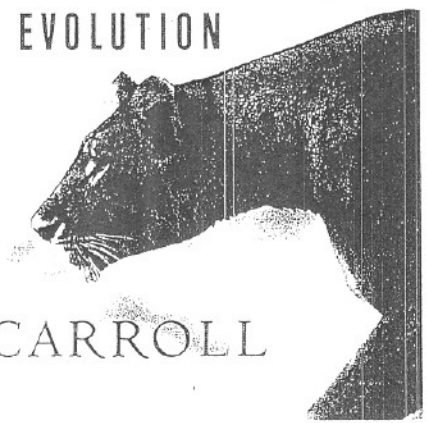
THE MAKING OF THE FITTEST

DNA AND THE ULTIMATE FORENSIC RECORD OF EVOLUTION



SEAN B. CARROLL

Chapter 3

Immortal Genes: Running in Place for Eons

To be sure, everything in nature is change but behind the change there is something eternal.

—Johann Wolfgang von Goethe.

HE WASN'T LOOKING FOR A NEW KINGDOM.

Microbiologist Tom Brock and his student Hudson Freeze were prowling around the geysers and hot springs of Yellowstone National Park one day late in the summer of 1966. They were interested in finding out what kinds of microbes lived around the pools and were drawn to the orange mats that colored the outflows of several springs.

They collected samples of microbes from Mushroom Spring, a large pool in the Lower Geyser Basin whose source was exactly 163 degrees F, thought at the time to be the upper temperature limit for life. They were able to isolate a new bacterium from this site, a species that thrived in hot water. In fact, its optimal growth temperature was right around that of the hot spring. They dubbed this "thermophilic" creature *Thermus aquaticus*. Brock also noticed

some pink filaments around some even hotter springs, which raised his suspicion that life might occur at even higher temperatures.

The next year, Brock tried a new approach to "fishing" for microbes in the hot springs of Yellowstone. His fishing tackle was simple: he tied one or two microscope slides to a piece of string, dropped it in the pool, and tied the other end to a log or a rock (don't try this on your own—you will be arrested and quite likely scalded or worse). Days later, upon retrieving the slides, he could see heavy growth, sometimes so much that the slides had a visible film. Brock was right that organisms were living at higher temperatures than had previously been thought, but he did not imagine that they were living in *boiling water*. And they weren't just tolerating 200 degrees F or more—these organisms were thriving in smoky, acidic, boiling pots such as Sulphur Cauldron, in the Mud Volcano area of the park. Brock's Yellowstone explorations opened eyes and minds to the extraordinary range of life's adaptability, identified bizarre but important new species such as *Sulfolobus* and *Thermoplasma*, and launched the scientific study of what he called "hyperthermophiles," lovers of superheat.

Brock's discovery of hyperthermophiles would, in time, lead to three more discoveries with profound impacts on biology. Brock lumped all of his new species into the classification "bacteria." Under the microscope, they did appear a lot like ordinary bacteria (figure 3.1). But, a decade later, Carl Woese and George Fox at the University of Illinois discovered that various sulfur-, methane-, and salt-loving species actually formed an entire kingdom unto themselves. They were as different from bacteria as bacteria are from eukaryotes (the division of life to which we animals belong, as well as plants, fungi, and protists). This new third domain, or division, of life is now referred to as the Archaea.

The second discovery from Brock's world was a practical one. A heat-stable enzyme that could copy DNA at high temperatures was isolated from *Thermus aquaticus*. This enzyme led to the invention of a new, efficient, and very fast technique for the study of genes in any species. This technique catalyzed a vast expansion in the amount and diversity of DNA information that could be obtained from nature, as

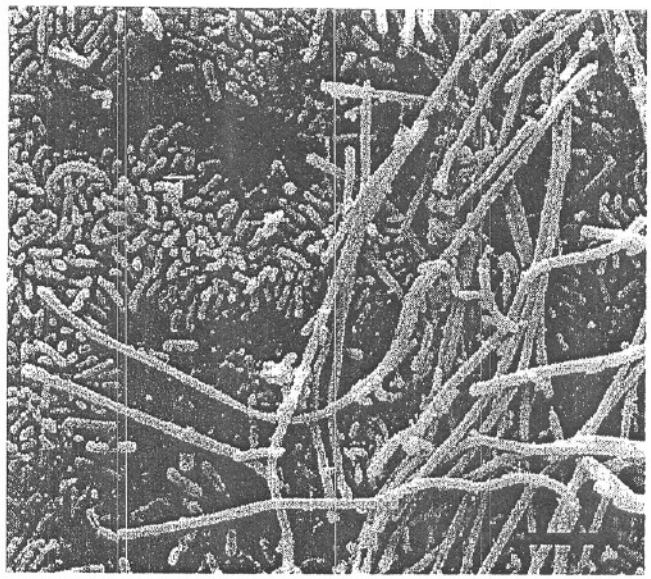


FIG. 3.1. **A sample of microbes from a hot spring.** This scanning electron micrograph reveals the growth of a variety of microbes on a slide immersed in the Obsidian Pool of Yellowstone National Park. *Figure from P. Hogenholtz et al. (1998),* Journal of Bacteriology *180:366.*

well as the creation of a multi-hundred-million-dollar market in DNA diagnostics and forensics.

The third and most recent discovery has emerged from the study of archaean genomes. Scrutiny of archaean genes has revealed critical clues about the making of our own eukaryotic ancestors nearly 2 billion years ago. Still preserved in the DNA of these primitive organisms are many pieces of DNA code that also exist in humans and all other eukaryotes. This shared text forms the remaining traces of an early event that gave rise to the first eukaryote, and is crucial evidence that an archaean was one of our original genetic parents.

In this chapter, we are going to examine some of the oldest DNA text on Earth. The fact that such ancient text has endured over eons of time, against the steady bombardment of mutations that could have

erased it many times over, is itself remarkable. But these "immortal" genes are also powerful evidence of two key elements of the evolutionary process—the power of natural selection to preserve the DNA record and the descent of life from common ancestors.

The immortal genes vividly reveal the evidence for one very important but somewhat underappreciated face of natural selection. More thought and attention has been directed to the "creative" dimension of natural selection and how new traits evolve, but this is only one aspect of the evolutionary process. Natural selection also acts to remove, in Darwin's words, "injurious change." I will explain how the effect of the removal of harmful mutations by natural selection is manifest in the DNA records of species, in the form of hundreds of genes that have been preserved across kingdoms of life for more than two billion years. In these immortal genes, the steps of evolution we see are just a matter of "running in place" as the gene's text changes only within narrow limits set by natural selection.

The survival of individual genes over vast geological periods provides more than unimpeachable evidence of the preservative force of natural selection. They are clues to the history of life's evolution from ancient ancestors, a new kind of evidence that Darwin could never have imagined. I will show how these immortal genes are powerful genealogical records that reflect the degree of relatedness among kingdoms and help us retrieve and reconstruct events in the history of life that are not visible in the fossil record.

Looking at DNA and Reading the Code

The skyscrapers of DNA sequences that we now own contain a lot of text, about forty thousand volumes, at a million characters each. The records of some species, such as humans, require a whole encyclopedia of about three hundred volumes while others, such as a bacterium, just a three- or four-volume set. No matter which volume one looks in, the text at first looks pretty much the same, like this:

ACGGCTATGGGCTACCAAGGGCTACCAACTACCAAAGTTACGGCTAATCGACAT
AATTGGCTACCAAGACATAACCTGGCTACCAATTACTATGGACGGCCTACGGCG
TCCGCTAATCGACATAACCTTTACTATGGCTACCAAAGTGACATAACCTTTACT
CATAACCTGGCTACCAACCAAGGGCTACCAACTACCAAAATTACTATGGGACAT
TAATCGACATAACCTTTACTAACCTGGCTACCAATTACTATGGACGGCCAATCG

etc., for hundreds of pages.

How can such a monotonous text composed of just four different characters encode the instructions for a complex creature? Moreover, how the heck do we read this stuff?

To make sense of the language of DNA, we need to learn how to look at genomes and genes and how to read DNA code. We can then make comparisons between species at many different scales, from very close relatives to vastly different life-forms whose lines split off from one another early in life's history. The clues to evolution emerge from *understanding the meaning of the similarities and differences* we find.

In order to decipher the natural history that resides in the DNA record, we have to have a firm grasp of the language of DNA, and of how DNA information is decoded in making the working parts of living organisms. Don't be intimidated—you can learn the language of DNA. It has a very small alphabet and a very limited vocabulary, and its rules of grammar are simple. The payoff for learning about the DNA code is being able to see, and therefore so much better understand, the process of evolution at its most fundamental level. I understand that new terms can get confusing, so you might want to bookmark this short section for future reference.

