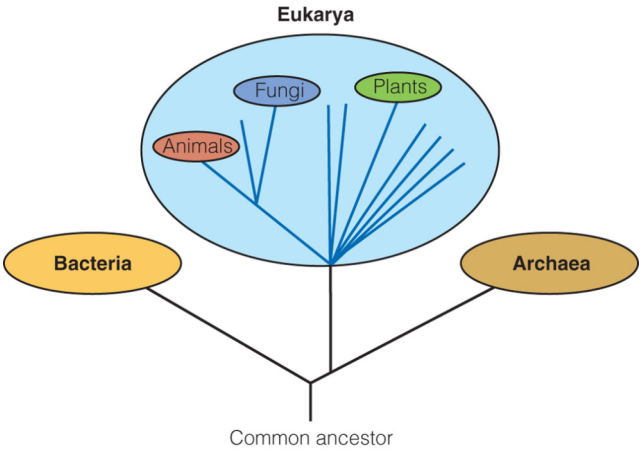
Chapter 18

Evolution and Systematics



KEY CONCEPTS

1. Linnaean taxonomy organizes and groups organisms into a hierarchy of taxa (species, genera, families, orders, classes, phyla, kingdoms, and domains) on the basis of comparisons of traits.

2. Life evolves because DNA can be altered at random by mutation and recombination, leading to organisms with varied characteristics. Some of the variants reproduce more than others because they respond better to environmental factors that impose natural selection.

3. A species can split into two species (speciation) by dividing into populations that evolve separately.

4. By comparing traits in modern and fossil species, phylogenetic systematists seek the paths of evolution (phylogeny) that led to modern taxa. The paths are expressed in diagrams called phylogenetic trees or cladograms.

5. Cladistics provides quantitative, computerized methods for deducing evolutionary relationships among taxa. The cladistic ideal of limiting formal names to monophyletic groups helps make the taxonomic system more logical and predictive.

18.1 THE DIVERSITY OF LIFE

Ranging from oak trees to bacteria and from whales to mushrooms, life takes on a colossal number of forms. Scientists have named about 1.5 million species, which may be less than 10% of the species currently alive. Fossils suggest that a hundred times that many species arose, flourished, and died out in the past. The multitude of species makes up the diversity of life, or **biodiversity**.

How did so many species of life come into being? How can we sort them for study? Did they descend from common ancestors? If they did, how are modern species related? We have many answers, but much is still unknown. Currently, a young science called **cladistics** is reshaping old ideas and bringing new insights. This is a great time to be studying the diversity of life.

The most relevant fields of study regarding the diversity of life include taxonomy, evolution, and systematics. **Taxonomy** is the process of sorting and naming the multitude of life forms--a practical task of great importance, for without an orderly system of names, biologists could not tell each other which organisms they study. **Evolution** is the process by which living species change and new species come into being. **Systematics** is the effort to find how modern life forms are related. Today, most systematists look for the evolutionary steps that led from ancient to modern forms of life. Those steps are called **phylogeny**, a term that means "the origin of groups." This chapter introduces all these

fields of study, beginning with the question of how we distinguish between species in the first place.

18.2 DEFINING SPECIES

If you spend time in a wild area, you find that organisms are not infinitely variable--there are recognizable kinds, such as the orange poppies in Figure 18.1. Each kind is a distinct species--a group of organisms that are more closely related to one another than to organisms of any other kind. Members of a given species may also look more like one another and interbreed more freely with one another than with organisms out side the group.

Figure 18.1. Several kinds of wildflowers are easy to distinguish in this California field. Each kind is a different species.

Most currently known species were defined by combinations of traits, or characters--aspects of organisms that range from shapes and colors of body parts to the presence of specific molecules, such as the oils that give mustard its acrid flavor. Among the characters, details of DNA have become increasingly important in the last few decades.

Species that are defined by combinations of traits are called **phenetic species**. Citrus trees are an example. Based partly on distinctions between their fruits, orange trees are assigned to the species *Citrus sinensis*, lemon trees to the species *Citrus limon*, and grapefruit trees to the species *Citrus paradisi*. Few people would mistake one of the fruits for another; they differ in characteristic ways, as do the trees that bear them.

With *Citrus* species as examples, we can discuss the difficulties in defining species. Let us begin with a practical problem: given an individual organism, how can we tell whether it belongs to a particular species? One way is to compare the organism with a specimen that a trained taxonomist identified and placed in a museum or botanical garden. The final authority is a *type specimen* that was placed on file when the species was first named. But your specimen will differ somewhat from the museum sample. How much can organisms differ without being separate species? The answer depends on how much natural variation occurs within the species. Differences occur with age, and environmental factors such as climate can affect body form. Individuals of the same

species can even differ slightly in DNA, leading to variations such as those that distinguish one human being from another. Thus, to define a species we need many specimens, of various ages and from varied environments, to display the range of variation within the species.

The *Citrus* example illustrates yet another problem in defining species. The boundaries drawn depend on the traits used for comparison. Yet there is no specific rule about what kinds of traits must be considered and how many traits must differ between two populations to consider them as separate species.

To avoid these difficulties, many biologists prefer to define species by means of a *mating test*: if organisms form two populations mate and produce fertile progeny (offspring) under natural conditions, then the two populations belong to the same species. Species defined by the mating test are called **biological species**. The mating test can have surprising results at times. Two populations may turn out to be different biological species even though they look alike, because they cannot mate and produce fertile offspring. In other cases, two populations belong to the same biological species despite having many differences. This is true of the many breeds of dogs, all of which belong to the same biological species

But the mating test has its own problems. Most importantly, it does not apply to organisms that lack sexual reproduction. Examples include the thousands of species of bacteria and fungi. By the mating test, each individual in an asexual population would be a separate species, a conclusion that would defeat the purpose of taxonomy. Then, too, extinct organisms that are known only from fossils cannot be subjected to a mating test. For life forms that are extinct or entirely asexual, species can only be defined phonetically, by comparing combinations of characters.

A second problem with the mating test is that many plant species can interbreed with closely related species, producing progeny that are weakly fertile. How fertile must they be to pass the mating test? Any dividing line that might be set would be arbitrary and controversial.

Because of such difficulties, the assignment of life forms to species is always open to debate and change. Nevertheless, the great majority of named species are well accepted. They provide a sound basis for exploring life's diversity and how it came about.

18.3 TAXONOMY: NAMING LIFE FORMS

People have always grouped familiar organisms into species that are given informal names, such as "cats." But common names vary, so that a certain name may refer to different species in different locations. For example, the bird called the robin in the United States is a completely different species from the robin of England. For scientific communication, we need names that everyone accepts, and a formal system for assigning new names. This is the function of taxonomy.

The Linnaean System Is Hierarchical

The taxonomy in common use today is based on a hierarchy, meaning that it has levels, and the groups at one level are nested within groups at the next level. The taxonomic hierarchy began with the work of a Swedish physician and botanist named Carolus

Linnaeus. In 1753, Linnaeus published a book in which he named about 6,000 species of plants and assigned them to 1,000 groups called *genera* (singular, *genus*). A **genus** is a group of species that are similar enough to be obviously related, as in the *Citrus* example.

For each species, Linnaeus wrote a short descriptive phrase in Latin. He regarded this phrase as the formal species name, but for convenience, he also wrote a single word in the margin that could be combined with the genus to give an abbreviated two-word name, the *binomial*. Currently, every species is given a binomial, or **species name**, which is always printed in italics. The first word (always capitalized) is the genus. The second word (never capitalized) is the **specific epithet**.

Linnaeus also saw that two or more genera can resemble one another more than others, making it possible to group genera into larger sets. Genera with similar traits make up a **family**. For instance, all the plants that have rose-like flowers belong to the family *Rosaceae*. This family includes garden roses, cherry trees, almond trees, and many others. Modern taxonomists group families into **orders**, orders into **classes**, classes into **phyla** (singular, *phylum*, also called **divisions**), phyla into **kingdoms**, and kingdoms into **domains**. Table 18.1 illustrates these categories. There you see that the names at certain levels have standard endings.

Table 18.1 Linnaean Classification Illustrated by the Common Garden Nasturtium (*Tropaeolum majus*)

Linnaean Rank	Name	Ending
Domain	Eukarya	*-a*
Kingdom	Plantae	*
Phylum (division)	Magnoliophyta	*-ophyta*
Class	Magnoliopsida	*-opsida*
Subclass	Rosidae	*-idae*
Order	Brassicales	*-ales*
Family	Brassicaceae	*-aceae*
Genus	*Tropaeolum*	*
Species name	*Tropaeolum majus*	*
Specific epithet	*majus*	*

*Names at this level do not have a consistent ending.

You might think the levels in Table 18.1 would be enough to classify all the forms of life, but taxonomists need extra levels to divide up the multitude of species. For example, a family may be divided into subfamilies, and several families may be groups into a superfamily. A species can be divided into subspecies, varieties (or races, among animals), and forms. Subdivisions below the species level are important to cultivated plants, where breeding programs have led to variants called cultivars. The term *cultivar* is equivalent to

variety, but it is only used to describe products of human selection within a species. Cabbage, cauliflower, Brussels sprouts, and broccoli are cultivars of the same species, *Brassica oleraceae*.

In discussing taxonomy and systematics, there often is a need to speak of taxonomic groups in general. Here the term **taxon** (plural, *taxa*) is useful. A taxonomic group at any level is a taxon. For instance, a species is a taxon, and a kingdom is another taxon.

Biologists Define Domains and Kingdoms of Life

The taxonomic system is a work in progress, changing every year as new discoveries challenge old boundaries. Nothing illustrates this point better than the kingdom concept.

Until the nineteenth century, scientists placed all forms of life into two kingdoms: animals, which move actively and consume prey; and plants, which do not. But then biologists found that some microscopic organisms combine animal-like mobility with the plantlike ability to take energy from sunlight. Also, studies of cell structure and metabolism showed that fungi (mushrooms and their relatives) have more in common with animals than plants, even though fungi traditionally had been grouped with plants. Finally cells of plants resemble those of animals much more than they resemble bacteria, even though bacteria traditionally had been put in the plant kingdom. In response, early twentieth-century biologists divided the old plant kingdom into four new kingdoms. Bacteria and similar organisms were assigned to kingdom *Monera*, fungi went into kingdom *Fungi*, the green plants remained in kingdom *Plantae*, animals in kingdom *Animalia*, and all other forms of life were placed in a catchall kingdom *Protista*.

In the mid-twentieth century, studies with the electron microscope confirmed the wisdom of giving bacteria their own kingdom, showing that bacteria have much simpler cell structures than other life forms. Among other complexities, the cells of plant, animals, fungi, and protests are *eukaryotic*, meaning that they enclose most of their DNA in a membranous envelope, forming a true nucleus. Cells of bacteria have no envelope around their DNA and are *prokaryotic*.

In the late twentieth century, biologists toppled kingdom Monera when they began to compare the DNA of various organisms. Led by American bacteriologist Carl Woese, they found that prokaryotes include two distinct groups of organisms, different enough to have evolved separately since the early days of life (Fig. 18.2). The main evidence came from studies of the DNA that tells cells how to make ribosomes, the organelles that build proteins. Each ribosome contains several molecules of RNA, and the DNA that specifies ribosomal RNA is called rDNA. Woese's group compared rDNA from many organisms and found far more variation within each group of prokaryotes than among all plant, animals, and fungi (Fig. 18.2).

Biologists had long known that prokaryotes vary in metabolic abilities far more than animals do. That fact, and the variation in DNA, suggested that each of the two prokaryotic groups should have the same taxonomic status as all eukaryotes put together. But the kingdom concept was already firmly associated with animals, plants, and fungi--groups within the eukaryotes. Thus, to express the diversity of prokaryotes, biologists had to invent a new, higher category above the level of the kingdom: the domain, a group that contains one or more kingdoms.

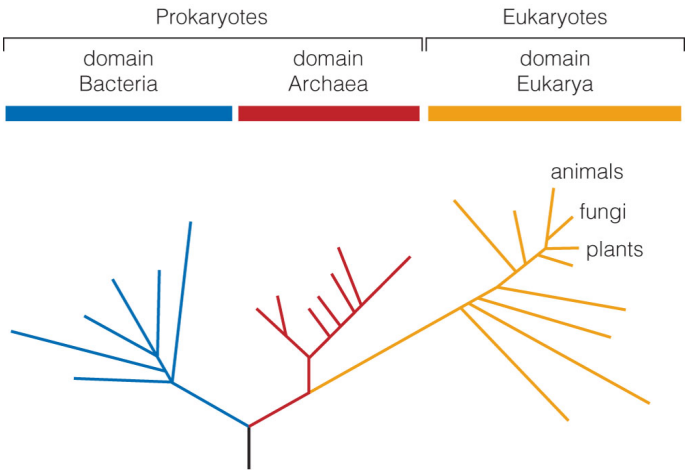


Figure 18.2. Variation in rDNA between prokaryotes and eukaryotes, supporting the domain concept. Animals, plants, and fungi are at the short line segments as labeled. Other terminal lines represent other groups included in the study. Their unfamiliar names are omitted for simplicity. The difference in rDNA between two groups is proportional to the length of line segments between them. Bacteria (*blue*) and Archaea (*red*) are prokaryotic. Eukarya (*orange*) are eukaryotic.

Currently, biologists divide life forms into two prokaryotic domains (*Bacteria* and *Archaea*) and one eukaryotic domain (*Eukarya*). Animals, plants, and fungi are well-established kingdoms within the Eukarya, but the old "kingdom" Protista has fallen apart, leaving many smaller groups that do not fit into the three established kingdoms (Fig. 18.3). Scientists are currently debating the status of the smaller eukaryotic groups. Two of them, the *Heterokonts* (also called *Stramenopila*) and the *Alveolates* (see Chapter 21), have been proposed for kingdom status, but the issue is not yet settled. Likewise, it remains to be seen how the domains Bacteria and Archaea will be divided into kingdoms.

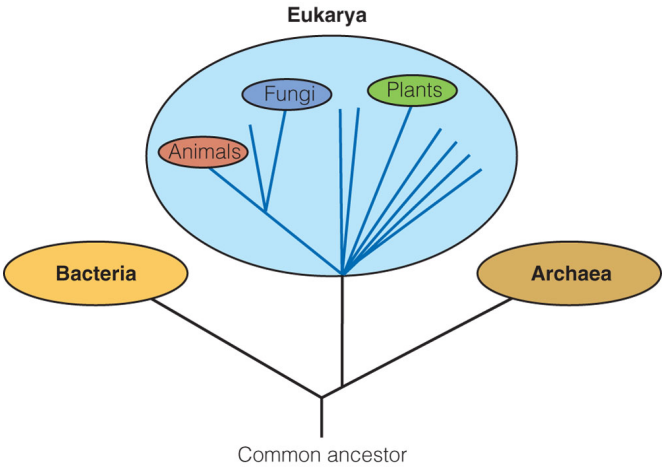


Figure 18.3. The domains of life (Bacteria, Archaea, Eukarya). Domain Eukarya contains three accepted kingdoms (Animals, Plants, Fungi) and many smaller groups that are not currently given kingdom status (*blue lines*). These smaller groups were previously assigned to an artificial kingdom called Protista and are still called "Protists" for convenience, even though the kingdom Protista is no longer recognized.

