The Mechanisms of Evolution

The rapid advances of molecular genetics over the past two decades have accounted for the origin of mutations and have revealed that the variation within species is much greater than Darwin postulated

by Francisco J. Ayala

In the 119 years since the publication of *On the Origin of Species* Darwin's basic principles have been progressively refined. According to Darwin, the basis of evolution is the occurrence of random heritable modifications in the individuals of a population. The advantageous modifications are then adopted and the disadvantageous ones are discarded through natural selection: the differential survival and reproduction of genetically variant individuals. In this way evolutionary adaptation involves a mixture of variation and selection, of chance and necessity.

Darwin thought of variation as a transient phenomenon. Because a population of organisms is closely adapted to its environment, he argued, the vast majority of modifications will be disadvantageous and the modified individuals will accordingly be eliminated by natural selection. In the rare event that a modification is advantageous it will render the individual more likely to survive and reproduce. As a result the advantageous modification will gradually spread to all the members of the population over the generations, ultimately replacing the type that was formerly dominant.

Darwin's theory implies that natural populations are made up of a more or less common genetic type with a few rare variants. In recent years this assumption has been contradicted by evidence that natural populations possess an enormous reservoir of genetic variation, suggesting that the role of chance in the evolutionary process is subtler than Darwin supposed. The advances in molecular biology, together with the statistical approach to evolution provided by population genetics, have enabled biologists to better understand where genetic variation comes from, how it is maintained in populations and how it contributes to evolutionary change.

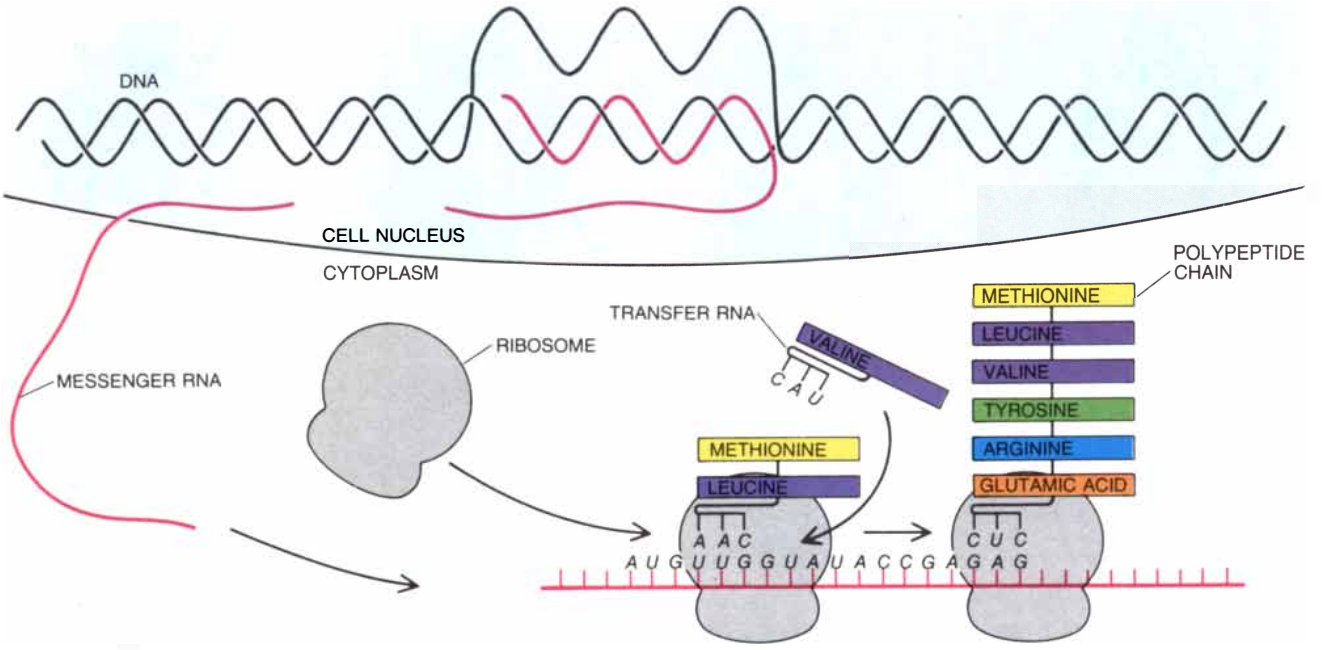
In Darwin's day the science of genetics had not yet been born. The discrete units of heredity called genes were first identified by Gregor Mendel in Darwin's lifetime but did not become widely known until the 20th century. Darwin's vague but prescient notion of random fluctuations in the hereditary material nonetheless turned out to be an approximation of Mendel's more precise concept of genetic variation, and so Mendelian genetics could be incorporated into the theory of natural selection without too much difficulty. The fusion of the two disciplines from the early 1920's through the late 1950's is often referred to as Neo-Darwinism or the modern synthesis.

The dramatic discoveries of molecular genetics over the past 20 years have led to yet another synthesis, encompassing an understanding of evolutionary processes at the molecular level. A gene is now known to be a segment of one of the extremely long DNA molecules in the cell that store the organism's genetic information in their structure. The sequence of four kinds of nucleotide base (adenine, cytosine, guanine and thymine) along each strand of the DNA double helix represents a linear code. The information contained in that code directs the synthesis of specific proteins; the development of an organism depends on the particular proteins it manufactures. Proteins are made up of long chains of amino acids, and the specific properties of each protein are determined by the sequence of amino acids in the chain. This sequence is in turn specified by the sequence of nucleotide bases in the DNA of the genes.

The genetic information stored in the DNA molecule is expressed in two steps. In the first process, called transcription, the sequence of nucleotide bases along one of the DNA strands is copied onto a complementary strand of RNA (which is made up of the same nucleotide bases as DNA except that thymine is replaced by the closely related uracil). In the second process, called translation, the genetic program of the organism is "read" from the RNA in codons, or successive groups of three nucleotide bases. The four RNA bases form 64 different codons that specify the 20 common amino acids in proteins. (The discrepancy between the 64 codons and the 20 amino acids is due to the redundancy of the genetic code and the fact that certain codons represent instructions such as "Start" and "Stop.")

In protein synthesis the amino acids specified by the sequence of codons along the gene are added one by one to the growing chain. Once the protein has been assembled it spontaneously assumes a specific three-dimensional form and begins to function as an enzyme, as a structural component or in some other biological role. The characteristics and behavior of organisms depend ultimately on the sequences of amino acids in

GENETIC VARIATION WITHIN A SPECIES is apparent in the color patterns on the elytra (wing covers) of the Asiatic lady beetle *Harmonia axyridis*, as is illustrated in the painting on the opposite page. A species indigenous to Siberia, Japan, Korea and China, *H. axyridis* occurs in a number of discrete variant forms with different geographical distributions. Variant *19-signata* (*top three rows*) has many patterns of black spots on a yellow field and even a few solid-black individuals, variant *aulica* (*fourth row*) has a large pair of yellow spots on a black field, variant *axyridis* (*fifth row*) has spots that may range in color from orange-yellow to pale orange and variant *spectabilis* (*sixth row*) has red spots on a black field. The geographical distribution of the populations of this species is quite sharp: west-central Siberia is occupied by a population nearly uniform for the black-background *axyridis* pattern. Farther eastward the populations become more variable, with the yellow-background forms such as *signata* increasing in frequency. The red-on-black *spectabilis* pattern is found only in the Far East. The diverse color patterns are believed to be determined by a series of variant forms of the same gene. Although a discrete and striking variation of this type, called a polymorphism, is rare, subtler types of variation are seen in all living species, including man. In addition natural populations harbor large reservoirs of hidden variation, enabling them to adapt to changing environments.



"CENTRAL DOGMA" of molecular genetics states that genetic information flows from DNA to messenger RNA to protein. Genes are relatively short segments of the long DNA molecules in cells. The DNA molecule comprises a linear code made up of four types of nucleotide base: adenine (*A*), cytosine (*C*), guanine (*G*) and thymine (*T*). The code is expressed in two steps: first the sequence of nucleotide bases in one strand of the DNA double helix is transcribed onto a single complementary strand of messenger RNA (which has the same bases as DNA except that thymine is replaced by the closely related uracil, or *U*). The messenger RNA is then translated into protein by means of complementary transfer-RNA molecules, which add amino acids one by one to the growing chain as the ribosome moves along the messenger-RNA strand. Each of the 20 amino acids found in proteins is specified by a "codon" made up of three sequential RNA bases.

their proteins, and evolution consists largely in the progressive substitution of one amino acid for another.

The new understanding of the chemical nature of the gene has provided a view of mutation at the molecular level. A mutation can be considered an error in the replication of DNA prior to its translation into protein. Such an error is often confined to the replacement of one nucleotide-base pair by another (a point mutation), and it may lead to the replacement of one amino acid by another in the protein specified for by that gene. Point mutations that result in the substitution of an amino acid are called missense mutations; those that convert the codon for an amino acid into a "stop" codon are called nonsense mutations. Other mutations may involve the insertion of a nucleotide into the DNA mole-

FIRST RNA NUCLEOTIDE BASE	SECOND RNA NUCLEOTIDE BASE				THIRD RNA NUCLEOTIDE BASE
	U	*C*	*A*	*G*	
URACIL *(U)*	PHENYLALANINE	SERINE	TYROSINE	CYSTEINE	*U*
	PHENYLALANINE	SERINE	TYROSINE	CYSTEINE	*C*
	LEUCINE	SERINE	STOP	STOP	*A*
	LEUCINE	SERINE	STOP	TRYPTOPHAN	*G*
CYTOSINE *(C)*	LEUCINE	PROLINE	HISTIDINE	ARGININE	*U*
	LEUCINE	PROLINE	HISTIDINE	ARGININE	*C*
	LEUCINE	PROLINE	GLUTAMINE	ARGININE	*A*
	LEUCINE	PROLINE	GLUTAMINE	ARGININE	*G*
ADENINE *(A)*	ISOLEUCINE	THREONINE	ASPARAGINE	SERINE	*U*
	ISOLEUCINE	THREONINE	ASPARAGINE	SERINE	*C*
	ISOLEUCINE	THREONINE	LYSINE	ARGININE	*A*
	START / METHIONINE	THREONINE	LYSINE	ARGININE	*G*
GUANINE *(G)*	VALINE	ALANINE	ASPARTIC ACID	GLYCINE	*U*
	VALINE	ALANINE	ASPARTIC ACID	GLYCINE	*C*
	VALINE	ALANINE	GLUTAMIC ACID	GLYCINE	*A*
	VALINE	ALANINE	GLUTAMIC ACID	GLYCINE	*G*