Here we go.

Proteins are the molecules that do all of the work in every organism—from carrying oxygen, to building tissue, to copying DNA for the next generation. The DNA of each species carries the specific instructions (in code) necessary for the building of these proteins.

DNA is made of two strands of four distinct *bases*. These chemical building blocks are represented by the single letters A, C, G, and T. The strands of DNA are held together by strong chemical bonds

between pairs of bases that lie on opposite strands—A always pairs with T, C always pairs with G—as shown here:

```
~AGTCAGTC~
||||||||
~TCAGTCAG~
```

so, if we know the sequence of one strand of DNA, we automatically know the sequence of the other strand. It is the unique order of bases in a sequence of DNA (ACGTTCGATAA, etc.) that forms the unique instructions for building each protein. The most amazing fact about DNA is that all of life's diversity is generated through the permutations of just these four bases. So, if we want to understand diversity, we have to crack the code.

How are proteins built and how do proteins know what their job is? Proteins themselves are made up of building blocks called *amino acids*. Each amino acid is encoded as a combination of three bases or a *triplet* (ACT, GAA, etc.) in the DNA molecule. The chemical properties of these amino acids, when assembled into chains averaging about 400 amino acids in length, determine the unique activity of each protein. The length of DNA that codes for an individual protein is called a *gene*.

The relationship between the DNA code and the unique sequence of each protein is well understood because biologists cracked the genetic code forty years ago. The decoding of DNA in the making of proteins occurs in two steps, which I will now describe. In the first step of decoding DNA, the sequence of bases on one of the strands of the DNA molecule is *transcribed* into a single strand of what is called messenger RNA (mRNA). Then, in the second step, the mRNA is *translated* into the amino acids that build the protein. In the cell, the genetic code is read (from the mRNA transcript) three bases at a time, with one amino acid determined by each triplet of bases (a short example is shown in the right half of figure 3.2).

There are sixty-four different triplet combinations of A, C, G, and T in DNA, but just twenty amino acids. Multiple triplets code for particular amino acids (and three triplets code for nothing, and mark the stopping point in the translation of mRNA and the making of a protein—as periods mark the end of a sentence). Much to our convenience, but also of profound evolutionary significance, this code is, with few minor exceptions, the *same* in every species (this is why bacteria can be used to produce human proteins for pharmaceutical use, such as insulin).

Thus, given a specific DNA sequence, it is easy to decipher the protein sequence that that DNA encodes. However, not every base in DNA is part of a message for a protein. In fact, a large portion of DNA is "noncoding." The first challenge biologists have when given long reams

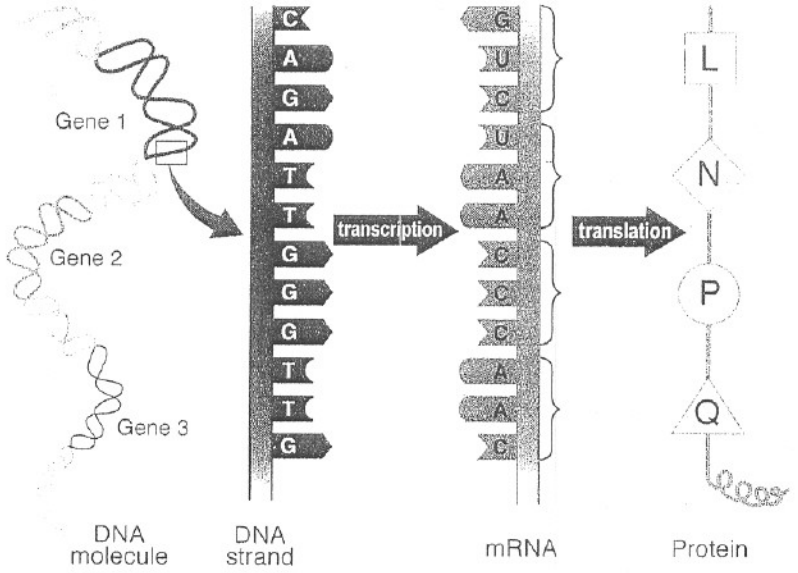


FIG. 3.2. **The expression and decoding of DNA information.** An overview of the major steps in decoding DNA into a functional protein. *Left,* long DNA molecules contain many genes. The decoding of a portion of one gene is shown in two steps. First, the complement of one DNA strand is transcribed into mRNA. Then, the mRNA is translated into protein, with three bases of the mRNA encoding each amino acid of the proteins (shown as the letters L, N, P, and Q here). In mRNA, the base U is used in place of the T in DNA. *Figure by Leanne Olds.*

of DNA text is to figure out where the "coded" messages begin and end. With whole genome sequence data this, thankfully, is now carried out largely on computers using algorithms that are really good at searching for, and finding, the needles in the haystacks of DNA sequence.

The coding sequence of an average gene is about 1200 base pairs in length. In some species—particularly, microbes such as bacteria or yeast—genes are very closely packed with relatively small spaces of noncoding DNA between the thousands of genes in the entire genome. In humans, and many other complex species, genes occupy only a small fraction of all of the DNA, and are separated by long intervals of noncoding DNA. Some of this noncoding DNA functions in the control of how genes are used, but a lot of it is what is called "junk." This junk accumulates by various mechanisms and often contains long repetitive tracts with no informational content; it is not purged unless it has adverse effects. I will generally ignore this junk, but it is worth mentioning in order to have a picture of the structure of our genomes as archipelagoes of islands (genes) separated by vast areas of open sea (junk DNA).

The Fates of Genes: The Immortal Core

When scientists look at entire genomes, their first aim is to locate all of the genes within the entire DNA sequence. This allows them to take an inventory of a species' genes that includes the total number of genes and a list of every individual gene. Because biologists have been studying the genes and proteins from species for a while now, we can sort genes and the proteins they encode into categories based upon their function and resemblance to existing genes and proteins.

The most interesting fact from the comparison of genomes is that while the number and kinds of genes differ considerably both between and within the three major divisions of life, great increases in complexity do not require proportionate changes in gene number. As shown in table 3.1, most bacteria possess on average around 3000

Table 3.1. The number of genes in genomes

Bacteria	
Aquifex aeolicus	1560
Neisseria meningitidis	2079
Vibrio cholerae	3463
Staphylococcus aureus	2625
Escherichia coli K12	4279
Salmonella typhi	4553
Archaea	
Sulfolobus solfataricus	2977
Methanocaldococcus jannaschi	1758
Halobacterium sp.	2622
Eukaryotes	
Saccharomyces cerevisiae (yeast)	6338
Drosophila melanogaster (fruit fly)	13,468
Caenorhabditis elegans (nematode worm)	20,275
Tetraodon nigroviridis (fish)	20–25,000
Mus musculus (mouse)	20–25,000
Homo sapiens (human)	20–25,000
Arabidopsis thaliana (plant)	25,749

genes, with the smallest genome of a free-living species containing about 1600 genes. Any two bacterial species may differ in size, however, by as many as 3000 genes. Animals possess roughly 13,000 to 25,000 genes, with some animals differing by many thousands of genes. Note that complex creatures, such as a fruit fly, have only roughly twice as many genes as a single-celled brewer's yeast, and that humans have almost twice as many genes as a fruit fly. But we humans have just about the same number of genes as a mouse.

However, gene number is just a raw figure. More detailed clues about evolution emerge from the direct comparison of the fates of individual genes. The differences in gene number tell us that certain genes must be present in some species and absent from others. Before I discuss some specific comparisons, it is important to think about what we might find when we compare the genes of species that belong to different groups. How similar or different should we expect the genes of different species to be?

Before DNA sequencing was possible, some of the great minds of evolutionary biology in the mid-twentieth century contemplated this question. They knew a bit about mutation and concluded that, over geologic time, mutation would eventually change just about every base pair in a genome. For example, with a mutation rate of about 1 mutation per 100 million base pairs per generation, in 100 million generations, most sites in a gene would be mutated at least once, on average. Given the very short generation times of microbes (on the order of hours), and the modest generation times of plants and small animals (a year or less), then one might expect little trace of similarity between the gene of any two species whose lineages diverged 100 million years ago. Indeed, in his 1963 book *Animal Species and Evolution*, the great biologist Ernst Mayr remarked, "Much that has been learned about gene physiology makes it evident that the search for homologous genes [the same gene in different species] is quite futile except in very close relatives."

But, when we compare different kinds of bacteria with one another, or different animals (whose ancestors diverged well over 100 million years ago) with each other, we find extensive similarities in their genes. For example, when the genome of the infamous delicacy the puffer fish is compared with the genome of the gourmand stupid enough to eat this deadly creature (the human), at least 7350 genes are found that are clearly shared between the two species. Furthermore, the proteins encoded by these genes are on average 61 percent identical. Since the evolutionary lines of fish and other vertebrates (including humans) separated about 450 million years ago, this is a much more extensive similarity than would be expected if mutations were simply allowed to accumulate over time.

