copies (e.g., He and Zhang 2005; Marcussen et al. 2010).

NEOSUBFUNCTIONALIZATION OF A DUPLICATED PANCREATIC RIBONUCLEASE GENE Colobine monkeys are unique among primates in their use of leafy matter as their primary food source. Similar to ruminants, the colobines have digestive systems in which the leaves are fermented by bacteria in the foregut, and the colobines recover nutrients by breaking up and digesting the bacteria with various enzymes. Here we describe the evolution of pancreatic ribonuclease, which is secreted from the pancreas and transported into the small intestine to degrade RNA. For the colobines, one reason to use pancreatic ribonuclease as a digestive enzyme may be related to the fact that rapidly growing bacteria have huge amounts of nitrogen in their RNA, and high concentrations of ribonuclease are needed to break down bacterial RNA so that nitrogen can be acquired.

Zhang et al. (2002) found that while 15 noncolobine primates have a single pancreatic ribonuclease gene (*RNASE1*), one colobine monkey, the leaf-eating douc langur (*Pygathrix nemaeus*), has a duplicate gene (*RNASE1B*). The unequal crossing over that gave rise to the two genes was inferred to have occurred 2.4–6.4 million years ago. There are 12 nucleotide differences between the coding regions of the two genes, and all 12 nucleotide substitutions giving rise to these differences occurred in the *RNASE1B* lineage. Of the 12 substitutions, 2 were synonymous and 10 were nonsynonymous, a strong additional indication that the *RNASE1B* gene evolved under positive Darwinian selection. Of the 10 nonsynonymous substitutions, one occurred in the signal peptide and 9 occurred in the mature protein. Of these, 7 involved charge changes. Interestingly, all 7 charge-altering substitutions increased the negative charge of the protein (5 from positively charged amino acids to uncharged amino acids, and 2 from uncharged amino acids to negatively charged amino acids).

The charge-altering substitutions reduced the net charge of the *RNASE1B*-encoded protein and the isoelectric point from 9.1 to 7.3. Because RNA (the substrate) is negatively charged, the net charge of the *RNASE1B*-encoded protein should influence its interaction with the substrate and its catalytic performance. Indeed, the optimal pH for the douc langur *RNASE1*-encoded enzyme was 7.4; the optimal pH for the *RNASE1B*-encoded enzyme was 6.3. At the lower pH, the *RNASE1B*-encoded enzyme digested RNA six times as fast as its *RNASE1* counterpart. These results suggest that the amino acid substitutions in one of the duplicates, *RNASE1B*, were driven by positive selection for enhanced enzymatic activity in the much more acidic environment of the colobine small intestine.

Interestingly, the adaptive evolution of RNASE1B came at a price. RNASE1 has the ability to degrade double-stranded RNA (dsRNA) as well. The ability of *RNASE1B* to degrade dsRNA is almost completely lost (merely 0.3% of the activity of RNASE1). As it happens, the nine substitutions that enabled the *RNASE1B*-encoded enzyme to be active in a highly acidic environment have concomitantly reduced its dsRNA enzy-