Modern biologists widely agree that all life probably evolved from a single common ancestor, and the most logical classification would group organisms by their evolutionary relationships, just as we group humans into families on the basis of parent-child relationships. We now have tools (see later in the chapter) to deduce the paths of evolution that led from the original ancestor to the millions of modern living species. Because evolution is the cause of all this diversity, the next section discusses the study of how and why life evolves.

18.4 EVOLUTION: THE CAUSE OF LIFE'S DIVERSITY

Through most of recorded history, scholars viewed life's diversity as a divine plan of creation, in which the Creator established an unchanging ideal body form for each species. One belief was that growing organisms strive to match the ideal patterns, and they differ because the physical world is imperfect. The scholar's goal was to deduce the ideal pattern for each species by examining the common features of individuals.

This way of thinking began to recede in the nineteenth century as scientists explored fossils, relics of life such as bones and leaves that are embedded in rock. Observing the way rocks form today, scientists concluded that more deeply buried fossils were formed earlier in history and thus were older. Older fossils differ from more recent ones, challenging the view that each species is unchanging. The difference between dinosaurs and modern reptiles is a familiar example, but equally great changes have taken place among plants. For instance, all the modern marsh plants called horsetails (genus *Equisetum*) are herbs. But 300 million years ago, there were tree-sized horsetails with secondary growth and wood; and 600 million years ago, there were no land plants at all. The only photosynthetic organisms were prokaryotes and algae.

Darwin Proposed a Model for Evolution

By the 1850s, fossil studies had made a convincing impression that life's history emphasizes change rather than constancy. Consequently, many scientists were ready to believe that the hereditary characteristics of species could change, or evolve, over the course of many generations. All that scientists needed was a clear idea of how such evolution might occur. Charles Darwin and Alfred Wallace, who were English naturalists, came up with the required idea at about the same time. Darwin's ideas about the mechanism of evolution began to take shape during a trip around the world, which began when he was 22 years old. Many sights, such as fossils and an earthquake in South America, influenced Darwin's thinking, but a brief stop at the Galápagos Islands, 900 km off the coast of Ecuador, made the strongest impression. There he found 14 species of birds that resembled finches he has seen in Ecuador, but they also differed in many ways. It occurred to Darwin that finches could have migrated from the mainland to the islands long ago, where they encountered new conditions that somehow altered their heredity. Here was born the important idea that the environment could change the patterns of heredity in a species, shaping life into new forms.

Darwin resolved to keep his thoughts to himself until he could present a convincing argument, and many years passed as he carefully built has case. Meanwhile, Alfred Wallace built up similar ideas in the rain forests of Southeast Asia. When Darwin received a letter from Wallace that revealed the coincidence, Darwin knew the time had come to publish. He presented his ideas and arguments in *The Origin of Species,* published in 1859. In his book, Darwin proposed a mechanism of evolution based on the following four assertions:

1. Changes in heredity occur in the individuals of a population, leading to varied progeny.

2. Populations produce more progeny that the environment can support. This leads to competition among the progeny.
3. The progeny that are best adapted to the environment will reproduce most abundantly. **Natural selection** was Darwin's term for this environmental effect.
4. Repeated over many generations, the preceding three factors could lead to great changes in heredity and hence, great changes in the forms of life.

Darwin's arguments were so convincing that most biologists adopted his view within a few decades. His ideas suggested that there is no ideal body form for each species, and that forms can change as the environment changes. Among scientists, the quest for perfect patterns was replaced by a search for paths of evolution that led to modern forms. It was a revolutionary change in the way scientists thought about the diversity of life. And as knowledge grew, scientists refined Darwin's ideas to build the more complete view of evolution that is present below.

Mutations and Recombination Alter Heredity

To explore current ideas about evolution, this section begins with how and why hereditary variations occur within every species. Darwin sharpened his own ideas by experiments with pigeon breeding, which clearly showed that organisms of the same species can vary in ways that are passed on to future generations, and that new variations arise from time to time. Little was known about heredity in those days, so Darwin could only guess at how variations might occur. Now we know that hereditary information is stored in molecules of DNA, and we know how cells copy DNA and pass it on to progeny. We also know how information in DNA can be changed--indeed, scientists change it daily to create new hereditary treaits in the quest for knowledge and more productive plants and farm animals. The two main sources of change in DNA, called mutation and recombination, are discussed in the following sections.

MUTATION Many events can cause changes at random points along a DNA molecule. These random changes in DNA, called **mutations**, are vital to evolution. They are the primary source of new hereditary information. To understand mutation, a brief review of the way DNA stores information is necessary.

DNA is a giant molecule consisting of two paired chains of many subunits called nucleotides. There are four kinds of *nucleotides,* which differ in one of their components, called a DNA base. Information is stored in the sequence of bases along DNA, just as a book stores information in the sequence of letters. The information in a book is organized as a series of messages. The same is true of DNA, where each message is a segment of DNA called a *gene*. Hundreds of thousands of genes occur along each DNA molecule. The messages stored in genes are concerned with making protein molecules that build the body.

With this background, we can explore the way mutations occur. Some mutations take place when cells copy their DNA before dividing. The copying process is extremely accurate, but heat energy keeps the molecules in motion, causing collisions that sometimes lead to insertion of the wrong base at some point along the copy. Such

changes, called base substitutions, are much like errors you would make if you typed a page from a book (Fig. 18.4).

Figure 18.4. A base substitution, the simplest kind of mutation. For simplicity, only the bases along each segment of DNA are shown, with the letters A,C,G, and T. In the copy base G substitutes for the original base A.

The proteins that copy DNA can correct many base substitutions, but some are missed. As a result, one out of every billion nucleotide pairs is miscopied each time DNA is copied. The errors occur at random locations, so any one spot in DNA is only rarely miscopied. But a human cell contains about 6 billion nucleotide pairs, so there are likely to be several new mutations every time a human cell copies its DNA. When such errors occur in the DNA of reproductive cells, the altered gene may produce new hereditary characteristics in the progeny (Fig. 18.5).

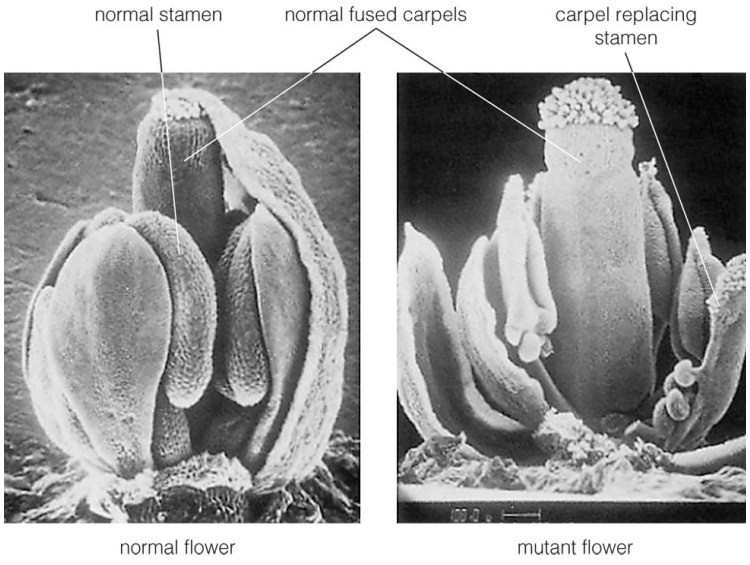


Figure 18.5. Flowers from normal and mutant plants of the species *Arabidopsis thaliana*, which is widely used in research on plant gene control. Outer parts of flower buds were removed to show female structures (carpels) and male structures (stamens). In the mutant, carpel-like structures occur where stamens would normally form. The mutant trait resulted from changes in a single gene, which was altered by treating parental plants with a chemical mutagen.

Agents that cause mutations are called **mutagens**. For the base substitutions described previously, body heat was the mutagen. But other mutagens, coming from the environment or the body's metabolism, can greatly increase the frequency of copy errors, or cause more extensive changes. Some mutagens are high-energy *radiations*, such as dental X-rays, ultraviolet light from the sun, and high-energy particles that are released when nuclei of radioactive elements decay. Radiations strike DNA at random locations and sometimes break the DNA or cause bases to fuse together. Cells have proteins that repair some of the changes, but sometimes the repair is imperfect and leads to extra or

missing segments of DNA or wrong connections when multiple breaks occur and the ends are reconnected. Other mutagens are *chemicals* that occur in foods, industrial products, and the natural environment. An example is benzopyrene, which occurs in smoke and causes errors when DNA is copied. Even normal metabolism makes mutagens, such as small amounts of highly unstable molecules called reactive oxygen species. In summary, mutagens strike at random locations along the DNA, and they normally strike any one gene only rarely. But mutagens are always present; therefore, a few mutations may occur almost every time DNA is copied.

Occasionally, a mutation helps the organism and spreads through the population, contributing to evolution. Without these rare positive mutations, all organisms would still be nearly identical to the original forms of life on Earth. A familiar example of positive mutation is the appearance of antibiotic resistance in bacteria that cause human disease. When an antibiotic is applied, any resistant bacteria rapidly replace more vulnerable bacteria. This example shows that the value of a mutation may depend on the environment, because bacteria that spend energy on making resistance genes would probably die out in a world that lack antibiotics.

Although some mutations are beneficial, most mutations have little or no effect on evolution because they cause damage that leads to their elimination. Most genes have been perfected by millions of years of evolution, so random changes in a gene usually detract from the quality of the message, just as random typographical errors detract from the information in a book. Sometimes changing a single DNA base has no effect at all. But if the change occurs at a critical point in the gene for a vital protein, the mutated cell will make bad copies of the protein, leading to cell death. Worse yet, when mutations damage genes for proteins that control cell division, the cells may multiply without limit and produce tumors and, in animals, cancers.

RECOMBINATION Large changes in the forms of life, such as the evolution of flowering plants from mosses, take an immense length of time because they require positive changes in many genes--and positive mutations are rare. But evolution would be even slower without **recombination**, a process that creates new combinations of genes by joining parts of DNA molecules from separate organisms. Recombination can quickly produce valuable combinations of genes that would take thousands of generations to form by mutation alone.

Recombination occurs in several ways. In *transduction*, viruses sometimes carry DNA of one host organism to another. In *transformation*, bacteria take up segments of DNA that are released from decaying organisms, and enzymes insert compatible portions of the foreign DNA into the cell's own DNA. In *conjugation*, some bacteria pass a copy of their own DNA into another bacterium of the same species, and enzymes exchange parts of the host's own DNA for some of the transferred DNA. But in cells of eukaryotes, *sexual reproduction* is the most common source of recombination.

Figure 18.6 briefly outlines the way sexual reproduction recombines genes. Sexual reproduction brings together two haploid reproductive cells (*gametes*), each carrying one complete set of chromosomes from a parental organism. Two gametes fuse to become a diploid cell (the *zygote*) with two sets of chromosomes. The zygote grows and divides to make an adult body with diploid cells. When the diploid organism prepares to reproduce, some of its cells go through a special division process called *meiosis*, during which the

chromosomes first double and line up side by side. With the DNA of both parents side by side, enzymes break the chromosomes at precisely the same point and reattach the broken ends in such a way that the chromosomes exchange segments. This event, called *crossing over*, happens at many random points along most chromosomes. As a result, each chromosome emerges with the proper set of genes, but some segments come from one parent and some from the other parent. Then two successive divisions produce four haploid cells that become gametes with new combinations of genes. Because the multiple crossover points are randomly located, no two gametes are likely to have the same combination of parental chromosome segments.

Even if there were no such thing as crossing over, meiosis would yield considerable recombination, because each pair of chromosomes lines up at the cell's equator before meiosis without regard to the orientation of parental chromosomes. As a result, any two haploid cells emerging from meiosis are likely to have different combinations of parental chromosomes. However, crossing over yields far more recombination.

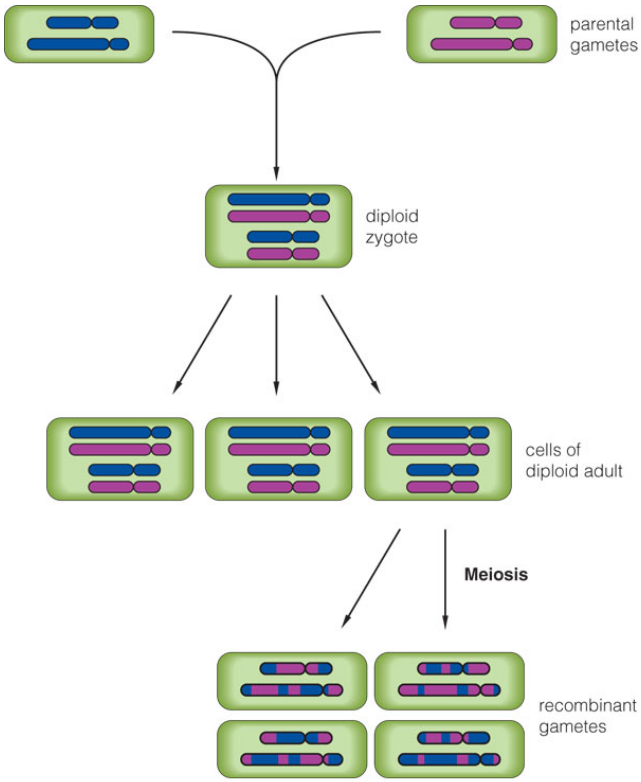


Figure 18.6. How sexual reproduction recombines DNA from two organisms, illustrated with a species that has two kinds of chromosomes. To keep track of parental origins of DNA segments, chromosomes from the two parental gametes are shown with different colors.

HYBRIDIZATION As an agent for bringing together useful genes, sexual reproduction is limited in that many organisms mate only with members of their own species. This is particularly true in animals. But among plants, mating often occurs between members of different species. When two species mate successfully, the process is called **hybridization**, and the progeny are **hybrids**. Usually, only species of the same genus (such as *Citrus*) can hybridize; less closely related species, for instance from different genera, usually differ too much for hybrid progeny to survive.

Hybrid plants often cannot reproduce sexually because the mismatch between chromosomes disrupts meiosis. Nevertheless, hybrids can be vigorous and may multiply by asexual reproduction. Where two related species coexist, many hybrids may occur because hybridization occurs repeatedly.