More stunning, when we compare the genomes of Archaea, bacteria, fungi, plants, and animals, we find about 500 genes that exist in all domains of life. We know from the fossil record that eukaryotes are at least 1.8 billion years old and the Archaea and bacteria well over 2 billion years old. The genes these organisms all share have withstood more than 2 billion years of the steady bombardment of mutation and stand out as threads of text whose sequence and meaning have not changed significantly despite the vast differences among the species that carry them. These genes are *immortal*.

The functions of immortal genes are central to fundamental, universal processes in the cell, such as the decoding of DNA and RNA and the making of proteins. All forms of life have depended upon these genes since the origin of complex DNA-encoded life early in Earth's history. These genes have survived through an immense arc of time, and life will continue to depend upon this core set of genes as it evolves in the future.

Immortal genes have survived not because they avoid mutation—they are as vulnerable to mutation as all other genes. The genes are immortal in the sense that the gene as a unit endures; however, not every letter of the gene's code endures. This fact can be seen upon more detailed inspection of their DNA sequences and of the sequences of the proteins they encode, and it is a key demonstration of one aspect of the process of natural selection.

The new wealth of data from genomes offers unique insights into the deep past that could not be deciphered from any other source. I will close this chapter with the story of the evolution of the domain we belong to (eukaryotes), and the unique contributions that archaea and bacteria appear to have made to our ancestry.

The Making of Eukaryotes: A Marriage of Two Very Different Parents?

> The time will come, I believe, though I shall not live to see it, when we shall have fairly true genealogical trees of each great Kingdom of Nature.
>
> —Charles Darwin, letter to T. H. Huxley, September 26, 1857

Our understanding of the structure of nature has come a long way since Darwin's time. In his day, the living world was divided into just the plant and animal kingdoms. This dual system had been recognized since Aristotle and was formalized by Carl von Linné in 1735. Ernst Haeckel, in 1866, with his remarkable studies of protists, added a third kingdom to life's tree. The bacteria and fungi were not added as full-fledged kingdoms until the twentieth century.

Within this five-kingdom scheme, a higher primary division was also recognized, based upon fundamental differences in the types of cells found in different kingdoms. In 1938, French biologist Edouard Chatton proposed the names "prokaryote" and "eukaryote" based solely on the absence and presence, respectively, of nuclei. These two "superkingdoms" encompassed all of the known living world, until Carl Woese started studying the genes of the kinds of species Tom Brock found in Yellowstone.

Woese believed that bacterial taxonomy was a mess and needed more objective means of determining the evolutionary relationships among species than their appearance or physiological characteristics. He turned to molecules. The possibilities for building species trees based upon DNA, RNA, and protein sequences were quickly recognized by scientists (such as Francis Crick, Emile Zuckerkandl, and Linus Pauling), as soon as protein sequencing began to reveal the similarities and differences in proteins shared among groups of species. The basic idea is quite straightforward. Sites in the text of DNA, RNA, or protein sequences that differ among a group of species, but are shared among subsets of these species, reflect their degree of kinship. Just as we build family trees based upon degrees of genetic relationships, we build species trees based upon their genetic kinship. But, as I will explain, sometimes there is a marriage that really confuses the family tree.

Woese used an abundant type of RNA molecule to make trees of bacteria. But when he included the thermophilic, methane-producing species with conventional bacteria, he found that "these 'bacteria' appear to be no more related to typical bacteria than they are to eukaryotic cytoplasms." He proposed there existed a third superkingdom which, because of the adaptation of these species to the sorts of extreme environments that were presumed to exist early in Earth's history, might be the original or *ur*-kingdom, so he suggested calling this new superkingdom "archaeabacteria." This name was modified later to Archaea, in part to avoid confusion with the bacteria, and the superkingdom category was renamed "domain."

While the division of life into three domains—Eukarya, Archaea, and Bacteria—has held up, the relationships between these three groups has been challenging to sort out. Darwin described the genealogy of species as trees, with speciation producing ramifying branches. But in the world of microbes, unknown to Darwin, some events happen that violate the pattern of treelike evolution. Microbes exchange genes, and some microbes live within other host species in a process called endosymbiosis. These processes enable the transfer of genes between very distant relatives, and thus confuse the family tree. In

order to figure out the relationships between eukaryotes, archaea, and bacteria, biologists have to sort out the history of lots of genes, not all of which may have the same family resemblances.

For example, some of the first studies of archaean molecules turned up some striking resemblances between some archaea and eukaryotes. Proteins that archaea use to package their DNA in chromosomes, to transcribe DNA, and to decode messages bear such similarity to those in eukaryotes that it suggested to many that eukaryotes evolved from some archaean. Some of these provocative similarities are in short "signature" sequences in proteins that are shared among some archaeans and eukaryotes, and no other group. For example, there is a short insertion of eleven amino acids in one of the immortal proteins involved in decoding messages. Table 3.2 shows the sequence of this insertion in different eukaryotes and archaea:

Table 3.2. Insert sequences

Eukaryotes	
Human	GEFEAGISKNG
Yeast	GEFEABISKDG
Tomato	GEFEAGISKDG
Archaea	
Sulfolobus	GEYEAGMSAEG
Pyrodictum	GEFEAGMSAEG
Acidianus	GEFEAGMSEEG
Bacteria	(Absent)

The existence of this sequence in two domains but not in the third would be most logically explained by the archaea and eukaryotes being more closely related to each other on life's tree than to bacteria. The resulting picture of the tree would posit that there was a common ancestor of all three domains (the last "universal" common ancestor, or LUCA) that then split into two domains, the Bacteria and Archaea,

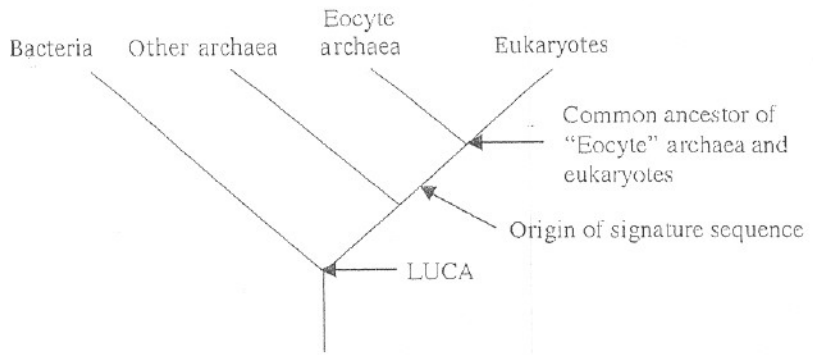


FIG. 3.4. A **"conventional" tree of life.** The tree depicts all domains arising by the splitting of lineages. *Figure by Jamie Carroll.*

and the eukaryotes arose later from a branch of the Archaea. The tree of life would then be as shown in figure 3.4.

However, sequencing of whole archaean and bacterial genomes revealed, somewhat unexpectedly, that the majority of archaean genes show the greatest similarity to bacterial counterparts. Then, as more eukaryotic genomes were sequenced, their analysis suggested that many eukaryotic genes were more related to those of bacteria than of archaea. The story was taking on the nature of one of those riddles like "If your sister is also your aunt, then who is your father?" In short, the answer to the question of which groups are most closely related was muddied.

The resolution of the riddle stems from the pursuit of a key observation. It was noted that most of the similarities between archaea and eukaryotes were in so-called informational genes whose products dealt with the copying and decoding of DNA. Furthermore, most of the similarities between eukaryotes and bacteria were in operational genes involved in the metabolism of various nutrients and basic cellular materials. From the viewpoint of eukaryotes, it appeared as though they got their "brains" (informational genes) from one parent, and their "looks" (operational genes) from another.

This raised the specter that eukaryotes were the product of a mixed

marriage—a genetic fusion of archaean and bacterial parents. The notion of a fusion between vastly different species is not new. In 1970, Lynn Margulis proposed that mitochondria and chloroplasts, two key energy-producing organelles in eukaryote cells, arose from bacteria living within eukaryotes (this fusion process is endosymbiosis). This view is now widely accepted.

But what about the making of a eukaryote from an archaean and bacterial ancestor? Maria Rivera and James Lake of UCLA have concluded that, indeed, eukaryotes are of dual origin, from parents belonging to different branches of life. Rivera and Lake analyzed bacterial, archaean, and eukaryotic genomes for the sets of genes shared in all, all but one, all but two, all but three, etc., of the major divisions within these three domains. Comprehensive analysis of these patterns of shared genes indicate that the eukaryote genome is the product of a fusion between a relative of a type of archaean and a type of bacterium. Because symbiotic relationships are common among organisms living together (for instance, Yellowstone's *Thermus aquaticus* gets its energy from photosynthetic cyanobacteria that color those bacterial mats), and this occasionally leads to endosymbiosis, the likely explanation for eukaryote origins was as a product of the fusion of genomes between an endosymbiont and its host. The resulting base of the tree of life is then not a trunk, but a ring from which our tree ascends and branches (figure 3.5).