NEUTRAL AROMATIC BASIC ACIDIC SULFUR-CONTAINING

DICTIONARY OF THE GENETIC CODE is tabulated here in the language of messenger RNA. The code is universal: all organisms, from the lowliest bacterium to man, use the same set of RNA codons to specify the same 20 amino acids. In addition *AUG* serves as a "start" codon to signal the beginning of the messenger-RNA transcript, and *UAA, UAG* and *UGA* serve as "stop" codons that signal the end of the transcript and cause the completed protein to be released from the ribosome. The code is highly redundant in that several codons specify the same amino acid. Nevertheless, certain point mutations (single substitutions of one nucleotide-base pair for another in the DNA molecule) may change a codon so that it specifies a different amino acid. In the table amino acids with similar chemical properties have been grouped together by color. Point mutations that result in the substitution of one amino acid for another of the same group ("conservative" mutations) usually lead to subtle changes in the structure and function of a protein. In contrast, point mutations that result in the substitution of one amino acid for another of a different group may lead to drastic changes in the protein. Because of the clustering of amino acids of similar type most point mutations lead to conservative substitutions and hence to minor changes in proteins.

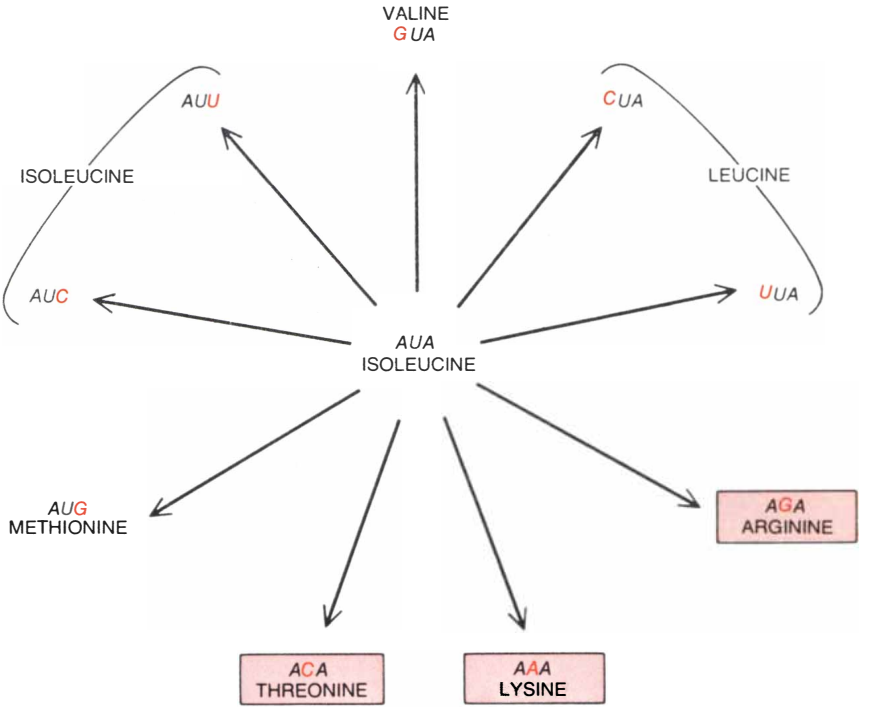
cule or the deletion of a nucleotide from it; such mutations may have more pervasive effects by shifting the "frame" in which the nucleotide sequence is read, and they may lead to several missense or nonsense substitutions. If these DNA mutations occur in the germ cells of the organism, they will be passed on to the next generation.

In addition to changes in the structure of the genes by mutation, evolution involves changes in the amount and organization of the genes. A human being has in each cell many times more DNA than our single-cell ancestors of a billion years ago had. Evolutionary increments (or decrements) in the hereditary material occur largely by means of duplications (or deletions) of DNA segments; the duplicated segments can then evolve toward serving new functions while the preexisting segments retain the original function.

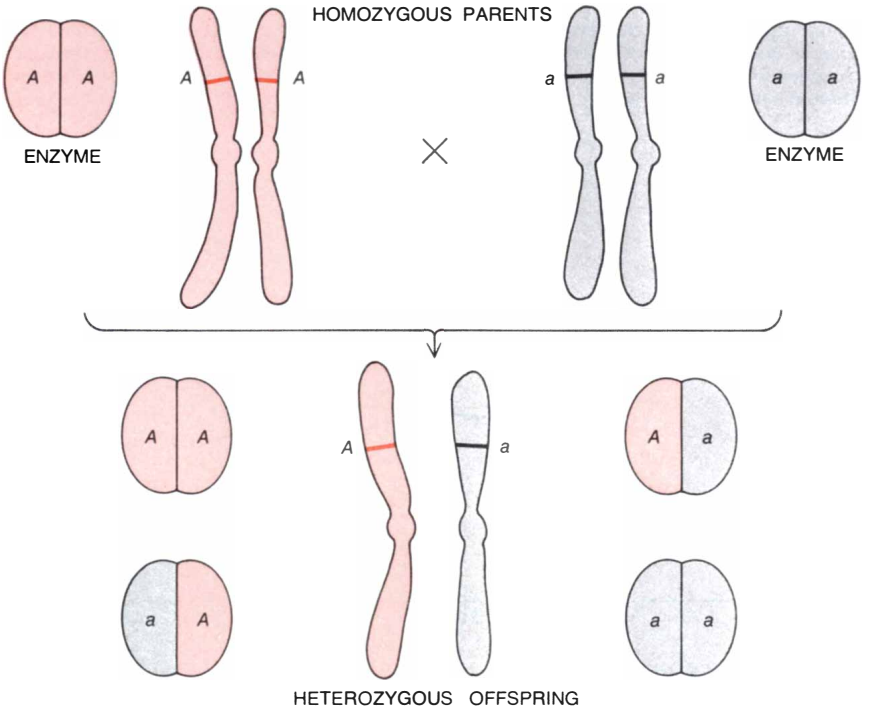
The forces that give rise to gene mutations operate at random in the sense that genetic mutations occur without reference to their future adaptiveness in the environment. In other words, a mutant individual is no more likely to appear in an environment in which it would be favored than in one in which it would be selected against. If a favored mutation does appear, it can be viewed as exhibiting a "preadaptation" to that particular environment: it did not arise as an adaptive response but rather proved to be adaptive after it appeared.

A population consisting of several million individuals is likely to have a few mutations per generation in virtually every gene carried by the population. Mutations that give rise to substantial changes in the physical characteristics of the organism, however, are unlikely to be advantageous. Since a population is usually well adapted to its environment, major changes are usually maladaptive, just as a large random change in the construction of a clock (the removal of a spring or the addition of a gear) is not likely to improve its functioning. Most evolutionary changes seem to occur by the gradual accumulation of minor mutations (analogous to the tightening of a screw) accompanied by slow transitions in the physical characteristics of individuals in the population.

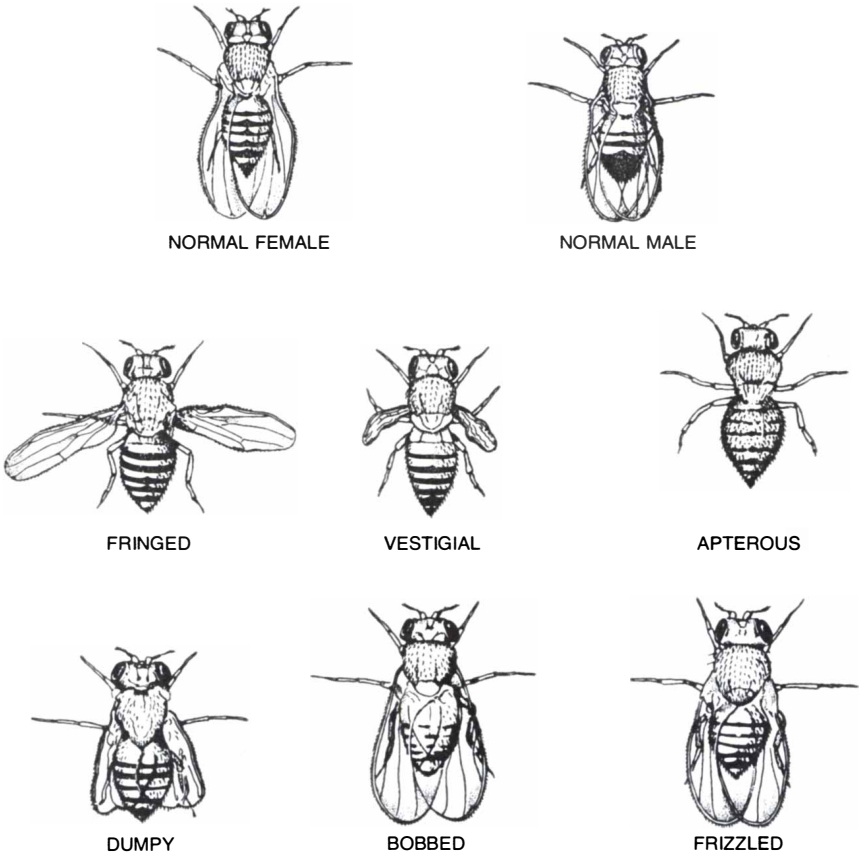
The DNA molecules in the nucleus of higher cells are associated with protein and packed into the dense bodies called chromosomes. The number of chromosomes in the cell nucleus differs from species to species: eight in the fruit fly *Drosophila,* 20 in corn, 24 in the tomato, 40 in the house mouse, 46 in man, 48 in the potato. A substantial reorganization of the hereditary material can result from transpositions of chromosomal segments, each of which comprises hundreds or thousands of nucleotide bases. The total number of chromosomes can be increased by duplication