Hybrid pants sometimes transfer genes between the two parent species, a process called **introgression**. Such transfer requires **back-crossing**, in which a hybrid plant mates

with a member of one parent species. Back-crossing is possible if some of the hybrid's meiotic divisions produce viable gametes that are similar enough to those of the parent species. When that happens, the parent species acquires genes the hybrid brought in from the other species.

Biologists agree that hybridization has been important in the evolution of some plant species, but it is uncertain how often it occurs among plants on the whole. One recent study in five parts of the world revealed hybrid populations in only 6% to 16% of wild plant genera. However common or uncommon it is, hybridization and introgression can be important because of fears that it may allow genes from genetically engineered plants to escape into wild populations. Also, plant breeders use hybridization and backcrossing to improve crops. For example, commercial varieties of tomatoes produce abundant fruit but have lower resistance to certain diseases than wild species of the same genus from Peru. Tomato breeders have been hybridizing and back-crossing these species to make highly productive tomato plants that also have high disease resistance.

ENDOSYMBIOSIS Strong barriers, such as flight responses or chemical antagonisms, usually prevent very different organisms from merging their DNA. These mechanisms avoid wasting reproductive resources on progeny that have fatal conflicts between parental genes. However, DNA from very different organisms may come together in an association called **endosymbiosis**, in which cells of one species reside inside cells of another species. For example, certain bacteria live inside cells of host organisms. In most cases, both organisms retain their genetic identity. But if endosymbiosis lasts for many generations, DNA can pass from the guest to the host, adding to the host's nuclear DNA and leaving the guest as a dependent organelle. This can give the host cell important new capabilities.

Strong evidence indicates that endosymbiosis led to the energy-processing organelles called *mitochondria* and *chloroplasts*. Mitochondria occur in nearly all eukaryotic organisms, where they release food energy by using respiration to oxidize fuel molecules. Chloroplasts, found in plants and some "protists," make sugars by using light energy in *photosynthesis*.

Biologists have concluded that mitochondria and chloroplasts evolved from endosymbiotic bacteria that were capable of respiration and photosynthesis. Mitochondria may have evolved as shown in Figure 18.7. The process began when free-living bacteria, which had evolved respiration, took up residence inside primitive eukaryotic cells that could draw energy only from the less efficient process of fermentation. The bacteria paid for shelter by giving off energy-rich molecules that the host could use. In time, most (but not all) of the bacterial DNA was transferred to the nucleus. The transfer left the residents unable to survive outside the host but able to reproduce with help of the nucleus. The bacteria had become mitochondria.

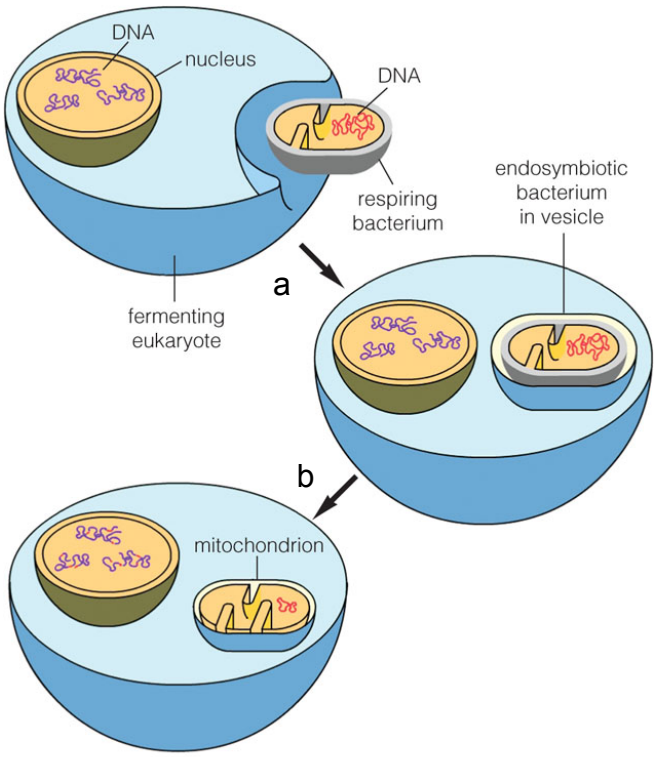


Figure 18.7. Proposed endosymbiotic origin of mitochondria. (a) A respiring bacterium took up residence inside a primitive eukaryotic cell, where it divided along with the host cell. (b) Descendants of the respiring guest bacterium lost their walls and most of their DNA, to become mitochondria. Some of the lost DNA was transferred to the nucleus. Similar events led to the endosymbiotic origin of chloroplasts.

Later, some of the respiring eukaryotic cells engulfed photosynthetic bacteria, giving the host cells an ability to use light energy. These bacteria transferred most of their DNA to the nucleus and became chloroplasts. The origin of mitochondria and chloroplasts from bacteria is called *primary endosymbiosis*. Still later, other eukaryotic predators gained chloroplasts through endosymbiotic partnership with eukaryotes that already had chloroplasts. This *secondary endosymbiosis* lad to brown algae and certain other protists (see Chapter 21).

By merging DNA from verydifferent organisms, endosymbiotic associations can open the way to major innovations and great increases in the diversity of life. Mitochondria and chloroplasts demonstrate this point. Mitochondria gave eukaryotes the respiratory energy needed to grow large and complex, and chloroplasts allowed some eukaryotes to evolve into algae and land plants.

Natural Selection Guides Evolution

For large changes in the forms of life, such as the evolution of trees from moss-like ancestors, evolution needs a guiding agent that is much less random than mutation and recombination. Darwin and Wallace proposed that selective agents in the environment guide evolution by influencing the reproductive success of variant organisms within each population.

To prove that selection can lead to evolution, Darwin cited the artificial selection that farmers practice to improve livestock and crop plants. For thousands of years, farmers have used the most productive animals and plants for breeding. The farmer acts as a *selective agent*, and the favored trait is high productivity. Darwin proposed that the environment has its own selective agents, which impose natural selection on wild species.

Natural selective agents include nonliving (*abiotic*) factors such as climate, water supply, and light, as well as living (*biotic*) factors such as competing organisms, predators, and prey. Consider a field where tall and short plants grow in the presence of grazing animals. Tall herbs attract attention and often are bitten, whereas shorter plants often escape notice. In this way, the selective agent, the grazing animal, gives small size a selective advantage.

Multiple selective agents may act at the same time, sometimes affecting the same trait in opposite ways. For example, tall plants capture more sunlight and spread seeds farther. Here the selective factors include light and agents that disperse seeds and affect seedling survival. Their influence favors tall size and offsets the negative effect of attracting grazers. We can expect most cases of natural selection to involve such compromises among effects of selective agents. It is therefore unwise to settle on any simple idea of why a given trait evolved without conducting careful experiments to explore the effect of many selective agents.

DIRECTIONAL SELECTION Because of natural selection, prolonged exposure to a stable environment can cause a species to accumulate hereditary traits that enhance success in that environment. Such favorable traits are called **adaptations**. But once the population is fully adapted, any further changes--such as new mutations--are likely to have a negative impact, lessening the degree of adaptation. Now the same agents that led to progressive evolution tend to keep the species stable. To acknowledge these facts, biologists distinguish between **directional selection**, which leads to new adaptations, and **stabilizing selection**, which maintains existing adaptations.

Directional selection can be illustrated by events that may have happened as cacti were adapting to life in deserts. Modern cacti have spines in positions were other plants have leaves, suggesting that spines evolved by modifying genes that originally called for broad leaves (Fig. 18.8).

To illustrate the components of directional selection, Figure 18.8 shows 1 year of an adaptive process that would have taken many generation to be complete. Figure 18.8a shows haw leaf width varied in the parental population. In Figure 18.8b, mutation and recombination have led to young plants with more variation. But in the desert, a selective agent, very dry air, acted against progeny with broad leaves, which lose much water. Progeny that reached reproductive age under these conditions had narrower average leaf widths than their parents (Fig. 18.8c). Repeating these events over many generations, this directional selection could have narrowed the leaves to the width of spines.

You may wonder why adaptation to the desert did not eliminate leaves entirely. Part of the answer is that spines deter animal attack. Recent studies also suggest that spines help a cactus to collect rainwater by channeling the flow of water. These ideas suggest that Figure 18.8 shows just one of several simultaneous selective effects that led to the conversion of cactus leaves to spines. Also, since spines lack photosynthesis (that are dead at maturity), all these changes would have been fatal without the previous or simultaneous transfer of photosynthesis from green leaves to the stem.

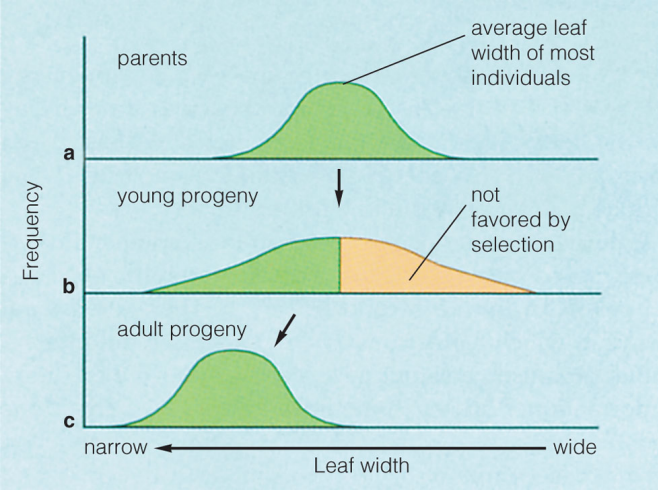


Figure 18.8. Directional selection, illustrated by adaptive narrowing of cactus leaves (ultimately converting leaves to spines) in response to desert conditions. (a) Variation in leaf width among parental plants. (b) Variation in leaf width among young progeny. Mutations and recombination expanded the range of variation, but directional selection in the desert acts against progeny with broad leaves (*orange*). (c) Variation in leaf widths among progeny that reach reproductive age. The average leaf width is less than in the preceding generation.

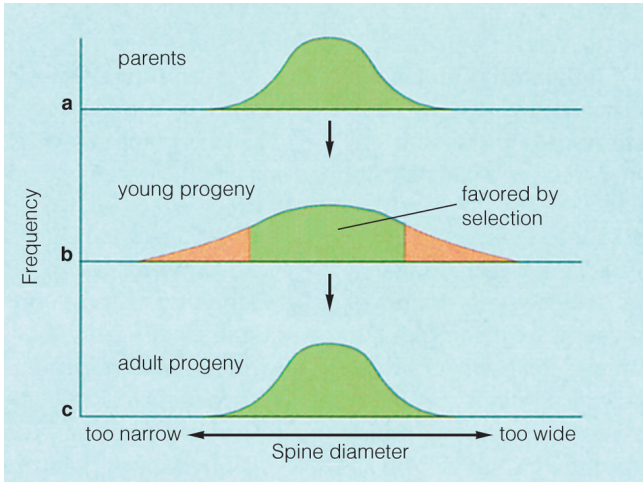


Figure 18.9. Stabilizing selection, illustrated by preservation of the most adaptive spine diameter in cactus plants. (a) Variation among parental plants. (b) Variation among young progeny. Mutation and recombination expanded the variation, but natural selection favors progeny with spine diameters most like the parents (*green*). (c) Variation among progeny that reach reproductive age. The final variation in spine diameter is the same as in the preceding generation.

STABILIZING SELECTION When directional selection has brought a trait to the best condition for the prevailing environment, the same forces that previously caused change will now exert stabilizing selection, which holds the trait in its present condition. For example, suppose a cactus population in the desert already has highly adapted spines. Narrower spines are too weak to protect the plant, and broader spines waste matter and energy on excessive strength. In every generation, mutation and recombination produce some offspring with spines that are too narrow or too broad, but natural selection acts against them. Since any change from the current mean is harmful, selective forces now act equally against variants on both sides of the mean, leaving the next generation of adults with the same average spine diameter as the generation before (Fig. 18.9).

DIVERSIFYING SELECTION Both directional and stabilizing selection tend to reduce the variety within a population. But too much loss of variety can be costly. This is well known in agriculture, where some crops and so genetically uniform that local diseases can easily become major epidemics. A fine example is the black Sigatoka disease of banana. Caused by a fungus that infects both leaves and fruits, the disease was first seen in Fiji in 1963, but

it spread worldwide by the 1970s because most commercial banana trees are of a single uniform cultivar with low resistance to the disease.

In nature, the very fact that genetic uniformity results in epidemics gives a selective advantage to variants that are not susceptible. The result is a form of natural selection that increases genetic diversity in a population--the opposite of the effect that comes from directional and stabilizing selection. This **diversifying selection** is defined as natural selection that increases genetic variation. It can be caused not only by disease agents, but also by other factors that favor two or more distinct types in a population.

Diversifying selection has been well studied in a grass (*Agrostis tenuis*) that grows on lead-mine tailings in Wales. The mine tailings are piles of rock that are rich in lead and zinc. *A. tenuis* plants occur on mine tailings and also on surrounding normal soil. Plants on mine tailings and plants on normal soil belong to the same biological species; they can, and do, exchange pollen and produce fertile progeny. But the progeny vary widely in tolerance for lead and zinc. Plants that thrive on normal soil have low tolerance; they fail when transplanted to mine tailings. Likewise, plants that thrive on mine tailings grow poorly on normal soil. Only the presence of mine tailings beside normal soil permits tolerant and intolerant plants to persist simultaneously in the population.

The mine tailings have existed for less than 200 years, so we infer that before then, all the plants grew only on normal soil where stabilizing selection kept their metal tolerance low. From that starting point, Figure 18.10 shows how diversifying selection may have increased metal tolerance within the population.

For simplicity, Figure 18.10 compresses adaptive evolution that probably spanned several decades into one generation. Before adaptation is complete, diversifying selection is a special case of directional selection, differing only in that selective agents favor more than one type. There are two results: the adult progeny are more diverse than the parents, and their mean metal tolerance is grater than in the parents. Figure 18.10 does not show what happens when the population is fully adapted, but it can easily be inferred. If mine tailings persist together with normal soil, diversifying selection becomes a special case of stabilizing selection. Mine tailings and normal soil will continue to favor two tolerance levels, trimming extremes and

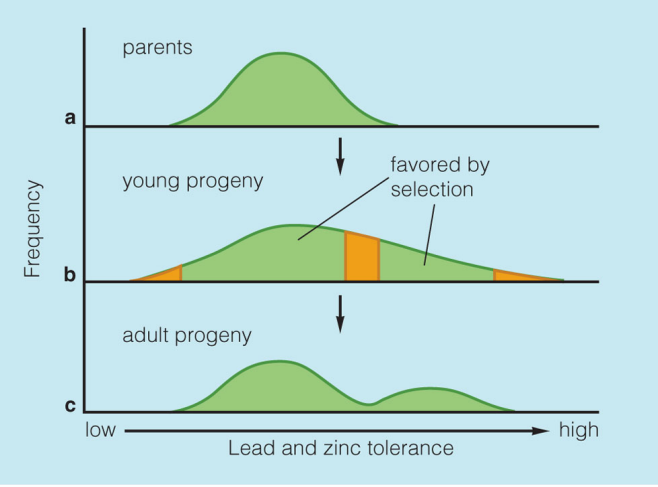


Figure 18.10. Diversifying selection, illustrated by the evolution of metal tolerance where lead-mine tailings cover some normal soil. (a) Variation among parental plants before mine tailings were present. (b) Variation among young progeny. Mutation and recombination broadened the range of variation, but growth on mine tailings selects for plants with high metal tolerance, whereas growth on normal soil selects for plants with low tolerance. (c) Variation among progeny that reach reproductive age. The progeny vary more widely than the parents.

intermediate types. The adapted population's broad variation in metal tolerance will be maintained from one generation to the next.