So, if you happen to visit magnificent Yellowstone, don't turn away from the stinking, boiling soup of those hot cauldrons or be revolted by the colorful strings and mats of slime that ooze around their edges. That's no way to respect one's relatives, no matter how distant. Ponder the amazing fact that you share hundreds of genes with members of this community. And, in this sort of community, somewhere an unfathomably long time ago, perhaps along a deep sea vent, in a belch of methane there emerged the ancestor of all of the familiar and visible kingdoms on Earth.

Of course, if all natural selection did thereafter was to maintain the status quo within very strict limits, life would be uniform and

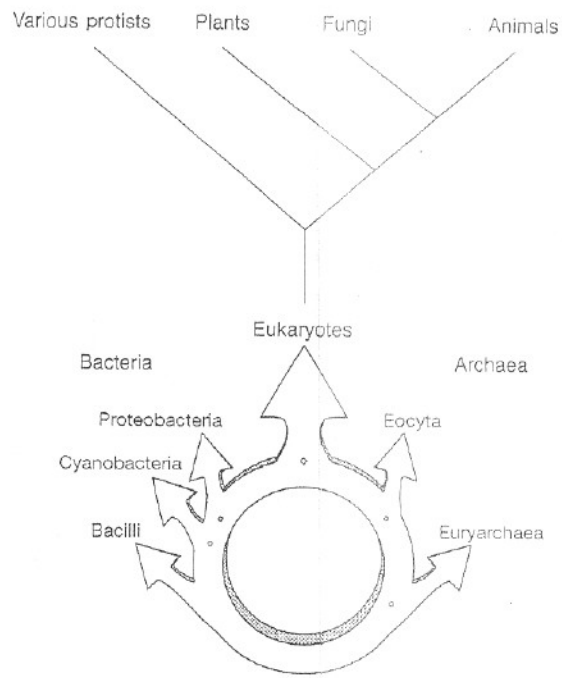
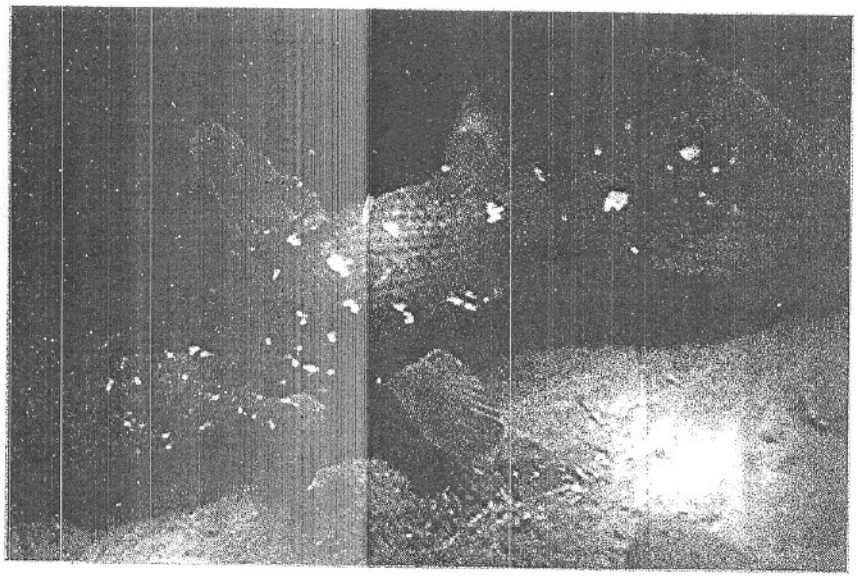


FIG. 3.5. **A new picture of the tree of eukaryotes.** The DNA record indicates that an ancient fusion of some type of archaea with some type of bacteria contributed to the origin of eukaryotes. The base of the tree is a ring, not a conventional trunk. *Adapted from M. Rivera and J. Lake (2004),* Nature *431:152.*

unchanging, and not the riot of diversity we see in the world today and in the past 3 billion years of the fossil record. The figures on gene number in table 3.1 tell us that vast differences exist in gene content between life-forms. Beyond the core of 500 or so immortal genes, species vary widely in gene number. The differences in gene number tell us that, in the course of evolutionary time, new genes must be born. They are indeed, and that creative dimension of evolution will be the focus of the next chapter. It is also a hint that genes might also die. They do die, and I will take up that twist and what it can tell us about evolution in chapter 5.



Coelacanth. *Photograph from JAGO submersible, Jurgen Schauer and Hans Fricke.*

Chapter 5

Fossil Genes: Broken Pieces of Yesterday's Life

Nature is, after all, the only book that offers important content on every page.

—Johann Wolfgang von Goethe

IT WAS A MAGNIFICENT EARLY CHRISTMAS PRESENT.

Midmorning on December 22, 1938, Marjorie Courtenay-Latimer received a message from the manager of the local fleet—the *Nerine* had docked and it might have some fish for her collection. Miss Latimer was the first curator of the East London Natural History Museum, in the Cape Province of South Africa, and she was busy that day trying to put together a dinosaur skeleton she had excavated, not to mention getting ready for the holiday.

The fleet manager did not call very often, so she decided to set aside her work and go down to the dock. Hitching up her cotton dress, she boarded the trawler and surveyed the stinking pile of sharks, sponges, and other familiar creatures lying out in the heat of the sun. She was about to return to the museum when it caught her eye. As

she pulled away a bunch of carcasses she saw "the most beautiful fish I had ever seen. . . . It was five feet long and a pale mauve-blue with iridescent markings."

It was also unlike any other fish she had ever seen. It was covered in hard scales, had four limblike fins, and a strange puppy-dog tail. She knew it had to be preserved. The fish weighed 127 pounds, so getting the dead and decomposing creature back to the museum was no small task. It took quite a bit of persuasion for a taxi driver to allow it into his trunk.

Once back at her post, she showed off her prize to the museum's director. He promptly dismissed it as a rock cod. Miss Latimer, virtually self-taught in natural history, thought differently, but none of her reference books helped her to identify the rotting hulk on her examination table. Miss Latimer decided to seek outside assistance in the form of Dr. J. L. B. Smith, a chemistry lecturer and amateur ichthyologist at Rhodes University, one hundred miles away. When she could not reach Smith by phone, she sent him a letter the very next day, enclosing a description and a drawing of the fish.

Smith did not get the letter until after the new year. He was recovering from an illness, and when he finally had the chance to study it he was bewildered. "And then a bomb seemed to burst in my brain, and beyond that sketch and the paper of the letter I was looking at a series of fishy creatures that flashed up as on a screen, fishes no longer there, fishes that had lived in dim past ages gone, and of which only fragmentary remains in rocks are known."

Smith immediately wired Miss Latimer: MOST IMPORTANT PRESERVE SKELETON AND GILLS FISH DESCRIBED.

Miss Latimer had managed in the meantime to get a taxidermist to preserve what he could.

Smith was positively riled by a possibility that his brain kept telling him was impossible. Yet the sketch, and then some scales he received later, told him that this fish was a coelacanth, a member of a group of fish with paired fins thought to be closely related to the first four-legged vertebrates *and believed to be extinct since the end of the Cretaceous period 65 million years ago.*

Smith finally got the chance to see the fish in person, which removed all doubt—as well as his faculties! He wrote, "I forgot everything else and just looked and looked, and then almost fearfully went close up and touched and stroked [it]."

Smith named the fish *Latimeria chalumnae*, in honor of Miss Latimer, and the river near where it was caught. It would be fourteen years before Smith or anyone else saw another coelacanth (and on that second occasion, Smith wept). Many more coelacanths have since been found in recent decades, including a second species discovered off Indonesia.

The coelacanth holds a special place in natural history. It is the only living member of an ancient tribe, with body features that link it to distant ancestors that lived 360 million years ago. The coelacanth has thus been dubbed a "living fossil."

In this chapter, we are going to uncover a different kind of fossil, one found in living species, that provides links to distant ancestors and former ways of life. These are *fossil genes*.