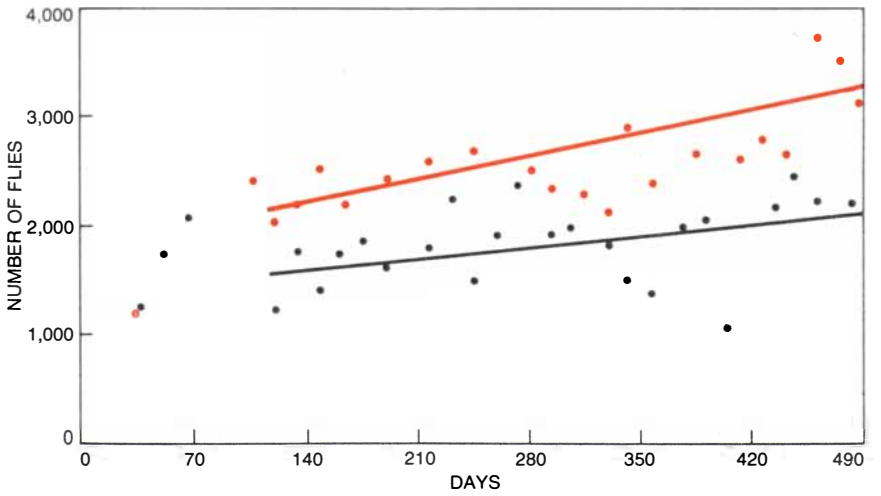
POINT MUTATIONS occur randomly during the replication of DNA molecules. They can be induced by ionizing radiation, elevated temperatures and a variety of chemical reagents or can arise naturally through other processes. This diagram shows that substitutions at the first, second or third position in the messenger-RNA codon for the amino acid isoleucine can give rise to nine new codons that code for a total of six different amino acids. (Because of redundancy of the genetic code some point mutations cause no change in amino acid.) Codons in boxes specify amino acids with chemical properties that differ sharply from those of isoleucine.



ALLELES, or variant genes, are carried on chromosomes at specific positions termed loci. In this diagram the individual at the left has allele *A* at a given locus on two homologous, or corresponding, chromosomes. The individual at the right has a different allele *a* at the same locus on two homologous chromosomes. Because these individuals possess two copies of the same allele, they are called homozygotes. When they are crossed, their offspring will possess one copy of each allele, making them heterozygotes. Because each allele codes for a slightly different protein, heterozygosity can be inferred from the presence of two variants of a given protein in a single individual. For example, here the enzyme coded for by the locus is made up of two identical protein chains that combine spontaneously. Each homozygote will manufacture either the *AA* or the *aa* form of the enzyme, whereas the heterozygote will manufacture *AA, aa* and *Aa.*



DISADVANTAGEOUS RECESSIVE ALLELES in a population of the fruit fly *Drosophila melanogaster* gave rise to the gross anatomical defects shown here. These disadvantageous alleles (originally created in the laboratory by ionizing radiation) are expressed only when they are homozygous; in the heterozygous state they are usually concealed. Their existence was revealed by inbreeding closely related individuals so that many disadvantageous alleles became homozygous in the progeny and were therefore expressed. Such alleles are maintained in a population at low frequencies; they may become advantageous when the environment changes.



EFFECT OF GENETIC VARIATION on the rate of evolution was demonstrated by the author in experiments performed on *Drosophila serrata*. Two populations of the species were studied: one derived from a single strain and the other derived from crossbreeding two different strains, so that it had about twice as much genetic variation. Both types of population were then placed in closed bottles for 25 generations under conditions of intense competition for food and living space, which fosters rapid evolutionary change. Although both the single-strain and the mixed-strain populations became increasingly adapted to the laboratory conditions, as evidenced by a rise in population over a period of time, the average rate of increase in the mixed population (*color*) was about twice that of the single-strain population (*black*). The greater the variation stored in a population is, the more readily it is able to adapt to a new environment.

or reduced by fusion. A segment of a chromosome can be deleted, an extra piece can be inserted or a segment can be removed, inverted and put back. A segment from one chromosome can be transferred to another, or noncorresponding pieces can be exchanged. All these chromosomal aberrations alter the organization of the genes and contribute new raw material for evolutionary change.

Of the 46 chromosomes in every human cell, 23 are copies of those originating in the sperm of the father and the other 23 are copies of those originating in the egg of the mother. The genes thus occur in pairs, one on a maternal chromosome and the other on the homologous, or corresponding, paternal chromosome. The two genes in a pair are said to occupy a certain locus, or position, on each of the homologous chromosomes. For example, there is a locus on one pair of homologous chromosomes that codes for eye color. Each chromosome may comprise many thousands of gene loci.

A gene at a given locus may have variant forms known as alleles. There may be several alleles at a locus in a large population, although there can be only two in any one individual. Each allele arises by mutation from a preexisting gene and may differ from it at one or several parts of its nucleotide-base sequence. When the two alleles at a certain locus on the homologous chromosomes in an individual are identical, the individual is said to be homozygous at that locus; when the two alleles are different, the individual is said to be heterozygous at that locus.

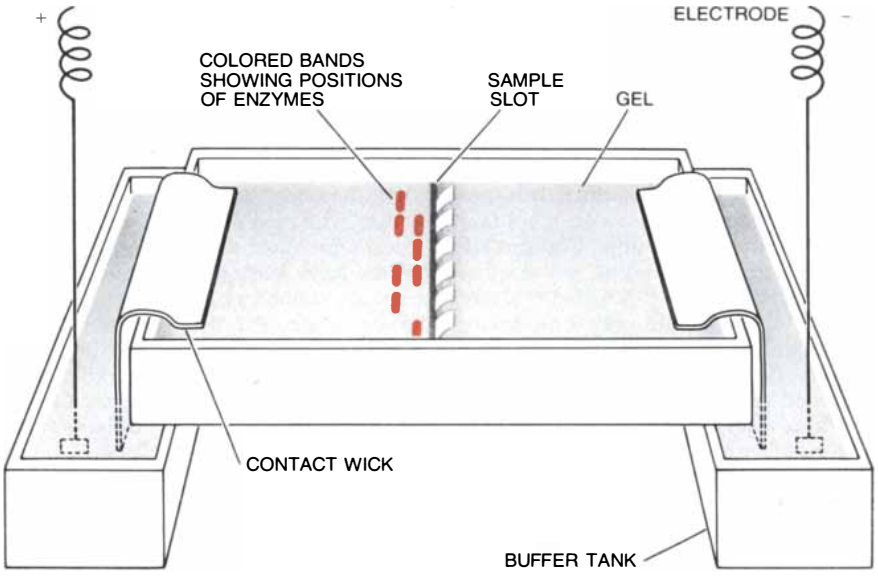
Hereditary variation, as reflected in the existence of multiple alleles in a population, is clearly a prerequisite for evolutionary change. If all the individuals in a population are homozygous for the same allele at a given locus, there can be no evolution at that locus until a new allele arises by mutation. If, on the other hand, two or more alleles are present in a population, the frequency of one allele can increase at the expense of the other or others as a consequence of natural selection. Of course, the selective value of an allele is not fixed. The environment is variable in space and time; under certain conditions one allele is favored and under different conditions another allele is favored. Hence a population that has considerable amounts of genetic variation may be hedged against future changes in the environment.

Laboratory experiments have demonstrated that the greater the amount of genetic variation in a population, the faster its rate of evolution. In one experiment two populations of the fruit fly *Drosophila* were bred so that one population had initially about twice as much

genetic variation as the other. The populations were then allowed to evolve in the laboratory for 25 generations with intense competition for food and living space, conditions that tend to stimulate rapid evolutionary change. Although both types of population evolved, gradually becoming better adapted to the laboratory environment, the rate of evolution was substantially higher in the population that had the greater initial variation.

The question of how much variation exists in natural populations is therefore of central interest to biologists, since it determines to a large extent the evolutionary plasticity of a species. The task of estimating genetic variation, however, is a difficult one since much genetic variation is hidden in each generation and is not expressed as manifest traits. The reason is that at a given locus in a heterozygous individual one allele is usually dominant and the other is recessive, that is, only the dominant allele is expressed in the heterozygous state. If a human being has a dominant allele for brown eye color and a recessive one for blue eye color, his eyes will be brown and the fact that he carries a gene for blue eye color will be concealed.

Such hidden variation can be revealed by breeding experimental organisms with their close relatives. When this inbreeding is done, some of the recessive alleles that have been concealed in the heterozygous state will become homozygous and will then be expressed. For example, intensive inbreeding of fruit flies has revealed they possess several recessive alleles that when the locus is homozygous result in the expression of grossly abnormal traits such as extremely short wings, deformed bristles, blindness and other serious defects.



GEL ELECTROPHORESIS is a method for estimating the genetic variation of natural populations by examining the variant proteins manufactured by different individuals. First a tissue sample from each of the organisms to be surveyed is homogenized to release the proteins in the tissue; the proteins are placed on a gel of starch, agar or polyacrylamide. The gel with the tissue samples is then subjected to an electric current, usually for a few hours. Each protein in the sample migrates through the gel in a direction and at a rate that depend on its net electric charge and molecular size. After the run is over the gel is treated with a chemical solution containing a substrate that is specific for the enzyme under study and a salt. The enzyme catalyzes the conversion of the substrate into a product, which then couples with the salt to give colored bands at the points to which the enzyme had migrated. Because enzymes that are specified by different alleles may have different molecular structures and charges (and hence different mobilities in an electric field) the genetic makeup at the gene locus coding for a given enzyme can be established for each individual from the number and position of the electrophoretic bands.

Another indication of the magnitude of genetic variation in natural populations has been provided by artificial-selection experiments. In such experiments those individuals of a population that exhibit the greatest expression of a particular commercially desirable trait are chosen to breed the next generation. If a plant breeder wants to increase the yield of a variety of wheat, he will select those plants with the greatest yield at each generation and utilize their seed to grow new progeny. If the selected population changes over the generations in the direction of the applied selection, then it is clear the original plants possessed a reservoir of genetic variation with respect to the selected trait.