DIVERGENT EVOLUTION The concept of directional selection helps to explain the great diversity of life on Earth. If a population moves to a new environment, directional selection will gradually change the population's hereditary traits in ways that better fit the new environment. The population will come to differ from related populations that remained in the old environment. This increase in genetic differences among groups is called **divergent evolution**. The cacti and their close relatives provide an example. Where climatic change converted moist habitats to deserts, broad-leafed plants evolved water-saving adaptations such as fleshy green stems and leaves that are reduced to spines (Fig. 18.11a). These desert-adapted cactus plants diverged strongly from broad-leafed relatives that still thrive in moist tropical forests.

With directional selection in differing local environments, countless cases of divergent evolution have occurred in the billions of years that life has existed on Earth. The net result is the great diversity of life forms on this planet.

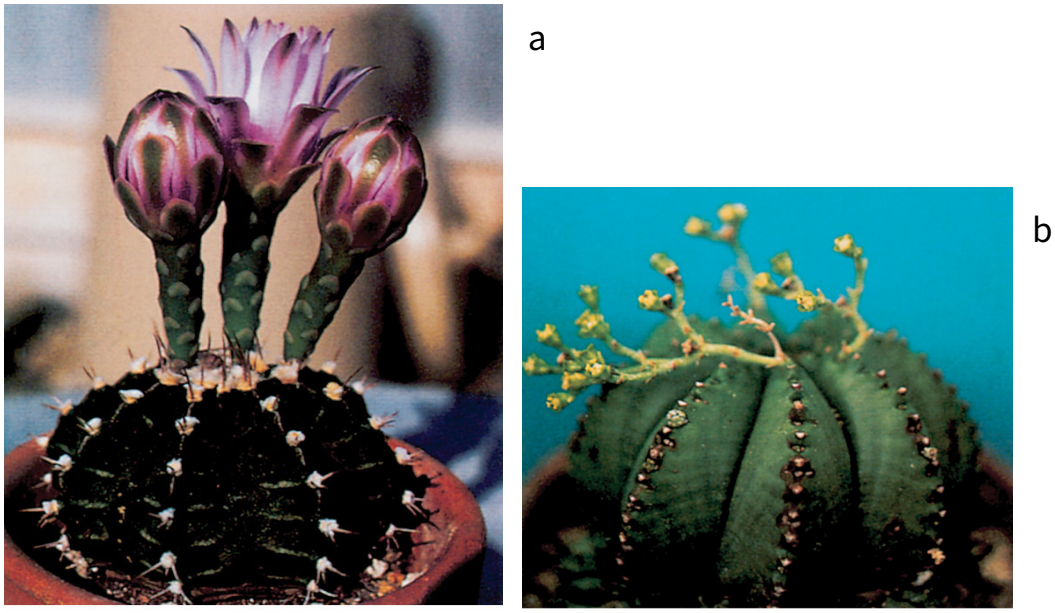


Figure 18.11. Desert plants of the (a) southwest United States and (b) Africa. The similarities are attributed to convergent evolution. On the basis of floral and anatomical features, confirmed by DNA comparisons, these plants are placed in separate families: (a) Cactaceae, and (b) Euphorbiaceae. The two families also include plants that do not have desert adaptations and are much less similar than those shown here.

CONVERGENT EVOLUTION Directional selection can also cause quite different species to evolve similar traits. Such an increase in similarity between two taxa is called **convergent evolution**. It occurs when differing populations are exposed to similar environments over many generations. For example, some plants in African deserts resemble cacti of American deserts (Fig. 18.11). The plants in both areas have spines, fleshy green stems,

and reduced leaves. This might tempt us to put the African and American species in the same family. But other traits, including differences in flowers and DNA, show that the African and American plants belong to different families, the Euphorbiaceae and the Cactaceae. Desert members of the families evolved similar traits because they experienced similar directional selection for a long time.

COEVOLUTION Many species interact as competitors, predator and prey, or symbiotic partners. In such interacting pairs, each species may exert selective pressure on the other. The result is simultaneous evolution of both species, a process called **coevolution**. In many cases, coevolution leads to a close match between the two species. For examples, plants that are pollinated by moths often produce nectar at the base of long, slender tubes or spurs, beyond the reach of other pollinators but ideal for the long tongues of moths. An exclusive pollinating relationship helps both partners: pollen transfer is more efficient if the pollinator visits just one plant species, and the pollinators get a private food supply. The mutual benefit suggests that moth pollination favored the evolution of long spurs in the flowers, as well as long tongues in the moths.

An extension of the preceding example shows how coevolution can increase the diversity of life. Competition for pollinators reduces a plant's reproductive success, leading to directional selection for new traits that reduce competition--new flowering dates and new flower colors and shapes. As new kinds of plants appear, they become new resources for pollinators that can exploit them. The result is directional selection for pollinators that have matching changes. Over time, this coevolution can result in many new species of flowering plants and pollinators.

Population Genetics Reveals Evolution and its Causes

Darwin's theory of natural selection was a vital first step in building the modern theory of evolution, but his ideas were handicapped by lack of knowledge about how hereditary information is stored, altered, and passed to progeny. Not until the twentieth century was genetics advanced enough to show the molecular basis of evolution. By then, geneticists had adopted Mendel's idea that hereditary information is carried on particles (now called genes), but they wondered why different versions of the same gene (alleles) persist in a population, even though one allele is more abundant or is expressed more strongly than the other. It is easy to see the relevance of these questions to evolution, for evolution in its most basic form consists of changes in the relative abundances of alternative alleles. Such questions attracted scientists who love to create mathematical models of real events. In 1908, the English mathematician G.H. Hardy and the German physician G. Weinberg simultaneously published the same model to answer such questions. Their analysis postulated an ideal population in which the following five conditions apply:

1. Mutations do not occur.
2. Organisms do not migrate between populations.
3. Reproduction is limited to random sexual mating.
4. There is no natural selection.
5. The population is very large.

The analysis by Hardy and Weinberg showed that under such conditions, two alleles for the same gene will remain indefinitely in the population in a fixed ratio, even if one allele is dominant over the other. This conclusion, called the **Hardy-Weinberg equilibrium**, became the basis of a new discipline known as population genetics--a field of study that integrates genetics and evolution. Because the mathematical details are not necessary to understand what this text says about evolution, this chapter does not describe the mathematics. However, it shows how Hardy-Weinberg equilibrium can be used to determine whether and how evolution is taking place in a population.

To determine whether a wild population is at Hardy-Weinberg equilibrium with respect to a pair of alleles, the researcher takes some organisms from the population and determines the ratio of alleles they collectively contain, then measures the same ratio in the next generation and compares the two results. If the ratios are similar enough that any differences can be explained by random errors, the pair of alleles is at Hardy-Weinberg equilibrium. Many genes and alleles in wild populations have been studied in this way. In many cases, some alleles are at equilibrium, whereas others are not. Because evolution consists of changes in allele ratios, those that are not at equilibrium are evolving. In this way, the Hardy-Weinberg equilibrium yields clear evidence that much evolution is taking place on a micro-scale today, and that one gene can be stable while another gene in the same species is evolving.

Studies related to the Hardy-Weinberg equilibrium can identify factors that are causing evolution. This is true because all the conditions specified for Hardy-Weinberg equilibrium relate to causal factors. To show this, the following list considers the five conditions in turn:

1. Mutations convert one allele to another, and therefore alter the ratio of alleles, unless forward and reverse mutations exactly balance.
2. If many individuals enter or leave the population, the allele ratio will change unless the migrating individuals have alleles in exactly the same ratio as the overall population.
3. If mating is not random, some allele combinations may be reproduced disproportionately often.
4. Natural selection favors the reproduction of individuals with a certain allele combination over others.
5. If the population is very small, chance can determine which individuals reproduce.

Thus, failure to meet any one of the Hardy-Weinberg conditions can result in evolution (disequilibrium). All five causes of disequilibrium have been measured many times, revealing that natural selection typically has much larger effects than mutation, migration, and nonrandom mating. With such information, many evolutionists use population genetics as a tool to predict changes and to explore the cause of evolution.

EFFECTS OF CHANCE ON SMALL POPULATIONS Darwin attributed all evolution to natural selection, but population genetics predicts that chance can affect evolution in small populations. The prediction has been verified in field studies. In small populations, the best-adapted individuals do not always leave the most offspring. If a population is so small that only a few individuals have a certain valuable trait, a random accident such as a fire

or epidemic may accidentally eliminate all the individuals that have the best allele, whereas some of those with other alleles are spared. Such a random change in the allele ratio is called **genetic drift.**

Another effect of chance on small populations, the **founder effect**, occurs when a few individuals from a large population establish a small, isolated population. Chance may determine which of the main population's alleles are present in the founders. As a result, the founders may have a combination of traits that is uncommon in the old population. They may start the new population on a new path of evolution. The founder effect often is seen in studies of oceanic islands, where wind, water, and birds occasionally bring seeds from mainland plants. The island plants are related to mainland species, but their traits often differ in many ways.

Speciation Multiplies Species

Natural selection by itself does not increase the number of species; it changes species that already exist. Nevertheless, fossil studies show many cases where the number of species increased with time. Evolutionists explain such multiplication by a process called **speciation**, which splits one species into two.

For one species to become two, an original population must divide into two populations that encounter different directional selection, resulting in divergent evolution. But even with different patterns of selection, the populations will remain a single species if they exchange genes through sexual reproduction. Thus, in sexual life forms, speciation can occur only if some factor prevents the populations from exchanging genes. Such a block to gene exchange is called **reproductive isolation**. Common causes of reproductive isolation are described in the next sections.

GEOGRAPHIC ISOLATION In many cases, speciation begins when geographical barriers prevent populations from meeting to exchange genes. Such a separation is called **geographic isolation**. For example, a storm may blow individuals from a mainland population to a distant land with new environmental conditions and selective agents. In time, the resulting divergence can become so great that individuals from the two populations no longer crossbreed even if they are brought back together. This may have happened in the case of Darwin's finches on the Galápagos Islands. Geographic isolation also can result from slow geological events, such as the creation of a mountain range--events that leave parts of a population on opposite sides of an impassible barrier.

POLYPLOIDY The possession of more than two chromosome sets per cell--a condition called **polyploidy**--is an important source of new species in plants. Comparisons of chromosomes suggest that up to 70% of flowering plant species have a polyploid origin. Wheat, potatoes, and cotton are examples. Until the time comes for making gametes, reproductive cells in seed plants remain diploid by dividing every time they duplicate their chromosomes. But occasionally, a cell with duplicated chromosomes fails to divide. This cell and its descendants have four sets of chromosomes instead of the normal two sets. If polyploid cells give rise to a stem that forms roots, an independent polyploid plant is formed.

A new polyploid plant with four sets of chromosomes per cell is instantly unable to exchange genes with its diploid relatives because mating with its diploid ancestors will not produce fertile offspring (Fig. 18.12). Gametes from the polyploid plant have two sets of chromosomes, whereas gametes from a diploid plant have one set of chromosomes. If the two kinds of gametes fuse, the resulting triploid plant may be vigorous, but its gametes are defective because complete pairing in meiosis requires an even number of chromosome sets. Thus the new polyploid plant is reproductively isolated, but if it has perfect (bisexual) flowers, it is self-fertile and can be considered a new species.

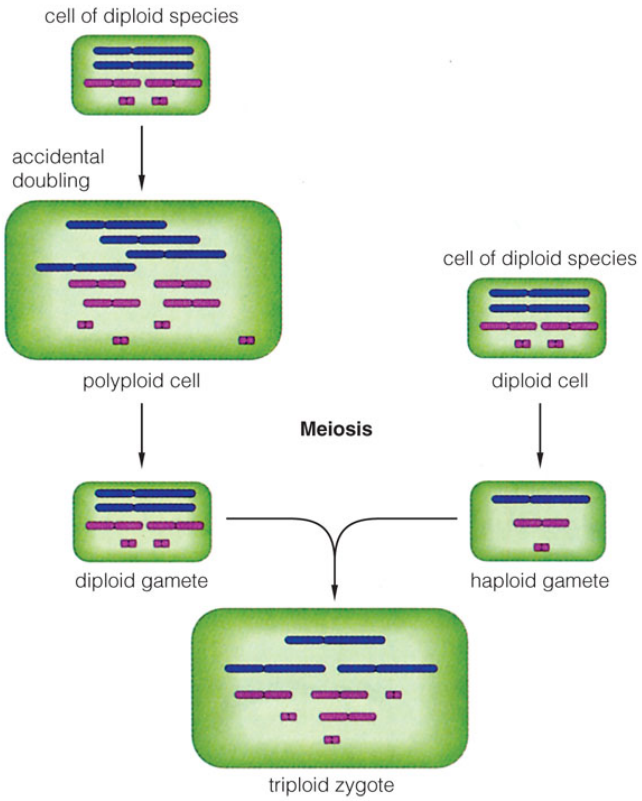


Figure 18.12. Polyploidy can lead to reproductive isolation and thereby to a new species. The diploid species has two sets of chromosomes. Failure to divide after doubling chromosomes results in a polyploid cell with four sets of chromosomes. When a polyploid cell goes through meiosis, it makes gametes with two sets of chromosomes, whereas gametes of the originating diploid plants have one set. A joining of diploid and haploid gametes creates a triploid plant, which is sterile because meiotic pairing requires an even number of chromosome sets.