We have seen that shifts in species lifestyles, from land to water, from seeing visible hues to ultraviolet colors, and from eating fruits and insects to ruminating on leaves, involve the formation and fine-tuning of new genes. Here, we will see that such shifts also leave their traces in the form of genes whose use and function have been abandoned. Fossil genes reside in DNA much in the same way that fossils reside in sedimentary rock, and the text of the genes similarly breaks apart and erodes away over time. For information in DNA, the cardinal rule illustrated by fossil genes is "use it or lose it." The decaying text of fossil genes is evidence of the relaxation of natural selection and is very specific to individual genes and to particular species. We'll see how these broken pieces of yesterday's code reflect the adaptation of species, including humans, to new ways of life. I'll begin with coelacanth DNA and fossils of some now very familiar genes, and then work my way toward some examples of gene fossilization on a massive scale.

Shifting Habitats and Fossil Opsin Genes

Great fascination with the coelacanth has inspired expeditions to observe the animal in its native habitat. It has been seen from submersibles and by divers in deep underwater caves off the Comoros Islands and in waters around South Africa. The coelacanth retreats to these caves during the day and cruises slowly over the ocean floor at night to feed. At a depth of 100 meters or more, only dim blue light reaches the coelacanth in its native environment.

The coelacanth's lifestyle and its unique status have prompted interest in its visual system and opsin genes. Curiously, the coelacanth has a dim-light rhodopsin but no MWS/LWS opsin genes, the type other fish and we humans use for red-green vision. Because fish, mammals, and most other vertebrates have at least one version of this opsin, we know that coelacanth ancestors also had this gene, so somewhere along the line of coelacanth evolution the MWS/LWS opsin gene was lost. The loss of this gene raises a very general question: How and why is a gene that is so useful to some species lost in others? We can get a very good picture of the process of gene loss from another opsin gene that, while still in the coelacanth's DNA, is slowly eroding away.

The coelacanth has one SWS opsin gene. Remember that this is the short wavelength opsin used in humans and birds to detect the color violet, and in various species to see in the ultraviolet range. However, the code of the coelacanth SWS opsin gene contains many changes that disrupt its text. For instance, at positions 200–202 in the DNA code, where the mouse and other species have the three DNA bases CGA, the coelacanth has TGA. This change from a C to a T may seem like a tiny difference, but in this case it is a whopper. The letters TGA are a "stop" triplet that functions as a period to terminate the translation of the remainder of the SWS opsin text. This abolishes the coelacanth's ability to make a functional SWS opsin protein. Elsewhere in the code of the gene there are other deletions and changes that severely disrupt the opsin text. The many disruptions in the coelacanth opsin code tell us that it has become nonfunctional—it is a fossil gene (biologists call these "pseudogenes," but I will stick with the "fossil" label). It was functional in ancestors of the coelacanth, but no longer works in the fish today. The gene is still recognizable by the fragments of its parts. But, because it is not functional, it will continue to accumulate additional mutations and deletions that will erode it further, and will eventually erase it from the DNA forever, just as the coelacanth's MWS/LWS opsin was erased (and one of the globin genes was erased in the icefish ancestor I described in chapter 1).

Dolphin	TTT	*TT	CTG	TTC	AAG	AAC	AT*	***	TTG
Cow	TTT	CTT	CTG	TTC	AAG	AAC	ATC	TCC	TTG

FIG. 5.1. **A fossil opsin gene in a dolphin.** Short portions of the sequences of a dolphin and a cow SWS opsin gene. The shaded regions show the positions of deleted bases (asterisks) that disrupt the code of the dolphin gene. *Figure by Jamie Carroll.*

You are probably wondering by now why a good gene would be allowed to decay. Are fossil genes a rare kind of mistake just found in weird animals like coelacanths and icefish? Before I explain further, I will mention one more illuminating example.

Inspection of the SWS opsin gene of dolphins and whales reveals that, just as in the coelacanth, these cetaceans' SWS opsin gene has become a fossil. For example, the SWS gene of the bottlenose dolphin is missing one base at one position and four bases at another position near the beginning of the text (figure 5.1). The missing bases throw off the three-base-at-a-time decoding of the gene's text, shifting the reading frame and making the gene nonfunctional. Examination of other cetaceans reveals a number of changes in their SWS opsin gene that have rendered them nonfunctional. All dolphins and whales have a fossil SWS opsin gene.

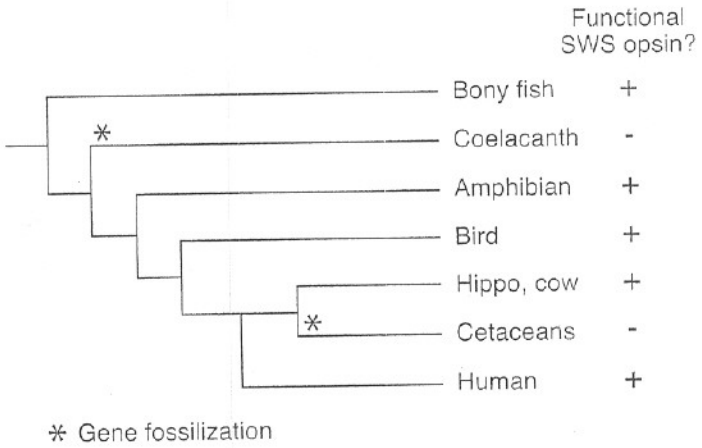


FIG. 5.2. **The same opsin gene has been fossilized twice.** The distribution of different mutations found in the coelacanth and cetacean SWS opsins and the evolutionary relationship of these species indicates that the SWS opsin was fossilized at least twice (asterisk). *Figure by Jamie Carroll.*

So, here, we have two examples of fossil SWS opsin genes, in the coelacanth and cetaceans. Is there anything in common or any connection between these animals that could explain why their SWS opsins are fossilized?

The first thing we can say is that the genes were fossilized independently. We know this from where these animals sit in the evolutionary tree of vertebrates (figure 5.2). The coelacanth belongs to a primitive group of fish that split off from a line that gave rise to four-legged vertebrates. Because amphibians, reptiles, birds, and many mammals have intact SWS opsin genes, we know that the fossilization of the coelacanth gene occurred during the evolution of the coelacanth lineage. Because hippos and cows, two close relatives of dolphins and whales, have functional SWS opsins, while all cetaceans do not, we can deduce that the function of the cetacean SWS opsin was lost in an ancestor of all dolphins and whales. The lack of functional SWS opsin in modern dolphins and whales is due to inheritance of this fossil gene from ancestors that lived more than 40 million years ago.

The best explanation for why the gene became fossilized comes from consideration of the animals' ecology. Surely, there must be some link to not using SWS opsin and these animals living in a marine habitat. Dolphins and whales are fully aquatic and belong to the only order of mammals that lacks the potential for any form of color vision (because its members have only one cone opsin, whereas most other mammals have two). As we saw in the last chapter, dolphins have tuned their dim-light-type rhodopsin to the bluer range of the light spectrum. The coelacanth is also a deep-dwelling animal. It, too, apparently has no use for color vision. The ecological rationale for the loss of the SWS opsin function is that it became dispensable to the ancestors of these species.

The dispensability of SWS opsin explains what we see happening in the DNA code. If the opsin is no longer needed, then natural selection, which would normally preserve the opsin's text, is *relaxed*. When natural selection is relaxed, there is no mechanism for purging genes of mutations that disrupt their function. The random process of mutation assures that *all* genes experience mutation. Most of the time, disruptive mutations are purged by the competitive process of natural selection because individuals and offspring bearing them are less fit. But when a trait is no longer under selection, in this case because of a shift in habitat, the genes that were essential in one lifestyle can become dispensable, and mutations can accumulate in them.

Use it or lose it.

In more formal terms, *fossil genes are exactly what we would predict to evolve as a consequence of the continuing action of mutation, over time, in the absence of natural selection*. The unmistakable signature of genes' dispensability is the accumulation of text-disrupting mutations. The resulting fossil genes are thus marks of changes in lifestyle from those of ancestors, and when we can spot and track fossil genes, these are valuable clues to reconstructing natural history.

I will highlight several more examples of opsin gene fossils linked to other kinds of changes in lifestyle and then expand the discussion of fossil genes to more genes, and many more species.

Living in Darkness

One leading theory to explain the reduction in the number of opsin genes and the loss of full color vision that occurred in the ancestors of mammals is that early mammals were nocturnal, and that color vision was dispensable for these rodentlike creatures' lifestyles. Nocturnality has evolved repeatedly in mammals, so one way of testing this theory, and the general idea of gene fossilization linked to shifts in lifestyle, is to examine more recent species with markedly different lifestyles.

The owl monkey, for example, is the only nocturnal species among higher primates (figure 5.3). And sure enough, examination of its SWS opsin gene reveals that it, too, has accumulated mutations that render it nonfunctional. Just sixty bases into the code for the opsin protein, the owl monkey has a mutation that changes a "TGG" into "TGA," another case of a stop triplet that terminates the translation of the remainder of the gene's text. All of the diurnal (daylight-active) relatives of the owl monkey have an intact SWS opsin gene, so this is pretty good evidence that the shift to nocturnality relaxed selection on the SWS opsin gene.