Indeed, the changes obtained by artificial selection are often enormous. In one flock of White Leghorn chickens egg production increased from 125.6 eggs per hen per year in 1933 to 249.6 eggs per hen per year in 1965: an increase of nearly 100 percent in 32 years! Selection can also be successfully practiced in opposite directions. For example, selection for high protein content in a variety of corn increased the protein content from 10.9 to 19.4 percent, whereas selection for low protein content reduced the protein content from 10.9 to 4.9 percent. Artificial selection has been successful in creating large numbers of commercially desirable traits in domesticated species such as cattle, swine, sheep, poultry, corn, rice and wheat, as well as in experimental animals such as fruit flies, in which artificial selection of more than 50 different traits has been accomplished. The fact that artificial selection works almost every time it is attempted indicates there is genetic variation in populations for virtually every characteristic of the organism.

This kind of evidence suggested to biologists that natural populations do have large stores of genetic variation. Yet until quite recently the limitations of traditional genetic analysis prevented investigators from determining precisely how much variation there is. Consider what would be required to find out what proportion of the genes of an individual are heterozygous. It is almost impossible to study every gene locus because of the scale of the task, but if one could obtain an unbiased sample of all the genes of an organism, it would be possible to extrapolate the values observed in that sample to the population as a whole. Indeed, opinion pollsters are able to predict with fair accuracy how millions of people will vote in a U.S. Presidential election on the basis of a representative sample of about 2,000 people: .001 percent of the population. The fact remains that with Mendelian techniques it is impossible to obtain an unbiased sample of all the genes in an individual because classical genetic analysis (involving crossbreeding of individuals exhibiting different traits) detects only those loci that are variable (that have different alleles). Since there is no way to detect invariant loci, it was impossible to obtain a truly random sample of all the genes.

The way out of this dilemma was provided by the molecular biological revolution of the past two decades. Since many genes code for proteins, one can infer variation in the genetic material from variation in the proteins manufactured by individuals. If a certain protein is invariant among the individuals of a population, the gene coding for that protein is probably also invariant; if the protein is variable, then the gene too is variable. By selecting a number of proteins that represent an unbiased sample of the genes in an organism it is there-

fore possible to estimate the number of alleles in a population and the frequencies at which they occur.

Biochemists have known since the early 1950's how to determine the amino acid sequence of proteins, but several months or years are usually required to sequence one protein, let alone the thousands that would be needed to obtain a statistically valid sample. Fortunately there is a simple technique, gel electrophoresis, that makes it possible to study protein variation with only a moderate investment of time and resources. Since the late 1960's this technique has been exploited to estimate the genetic variation in several natural populations.

In gel electrophoresis ground tissue or blood from several individuals is inserted into a gel consisting of starch, the synthetic polymer acrylamide or some other substance providing a homogeneous matrix. When an electric current is passed through the gel, the proteins in the tissue migrate at a rate that is determined primarily by the electric charge on their constituent amino acids (although the size and conformation of the protein may also influence the migration). Electrophoresis is so sensitive that it can detect proteins that differ by a single amino acid out of a total of some hundreds—provided that the substitution of one amino acid for another results in a change in the total electric charge on the molecule.

The proteins manufactured by different individuals in a population are compared by running them side by side in a gel for a certain time interval. The positions of the proteins after they have migrated are determined by applying a stain specific for the protein under study, which is usually an enzyme. Because each amino acid chain in a protein (some proteins have more than one chain) is the product of a single gene this approach enables the investigator to estimate how many loci in the population have multiple alleles and how many are invariant. To obtain a rough survey of variation in natural populations about 20 loci are usually examined. One useful measure of variation is heterozygosity: the average proportion of loci at which an individual in the population possesses two alleles.

Electrophoretic techniques were first applied to estimating genetic variation in natural populations in 1966, when three studies were published, one dealing with man and the other two with *Drosophila*. Since then numerous populations have been surveyed and many more are studied every year. One recent survey concerned the krill *Euphausia superba,* a shrimplike organism that thrives in the waters near Antarctica and is a major food source of whales. A total of 36 gene loci coding for different enzymes were examined in 126 krill individuals. No variation was detected at 15 loci, but at each of the other 21 loci two, three or four allelic gene products were found in the population. In other words, approximately 58 percent of the loci in this krill population had two or more alleles. On the average each krill individual was heterozygous at 5.8 percent of its loci.

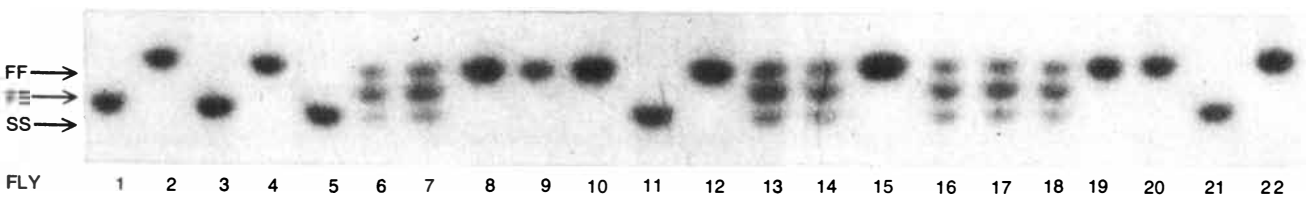
Large amounts of genetic variation have been found in most natural populations studied, including 125 animal species and eight plant species. Among animals, invertebrates generally show more genetic variation than vertebrates, although there are some exceptions. The average heterozygosity for invertebrates is 13.4 percent; the average for vertebrates is 6.6 percent. The heterozygosity for man is 6.7 percent, close to the vertebrate average. Plants have a great deal of genetic variation: the average heterozygosity for eight species is 17 percent.

These estimates become even more dramatic when it is taken into account that electrophoresis underestimates genetic variation. One reason is the redundancy of the genetic code: not all mutations or substitutions in the DNA result in changes in the amino acid sequence of proteins. Moreover, since electrophoresis distinguishes proteins that have different amino acid compositions by their differential migration in an electric field, if a mutation does not alter the electrical properties of the molecule, it will not be detected. For example, if a positively charged amino acid (say glutamic acid) is replaced in a variant protein by another positively charged amino acid (say aspartic acid), the two proteins may be indistinguishable by electrophoretic criteria. Although it is clear that the estimates of variation in natural populations obtained by electrophoresis are underestimates, the degree of underestimation is not yet known. Several laboratories are now attempting to solve this problem so that genetic variation can be more precisely estimated.

In any case the extent of the variation observed in natural populations is vastly greater than that predicted by classical Darwinian theory. Instead of being homozygous for a dominant allele at most loci, individuals are heterozygous at a large proportion of loci. This fact has important consequences, particularly for animals that reproduce sexually.

Sexual reproduction involves the fusion of two germ cells (the sperm and the egg in animals), each of which possesses only one set of chromosomes instead of the two homologous sets possessed by each tissue cell. The germ cells are formed by the process of meiosis, or reduction division, in which the normal complement of chromosomes is reduced by half. In the first stage of meiosis the chromosomes duplicate themselves and the homologous chromosomes then pair up. At this stage the paired chromosomes may break in several places and exchange pieces, the process called recombination. The resulting chromosomes are a mosaic of the homologous paternal and maternal chromosomes and hence have a new combination of alleles. In the second stage of meiosis each cell divides twice to yield four germ cells. During the second division the homologous chromosomes are randomly assorted, so that there is a mixture of maternal and paternal chromosomes in each germ cell.

The scrambling of the genes by recombination (which generates new combinations of alleles on the same chromosome) and random assortment

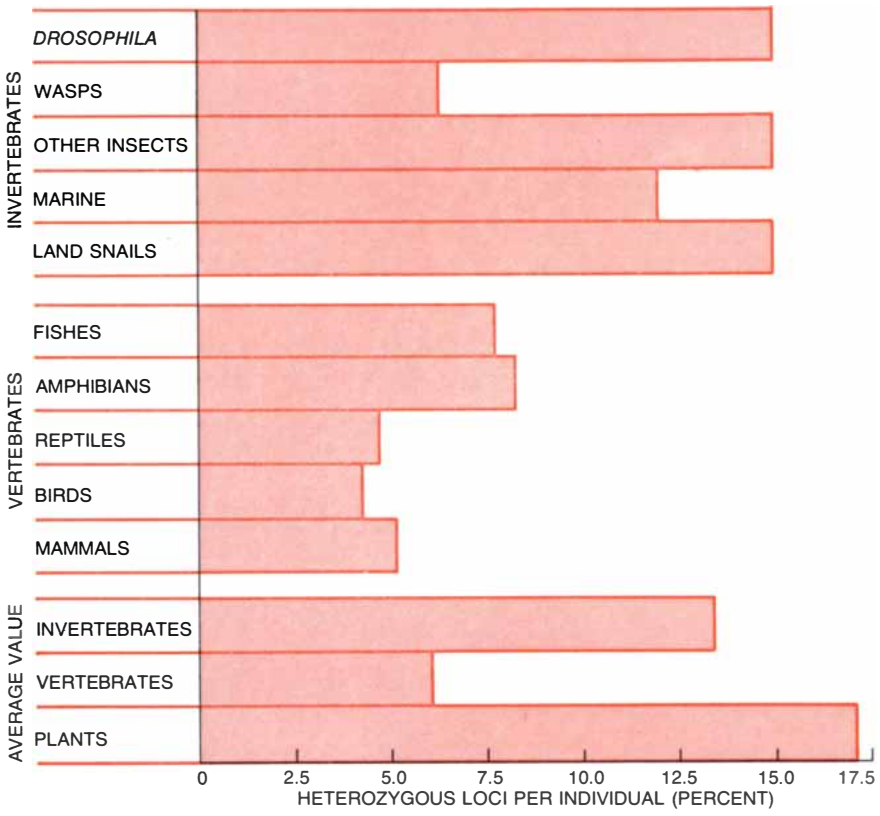


ELECTROPHORETIC GEL shown here was stained for malate dehydrogenase, an enzyme involved in the oxidation of foodstuff. The gel contains samples from 22 flies of the species *Drosophila equinoxialis*. Two variant polypeptides (the protein-chain products of two alleles) are apparent in this experiment: a fast-migrating polypeptide (designated *F*) and a slow-migrating one (designated *S*). Malate dehydrogenase consists of two polypeptide chains that combine spontaneously after they have been synthesized, so that homozygous individuals will make one form of the enzyme (either *FF* or *SS*), whereas heterozygotes will make three forms: *FF, SS* and *FS* (the last form has an intermediate electrophoretic mobility). Thus homozygotes exhibit only one band and heterozygotes exhibit three bands. This case, involving only two alleles in a population, is a simple one; some gene loci that code for proteins may have five alleles or more maintained in the population.