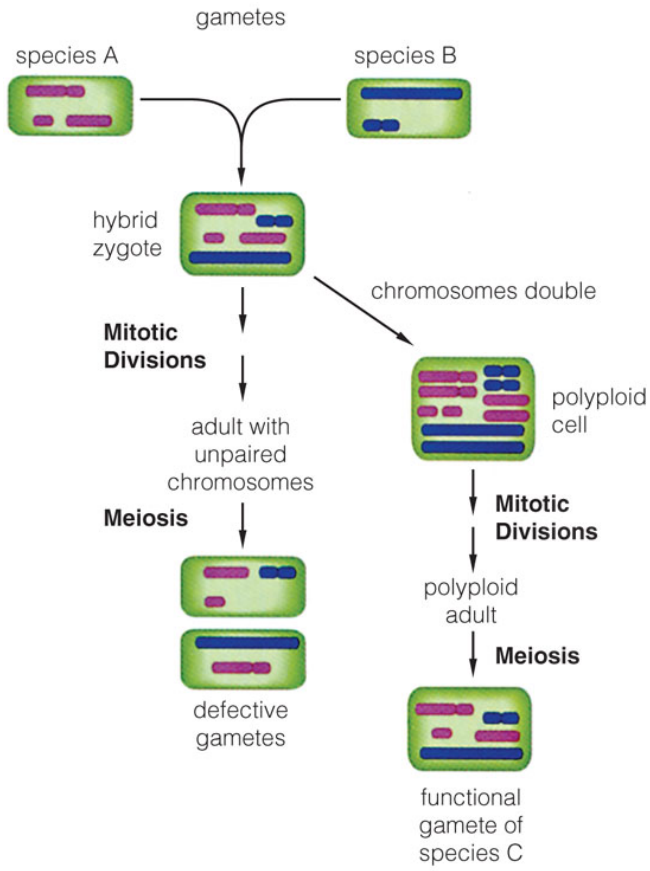


Figure 18.13. Why hybrids often are sterile and how a doubling of chromosomes (polyploidy) restores fertility. (Top) A hybrid forms by fusing gametes of two species with differing chromosome sets. (Left branch) The hybrid's chromosomes do not form matching pairs; hence, they are distributed abnormally in meiosis, yielding nonfunctional gametes. (Right branch) Chromosome doubling permits normal pairs in meiosis: gametes receive one chromosome of each kind and are fully functional.

HYBRIDIZATION Hybridization is another source of reproductive isolation that can lead to speciation. As discussed earlier, a hybrid arises by fusing the gametes of two species. If the species are close relatives, the hybrid may be vigorous and even fertile. An example is the widely planted London Plane tree (*Platanus X acerifolia*), which arose naturally when European and American species hybridized. Likewise, some grasses and some composites (relatives of the sunflower) have generated many species by hybridization.

New hybrids often are sterile, usually because the two parent species have different numbers or kinds of chromosomes (Fig. 18.13). If the chromosomes differ too much, meiosis fails because the chromosomes do not pair properly. But fertility can be restored if a cell at the tip of a hybrid plant becomes polyploid and initiates a polyploid shoot that forms gametes. When a polyploid cell undergoes meiosis, each chromosome has a partner, so pairing can occur and meiosis can be completed, thereby making gametes. The polyploid plant is reproductively isolated from both parent species because its gametes have a different number of chromosomes than gametes of either parent species. An example is *Triticosecale*, a human-made hybrid between wheat (*Triticum*) and rye (*Secale*) that combines the high productivity of wheat with the disease resistance of rye. The sterile hybrids became fertile when plant breeders doubled the chromosome number. Some crop plants, such as wheat, arose by hybridization followed by polyploidy.

Macroevolution Generates New Forms of Life

All evolution is based on changes in the information stored in DNA--information that guides development and leads to distinctive body forms, metabolic products, and behavior patterns. Based on the scale of change, biologists recognize two levels of evolution. **Microevolution** consists of changes too small to alter the fundamental nature of the species, such as alterations in flower color. **Macroevolution** consists of changes large enough to represent the emergence of a new life form, such as the evolution of flowering plants form moss-like ancestors. The difference is important when considering the kinds of research that are needed to explore the two types of evolution.

Microevolution is rapid, easy to observe, and easy to produce artificially in the laboratory. We are surrounded by everyday examples, such as in breeding turkeys with increased white meat or breeding corn (maize) for greater yields, and the increase in antibiotic resistance that has occurred in bacteria because of our use of antibiotics. Population genetics is concerned with microevolution.

Macroevolution can be the sum of many microevolutionary changes over long periods, or it may involve larger abrupt changes, such as chromosome rearrangements. Sudden radical changes usually are fatal, but occasionally they lead to a form that has great success.

Macroevolution is more difficult to observe than microevolution. Major changes in body form often require changes in many genes. Constructive changes are rare, so the time needed to accumulate many such changes is far longer than a human life. As a result, most ideas about macroevolution are based on indirect evidence. For instance, fossils show that ancient life forms differed greatly from modern forms. We also find clues about macroevolution by comparing cellular and molecular traits of modern organisms. It was a surprise when biologists found that plants and animals have similar cell structures. The discovery suggested that all life evolved from a common ancestor.

The idea gained strength when twentieth-century biologists found that all organisms use the same language to encode information in DNA.

To the great majority of modern biologists, these facts are evidence beyond a reasonable doubt that macroevolution generated all modern forms of life from microscopic organisms that first populated Earth some 3.8 billion years ago. With that point established, the current task is to explore the paths of evolution that produced today's millions of diverse species. These studies make up the discipline called **phylogenetic systematics**.

18.5 PHYLOGENETIC SYSTEMATICS

If all life evolved from the same original ancestor, then a diagram that puts all the paths of evolution together would resemble a tree that grew from a single seed, starting with one stem and adding millions of branches. Such a diagram is called a **phylogenetic tree**. The task of phylogenetic systematics is to find the most accurate tree.

Figure 18.14 shows the main features of a phylogenetic tree. Tips of branches are the most recent products of evolution along each branch, such as modern liverworts and flowering plants (*B* and *G* in Fig. 18.14). Each branch point is an act of speciation, where one species divided into two.

Today, phylogenetic systematics is one of the most active fields in science. Three developments late in the twentieth century made it so. The first was a new set of methods and concepts called cladistics, which provided a more orderly way to explore evolutionary relations among life forms. Equally important was the invention of fast, inexpensive computers that made it practical to analyze large amounts of data. The third big push came when molecular biologists invented quick ways to read information stored in DNA.

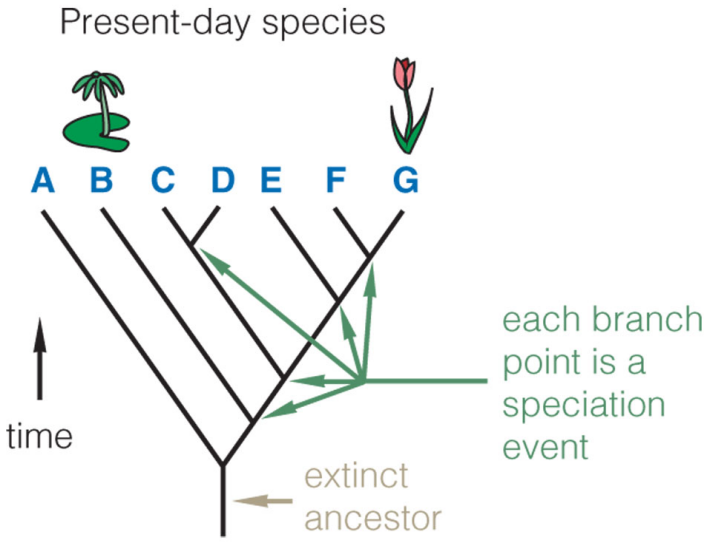


Figure 18.14. Basic components of a phylogenetic tree. Each branch point is an act of speciation, launching new taxa. The time axis is upward, and A through G represent the most recent products of evolution along each branch.

Phylogenetic Systematics Has Practical Value

Before exploring systematics, let us ask why the study is worth doing. The answer is that great practical rewards can come from knowing how evolution led to present-day species. Consider the search for new medicines: most pharmaceutical products, such as antibiotics, were first found in living organisms. Suppose a researcher finds that trees of a rare, slow-growing species make a compound that easily cures colon cancer. The

compound is too complex to make artificially. How could we find a better source? One way is to look for it in fast-growing relatives of the trees. But how would we recognize the relatives, which may barely resemble the trees? With the use of phylogenetic systematics, we would compare characteristics of a wide range of plants--combining structural, physiological, and molecular traits--to find those most likely to have shared recent ancestors.

Similarly, we may find ways to stop parasites that attach food plants by experimenting with relative of the parasites that can be grown without a host. Phylogenetic systematics can identify such relatives.

Cladistics Explores Clades by Means of Cladograms

Loosely defined, cladistics is a set of quantitative methods and concepts for exploring the evolutionary relations among taxa. A cladistic analysis compares many modern species to deduce the most probable point in evolution where each species branched off from another evolving group. The name cladistics, derived from the Greek *klados*, meaning "tree branch," reflects this focus on branch points. A **clade** is a branch on the tree of life, consisting of an originating taxon and all of its descendant taxa. Before cladistics was invented, scientists had no quantitative basis for deciding which groups of species make up true clades. Cladistics offered a more orderly method, and the phylogenetic trees it produces are called **cladograms** to reflect the method used in their development.

To illustrate the basic features of a cladogram, Figure 18.15 shows the relationships among three taxa, each represented by a different kind of fruit tree. The tip of each branch represents the most recent product of evolution along the branch, in this case, three kinds of fruit trees. New branches arise by speciation at the nearest branch point, which cladists call a **node**. At each node, an ancestral species splits to produce two new species, while the ancestor itself ceases to exist. The oldest node is called the **root** of the cladogram.

Cladograms rarely include more than a small sampling of species that evolved from the ancestor. Only species that contributed data to the study are listed. The cladogram would have many more nodes if all branch points and terminal species were included. For example, Figure 18.15 shows the "apple" branch evolving from the root node without further branching. But many other fruit trees (peach, pear, cherry, and so on) arose by separate speciation events along that branch. Thus, the word *apple* in the cladogram merely names a representative group that occurs on the branch. The same may be true of any terminal branch, such as the branch that ends with *orange*.

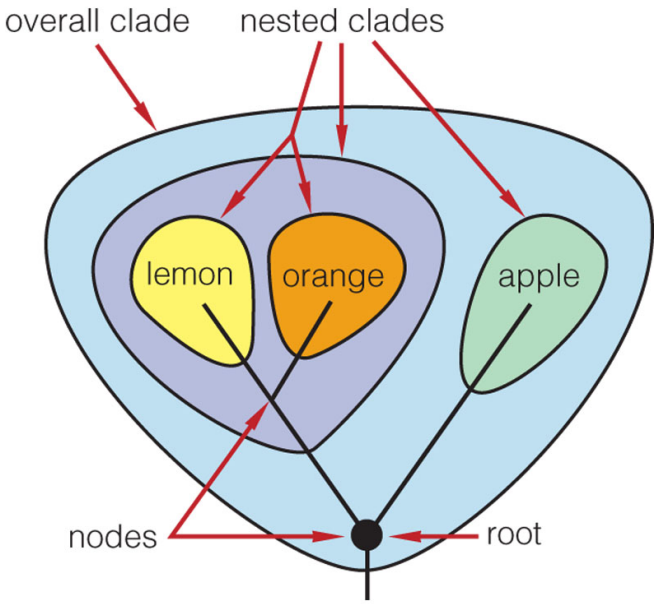


Figure 18.15. Basic features of a cladogram, illustrated with three kinds of fruit trees. Each of the five colored enclosures surrounds a clade (an ancestor and all of its descendants). Each branch point is called a node and is a speciation event. The oldest node is the root of the cladogram.

With this information, we are ready to identify clades in Figure 18.15. As stated earlier, a clade is an ancestral taxon and all of its descendants. Five groups in Figure 18.15 meet this definition: each is surrounded by a differently colored enclosure. The overall clade includes the ancestor at the root, all three named taxa, and many more taxa that are not included in the study. Four more clades are nested within the overall clade. The largest of these includes the ancestor of all *Citrus* species and its descendants, represented by orange and lemon. Each terminal branch (orange, lemon, apple) also is a clade, which may include many species that are not shown. An enclosure delimiting a terminal clade does not include the preceding node, for that node represents the ancestor of *both* branches. The ancestor of a terminal branch is one of the two species that *emerged* from the node.

There are several ways to draw cladograms for the same set of species (Fig. 18.16). The first three cladograms in Figure 18.16 are rooted, where Figure 18.16c is unrooted. The difference is important. **Rooted cladograms** identify the node in the cladogram that occurred first, thereby showing the direction of evolution throughout the clade. **Unrooted cladograms** do not show which node is closest to the root. As a result, they leave the direction of evolution between each pair of nodes unspecified.

In rooted cladograms, there are different ways to show branching. In one method, descendants diverge from a branch point like arms of the letter *Y* (Fig. 18.16a,b). The other method keeps the arms parallel by showing all the divergence from a branch point at once, then making a right angle bend in each arm, like the letter *U* (Fig. 18.16c). This method makes it easier to write names of taxa at the ends of the arms, especially if the cladogram is on its side (for example, in Fig. 18.16c). To indicate that a branch includes many species, the tip may be expanded into a triangle.

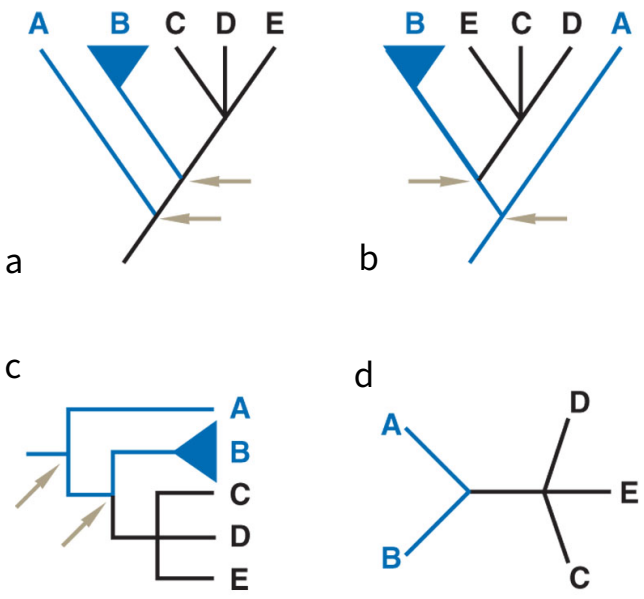


Figure 18.16. Four ways to draw a cladogram for the same taxa. (a-c) Equivalent rooted cladograms. All have the same number of nodes between any two taxa, illustrated for taxa A and B with arrows at the nodes. (d) An unrooted cladogram, which does not show a root node.

Alternative cladograms are equally valid as long as they agree on the number of nodes that separate any two taxa. Differences in orientation of branches are not important. Figure 18.16 illustrates this point by using arrows to show that each rooted cladogram has two nodes between taxa A and B. There is similar agreement for all other taxa; thus, the three rooted cladograms are equivalent. Because they differ in branch orientation, you cannot judge which taxa are closest relatives by looking at their positions along the ends of the branches. To determine evolutionary relationships, you must count the nodes that separate the taxa.