This idea about the link between nocturnality and SWS opsin can be further tested by examining the opsin genes of nocturnal prosimians. The prosimians are a primitive primate group that includes lemurs, tarsiers, bush babies, and the lorises. Lemurs include both nocturnal and diurnal species but the slow loris and bush baby are strictly nocturnal (see figure 5.3). Examination of their SWS opsin genes reveals, yet again, that the gene has become a fossil in these species. Each species gene has a big chunk of code missing near the beginning of the gene that obliterates the ability to make the opsin. Since the deletion is in exactly the same position and is exactly the

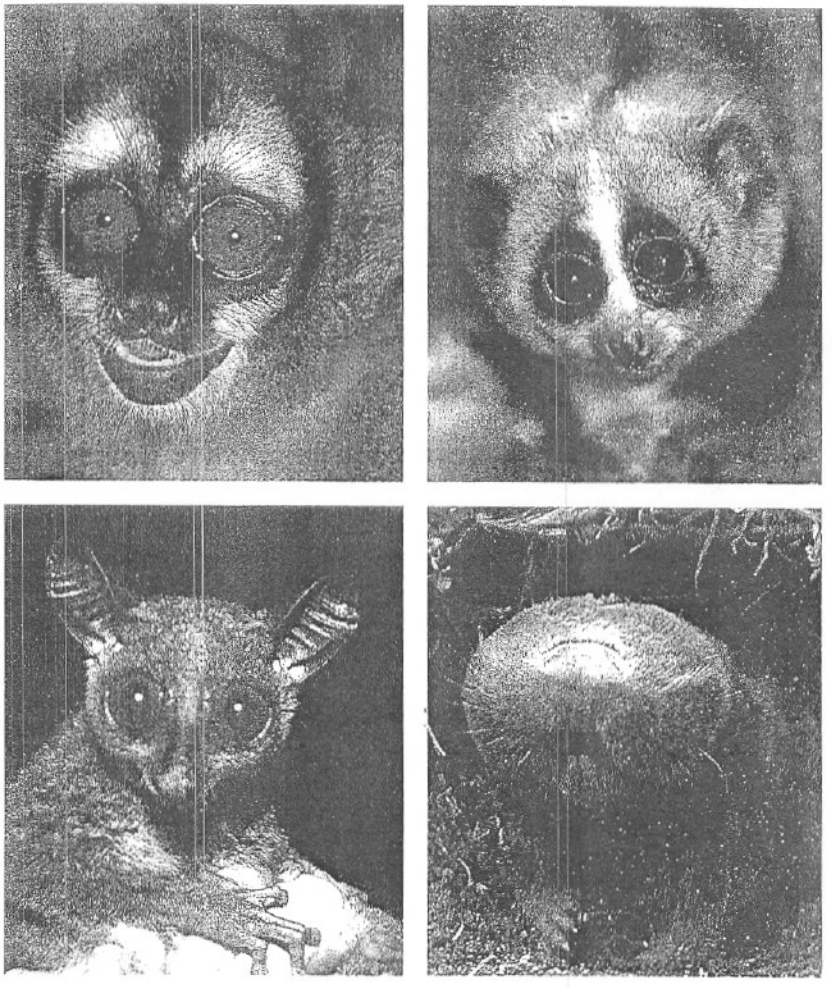


FIG. 5.3. **Nocturnal or subterranean mammals with fossilized opsin genes.** The owl monkey (top left; photograph by Greg and Mary Beth Dimijian), bush baby (bottom left; photograph by B. Smith, Cercopan, Nigeria), slow loris (top right; photograph by Larry P. Tackett, www.tackettproductions.com), and blind mole rat (bottom right; photograph by Tali Kimchi) all have fossilized SWS opsin genes, a result of their adaptation to nocturnal or subterranean lifestyles.

same size in each species, this indicates that the fossilization of SWS opsin first occurred in a common ancestor of the slow loris and bush baby, and was inherited by these species.

So far, so good. It appears that shifts in the light environments that species live in are well correlated with the loss, or retention, of color vision genes. Just one more test: How about when animals go underground?

The blind mole rat is a rodent that possesses the most degenerated eyes of any mammal (see figure 5.3). The fossil record suggests that this group of animals evolved from an aboveground ancestor that had normally proportioned eyes. The evolution of the mole rat's lifestyle has been accompanied by many changes in anatomy and physiology. Its eyes are so tiny that they cannot detect images. And even if the eyes themselves worked, seeing would still be difficult because they are located entirely under the skin and covered by a layer of fur. However, the blind mole rat can tell the time of day—its eyes have retinas that detect light and help maintain its circadian clock, which regulates its daily biorhythms.

Examination of the blind mole rat revealed two intact opsin genes, a red-shifted MWS/LWS pigment that is tuned to detect the light received through the subcutaneous eye and a dim-light rhodopsin. Clearly, despite its atrophied vision, selection remains on these genes, apparently in order to run the animal's biological clock. However, its SWS opsin gene is a fossil and contains numerous mutations that disrupt the text of the code for making the SWS opsin protein.

I have now described five separate cases of SWS opsin gene fossilization—in coelacanths, cetaceans, owl monkeys, the slow loris and bush baby, and the blind mole rat. In each case, the fossilization of the gene is correlated with the habitat in which the species live. In each case, the exact lesion in the SWS opsin gene is different. This fact, and the facts that the species belong to different parts of the evolutionary tree, and that close relatives of these animals have functional opsins, demonstrate that the fossilization of the SWS opsin has occurred independently and repeatedly by different mutations and at different times in history. This is overwhelming support for the fundamental prediction that relaxed selection on a gene will lead to its decay. Furthermore, in all these species, other kinds of opsins are intact and functioning, which demonstrates that the decay of genes is highly selective.

The frequent loss of SWS opsin and its correlation with shifts in lifestyle is a very loud hint that the fossilization of genes is a frequent signature of evolutionary change. Let's now turn to the human genome for some fossil signatures of how we are different from our ancestors.

Can't You Smell That Smell?

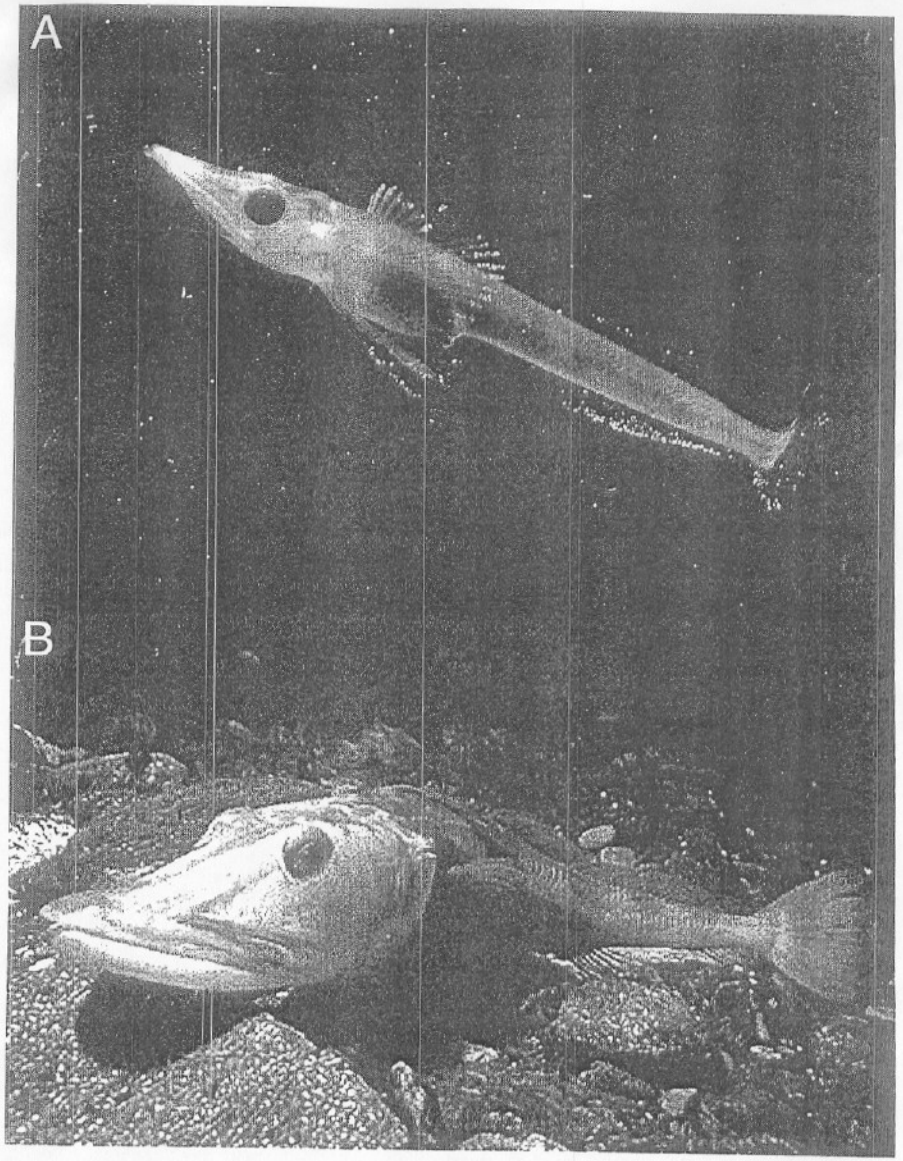
We have now seen many examples of how shifts in habitat are associated with the adaptation and loss of visual system genes. Other senses are also vital to animal behavior and survival, particularly the sense of smell. One walk in the park with a dog provides many examples of how their "view" of the world is shaped by their acute sense of smell.