(which results in new combinations of chromosomes in the germ cells) does not in itself alter gene frequencies or cause evolution. Indeed, as it was first independently postulated by the mathematician G. H. Hardy and the biologist W. Weinberg in 1908, recombination and random assortment cause no net change in the frequencies of alleles in a population. In the absence of selection gene frequencies will remain constant from generation to generation, a hypothetical situation that has been named the Hardy-Weinberg equilibrium. The effect of recombination and random assortment is merely to reshuffle the existing genes in a population so that new combinations of alleles are exposed to selection at each generation. Sexual reproduction therefore generates a large amount of genetic diversity, greatly increasing the possibilities for evolution and providing the population with an adaptability to a changing environment far beyond the reach of an asexual species. It may be for this reason that sexuality is virtually universal in the living world, except for organisms such as bacteria, which reproduce rapidly and exist in large numbers and so may incorporate mutations in short periods of time.

Clearly the greater the heterozygosity of individuals in a sexually reproducing population is, the larger will be the number of possible combinations of alleles in the germ cells and hence in the potential progeny. Consider man, with an average heterozygosity of 6.7 percent. If we assume that there are 100,000 gene loci in man, a human individual would be heterozygous for about 6,700 loci. Such an individual could potentially produce $2^{6,700}$ ($10^{2,017}$) different germ cells, a number vastly greater than the number of atoms in the known universe (roughly estimated as being 10^{80}). Of course, such a number of germ cells will never be produced by any human individual, not even all of mankind. It follows that no two human beings ever have been or ever will be genetically identical (with the exception of identical twins and other multiple births from the same zygote, or fertilized egg). Such is the genetic basis of human individuality. The same can be said of any other organism that reproduces sexually.

It therefore seems clear that, contrary to Darwin's conception, most of the genetic variation in populations arises not from new mutations at each generation but from the reshuffling of previously accumulated mutations by recombination. Although mutation is the ultimate source of all genetic variation, it is a relatively rare event, providing a mere trickle of new alleles into the much larger reservoir of stored genetic variation. Indeed, recombination alone is sufficient to enable a population to expose its hidden variation for many generations without the need for new genetic input by mutation.

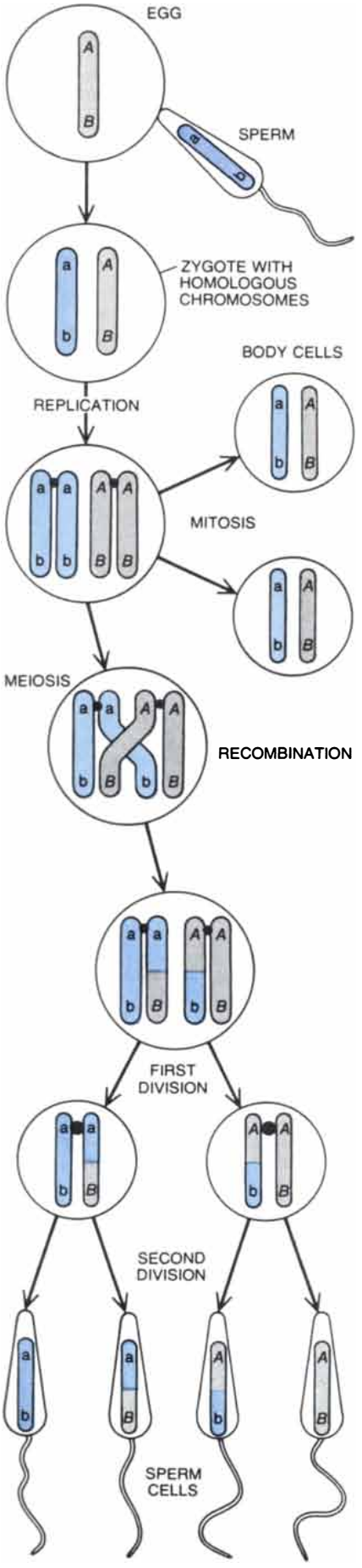


AMOUNT OF GENETIC VARIATION in natural populations, as estimated by gel electrophoresis, is surprisingly large. In general the invertebrates show more variation than the vertebrates, and the few plant species that have been studied show even more. The large numbers of alleles that are stored in the population, mostly at low frequencies, give it evolutionary flexibility.

One can conclude that large numbers of alleles are stored in populations even though they are not maximally adaptive for that time or place; instead they are maintained at low frequency in the heterozygous state until the environment changes and they suddenly become adaptive, at which point their frequency gradually increases under the influence of natural selection until they become the dominant genetic type. But how do natural populations maintain the large reservoirs of genetic variation needed to respond to a changing environment? When one allele is locally more adaptive than another, one would expect that natural selection would gradually eliminate the less advantageous alleles in favor of the more advantageous ones until every locus is homozygous. Hence the persistence of locally disadvantageous alleles in a population can be explained only by postulating mechanisms that actively maintain diversity in spite of the selective forces tending to eliminate it.

One such mechanism is heterozygote superiority. If the heterozygote *Aa* survives or reproduces better than either homozygote *AA* or *aa*, then neither allele can eliminate the other. The most striking example of the mechanism is sickle-cell anemia. This human disease, which is prevalent in tropical Africa and the Middle East, results from an allele that gives rise to a variant form of hemoglobin, the oxygen-transporting protein in red blood cells. Biochemical studies have shown that the trait is due ultimately to the substitution of one amino acid (valine) for another (glutamic acid) at one position along two of the four constituent chains (with a total of nearly 600 amino acids) in the hemoglobin molecule; the abnormal hemoglobin can be distinguished from the normal form by electrophoresis. The slight change in the structure of the variant hemoglobin has catastrophic effects: it causes the hemoglobin molecules inside the red blood cells to form long strands. As a result the cells collapse to the shape of a sickle, resulting in a severe form of anemia that is usually fatal before reproductive age.

Since the sickle-cell allele is obviously disadvantageous, why does it persist in the human population of tropical Africa at frequencies of as high as 30 percent? It turns out that individuals who are heterozygous for the sickle-cell trait are protected against the most lethal form of malaria, whereas normal homozygotes are not. Hence the heterozygote



individual is clearly superior over either homozygote: he is protected from malaria and does not suffer from sickle-cell anemia. As a result the heterozygotes preferentially survive and reproduce and the sickle-cell allele is maintained at high frequency in the population.

Selection may also act directly to maintain multiple alleles in a population. If the range of a species encompasses several different environments, natural selection will diversify the gene pool in such a way that several alleles will be optimally adapted to the different subenvironments. Indeed, recent investigations have shown that variant enzymes (coded for by different alleles) may differ in their catalytic efficiency, in their sensitivity to temperature, acidity or alkalinity and in their response to other environmental factors, thereby rendering them subject to natural selection. For example, some variants of the enzyme alcohol dehydrogenase in populations of the fruit fly *Drosophila melanogaster* have been found to be more resistant to heat than other variants; the heat-resistant variants are commoner in the fruit-fly populations of warmer environments than they are in those of cooler environments. This finding provides strong evidence that multiple alleles may be maintained at some loci by "diversifying selection" in populations that live in heterogeneous environments. Individuals that are heterozygous at a number of loci are also usually stronger and reproductively more successful than individuals homozygous at a large number of loci; the phenomenon is known as hybrid vigor. Perhaps the manufacture of slightly variant proteins and enzymes by the heterozygote enables it to adapt to a broader range of environmental conditions or to exploit marginal habitats.

A fourth mechanism by which multiple alleles can be maintained in a population is frequency-dependent selection, in which the fitnesses of two alleles are not constant but change with their frequency. If one allele is less advantageous than the other when it is at a high frequency but gains the advantage when its frequency declines to a certain level, then the frequency of that allele will tend to stabilize at about that level.

It is also possible that some of the variation observed in proteins represents insignificant changes at the functional level that do not alter the survival or reproductive success of the organism; such mutations would then be selectively neutral. For example, although some variant enzymes (such as the variants of alcohol dehydrogenase) have been found to have different functional characteristics, others may not. If this is the case, the few variant genes that are subject to natural selection might be scattered along a chromosome, together with other variant genes that are selectively neutral. Although some of the alleles would be selected for, the majority would merely be carried along without being tested. The extent to which evolution, particularly at the molecular level, is not subject to selection is a matter of continuing debate among evolutionary biologists.

Another controversy that has been aroused by the finding of large amounts of variation in populations is the problem of genetic load. If large numbers of less fit alleles are maintained in populations by heterozygote superiority, there will be a very high probability that at each generation a zygote will be homozygous at one or more loci for a disadvantageous allele. As a result a large number of less fit zygotes might be expected, which could be a burden of mortality and infertility too great for the population to bear. Yet it must be remembered that each locus is not subject to selection separate from the others, so that thousands of selective processes would be summed as if they were individual events. The entire individual organism, not the chromosomal locus, is the unit of selection, and the alleles at different loci interact in complex ways to yield the final product. Since alleles are more likely to be tested as members of groups than as isolated units, the cost of maintaining variation in a population is actually far lower than was originally believed.

In any case there can be no doubt that the staggering amount of genetic variation in natural populations provides ample opportunities for evolution to occur. Hence it is not surprising that whenever a new environmental challenge materializes—a change of climate, the introduction of a new predator or competitor, man-made pollution—populations are usually able to adapt to it.