A postulate of cladistics is that each act of speciation replaces one parent species with two new species. If a cladogram shows three or more species arising from the same point, it usually means that further studies are needed to determine which species branched off first. However, multiple species may occasionally branch off so closely in time that no study could reveal the branching sequence.

ALTERNATIVE CLADOGRAMS Any set of species might be related in a variety of ways, with a different cladogram for each--and the more species we include in the study, the more cladograms are possible. The number of possible unrooted cladograms depends only on the number of species. To illustrate, for any set of 5 species, there are 15 possible unrooted cladograms, 3 of which are shown in Figure 18.17. The cladograms differ as to which pairs of species are the closest relatives. In Figure 18.17a, only one node separates A and B, three nodes separate them in Figure 18.17b, and two nodes separate them in Figure 18.17c. Thus, alternative cladograms differ in how many steps of evolution stand between each pair of species. The cement this point, draw some of the 12 other unrooted cladograms that are possible. As you work, keep in mind that two cladograms are identical if one can be converted to the other by flipping it over, changing its branch angles, or rotating the diagram or its branches.

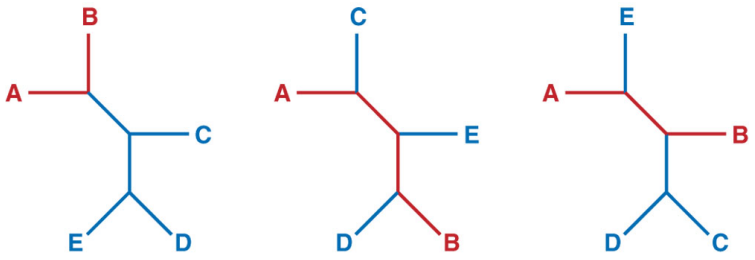


Figure 18.17. Three alternative unrooted cladograms that show ways in which any five species might be related. Twelve other cladograms are possible. Red is used to emphasize the differences in number of nodes separating species A and B.

Cladists find the Best Cladogram by Comparing Character States

Every cladogram is a hypothesis about evolution, and there are many alternative cladograms. How do cladists identify the cladogram that most likely reflects the real paths of evolution? The cladistic method compares species with respect to various characters. *Morphological characters* are related to body form, such as the number of flower petals, flower color, and growth habit (herb, tree, and so on). *Molecular characters* are chemical traits such as the detailed structure in a certain segment of DNA, or the ability to make a particular kind of molecule. To be useful in selecting the best cladogram, a character must occur in all the species being considered, and its details--called the **character states**--must differ among some of the species. For example, flower petals might be red in one species and white in another. Here the character is flower color, and the character states are red and white. Cladistics views a change in character state as the basic event in evolution.

To trace evolutionary relationships among species, systematists must compare traits that have a common evolutionary origin. Traits that arose from the same ancestral trait are said to be **homologous**. The wings of a bird and the forelegs of a horse are homologous, because both evolved from the bony forelimbs of a common ancestor. In cladistic terms, homologous traits are alternative states of the same character. Bird wings and horse forelegs are alternative states of the character "forelimb." By contrast, the wings of insects and the wings of birds are not homologous. Despite their similar function, they evolved from entirely different ancestral structures. They are **analogous**, meaning that they have a similar form or function, but evolved from different structures. Analogous structures are not alternative states of the same character. They are states of different characters.

Major errors can occur if a classification is based on comparisons of analogous structures. The character "has wings" is a bad choice for defining groups among animals, because it would promote the false conclusion that birds are related more closely to insects than they are to horses.

To compare characters among species, cladists list the character states in a table called a **character matrix**, such as shown in Figure 18.18. In the matrix, species are listed along the left margin. Characters are listed across the top, and the boxes show the state of each character for each species.

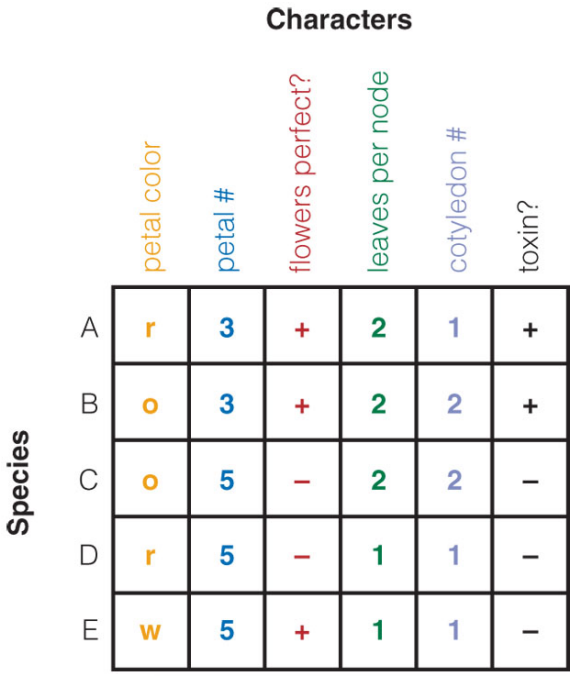


Figure 18.18. A character matrix. There are five species (A,B,C,D,E). Characters are listed across the top. Each entry in the matrix is a state for the character listed above it. Character states: Petal colors are red (r), orange (o), or white (w). Number of petals in each flower is 3 or 5. Flowers can be perfect (+) or imperfect (-). Plants can have 1 or 2 leaves at each node and 1 or 2 cotyledons in the embryo. Plants can have (+) or lack (-) toxin.

THE IMPORTANCE OF MOLECULAR CHARACTERS In the early days of systematics, biologists relied almost entirely on morphological characters to deduce paths of evolution. As biochemistry brought new information, molecular characters entered the picture--traits such as the ability to make a particular protein. But the choice of characters had the greatest increase when biologists learned how to read DNA. Systematists eagerly turned to DNA base sequences as a source of characters for comparing species. After all, evolution is based on changes in the information stored in DNA. Even morphological characters result from an organism following the instructions in DNA. By comparing DNA among species, we may hope to get the most detailed look at paths of evolution.

DNA offers a multitude of characters for comparison. Each nucleotide position along the DNA can be a character for cladistic analysis, and DNA contains millions of nucleotides. There are four possible character states at each position: the base can be A, T, G, or C. Most current cladistic studies include DNA data. Often, they confirm relationships that were traditionally accepted from morphological data. In other cases, they reveal new relationships.

To understand how cladistics works, there is no substitute for solving cladistic problems. The end note "IN DEPTH: Do a Cladistic Analysis of DNA" leads you from constructing a DNA character matrix through finding and rooting the best cladogram. Although you will not need a computer, you will appreciate why computers are important in systematics. To find the best cladogram, you must analyze all the possibilities--and as the number of species and characters increases, the number of alternatives rapidly becomes far too great to handle without a computer.

THE PRINCIPLE OF PARSIMONY With many possible cladograms, how do cladists choose the one that shows the real evolutionary relations among taxa? This is the key problem in cladistics. We were not present when the species evolved, so we can never know

definitively which cladogram is correct. But several statistical methods help to choose the cladogram that is most likely to be a true representation of evolutionary relationships. One popular method is based on the **principle of parsimony**, which postulates that the cladogram requiring the fewest evolutionary events is most likely to be correct. That cladogram is said to be the most parsimonious.

Although a cladogram may be most parsimonious, we can never be sure it is correct. It is simply a good hypothesis about the paths of evolution, and later studies with new evidence may prove it wrong. With the help of computer programs, cladists use parsimony and several additional methods to find candidates for the best cladogram. The methods often pick slightly different cladograms and a comparison of the choices reveals points on which all methods agree. Those points are said to be *strongly supported*. A compromise diagram, called a **consensus tree** or *consensus cladogram*, is drawn: it includes all the points of agreement while leaving points of disagreement unresolved as nodes from which more than two branches depart. Points of disagreement need further study with additional species and characters.

Rooted Cladograms Show the Sequence of Evolutionary Change

Finding the root of a cladogram is one of the most important tasks of cladistics, because it reveals the direction of evolution. To show its value, Figure 18.19 compares unrooted and rooted cladograms for the same five species, numbered **1** through **5**. Each numbered species represent the most recent product along its branch. This means recent evolution always runs toward each present species, as shown by the arrowheads in the unrooted cladogram (Fig. 18.19a). But the unrooted cladogram says nothing about the direction of evolution in the internal segments that lie between the three nodes. We do not know which node is the oldest and closest to the point where the ancestor started the evolution of the five species. In contrast, a rooted cladogram (Fig. 18.19b) shows that the common ancestor first split to form the ancestor of species **1** and the ancestor of all the remaining species. With that information, we suddenly know the direction of evolution between all nodes (orange arrows). With a root in place, the cladogram can be redrawn with the root at the bottom, current species at the top, and time flowing upward (Fig. 18.19c).

How do cladists locate the root? The answer is to include data on additional taxa, called **outgroups**, together with character data on the **ingroup**, the set of taxa that is the target of the study. Good outgroups are taxa that share many characters with the ingroup

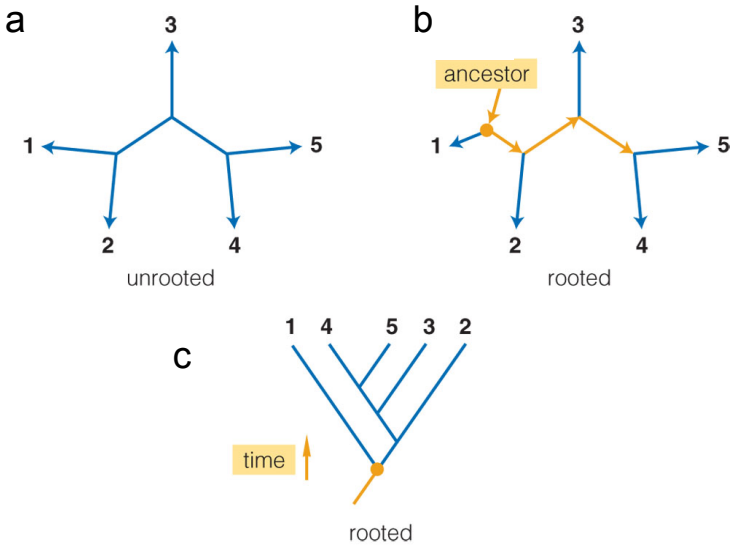


Figure 18.19. Rooting a cladogram reveals the direction of evolution. (a) An unrooted cladogram. Recent evolution (*arrows*) is toward all present species, but the direction between nodes is unspecified. (b) A rooted cladogram. The dot marks the root. Now we know the direction of evolution between all nodes (*orange arrows*). (c) A rooted cladogram redrawn with root at bottom and time flowing upward.

but differ too much to be part of the ingroup clade. Researchers often include several outgroups, in case some of them are poor choices. For example, green algae are good outgroups for a study of evolution among land plants. The researcher simply includes data on green algae in a character matrix together with data on land plants and then runs a computer program to find the best cladogram. Figure 18.20 shows a typical result. At the root node, an unknown ancestor split to launch one of the outgroups (left) and the ancestor of all the other species (right). The right-hand branch split again, launching the second outgroup (left) and the land plant clade (right). Among other things, the rooted cladogram shows that mosses arose before ferns, and ferns arose before conifers and flowering plants.

The results of one cladistic study can suggest good outgroups for more detailed studies. For example, there are thousands of fern species. How did they evolve? To answer, we could make a character matrix of ferns and include mosses as outgroups. For such a study, mosses are better outgroups than green

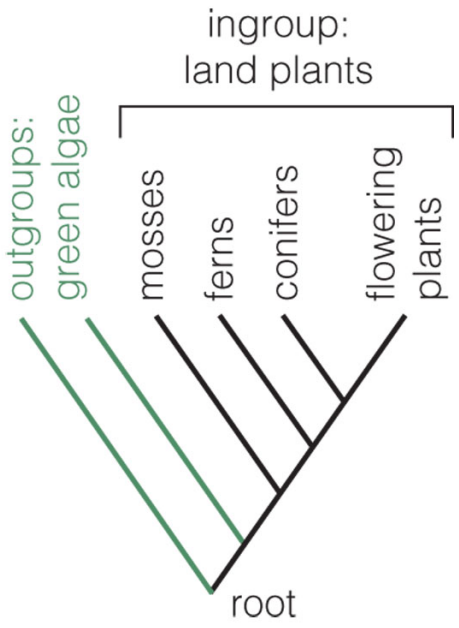


Figure 18.20. Use of green algae as outgroups to root a cladogram in which four kinds of land plants make up the ingroup. The rooted cladogram shows the sequence in which land plant groups evolved.

algae. This makes it possible to include many more characters in the matrix.

Because a rooted cladogram can show the sequence in which taxa arose, it also reveals the sequence in which important character states evolved. Cladograms often use labeled tick marks to show where such innovations arose (Fig.. 18.21). For example, all green algae and all land plants have the photosynthetic pigment chlorophyll *b*; therefore, green algae and land plants probably inherited this trait from their common ancestor. In contrast, all land plants form embryos as part of their life cycle, whereas no green algae have embryos. Thus, the ability to make embryos probably evolved just once in the lines that led to modern green organisms in the land plant clade before mosses branched off.

ANCESTRAL AND DERIVED CHARACTER STATES By showing where a character state first arose, a rooted cladogram allows us to distinguish between **ancestral character states**, which a clade inherited from its ancestor, and **derived character states**, which evolved later. Among seed plants the ability to make seeds and carpels illustrates the distinction (Fig. 18.22). The ability to make seeds evolved in the ancestor of all seed plants, and is therefore ancestral for the clade as a whole. In contrast, the ability to make carpels, the female parts of the flower, evolved later in the line that led to flowering plants. Thus, for the seed plant clade as a whole, the ability to make carpels is a derived character state.

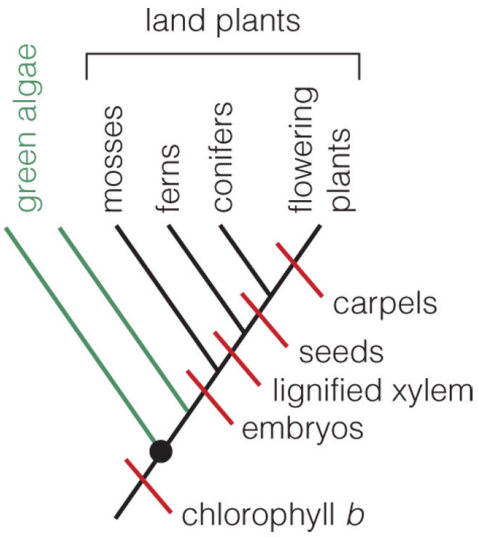


Figure 18.21. Locating innovations on a rooted cladogram. Labeled tick marks show where each innovation arose. They are located by finding the points where their origin accounts for the array of taxa with the fewest evolutionary changes. For example, all land plants have embryos, whereas green algae never do; the most parsimonious postulate is that embryo formation evolved in the land plant clade before mosses branched off.