Many other mammals also have powerful senses of smell, which are used to find food, identify mates and offspring, and detect danger. For a long time, it was a mystery how different odors could be detected and discriminated. In 1991, Linda Buck and Richard Axel discovered a family of genes that encoded odorant receptors. It turned out that the so-called olfactory receptor genes are the largest family of genes in mammal genomes. Mice have about 1400 of these genes in a genome of about 25,000 genes. It was discovered that the specificity of olfaction is due to each sensory neuron in the olfactory system producing just one, or in some cases a few, of these many olfactory receptors, each capable of detecting different groups of odorants. How a given chemical "scent" is perceived depends on the combination of receptors that detect it. For cracking the mystery of the genetics of smell, Buck and Axel were awarded the 2004 Nobel Prize in Physiology or Medicine.

The human olfactory genes have been studied in great detail, and it turns out that compared to the mouse, our olfactory genes are nothing to brag about. About half of all our olfactory receptor genes are fossilized and incapable of making functional receptors. The contrast between humans and other mammals is most striking for one class of receptors encoded by the *V1r* genes. The mouse has about 160 functional V1r receptors, while we have only 5 functional *V1r* genes out of more than 200 in our genome. Our repertoire of olfactory receptor genes has gone to pot.

The extraordinary proportion of fossilized olfactory receptor genes suggests that we no longer rely on our sense of smell to the degree that our ancestors once did. Two questions immediately come to mind. First, why have we abandoned the use of such a large fraction of odor receptors? And second, when in evolution did this happen?

Clues to the answers to both questions emerge from studying the fraction of fossilized odor receptors in other primates and mammals. Yoav Gilad and colleagues at the Weizmann Institute of Science in Rehovot, Israel, and the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, surveyed the olfactory gene repertoire in apes, Old World monkeys, New World monkeys, and lemurs and compared it to that of the mouse. They found a striking correlation between the proportion of fossilized olfactory receptor genes and the evolution of full color vision. In mice, lemurs, and New World monkeys that lack full color vision about 18 percent of olfactory receptor genes are fossilized. But in the colobus and Old World monkeys about 29 percent of receptor genes are fossilized, and in nonhuman apes such as the orangutan, chimp, and gorilla, this rises to 33 percent. Finally, in humans they found that 50 percent of this repertoire was fossilized. The fraction of fossil olfactory receptor genes is significantly higher in all species with full color vision. This suggests that the evolution of trichromatic vision—which allows these primates to detect food, mates, and danger with visual cues—has reduced their reliance on the sense of smell. Relaxed selection on the



A Juvenile icefish. The transparent appearance is due to evolutionary loss of scales and red blood cells. (Photograph by Flip Micklin.)

B Adult mackeral icefish, *Champsocephalus gunnari*.

R The dark form of the lion tamarin. The black body coat versus the orange form is determined by the *MC1R* gene. (Photograph by Claus Meyer.)

olfactory receptor genes in trichromatic species has allowed the genes' codes to decay. Conversely, in animals that rely heavily on their sense of smell, the fraction of intact genes is much higher.

There are other physical, behavioral, and genetic signs of this reduced reliance on the sense of smell in humans and other primates. The vomeronasal organ, a cigar-shaped sensory organ located toward the front of the nasal cavity, detects pheromones in most land vertebrates. But this again is greatly reduced in humans and higher primates in comparison with other species. The V1r receptors I mentioned above play a critical role in pheromone detection. So it appears that we are also less reliant on pheromones than other mammals, again perhaps because our ancestors relied more on visual signals in mating and other behaviors.

Because the vomeronasal organ and V1r receptors are so reduced in humans and other higher primates, we might also predict that other machinery involved in the transmission of information from the olfactory system would degenerate. This is exactly the case. Another gene that plays a specific role in vomeronasal organ function is called *TRPC2* and encodes a protein that regulates the trafficking of ions in sensory cells. In mice, TRPC2 is fully functional and required for normal behaviors in response to pheromones. But in humans, and all of the higher primates with trichromatic vision and larger numbers of fossilized olfactory receptors, the *TRPC2* gene contains a bunch of mutations that have rendered it a fossil gene.

The fact that different kinds of genes with different jobs in the nose are fossilized is a very striking and satisfying fulfillment of the prediction about the effect of relaxed selection on species traits. That is, when an entire organ or process falls into disuse, different genes responsible for different steps in a process may all experience relaxed selection and then undergo fossilization. The evolution of the vomeronasal organ and parts of its machinery raises the possibility that whole pathways of genes may become dispensable, decay, and eventually disappear. This is exactly what we see happening in some

species, sometimes on a massive scale. I'll describe two examples from other kingdoms that vividly illustrate how evolution throws away what is no longer useful.

Use It or Lose It

Yeasts and other fungi play important roles in human affairs. We use yeasts to ferment beer and wine and in bread-making, while fungi were the source of the first antibiotics. Because it is so easy to culture, the baker's and brewer's yeast *Saccharomyces cerevisiae* has been a favorite laboratory organism for many years. Through experiments with yeast, much has been learned about how cells grow and divide, how genes are used, and the biochemistry of life.

But there are many more yeast species than just good old baker's yeast. Under the microscope most of them look very similar. However, there are often detectable differences in species' abilities to metabolize nutrients and to grow in different environments. The breakdown of many nutrients into their useful components often occurs through a sequence of steps, or a *pathway*. One of the best studied nutrient pathways in any living organism is the galactose pathway of baker's yeast. Most organisms use the sugar glucose as an energy source. When it is not available, either stored sugars (starches) or alternatives must be used. Baker's yeast can utilize the sugar galactose as an alternative source because it can convert galactose into a usable form of glucose through a series of enzymatic steps. These steps require four different enzymes, encoded by four different genes. Furthermore, to ensure that the yeast makes these enzymes only when they are needed and galactose is available, three other proteins control the making of the enzymes. Altogether then, seven genes are devoted to running the galactose pathway in baker's yeast.

Most close relatives of baker's yeast can also utilize galactose, except for one. This species, *Saccharomyces kudriavzevii* (try saying that quickly several times), was discovered on decaying leaves in Japan, unlike the sugar-rich places where most other yeast species live in the field. When my graduate student Chris Hittinger looked at *S. kudriavzevii*'s seven genes of the galactose pathway, he quickly found out why this species can't utilize galactose—each gene was shot to hell. Each of the seven genes had various-sized chunks of code missing that obliterated the integrity of their text.

There is a very striking contrast between the state of the seven galactose genes and their immediate neighbors in the *S. kudriavzevii* DNA. The neighboring genes are perfectly intact, just as they are in baker's yeast and other related species. If one thinks of each gene's code as about one large paragraph of text, the text of each galactose gene's code in *S. kudriavzevii* is obliterated in many places, but the preceding and following paragraphs encoding other genes are untouched. This pattern reveals how exquisitely specific the fossilization of genes is. A gene that is no longer needed or used accumulates many mutations, while the codes of neighboring genes that are used are perfectly maintained. The fate of genes in *S. kudriavzevii* demonstrates how natural selection maintains what is needed, but it cannot maintain what is no longer needed. This species adapted to living on other sugar sources, and its galactose pathway was no longer needed and fell into disuse. Without the constant surveillance of natural selection to purge the galactose genes of inactivating mutations, the genes fossilized and are well on their way to being erased.

The selective decay of seven functionally related genes is a great example of how a substantial number of genes are abandoned by the relaxation of natural selection, but it pales in comparison to what has happened in some other microbes, such as the species *Mycobacterium leprae*, the pathogen responsible for the disease leprosy.