A dramatic recent example of such adaptation is the evolution by insect species of resistance to pesticides. The story is always the same: when a new insecticide is introduced, a relatively small amount is enough to achieve satisfacto-

ALLELES ARE RESHUFFLED during sexual reproduction. The germ cells are formed by meiosis, or reduction division, during which the homologous chromosomes exchange corresponding segments, the process called recombination. The homologous paternal and maternal chromosomes are also randomly distributed into germ cells, so that additional combinations of alleles are created. The greater the heterozygosity of two mating individuals, the larger the number of possible sets of alleles in the germ cells and hence in the potential progeny. Meiosis does not change gene frequencies; it exposes new combinations of alleles to selection at each generation.

ry control of the insect pest. Over a period of time, however, the concentration of the insecticide must be increased until it becomes totally inefficient or economically impractical. Insect resistance to a pesticide was first reported in 1947 for the housefly (*Musca domestica*) with respect to DDT. Since then resistance to one or more pesticides has been reported in at least 225 species of insects and other arthropods. The genetic variants required for resistance to the most diverse kinds of pesticides were apparently present in every one of the populations exposed to these man-made compounds.

The process of evolution has two dimensions: phyletic evolution and speciation. Phyletic evolution is the gradual changes that occur with time in a single lineage of descent; as a rule these changes result in greater adaptation to the environment and often reflect environmental changes. Speciation occurs when a lineage of descent splits into two or more new lineages and is the process that accounts for the great diversity of the living world.

In sexually reproducing organisms a species is a group of interbreeding natural populations that are reproductively isolated from any other such groups. The inability to interbreed is important because it establishes each species as a discrete and independent evolutionary unit; favorable alleles can be exchanged between populations of a species but cannot be passed on to individuals of other species. Since species are unable to exchange genes, they must evolve independently of one another.

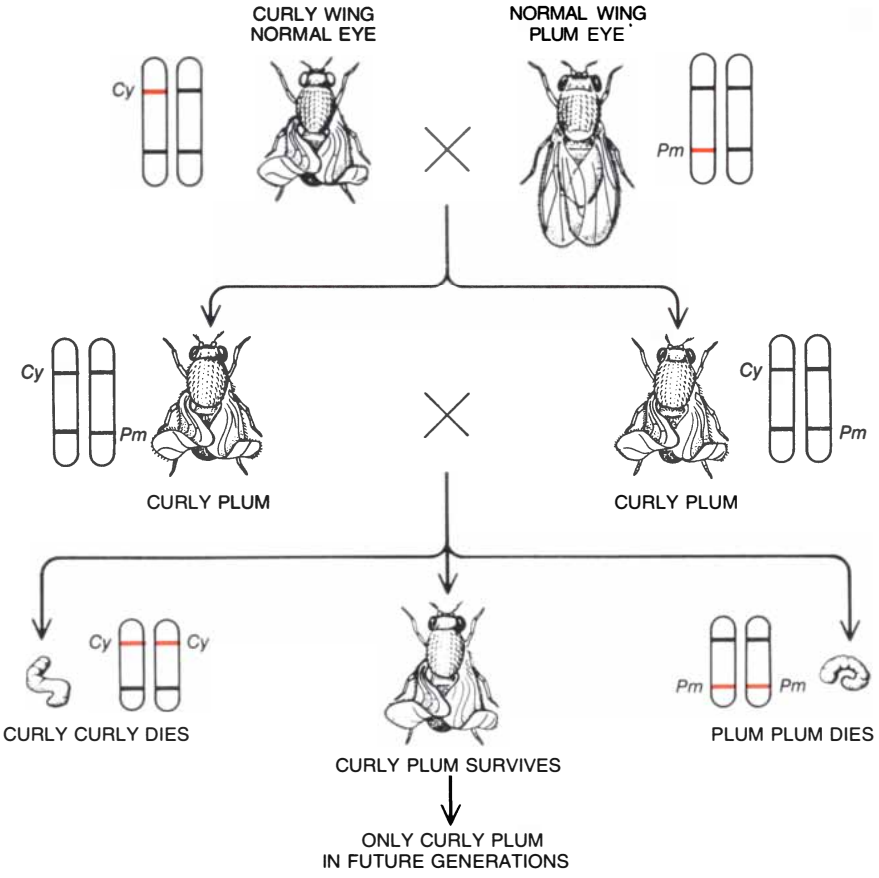
The reproductive isolation of species is maintained by means of biological barriers known as reproductive isolating mechanisms. These mechanisms are of two types: prezygotic mechanisms, which impede the mating between members of different populations and so prevent the formation of hybrid offspring, and postzygotic mechanisms, which reduce the viability or fertility of hybrid offspring. Both types of isolating mechanisms serve to forestall the exchange of genes between populations.

The prezygotic reproductive isolating mechanisms are of five major types: (1) ecological isolation, where populations occupy the same territory but live in different habitats and so do not meet; (2) temporal isolation, where mating in animals and flowering in plants occur in different seasons or at different times of day; (3) ethological isolation, where sexual attraction between males and females is weak or absent; (4) mechanical isolation, where copulation in animals or pollen transfer in plants is prevented because of the different size or shape of the genitalia or the different structure of flowers, and (5) gametic isolation, where the gametes, or male and female germ cells, fail to attract each other. The spermatozoa of male animals may also be inviable in the sexual tract of females or pollen inviable in the stigma of flowers.

The postzygotic isolating mechanisms are of three major types: (1) hybrid inviability, where the hybrid zygotes fail to develop or at least to reach sexual maturity; (2) hybrid sterility, where hybrids fail to produce functional gametes, and (3) hybrid breakdown, where the offspring of hybrids have reduced viability or fertility.

All these reproductive isolating mechanisms do not act simultaneously between two species, but two or more are usually operating. Temporal isolation tends to be commoner in plants and ethological isolation commoner in animals, but even among closely related species different sets of isolating mechanisms are often operating when different pairs of species are compared. The evolutionary function of reproductive isolating mechanisms is to prevent inbreeding, but how this end is achieved depends on the opportunism of natural selection acting in the context of the specific environmental circumstances and the available genetic variation.

Clearly the waste of reproductive effort is far greater for postzygotic isolating mechanisms than it is for prezygotic ones. If a hybrid zygote is produced that is inviable, two gametes have been wasted that could have been used in nonhybrid reproduction. Even worse, if the hybrid is viable but sterile, the waste includes not only the gametes but also the resources utilized by the hybrid during its development. The waste is still greater in the case of hybrid breakdown, which involves the resources utilized by both the hybrids and their offspring. Although gametic isolation also wastes gametes, and some other prezygotic mechanisms waste energy in unsuccessful courtship or failed copulation, in general prezygotic mechanisms are much less wasteful than postzygotic ones. For this

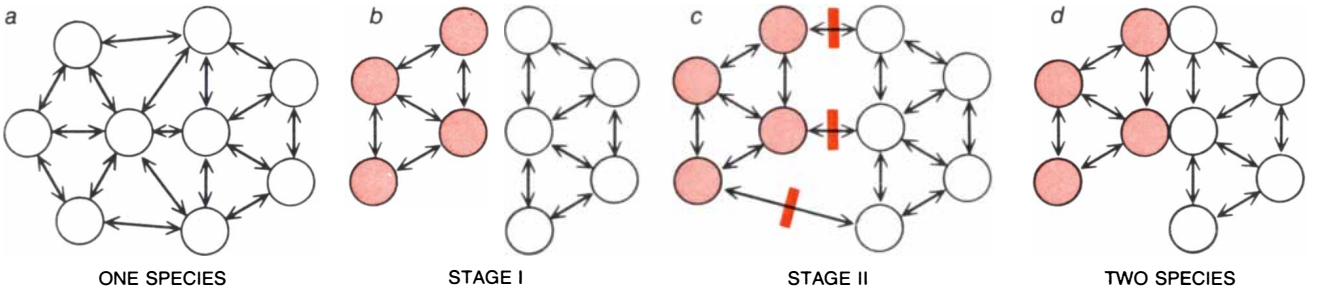


HETEROZYGOTE SUPERIORITY is one way natural selection can maintain disadvantageous alleles in a population. Shown here is a "balanced lethal" situation in *Drosophila* where homozygotes for either the "curly wing" allele or the "plum eye" allele die but the heterozygotes survive. As a result the two lethal alleles remain indefinitely in the population at frequencies of 50 percent each. A less extreme example of this mechanism in man is the case of the sickle-cell allele, which gives rise to an abnormal form of hemoglobin. Individuals who are heterozygous for the sickle-cell allele have a selective advantage over both homozygotes because they do not suffer from sickle-cell anemia (which afflicts homozygotes for the sickle-cell allele) and they are resistant to malaria (which afflicts homozygotes for the normal hemoglobin allele).