Use of the terms *ancestral* and *derived* requires care, because the judgment depends on the point of view, which must always be specified. For example, in discussing the seed plant clade as a whole, we concluded that the ability to make carpels is a derived character state, but if attention is limited to flowering plants, the ability to make carpels is ancestral, because all flowering plants inherited it from their ancestor.

Cladistics Reveals Convergent Evolution

The greatest problem in classifying organisms comes from assuming that a shared character state implies common ancestry. The assumption often is sound, but sometimes it leads to wrong conclusions. Errors occur because similar character states sometimes arise independently in two groups of organisms. This is convergent evolution, mentioned earlier in the chapter. It occurs when populations with different origins evolve under similar selection pressures and develop similar character

states. To minimize errors caused by convergent evolution, systematists include many characters in the analysis. Then, even though a few similarities result from convergent evolution, they affect the analysis much less than the differences that come from different ancestry.

With enough characters, a cladogram can reveal which characters arose through convergent evolution. Such insight occurs when a well-supported cladogram, based on many characters, places two taxa in separate clades even though they share a character state that is not found in other members of the clades. For example, in Figure 18.23, either the ancestor of all the taxa had desert adaptations and most taxa lost the adaptations, or the ancestor lacked the adaptation, and the two desert-adapted taxa evolved their adaptations independently. The latter explanation is the most parsimonious, considering how many taxa would have to change otherwise.

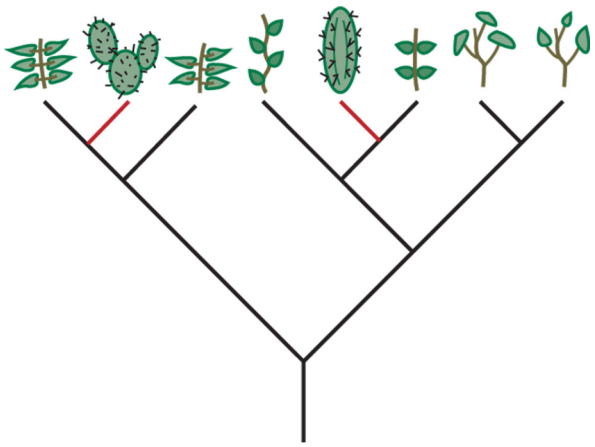


Figure 18.23. A cladogram can reveal convergent evolution. The two taxa on red lines have desert adaptations (fleshy green stems, leaves reduced to spines). Other taxa do not. The most parsimonious explanation is that the two desert-adapted taxa evolved their adaptations independently.

All Formally Named Taxa Should Be Monophyletic

Ever since the concept of evolution led to phylogenetic systematics, a major goal in taxonomy has been to group organisms by common ancestry. In cladistic terms, the goal will be accomplished when every formally named taxon is a true clade--an ancestor, all of its descendants, and nothing else. Such a taxon is said to be **monophyletic** (Fig. 18.24).

Each currently accepted domain and kingdom of life is believed to be monophyletic. But many traditional taxa at lower levels are still not monophyletic, despite more than 250 years of careful study. It takes time to correct taxonomic problems, for each systematic study is just one hypothesis, and further studies can lead to different views. The work is well worth the effort, though. As discussed earlier, a truly phylogenetic system of taxonomy has much more predictive value than the alternatives, and it can speed such goals as improving crops and discovering new cures for human disease.

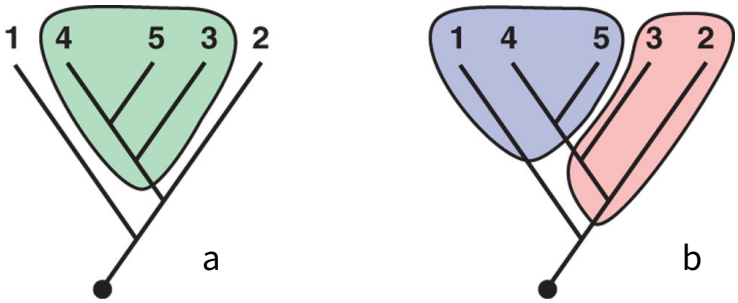


Figure 18.24. A monophyletic group and alternatives. (a) The outlined group is monophyletic. It includes an ancestor, all of the ancestor's known descendants, and nothing else. (b) Neither of the circled groups is monophyletic. One omits the group's ancestor and some of the ancestor's descendents; the other omits some of its ancestor's descendents.

KEY TERMS

adaptations
analogous (traits)
ancestral character state
back-crossing
biodiversity
biological species
character matrix
character states
clade
cladistics
cladograms
classes
coevolution
consensus tree
conserved sequence
convergent evolution
derived character states
directional selection
divergent evolution
diversifying selection
divisions
domains
endosymbiosis
evolution
family
founder effect
genetic drift
genus
geographic isolation
Hardy-Weinberg equilibrium
homologous (traits)
hybridization
hybrids
ingroup
introgression
kingdoms
macroevolution
microevolution
monophyletic
mutagens
mutations
natural selection
node
orders
outgroups
phenetic species
phyla
phylogenetic systematics
phylogenetic tree
phylogeny
polyploidy
principle of parsimony
recombination
reproductive isolation
rooted cladograms
speciation
species name
specific epithet
stabilizing selection
systematics
taxon
taxonomy
unrooted cladograms

SUMMARY

1. Linnean taxonomy groups organisms into a hierarchy of taxa, including species, genera, families, orders, classes, phyla (divisions), kingdoms, and domains. Organisms are assigned to a phenetic species by comparing combinations of characters and to a biological species by applying a mating test.

2. Biologists divide life forms between two prokaryotic domains (Bacteria and Archaea) and one eukaryotic domain (Eukarya). Domain Eukarya includes three monophyletic kingdoms (Plantae, Animalia, and Fungi) and many smaller groups that often are informally called protists, some of which are candidates for kingdom status. Domains Bacteria and Archaea are not yet subdivided into widely accepted kingdoms.

3. Studies of fossils, DNA, metabolism, cell structure, and reproduction suggest that all organisms evolved from a common ancestor through processes that alter the DNA in organisms and that permit some variants to reproduce more effectively than others (natural succession).

4. DNA carried by organisms in a population can be altered by mutations, which produce alternative versions (alleles) of genes; by recombination, which brings together DNA that arose in different organisms. For recombination, DNA from different organisms can be brought together by endosymbiosis, bacterial transformation and conjugation, viral transduction, and sexual reproduction within or between species.

5. Natural selection results from environmental interactions that allow some variants in a population to reproduce more effectively than others. Selective agents in the environment include food supply, climate, predation, and other factors. Natural selection can improve the adaptation of a species to the environment (directional selection), maintain the current adaptation (stabilizing selection), or increase the genetic diversity in a species (diversifying selection).

6. Population genetics integrates the fields of genetics and evolution and provides mathematical tools of great predictive value that help determine whether evolution is taking place and how it is caused.

7. In small populations, chance events can affect the direction of evolution independently of natural selection. Such changes produce genetic drift and lead to a founder effect when small colonies re established.

8. Speciation splits one species into two when part of a population becomes reproductively isolated from the rest, followed by divergent evolution in response to directional selection. Reproductive isolation may result from migration, formation of geological barriers, polyploidy, or hybridization.

9. Phylogenetic systematists seek the paths of evolution that led from a common ancestor to modern species. Their findings often are expressed in phylogenetic trees, which include cladograms. Given a set of species, many different cladograms can be drawn that reflect possible evolutionary relationships among the species.

10. The set of methods and concepts called cladistics provides an orderly, computerized way to select the cladogram that most probably reflects the real evolutionary relationships among taxa. To determine the direction of evolution, cladistic studies include outgroup taxa together with ingroup taxa. This reveals a root node in the cladogram, from which all evolution proceeded.

11. Cladistic analysis can determine the sequence in which character states and taxa evolved, distinguish between ancestral and derived character states, and detect instances of convergent and divergent evolution.

12. A tenet of phylogenetic systematics is that all named taxa should be monophyletic. By revealing cases where named taxa are not monophyletic, cladistic analysis points the way to changes that increase the predictive value of the taxonomic system.

Questions

1. How do biologists determine whether two organisms belong to different species?

2. What are the main taxonomic levels between the domain and species, and why do the levels constitute a hierarchy?

3. Describe the roles in evolution of mutagens, sexual reproduction, the founder effect, hybridization, and interactions between organisms and environment.

4. Under what circumstance is natural selection likely to be directional rather than stabilizing?

5. How does population genetics show whether a pair of alleles is undergoing evolution in a wild population?

6. How does polyploidy lead to new species?

7. What can be determined from a rooted cladogram but not from an unrooted cladogram?

8. Would insects be useful outgroups for rooting a cladogram of plant species? Why or why not?

9. How do outgroups increase the value of a cladistic analysis?

10. How does cladistics reveal cases of convergent evolution?

11. Copy the cladogram of Figure 18.14, and then draw a circle around a group of species to illustrate the concept of a monophyletic group. Why is it important for named taxa to be monophyletic?

12. To practice the methods used in cladistics, do the exercise in "IN DEPTH: do a Cladistic Analysis of DNA."

IN DEPTH*: Do a Cladistic Analysis of DNA*

This exercise presents a practice problem in cladistics, using concepts from this chapter. Your task is to determine how five given species most likely evolved from their common ancestor. To do it by the cladistic method, you must compare a set of characters that vary among the species. This exercise will take you from the selection of characters through the full cladistic analysis. It assumes that you have determined the sequences of an homologous gene from the five species

Select a Gene and Correct Any Displacement

In a cladistic analysis based on DNA data, the first step is to choose a gene that occurs in all the species you wish to study. Real genes contain thousands of bases along their length, and real studies compare many taxa. But you can learn how to generate a DNA character matrix more easily with a simpler case. We will examine how 5 species differ in the first 32 bases of an imaginary gene (Fig. 1).

Figure 1. The first 32 bases for the same gene in five species. Numbers at the top are nucleotide positions along the gene. A,C,G, and T indicate the DNA base at each position.

The array in Figure 1 already looks like a valid character matrix in which each numbered position is a character, and the base at that position is the character state for a given species. This view would be correct if genes only evolved by replacing one base with another. But sometimes genes lose or add nucleotides at various positions. If that happens, all nucleotides to the right of the alteration are displaced from the position they had in the ancestor of all the species. Unless we discover such displacements and correct them, comparisons of the DNA will lead to false conclusions about evolutionary relationships among the species. The error is the same as comparing analogous rather than homologous structures. If there is no displacement at a given position, the bases at that position in all species are homologous because they descended from the same position in the ancestor's DNA. If a nucleotide is deleted, its place is filled with a nucleotide that began elsewhere in the ancestor's DNA.

To find and correct displacements, look for lengthy regions in Figure 1 where a series of bases is the same in all the organisms under study, and color them for visibility. If a lengthy base sequence is the same in all the species, it is called a *conserved sequence,* because it probably descended from the ancestor with little or no change. The longer the conserved sequence, the more sure we are that it is truly a shared heritage rather than an accidental similarity among molecules. Our result is shown in Figure 2.

Figure 2. Three conserved sequences, with different colors for visibility. Loss or gain of nucleotides displaced the second and third consensus sequences (red, yellow) in some species.

TACA, the first conserved sequence in Figure 2, is already lined up in all five species, occupying position 1 through 4. This must be where the sequence occurred in the common ancestor. The other two sequences are not lined up, so nucleotides must have been removed or added in some species.

The next step is to line up the second and third conserved sequences by breaking the DNA and moving the broken parts to the right, leaving gaps. But how do you choose which species to break, and where to make the breaks? Inspection tells us the simplest choice is to break DNA in species 2 and 5. The first break must be in the black region between the first two conserved sequences--but where, exactly? With a trial-and-error approach, find a point where a shift to the right will reveal new conserved sequences. The more conservation your move reveals, the more likely it is that you found the real positions where DNA changed during the evolution of the species. Our result is shown in Figure 3.

Figure 3. To align the conserved sequences, the sequences in species 2 and 5 have been broken, and the right-hand parts have been moved to the right.

In species 2, we placed the break between position 4 ad 5, then shifted the right-hand part of the DNA one position to the right, leaving position 5 empty. In species 5, we placed the break between positions 11 and 12 and shifted the resulting DNA fragment two positions to the right. As shown by the blue regions in Figure 4, these choices reveal conserved sequences that were overlooked before.

Figure 4. Aligned DNA, with newly revealed conserved regions marked in blue.

Choose Characters and Build a Matrix

With the DNA aligned, Figure 4 shows that most DNA positions in all five species are conserved. Changes (shown in black) occurred at just 9 of the 32 positions along the gene. Forget the bases to the right of position 32; they arose outside the first 32 positions.

Now build a character matrix from the positions that were *not* conserved. These are all the positions, such as position 9, where the five species do not all have the same base. Evolution occurred at these locations, changing DNA after the species parted from the ancestor. Every non-conserved position is a character that may be added to the matrix. But to keep the workload down, only include the six positions where all five species have a base. Your character matrix should list the species along the left side, list the characters (DNA positions) along the top, and show bases (character states) from the aligned DNA at the intersections.

Our matrix is shown in Figure 5. We colored the six DNA positions that contribute to the matrix, and the resulting matrix is shown at the bottom.

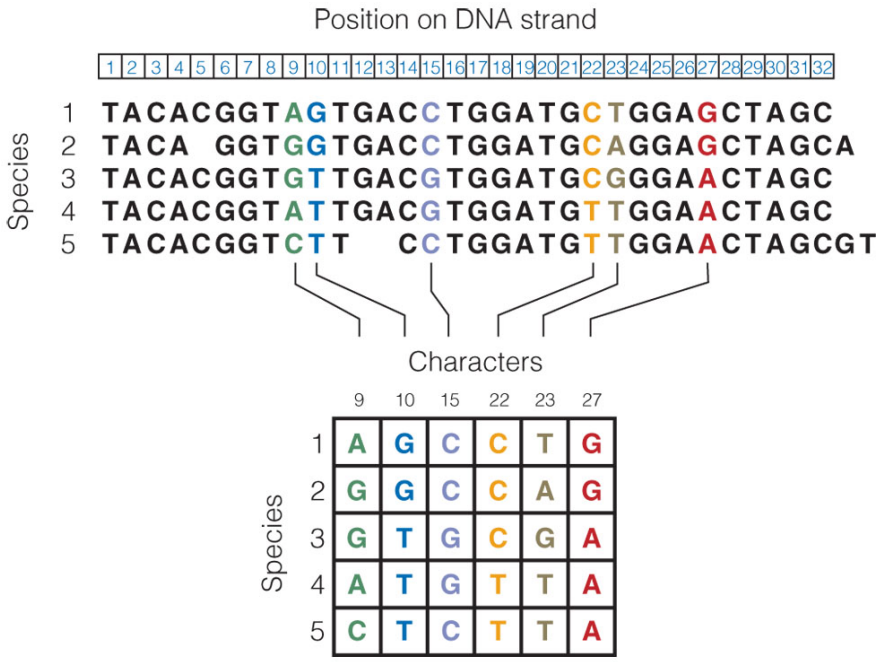


Figure 5. Building the character matrix. Colors indicate non-conserved positions in DNA where every species has a base, and their character states are entered into the matrix (bottom).