Sequencing of the *M. leprae* genome revealed that it contains about 1600 functional genes, and almost 1100 fossil genes—an enormous fraction of dead genes, and far greater than that of any other known species. *M. leprae* is closely related to *M. tuberculosis*, the species responsible for pulmonary tuberculosis. But *M. tuberculosis* has about 4000 intact, functional genes and only about 6 fossil genes. The differ-

ence between the two species reveals that *M. leprae* has fossilized or lost about 2000 genes in the course of its evolution. What explains the vast difference in the numbers of functional and fossil genes between the two organisms?

M. leprae has a very different lifestyle from its cousin. It can live only within cells of its host. It resides within cells called macrophages and infects cells of the peripheral nervous system, whose eventual destruction leads to the physical disfigurement typical of the disease. It is the slowest-growing bacteria of all known species (it takes about two weeks to divide, while the *E. coli* in our gut can divide every twenty minutes). Despite decades of effort, it has never been grown on its own in the laboratory. The specialization of living within host cells has allowed *M. leprae* to rely on the host for many metabolic processes. With the host cell genes doing much of the work, this has relaxed selection for the maintenance of many *M. leprae* genes. The massive decay of genes seen in *M. leprae* has also occurred in other intracellular parasites and pathogens. Species with vastly reduced numbers of functional genes demonstrate that a large fraction of all genes can become dispensable upon shifts in organisms' lifestyles.

The fossilization of individual genes, sets of genes in pathways, or larger groups of genes in species has important consequences for the future evolution of their descendants. Because decaying genes generally accumulate multiple defects, their inactivation cannot be easily reversed. This means that the loss of gene functions is generally a one-way street. Once gone, these functions will not return. Just as new species of icefish will not have or use hemoglobin, species that evolve from *S. kudriavzevii* will not be able to use galactose. Gene fossilization and loss imposes constraints on the future direction of evolution in lineages.

"Use it or lose it" is an absolute rule imposed by the fact that surveillance by natural selection acts only in the present—it cannot plan for the future. The downside to this rule is that if circumstances change, even over long periods of time, species that have lost particular genes will not have those genes available to adapt to new circumstances. This may be an important factor in the success or extinction of species. Keep in mind that biologists think that over 99 percent of all species that ever existed are now extinct.

Cause or Effect?

The prevalence of fossil genes offers powerful new means of viewing the process of evolution. However, it also raises questions about cause and effect. Is the fossilization of genes a cause of evolutionary change brought about by natural selection, or is fossilization largely an effect—a by-product of natural selection for other features? The answer appears to be that it can be either, depending upon the circumstances. I will explore this issue with examples of recently evolved fossil genes in a flowering plant and in humans that are most likely examples of cause and effect, respectively.

In plants, flower colors are often used to attract pollinators, particularly bees or birds. There are many well-documented cases of evolutionary shifts in pollinator species. It is easy to imagine how changes in climate or the abundance of pollinators could select for variations in flower colors. Furthermore, because flowers that provide nectar to hummingbirds or bees may be visited by unwelcome pests, there is selection on other structural features of flowers that fine-tune flower anatomy to different pollinators. For example, bird-pollinated species tend to produce larger amounts of nectar and have narrow floral tubes while bee-pollinated species produce small amounts of nectar and have broad floral tubes.

In the morning glory genus *Ipomoea*, the ancestral flower color was blue or purple. This group is typically pollinated by bees, but one species, *Ipomoea quamoclit*, has red flowers and is pollinated by hummingbirds. The red color appears to be an adaptation for attracting hummingbirds.

The production of red, blue, or purple flower color is determined by an enzymatic pathway in morning glories. Beginning with a com-

mon precursor, different sets of enzymes produce either the blue and purple or the red pigments. Recently, Rebecca Zufall and Mark Rausher at Duke University showed that the pathway for making blue and purple pigments has degenerated in the red *I. quamoclit*. One enzyme in the blue/purple part of the pathway appears to be completely impaired while a second is altered such that it can contribute to red but not blue/purple pigment synthesis.

Because the evolution of red flower color is adaptive, and its evolution in the morning glory is most likely directly due to changes in these two enzymes, it appears that gene inactivation in this instance is a cause of evolution. Natural selection may well have favored the inactivation of the blue/purple-promoting enzymes and the evolution of red coloration, as opposed to the enzymes' disuse evolving as a by-product of selection for some other trait.

More often, however, gene inactivation and fossilization is likely to be a consequence of relaxed selection on genes, where gene inactivation is among the last in a series of changes that have rendered a gene dispensable. This is probably the case for the opsin genes I have described and for a fascinating case of gene fossilization that occurred specifically in the human lineage, after our line split off from our last common ancestor shared with chimpanzees.

THE HUMAN GENE that was fossilized is called *MYH16*. In humans there is a two-base deletion in the *MYH16* code that disrupts the proper reading of its text (a relevant portion of the text is shown in figure 5.4; the deletion is indicated by asterisks). In chimps, gorillas, orangutans, and macaques the gene is perfectly intact.

In other primates, the MYH16 protein is made in a subset of muscles, particularly the very prominent temporalis muscle that extends over most of the area of each side of the temporal region of the skull. The large temporalis muscle is involved in movements of the large jaws in apes involved in chewing. In humans, the temporalis region

Human	ATG	ACC	ACC	CTC	CAT	AGC	**C	CGC
Chimp	ATG	ACC	ACC	CTC	CAT	AGC	ACC	CGC
Gorilla	ATG	ACC	ACC	CTC	CAT	AGC	ACC	CGC
Macaque	ATG	ACC	ACC	CTC	CAT	AGC	ACC	CGC

FIG. 5.4. **Fossilization of a human muscle gene.** A short portion of the sequence of the *MYH16* myosin muscle gene is shown. In humans, a deletion of two bases disrupts the code of the gene (asterisks) and is associated with the reduction of two muscles involved in chewing, which are massive in our ape relatives. *Figure by Jamie Carroll, based on data of H. H. Stedman et al. (2004),* Nature *428:416.*

and muscle are much reduced in comparison with gorillas, chimps, and macaques. The MYH16 protein is a myosin that forms part of the large fibers within muscles that generate their force. Human temporalis muscle fibers are much smaller than those in our relatives, and smaller fibers make for smaller muscles.

The intriguing correlation between a mutation in a protein affecting muscle and fiber size and evolutionary changes in the temporalis muscle raises the question of whether the mutation was a cause of the muscle's reduction or a by-product of the muscle's reduction that occurred by other means. This is difficult to say for certain, but we can consider some additional evidence in weighing the alternative explanations. It is known that mutations in this class of proteins can have severe effects on muscles. If a primate with a large jaw (such as that of apes and our ancestors) lost this muscle in one step, it would not be able to chew. In order for the *MYH16* mutation to have played a role in the evolution of the temporalis muscle, we should think of scenarios in which the muscle mass was not lost all at once. I think that the explanation for the fossilization of *MYH16* is more likely to

be similar to the history of globins in the icefish—that is, the muscle was becoming reduced by other genetic pathways and the fossilization of the *MYH16* was most likely a later event after the gene became dispensable.

Fossil Genes as Evidence Against "Progress" and "Design"

The examples described in this chapter demonstrate how the making of the fittest—whether speaking of ancient tribes of fish, magnificent dolphins, colorful flowers, slender-jawed humans, simple yeasts, or blind subterranean rodents—is not necessarily a "progressive," additive process. Modern species are not better equipped than their ancestors, they are mostly just different. They have often gained some coding information in their DNA and, as I have shown throughout this chapter, they have often lost some, or even many, genes and capabilities along the way.

The fossilization and loss of genes are powerful arguments against notions of "design" or intent in the making of species. In the evolution of the leprosy bacterium, for example, we don't see evidence that this pathogen was designed. Rather, we see that the organism is a stripped-down version of a mycobacterium, which still carries around over a thousand useless, broken genes that are vestiges of its ancestry. Similarly, we carry around the genetic vestiges of an olfactory system that was once much more acute than what we have today.

The patterns of gain and loss seen in species' DNA are exactly what we should expect if natural selection acts only in the present, and not as an engineer or designer would. Natural selection cannot preserve what is not being used, and it cannot plan for the future. The fossilization and loss of genes are exactly what is predicted to evolve in the absence of natural selection. Over time, chance mutations will accumulate, eventually disrupting the text of unused or unnecessary genes.

Furthermore, the repetition of gene fossilization in different ancestors of entirely different groups of animals is striking evidence that, when selection is relaxed on a particular trait, the same events will repeat themselves in DNA. The repetition of the independent fossilization of the SWS opsin gene described in this chapter is a profound demonstration of this principle. It is also a foretaste of the broader message of the next chapter, about the predictability and reproducibility of evolution in general, and the many amazing instances in which evolution has repeated itself.