Population Genetics: Reevaluation of Genetic Variation

Darwin's theory of evolution by natural selection appears to be consistent with most observed biological phenomena, and hence has won widespread acceptance in the scientific community. However, Darwin's theory has been criticized on the grounds that it is more often used to make retrodictions than predictions. It is difficult to devise experiments whose possible outcomes may be incompatible with this theory. Thus the advancement, about 6 years ago, of a theory of evolution (the neutralist theory) that leads to specific predictions and to experimental tests of those predictions was of considerable philosophical as well as scientific interest. But experimental tests of the neutralist theory have not yet provided conclusive evidence in accordance with those predictions.

Neutralist Theory

The neutralist theory of evolution leads to descriptions of the expected amounts and patterns of genetic variation in natural populations. After analyzing extensive studies of such genetic variation, many investigators now believe that the issue between the Darwinian and the neutralist theories is whether the neutralist theory is sufficient to explain most of the genetic variation, including evolutionary genetic substitution, in natural populations. Recent attempts to resolve this issue have led investigators to reevaluate the techniques used to measure genetic variation and to develop new mathematical models that may enable them to interpret genetic variation in terms of appropriate combinations of the two theories.

Proponents of the neutralist theory of evolution believe that many of the mutations that occur in the DNA of members of natural populations and eventually become fixed in those populations are adaptively neutral. An organism whose DNA contains a neutral mutation would experience neither enhanced nor diminished fitness to survive in a given environment. Over a period of time, random matings among members of a population would result in the elimination of some neutral mutations from the population and the establishment of other neutral mutations. This random elimination and establishment of neutral mutations is called genetic drift. Fundamental to the neutralist theory is the hypothesis that genetic drift plays an important role in the establishment of genetic changes in a population during evolution.

In order to assess the contribution of genetic drift to the evolution of species, investigators have compared both amino acid sequences of proteins (such as hemoglobin or cytochrome c) that have the same function in different species and entire DNA molecules from cells of different species. When similar proteins from different species were compared, they often varied greatly in their amino acid sequences. The amino acid sequence of a protein is a function of the DNA sequence of the gene that codes for that protein. Thus variations in specific proteins are measures of genetic variation.

When the entire DNA molecules from cells of various species were compared, they too differed greatly in both the amount of DNA in the cells of the species (cells from certain amphibian species, for example, have 10 to 20 times as much DNA as do cells from other, closely related, species) and in the similarity of the DNA sequences. The studies of specific proteins and of DNA molecules led investigators to propose that mutations accumulate at a steady rate in DNA during evolution. According to T. Jukes of the University of California at Berkeley, the large number of seemingly neutral mutations that are fixed in different species are not a consequence of the theory of natural selection and are strong evidence that the neutralist theory is important to explanations of the evolution of species.

Specific predictions about the distribution of genetic variations in natural populations, which follow from the neutralist theory, include the prediction that different populations which do not interbreed should tend to have different patterns of genetic variations. This follows from the hypothesis that a substantial amount of genetic variation is due to genetic drift. Another prediction is that most of the variation among those genes that do not interact with other genes can be described by a function that specifies the expected number of neutral mutations in a population. The function allows for predictions of numbers of neutral mutations in terms of the population size and the rate at which neutral mutations arise. Other predictions include expected distributions of genetic variants.

The neutralist theory was developed after R. Lewontin of Harvard University and others had observed large amounts of genetic variation in natural populations. Since that time many investigators have concerned themselves with developing and refining ways to measure this variation. Most measurements have been based on measurements of variations in the amino acid sequences of specific proteins.

Measurements of Variations

Most measurements of variations in specific proteins have been performed by the technique of gel electrophoresis, whereby proteins are separated on the basis of their directional movements in an electric field. Proteins that differ in charge or molecular weight move at different rates through a gel in an electric field and they are detected with a specific dye. Mutations that lead to an alteration in the protein's net electrostatic charge can then be detected.

Gel electrophoresis has been used to study genetic variation in natural populations of a wide variety of organisms, including the fruit fly (*Drosophila*), the horseshoe crab, the mouse, the bacterium (*Escherichia coli*), and man. All the organisms studied showed a large amount of electrophoretic variation. For example, S. Prakash of the University of Rochester estimates that populations of the fruit fly species *D. pseudoobscura* have electrophoretic variants at about 38 percent of their genetic loci.

Some biologists, extrapolating from the few genetic loci so far examined, conclude that a similar percentage of variation occurs at all loci. However, such extrapolations could be misleading. One possible source of error arises from the fact that only a restricted, and perhaps unrepresentative, class of genes has been studied. With few exceptions, the genes that were studied code for soluble enzymes (proteins), since these gene products are most

easily isolated and identified. Neither genes that are necessary for physiological control mechanisms nor genes that control the development of an organism's form or behavior have been studied.