Draw All the Possible Cladograms

Your next task is to draw all the possible cladograms. For any 5 species, there are 15 possibilities. Figure 6 shows two of them.

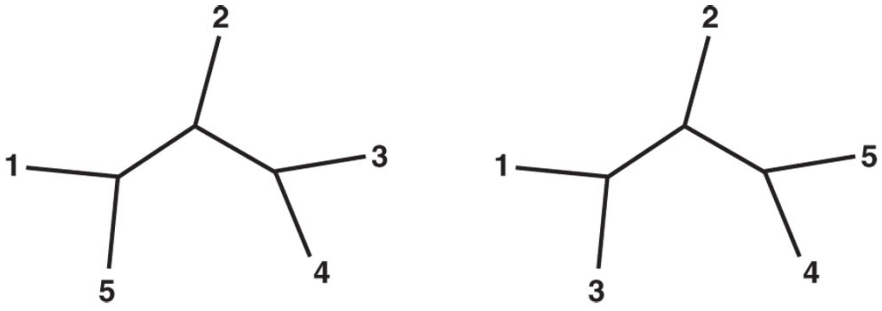


Figure 6. Two of the 15 unrooted cladograms that are possible for any 5 species.

As you draw the other 13 possible cladograms, remember that two cladograms are equivalent unless they differ in how many branch points separate any two species. You need all the possible cladograms to decide which one most likely reflects the evolution of the species.

Determine Which Cladogram Is Most Parsimonious

Cladistics gives a logical procedure to decide which cladogram most likely shows the real evolutionary relationships among taxa. To make the decision, a variety of methods exist. We will use a method based on the principle of parsimony. It assumes that the most probable cladogram is the one that needs the fewest evolutionary events to account for the differences among species.

To apply the principle, work with one cladogram and one character at a time. Figure 7 illustrates the method by showing how many events are needed to explain the evolution of the first character in the matrix (position 9) if the five species are related as in Figure 6a. A description of the steps follows the figure.

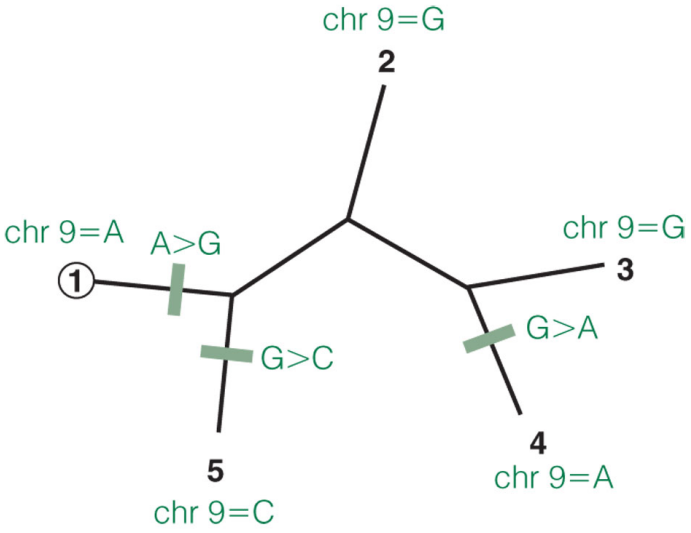


Figure 7. Possible evolution of character 9, assuming the ancestor was like species 1, and the five species are related as in Figure 6a. Current state of character 9 in each species is shown. Labeled tick marks show evolution events (e.g., A>G means A was replaced by G.)

Our first step in the analysis was to choose a point in the cladogram where the ancestor of the five species might have entered to initiate the group. We chose the ancestor to be exactly like species 1, entering at the position of species 1. How do we know this is the real starting point? The answer is that we do not know--but it does not matter, because the outcome of the parsimony analysis does not depend on the starting point you choose--as long as you make the same choice for all cladograms.

With species 1 as the starting point, we guessed that A changed to G before species 1 split, as shown by the tick mark labeled A>G. If that happened, then species 2 and 3 need no further change, because they got G from their ancestor. But species 4 and 5 differ, so two more changes are needed: G become A in the branch that ends at species 4, and G became C in the branch that ends at species 5. Altogether, it took three events (tick marks) to account for the states of character 9.

With this background, your task is to determine how many changes (tick marks) are needed to account for the other five characters in the matrix. Work with one character at a time, following the same approach as in Figure 7. For each character, start the analysis from species 1 so you can compare your result with ours, which is given in Figure 8. You will gain more from this exercise if you try it before checking our answer.

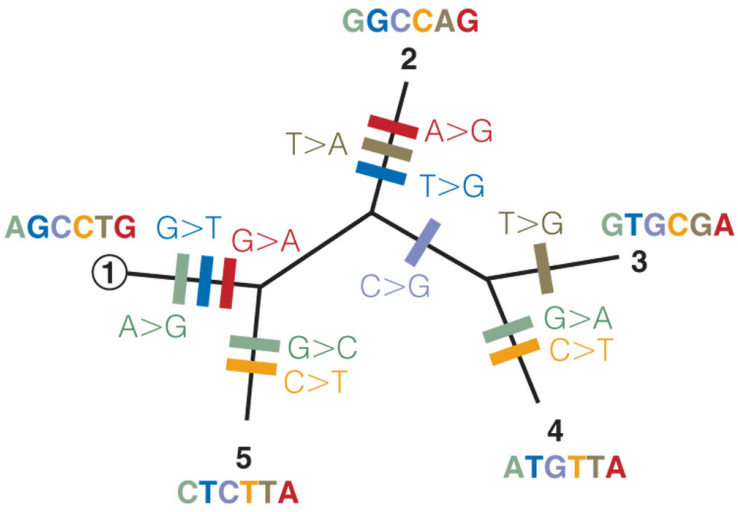


Figure 8. Possible evolution of all six characters in the matrix of Figure 5, if the species are related as in Figure 6a and the ancestor was like species 1. States of all six characters are listed beside each species. Colored tick marks show points where each character might have changed. Alternative points of change are possible.

By now you know that the cladogram in Figure 6a needed 12 character changes to explain the data matrix, if the ancestor was like species 1. But this is only 1 of 15 possible cladograms. To find which cladogram takes the fewest steps, you must conduct a similar analysis for all 14 of the remaining cladograms. To start, analyze the cladogram in Figure 6b. Remember that you must use the same starting point and ancestral character states that you used for Figure 6a.

Our results (Fig. 9) show that the cladogram in Figure 6b needs 10 steps to generate the present character states, if the evolution began with an ancestor identical to species 1. This makes Figure 6b more parsimonious than Figure 6a. Do you agree?

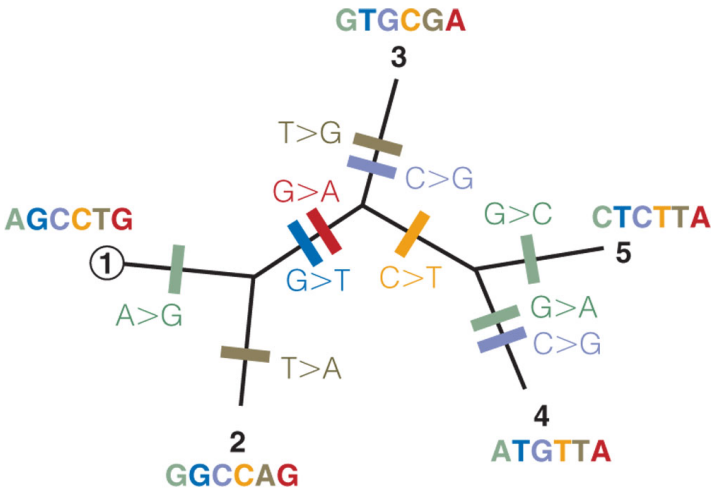


Figure 9. Possible evolution of all six characters in the matrix of Figure 5, starting from an ancestor identical to species 1, if the species are related as in Figure 6b. Other symbolism as described in the legend to Figure 8.

Now complete the search for the most parsimonious cladogram by analyzing the other 13 possible cladograms. Our own analysis showed that 2 of the 15 possible cladograms (Fig. 10a,b) take 10 steps, which is fewer than any of the other cladograms. Those two cladograms are equally parsimonious. When two or more cladograms are equally good, you must either do more studies with more data, or settle for a compromise, called a consensus tree (Fig. 10c).

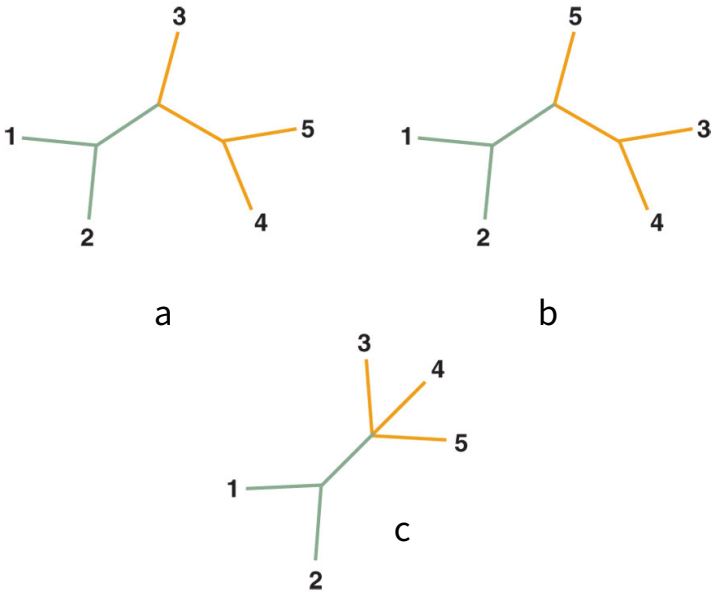


Figure 10. Preparing a consensus tree. (a,b) these two cladograms are equally parsimonious; each takes 10 steps. Nodes where both cladograms agree are shown in green; disagreements are shown in orange. (c) Consensus cladogram, with branches leading to species 3, 4, and 5 coming from a single point because a and b disagree on whether species 3 or 5 is more closely related to species 1 and 2.

A consensus tree (or consensus cladogram) is a diagram that has all the features shared by the equally parsimonious cladograms and leaves the conflicts unresolved (Fig. 10c). Like the two rival cladograms, the consensus tree shows species 1 and 2 arising from a single branch point. Cladists say species 1 and 2 are sisters, because they arose from the same node. But the two cladograms disagree about which of the three remaining species is most closely related to species 1 and 2. To reflect the disagreement, the consensus tree has species 3, 4, and 5 all branching from the same point. We normally expect just three lines to meet at each node. Therefore, the meeting of four lines is a signal that we have not resolved the sequence in which the three species evolved. Many such cases occur in published cladograms.

Root the Cladogram

The final task is to find the root of the chosen cladogram (or consensus tree). This is done by including *outgroups* in the character matrix (Fig. 11). Here we confess to a secret: among the species we have been studying, species 5 was included as an outgroup. Species 1, 2, 3, and 4 make up the *ingroup*, the target of the study (Fig. 11a).

Now rearrange the consensus tree (Fig. 11a) so it shows more clearly how the four ingroup species are related. To rearrange the tree properly, remember that species 5 is not the ancestor of the ingroup clade--it is a modern species, as are the four ingroup species. To represent that fact, the first step in rooting the cladogram is to put all five species at the top (Fig. 11b,c).

Next, locate the point where the ancestor of all five species enters the cladogram. Find that point by inspecting Figure 11a again. Species 5 joins the ingroup at the node where species 3 and 4 arise. Between that node and species 5, the ancestor must have split to give species 5 and the ancestor of the ingroup. Thus, figure 11b inserts a node between species 5 and the ingroup. That node is the root, the oldest node in the cladogram, represented by a dot. It is the point where the ancestor split to launch the evolution of all five species. The rest of branching in Figure 11b is the same as in Figure 11a, with adjustments in the orientation of the lines. Figure 11c shows the same thing with all lines straightened out.

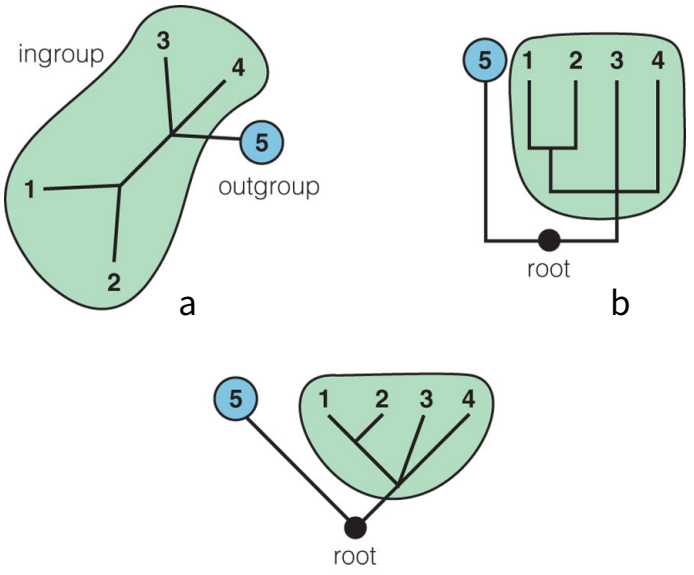


Figure 11. Rooting the consensus tree. (a) Species 5 was included in the study as an outgroup; species 1 through 4 are the ingroup clade (green) that is the focus of study. (b,c) Two ways to rearrange a, placing the oldest node (the root) at the bottom.

With the root in place and the outgroup at one side, the cladogram shows that species 1 and 2 are probably sisters, related to each other more closely than to any other species in the clade. But this consensus cladogram does not reveal the sequence in which species 3 and 4 separated from the ancestor of species 1 and 2. To resolve that point, we would need to repeat the study with more data to find which branch split off first.

For more information, see: http://www.ucmp.berkeley.edu/clad/clad1.html

Photo Credits

18.1 Michael G. Barbour
Table 1 Wikipedia
18.5 SEM by John L. Bowman
18.11 Tim Metcalf