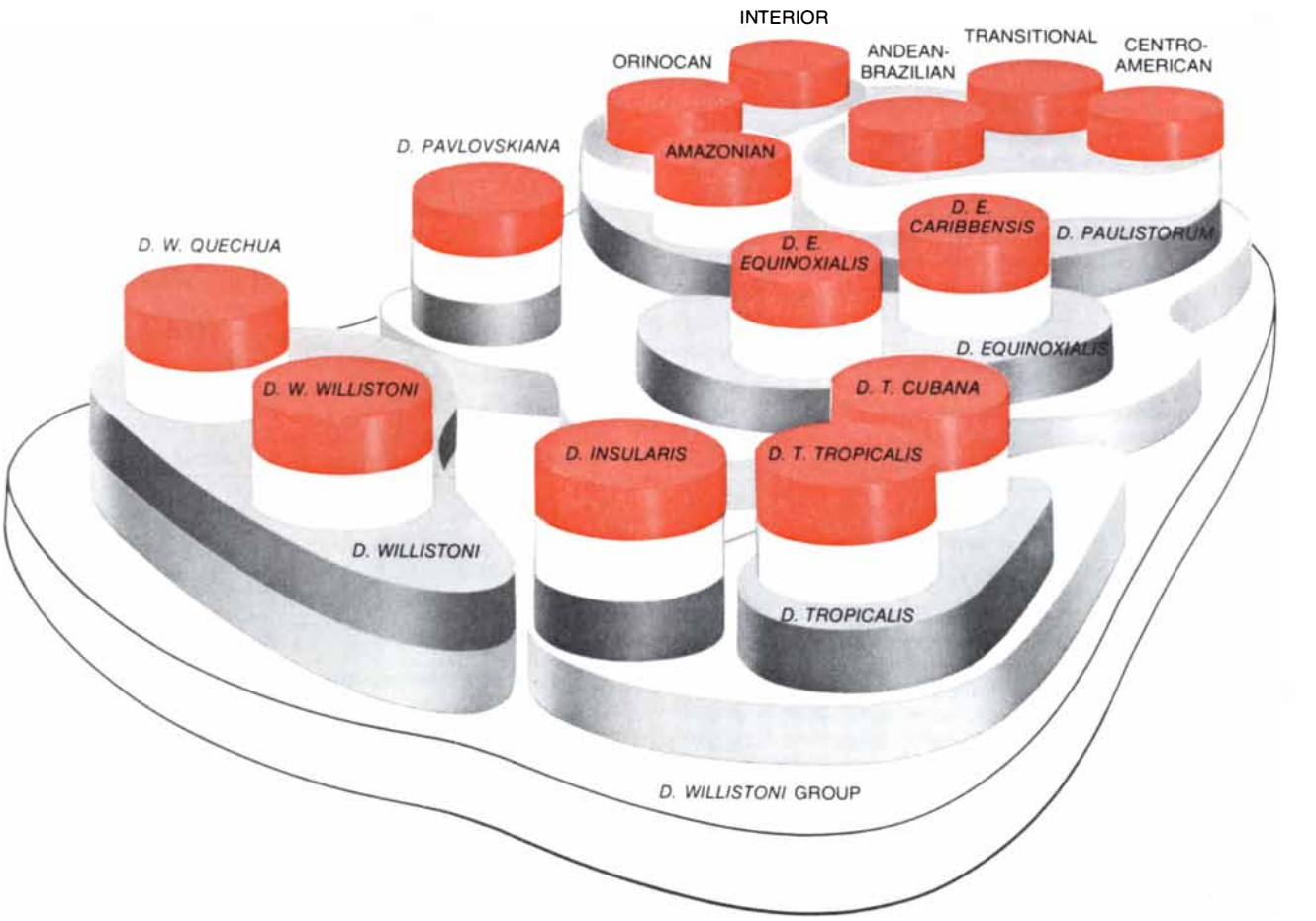
reason whenever two populations that have already been reproductively isolated by postzygotic mechanisms come in contact natural selection rapidly promotes the development of prezygotic isolating mechanisms.

Since species are reproductively isolated groups of populations, the question of how species arise is equivalent to that of how reproductive isolating mechanisms arise. Speciation commonly has two stages: a first stage in which reproductive isolation starts as an incidental by-product of the genetic divergence between two populations, and a second stage in which reproductive isolation is completed when it is directly promoted by natural selection.

The first stage of speciation requires that the exchange of genes between two populations of a species be interrupted, usually by means of a geographical separation (say the formation of a



GEOGRAPHICAL SPECIATION usually occurs in two stages. In *a* local populations of a single species are represented by circles; the arrows indicate that crossbreeding may occur when individuals migrate from one population to another. Stage 1 (*b*) begins when two groups of populations become geographically isolated, so that there is no further exchange of genes between them. The isolated groups adapt to local conditions and gradually diverge genetically. In Stage 2 (*c*) individuals from the two isolated populations again come in contact. Because of the genetic divergence between the two groups crossbreeding gives rise to unviable or sterile offspring. Natural selection therefore favors the development of less wasteful prezygotic isolating mechanisms, which prevent mating between the two groups. In *d* speciation is complete: the two groups of populations coexist in the same territory without ever exchanging genes and hence evolve separately.



FORMATION OF NEW SPECIES within the *Drosophila willistoni* group is represented by a series of cross sections in time through the diverging phylogenetic branches. The morphologically very similar subspecies of *D. willistoni* and *D. equinoxialis* are in the first stage of speciation. The six semispecies, or incipient species, of *D. paulistorum* exhibit prezygotic isolating mechanisms and hence are in the second stage of speciation. Where two or three semispecies of *D. paulistorum* inhabit the same locality speciation is virtually complete.

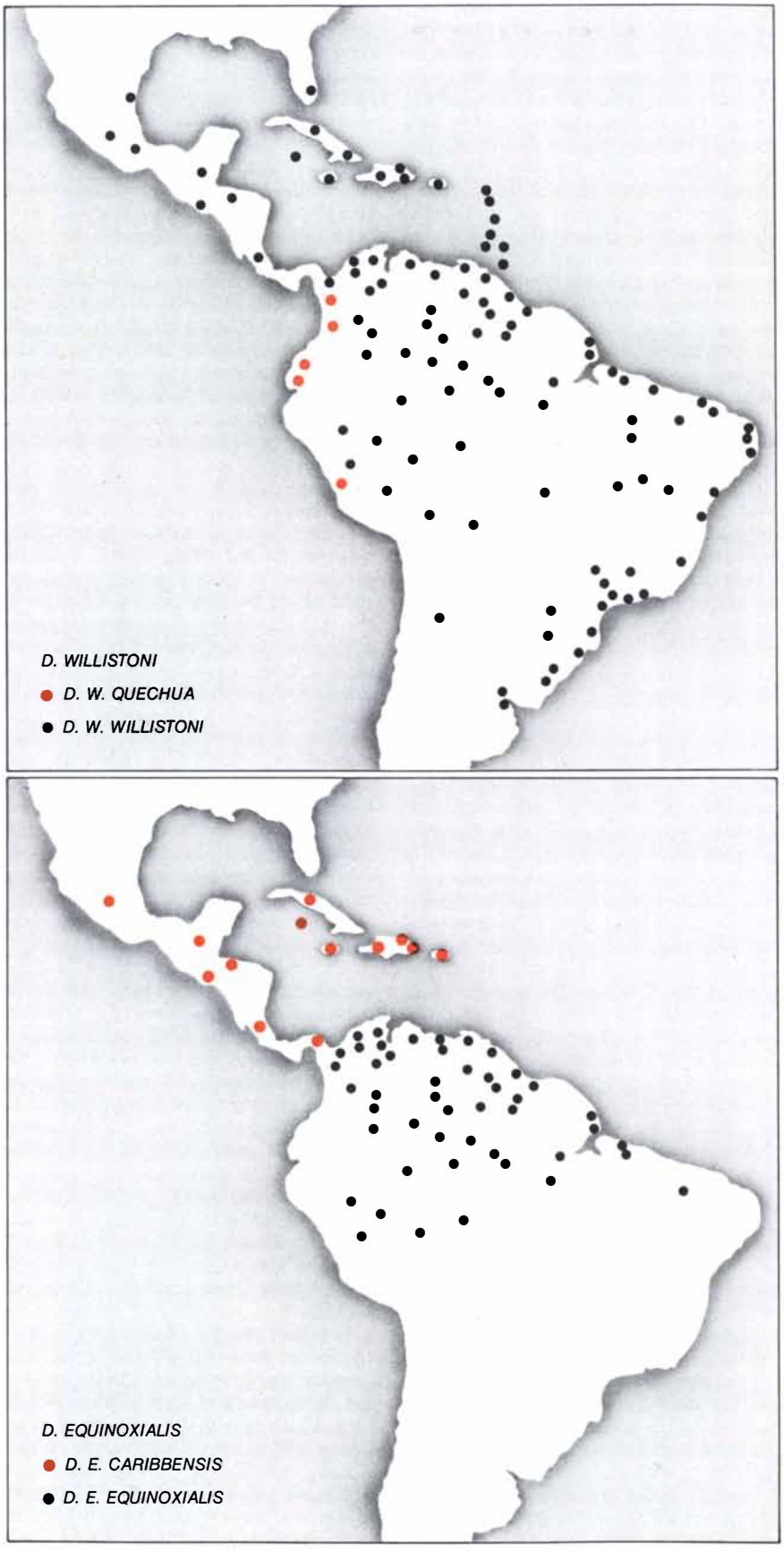
mountain range between them or the emigration of one of the populations to an island). The absence of gene exchange between the two populations makes it possible for them to diverge genetically, at least in part as a consequence of their adaptation to local conditions or ways of life. As the isolated populations become increasingly different genetically, postzygotic isolating mechanisms may appear between them because hybrid offspring would have disharmonious genetic constitutions and hence a reduced viability or fertility.

The first stage of speciation is usually a gradual process, and it is often difficult to decide whether or not two populations have entered it. Moreover, the first stage may be reversible: if two populations that have been geographically isolated for some time come to have overlapping ranges, it is possible for the two populations to fuse back into a single one if the loss of fitness in the hybrids is not too great. If, on the other hand, crossbreeding yields offspring with significantly reduced viability or fertility, the populations will undergo the second stage of speciation.

The second stage involves the development of prezygotic isolating mechanisms, a process that is directly promoted by natural selection. Consider the following simplified situation. Assume that at a certain locus there are two alleles: *A*, which favors matings within the population, and *a*, which favors crossbreeding with other populations. If postzygotic isolating mechanisms operate between two populations, *A* will be common in offspring of normal fitness and *a* will be common in hybrid offspring of low fitness. As a result the *a* allele will decrease in frequency from generation to generation. Natural selection therefore favors the development of prezygotic isolating mechanisms that avoid the formation of hybrid zygotes.

Speciation can take place without the second stage if gene exchange between two populations is prevented long enough for them to diverge genetically to a significant extent. For example, the ancestors of many plants and animals now indigenous to the Hawaiian Islands arrived there from the mainland several million years ago. There they evolved and became adapted to the local conditions. Although natural selection did not directly promote reproductive isolation between the species evolving in Hawaii and the species on the mainland, the reproductive isolation of many species has nonetheless become complete.