Preliminary evidence that the variation among genes that code for soluble enzymes is different from the variation among other kinds of genes was recently described by T. Yamazaki and T. Maruyama of the National Institute of Genetics in Mishima, Japan. Yamazaki and Maruyama report that the distribution of genetic variation among soluble enzymes is significantly different from the distribution among the blood group proteins.

Another problem with the measurement of genetic variation by gel electrophoresis is that it is difficult to evaluate the resolution of this technique. Since a mutation is only detected if it results in a net change in the electrostatic charge of a protein, mutations that fail to result in a net alteration of electric charge would be undetected.

The resolution of proteins by gel electrophoresis may be increased by the technique of isoelectric focusing. When an electric field is focused at the isoelectric point of a protein, the protein's sensitivity to directional movement in a gel is at a maximum. This new technique has enabled investigators to resolve additional protein variants from bands that appeared to represent single protein species when analyzed by gel electrophoresis. For example, D. Willner and M. Hayes of Cornell University Medical School in Ithaca, New York, found that isoelectric focusing permitted the resolution of 13 different protein variants from what appeared to be a single band obtained upon gel electrophoresis.

Another way to assess the resolution of the products of gel electrophoresis would be to make a comparison with other methods of measuring genetic variation. A recent attempt to do this provides evidence that estimates of variation based on gel electrophoresis may be unexpectedly low. L. Throckmorton, J. Hubby, and S. Bernstein at the University of Chicago used differences in heat stability to measure genetic variations in the enzyme xanthine dehydrogenase from *Drosophila*. In particular, they investigated enzymes from flies whose genes for xanthine dehydrogenase appeared to be identical when compared by gel electrophoresis. When the genes were studied by comparisons of heat stability of the gene product xanthine dehydrogenase, about twice as many genetic variants were detected as when they were studied with electrophoresis. The nine species of *Drosophila* studied by Throckmorton's group appeared to contain 11 variants of the xanthine dehydrogenase gene when the gene was studied with electrophoresis. Throckmorton's group detected 32 variants of the gene when they measured heat stabilities of the gene product.

Several investigators, including J. King of the University of California in Santa Barbara and M. Kimura and T. Ohta of the National Institute of Genetics in Mishima, Japan, have made a theoretical approach to the problem of assessing the technique of gel electrophoresis. For example, King has considered a model in which the probability of two different proteins moving to the same position during gel electrophoresis is a function of the total number of sites at which the amino acid residues of the proteins are different. Because these models are theoretical treatments of expected errors in electrophoretic measurements of genetic variation, they have immediate applications to other theoretical models that allow for the prediction of expected variation due to neutral genes.

The classical model of this type is one proposed by J. Crow of the University of Wisconsin in Madison and Kimura. The model proposed by Crow and Kimura is based on the assumption that every genetic change could be experimentally detected. With the many studies with electrophoresis in view, Kimura and his colleague Ohta have now incorporated an analysis of electrophoresis in a new model that allows them to predict that the number of detectable neutral genes should be fewer than they previously expected.

The models proposed by Crow and Kimura and by Ohta and Kimura have been criticized on several grounds. First, the models are based on the assumption that gene frequencies are distributed independently of each other. Many geneticists believe that models of interacting genes are more realistic than models in which genes do not interact. Those investigators who have suggested tentative models in which genes are assumed to affect each other's distribution have postulated numbers and distributions of neutral genes quite different from those postulated by Crow and Kimura or Ohta and Kimura.

A second criticism of the models of Crow and Kimura and of Ohta and Kimura is that both models contain the parameter $N\mu$. The term N, the size of the population, is generally a very large number. The term μ, the mutation rate, to neutral genes, is generally a very small number. Thus any observed amount of variation can be explained by appropriate choices for N and μ.

The problem of estimating the parameter $N\mu$ is circumvented in a statistical model developed by W. Ewens of the University of Pennsylvania. In Ewens's model, the neutralist theory is used to predict patterns of gene frequencies, and this prediction does not depend on either N or μ. Ewens evaluated published data on patterns of gene frequencies and concluded that the neutralist theory is not sufficient to explain at least one-third of those patterns.

In addition to leading to predictions of expected frequencies and patterns of distribution of neutral variants, the neutralist theory leads to descriptions of the geographical distributions of neutral variants of genes. Many investigators who have tested this prediction believe that it is not true.

Natural Selection

On the basis of his studies of genetic variation in natural populations of *Drosophila*, F. Ayala of the University of California at Davis argues that genetic drift is not sufficient to explain the distributions that he has observed. Ayala studied 70 natural populations of *D. willistoni* and found that, at most loci, the frequencies of genetic variants were the same in the different populations. This observation might be partially explained by geographical migration among the populations since, according to Maruyama and Kimura, only one migrant individual per generation, regardless of population size, is sufficient to keep populations essentially identical. However, Ayala has several arguments to answer this objection.

One argument is that, if certain of the populations studied interbreed, genetic recombination will be restricted. Genetic recombination is a process that results in the occurrence of progeny whose combinations of genes are different from those of their parents. This process is the means whereby the genes of the progeny of a randomly breeding population can become similar. Genetic recombination among the genes of the different populations that Ayala studied will be restricted because many of those populations had chromosomal inversions (reversals of

the usual linear order of genes in a chromosome). The inversions differed in organisms from different geographic areas. If a migrant from a population whose chromosomes contain an inversion mates with a member of a population whose chromosomes contain a different inversion, their viable progeny will not comprise any individuals whose chromosomes contain genetic recombinants of the inverted chromosomes. Thus, Ayala believes that migration could not account for the similarity of variants of genes of the inverted chromosomes in the different populations.

Another reason Ayala believes that migrations did not affect his results is that the geographical patterns of genetic variation he observed are inconsistent with the geographical patterns that genetic migration would produce. Ayala and his associate M. Tracey compared pairs of closely related species from among the 12 species he studied. For any two species, the frequencies of genetic variants were the same at about one-half of the genetic loci and were completely different at the remainder of the loci studied. Moreover, Ayala and Tracey found different sets of identical gene frequencies when they compared different pairs of species. If migrations affected the gene frequencies, those frequencies should be the same at all loci when pairs of species are compared.

Ayala believes that his work and the experimental work of many other investigators provide convincing evidence that a great deal of genetic variation may be maintained in a population as a result of natural selection (back to the Darwin theory). This belief has been questioned by others who point out that most empirical data can be interpreted so as to support either theory. For example, underlying Ayala's arguments is the assumption that he is studying randomly breeding populations and that most observed variations in genes are the result of mutations that took place a long time ago and have become fixed in the population. The validity of such assumptions is difficult to establish.

Ayala has, however, performed an experiment that lends support to one explanation of the roles of natural selection in maintaining genetic variation, namely, that genetic variation may enable a population to adapt to environmental fluctuations. In order to test the hypothesis, Ayala and his colleague J. MacDonald isolated 18 genetically identical populations of *Drosophila*. They subjected different populations to different amounts of environmental fluctuations. For example, some populations received one type of food, and some received two types of food. Some were maintained at a constant temperature and some were not. After 1 year (12 to 15 generations), Ayala and MacDonald measured the genetic variation of these flies by analyzing 20 genetic loci with gel electrophoresis. Increasing variation in the environment resulted in more genetic variation at those loci.

A recent theoretical study by J. Gillespie of the University of Pennsylvania complements Ayala's experimental work. Gillespie was concerned with the problem of explaining genetic variation and the prevalence of heterozygotes (individuals that have two different copies of a given gene) in natural populations. Most experimental studies indicate that the cells of individuals who are heterozygous for genes that code for enzymes synthesize enzymes whose functions are intermediate between the two variants.

Gillespie observed that enzyme functions may change in response to changes in the environment. His statistical model allows him to conclude that heterozygotes will be maintained in fluctuating environments. The intermediate functions of the gene products of individuals who are heterozygous would result in their being more able to adapt to an environment that fluctuates between extremes in which one of the two homozygotes (individuals who carry two identical copies of a gene) is at an advantage. Gillespie bases his model on the assumption that genes act in an additive fashion so that the fitness of a heterozygote can be described as one-half the sum of the fitnesses of the two associated genetic variants. He notes that this assumption is consistent with experimental data.

Ayala and Gillespie interpret their work as supporting the contention that natural selection may play a major role in the maintenance of genetic variation among members of natural populations. In view of the inherent difficulties in measuring genetic variation and in developing models that describe interacting genes, no firm conclusions about the roles of neutralism and natural selection can yet be made. However, the significance of the neutralist theory transcends the problem of whether or not it is a correct description of genetic variation and evolution. As Ayala and Crow point out, it has led to provocative experiments and continues to make a definite contribution to population genetics.—GINA BARI KOLATA

Additional Reading

1. T. Yamazaki and T. Maruyama, *Science* **183**, 1091 (1974).
2. J. L. King, *J. Mol. Evol.* **2**, 317 (1973).
3. F. Ayala and M. L. Tracey, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 999 (1974).
4. J. Gillespie and C. H. Langley, *Genetics*, in press.