The two stages of speciation are apparent in a group of closely related species of *Drosophila* that live in the American Tropics. The group consists of 15 species, six of which are morphologically very similar and so are termed sibling species. One of the sibling spe-



GEOGRAPHICAL ISOLATION of the subspecies of *D. willistoni* and *D. equinoxialis* is indicated on these maps. *D. willistoni willistoni* inhabits continental South America east of the Andes, Central America, Mexico and the Caribbean islands, whereas *D. willistoni quechua* inhabits South America west of the Andes. *D. equinoxialis equinoxialis* inhabits South America, whereas *D. equinoxialis caribbensis* inhabits Central America and the large Caribbean islands.

cies, *D. willistoni,* consists of two subspecies (races of a species that inhabit different geographical areas): *D. willistoni quechua,* which lives in continental South America west of the Andes, and *D. willistoni willistoni,* which lives east of the Andes and also in Central America, Mexico and the islands of the Caribbean. These two subspecies do not meet in nature; they are separated by the Andes because the flies cannot survive at high altitudes. Tests have shown that there is incipient reproductive isolation between the subspecies, particularly in the form of hybrid sterility, although the result depends on the direction of the matings. When a female *willistoni* is crossed with a male *quechua,* the male and female offspring are fertile. If, however, a male *willistoni* is crossed with a female *quechua,* the female offspring will be fertile and the males will be sterile. If these two subspecies came in geographical contact and crossbred, natural selection would favor the development of prezygotic reproductive isolating mechanisms because of the subspecies' partial hybrid sterility. The two subspecies are therefore considered to be in the first stage of speciation.

Drosophila equinoxialis is another species that consists of two geographically separated subspecies: *D. equinoxialis equinoxialis,* which inhabits continental South America, and *D. equinoxialis caribbensis,* which lives in Central America and the Caribbean islands. Laboratory crosses between the two subspecies always yield fertile female offspring and sterile male offspring, independent of the direction of the cross. Thus there is somewhat greater reproductive isolation between the two subspecies of *D. equinoxialis* than there is between the two subspecies of *D. willistoni*. Natural selection in favor of prezygotic reproductive isolating mechanisms would accordingly be stronger for *D. equinoxialis* because all the hybrid males are sterile. There is no evidence, however, of prezygotic isolating mechanisms among the subspecies of either *D. willistoni* or *D. equinoxialis,* and therefore they are not yet considered different species.

The second stage of the speciation process can also be found within the *D. willistoni* group. *Drosophila paulistorum* is a species consisting of six semispecies, or incipient species. As in *D. equinoxialis,* crosses between males and females of these semispecies yield fertile females and sterile males. In places where two or three semispecies have come in geographical contact, however, the second stage of speciation has advanced to the point where ethological isolation—the most effective prezygotic isolating mechanism in *Drosophila* and many other animals—is nearly complete. Semispecies from the same locality will not crossbreed in the laboratory but semispecies from different localities will; the reason is that the genes involved in ethological isolation have not fully spread throughout the populations. The semispecies of *D. paulistorum* therefore provide a striking example of the action of natural selection in the second stage of speciation. When ethological isolation is complete, the six semispecies will have become fully distinct species.

The final result of the process of geographical speciation can be observed in the species of the *D. willistoni* group. *D. willistoni, D. equinoxialis, D. tropicalis* and other species of this group coexist over wide territories without ever interbreeding. Hybrids are never found in nature, are extremely difficult to obtain in the laboratory and are always completely sterile.

Speciation is only one step, albeit the most fundamental one, in the diversification of the living world. Once reproductive isolation has been completed each newly formed species will take an independent evolutionary course; inevitably the species will become increasingly different as time passes. Since evolution is a gradual process, organisms that share a recent common ancestor are likely to be more similar to one another than organisms that share a remoter ancestor. This simple assumption is the logical basis of efforts to reconstruct evolutionary history by comparative studies of living organisms, which traditionally have been based on comparative morphology, embryology, cell biology, ethology, biogeography and other biological disciplines.

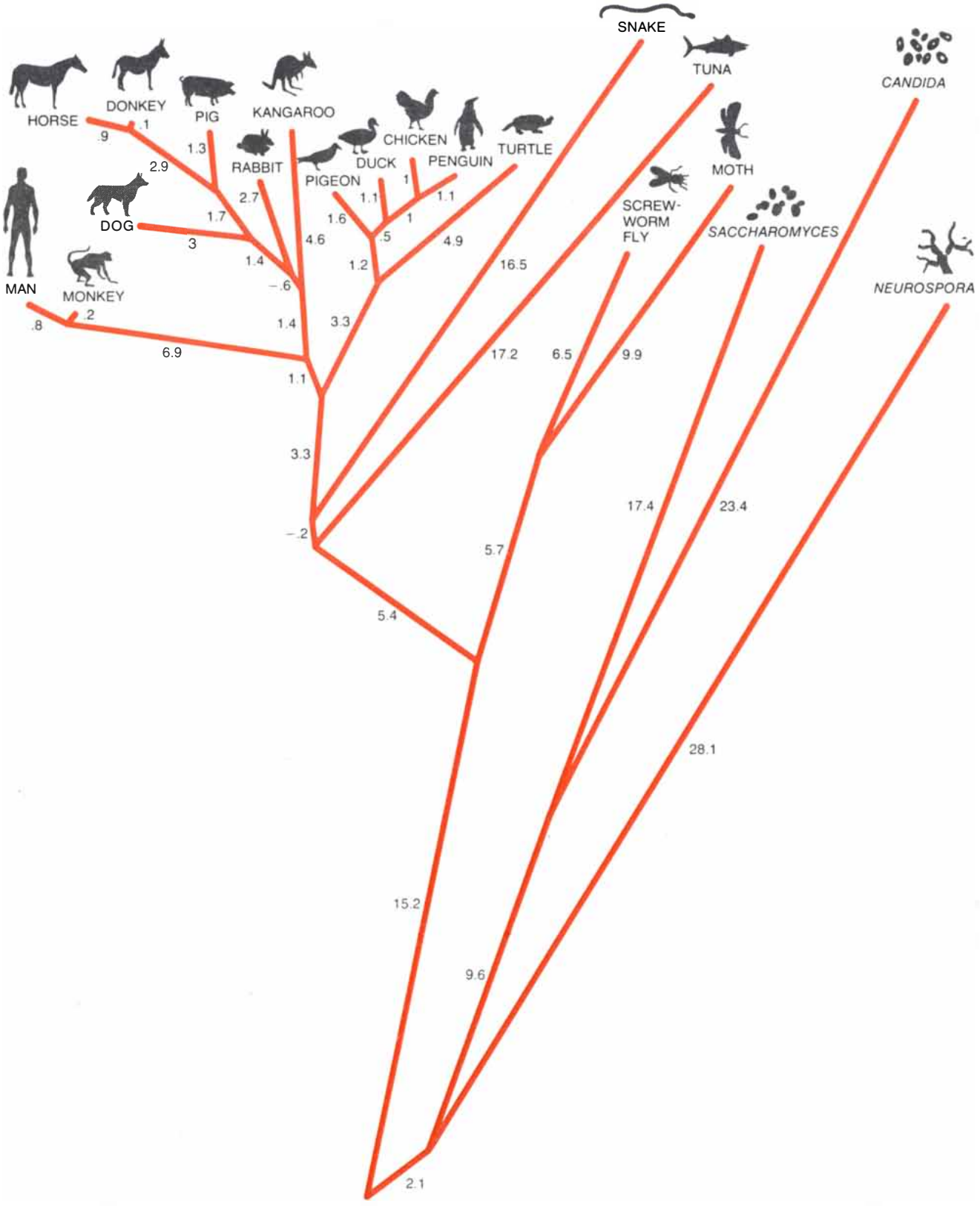
The task of reconstructing evolutionary history is far from simple: rates of evolutionary change may vary at different times, in different groups of organisms or with respect to different morphological features. Moreover, resemblances due to common descent must be distinguished from those due to similar ways of life, to the occupation of similar habitats or to accidental convergence. Sometimes the study of the fossil remains of extinct organisms provides clues to the evolutionary history of a group of species, but the fossil record is always incomplete and often altogether lacking.

In recent years the comparative study of nucleic acids (DNA and RNA) and proteins has become a powerful tool for the reconstruction of evolutionary history. These informational molecules retain a considerable amount of evolutionary information in their sequence of nucleotides or amino acids. Since at the molecular level evolution proceeds by the substitution of one nucleotide or amino acid for another, the number of differences in the sequence of an equivalent nucleic acid or protein in two species provides some indication of the recency of their common ancestry. One widely studied protein is cytochrome *c,* a protein involved in cell respiration; another is hemoglobin.

Investigations of evolutionary history at the molecular level have two notable advantages over comparative anatomy and other classical disciplines. One is that the information is more readily quantifiable: the number of amino acids or nucleotides that are different is readily established when the sequence of units in a protein or a nucleic acid is known for several organisms. The second advantage is that very different types of organisms can be compared. There is little that comparative anatomy can tell us about organisms as diverse as yeast, a pine tree and a fish, but there are proteins common to all three that can be compared readily.

For example, the amino acid sequence of cytochrome *c* has been determined for several organisms, from bacteria and yeast to insects and human beings. Since each amino acid substitution can involve one, two or three nucleotide substitutions in the corresponding DNA codon, one can calculate the maximum or minimum number of nucleotide changes that could have given rise to the observed amino acid substitutions. Taking the minimum number of possible nucleotide differences between the genes coding for cytochrome *c* as a basis of comparison for 20 different organisms, Walter M. Fitch and Emanuel Margoliash at Northwestern University were able to construct a phylogeny of these animals [*see illustration on opposite page*]. The overall relations agree fairly well with those inferred from the fossil record and other traditional sources. The cytochrome *c* phylogeny disagrees with the traditional one in several instances, including the following: the chicken appears to be related more closely to the penguin than to ducks and pigeons; the turtle, a reptile, appears to be related more closely to birds than to the rattlesnake, and man and monkeys diverge from the mammals before the marsupial kangaroo separates from the placental mammals.

In spite of these erroneous relations, it is remarkable that the study of a single protein yields a fairly accurate representation of the evolutionary history of 20 diverse organisms. A more accurate molecular phylogeny of these species and others should be obtained when the sequences of additional proteins and nucleic acids have been determined. The study of informational molecules from an evolutionary standpoint is a young science that was founded only about a decade ago. It is a powerful approach that should make increasingly important contributions to our understanding of biological evolution.



COMPUTER-GENERATED PHYLOGENY of 20 diverse organisms, based on differences in the amino acid sequence of cytochrome *c* from each species, was prepared by Walter M. Fitch and Emanuel Margoliash at Northwestern University. The phylogeny agrees fairly well with evolutionary relations inferred from the fossil record and other sources. The numbers on the branches are the minimum number of nucleotide substitutions in the DNA of the genes that could have given rise to observed differences in amino acid sequence.