The Nearly Neutral Theory of Molecular Evolution

Tomoko Ohta

Annual Review of Ecology and Systematics, Vol. 23. (1992), pp. 263-286.

Stable URL:
http://links.jstor.org/sici?sici=0066-4162%281992%2923%3C263%3ATNNTOM%3E2.0.CO%3B2-3

Annual Review of Ecology and Systematics is currently published by Annual Reviews.

Annu. Rev. Ecol. Syst. 1992. 23:263–86

THE NEARLY NEUTRAL THEORY OF MOLECULAR EVOLUTION

Tomoko Ohta
National Institute of Genetics, Mishima 411, Japan

KEY WORDS: molecular evolution and polymorphism, random drift, weak selection

INTRODUCTION

For a long time the study of evolution has been based on morphology; the long neck of a giraffe, the human brain, a bird's wing, and so on. Morphological change in evolution is explained by Darwin's theory of natural selection, but this theory is largely qualitative rather than quantitative. Population genetics started more than half a century ago as an attempt to understand evolutionary change quantitatively. Because evolution must take place in all individuals of a species, the change of gene frequency in the population has been analyzed. However, so long as the facts of evolution are based on morphological traits, evolutionary change is very difficult to connect with gene frequency change except in relatively few circumstances.

The remarkable progress of molecular biology has made it possible to apply population genetics theory to real data. We now know that genetic information is stored in linear sequences of DNA which are stably transmitted from generation to generation, and we can compare the linear sequences of DNA and amino acids among species. It is also possible to compare secondary and tertiary structures of proteins and nucleic acids from various sources.

Because of such progress, some aspects of traditional neo-Darwinism are beginning to need revision. The first step in such a revision is the neutral mutation-random drift hypothesis put forward by Kimura (47) in 1968. In the next year, King & Jukes (53a) published a similar idea, though from a more biochemical point of view than that of Kimura. This theory states that most evolutionary changes at the molecular level are caused by random genetic drift of selectively neutral or nearly neutral mutations rather than by natural selection. Because this theory was contrary to the neo-Darwinian view at that time, it provoked much controversy. A complete review of the theory is found

0066-4162/92/1120-0263$02.00

in Kimura's book (49), so I shall give only a brief outline here, together with the basic concepts of the stochastic theory of population genetics.

The second step in the revision is to clarify the interaction of natural selection and random drift at the molecular level. Natural selection cannot be so simple as to be "all or nothing." There are numerous types of mutations, whose behavior is influenced by both selection and random drift. In this article, theoretical studies of such "nearly neutral" mutations are reviewed, together with some recent findings on DNA sequence variation. In addition, some possible ways of connecting molecular and phenotypic evolution are discussed with special reference to the versatile nature of DNA, also clarified in recent years.

Population Genetics

The study of population genetics was started by a trio of pioneers, Fisher (23), Haldane (35), and Wright (112), who formulated the basic framework. Gene frequency changes in the population, caused by Darwinian natural selection, are essential in this theory. Through successive investigations, various detailed theories on the effects not only of selection but also of mutation, migration, and random sampling drift at reproduction have been worked out.

The results have been used for establishing the modern synthetic theory of evolution (20). In addition, human population studies (12) as well as plant and animal breeding (22) depend heavily on population genetics theory.

In the period from 1966 to 1980, many studies on genetic variation at protein producing loci within species were published. Lewontin & Hubby (60) and Harris (36) were among the first to use electrophoresis for studying protein variation from the standpoint of population genetics. It turned out that biochemical variation is rather abundant in human and in *Drosophila* populations. The typical individual is heterozygous at 5–15% of its protein loci. At about the same time, data on sequence comparisons of hemoglobins and cytochrome *c* between mammalian species became available (116). By examining such data, Kimura (47), who was an authority on stochastic processes of population genetics, proposed the neutral mutation–random drift hypothesis of molecular evolution. This hypothesis was based on the stochastic theory of population genetics and has been most stimulating to the further development of both experimental and theoretical population genetics at the molecular level.

Rate of Molecular Evolution

Since the advent of rapid DNA sequencing techniques, data on the primary structure of genes are accumulating amazingly fast, and now statistical studies of DNA sequences are quite popular. However, only 20 or so years ago, most

of the available data on molecular evolution were in the form of amino acid sequences (59, 4). Hemoglobin α of mammals consists of 141 amino acids, and it is one of the best-studied molecules. If one compares human hemoglobin α with that of the gorilla, all amino acids are identical except one, but 18 amino acids differ between human and horse. Such data on sequence divergence faithfully reflect the phylogenetic relationship. Since we know the approximate time of divergence of mammalian species, it is possible to estimate the rate of amino acid substitution. From a comparison of various species the rate of amino acid substitution of hemoglobins α and β is about 10^{-9} per amino acid site per year. It is quite impressive to find that almost the same value is obtained from any two species. Especially noteworthy is the fact that the rate seems to be almost the same in the line to living fossils whose morphological characters have hardly changed for tens of millions of years, and in the line leading to humans.

The apparent uniformity of evolutionary rates, as compared with phenotypic evolution, is a most remarkable characteristic of protein evolution, (116). By applying similar analysis to cytochrome *c* data, the rate seems again to be almost uniform for diverse organisms including plants, fungi, and mammals. However, the rate is much lower in cytochrome *c* than in hemoglobins. This is thought to be caused by stronger structural constraints on cytochrome *c* than on hemoglobins. This is another characteristic of molecular evolution, i.e. the stronger the constraint on the molecule, the lower is its rate of evolution. For some examples of the rates of protein evolution, see (49). It is now well known that fibrinopeptides have evolved rapidly with little constraint whereas histone IV has been evolving extremely slowly.

We attempted to estimate the variance of the evolutionary rate of a particular protein, in order to test the uniformity of the rate, and we found that the variance of amino acid substitution is larger than that expected if a simple Poisson process is assumed (96). This analysis may be flawed because it is based on paleontological estimates of divergence time. But the conclusion has been confirmed by Langley & Fitch (58), who used more data and statistics less dependent on paleontological estimates. Kimura (49) used data from mammalian orders that are believed to form a star phylogeny, i.e. simultaneous divergence of many mammals about 80 million years ago (198a). However, see Ref. 21a on some problems of this approach. For hemoglobins, myoglobin, cytochrome *c* and ribonuclease, the ratio, *R*, of the variance to the mean divergence turned out to be from 1.3 to 3.3, again confirming the previous result. Gillespie (28) extended such analyses for more proteins; hemoglobins, cytochrome *c*, insulin, prolactin, ribonuclease, LHP, albumin, cytochrome oxidase, ATPase 6 and cytochrome *b*, and he found that R takes values between 0.2 and 34.1. The extreme values are thought to be due to small sample size (three species with star phylogenies were examined).

Gillespie (28) argues that this pattern fits the "episodic" process in which a burst of amino acid substitutions is followed by a static phase.

Since methods of DNA sequencing have become available, more and more data on DNA sequences are accumulating and molecular evolutionary studies have shifted from analysis of amino acid sequences to that of DNA sequences. Several remarkable features of DNA evolution have emerged. The majority of genomic DNA of higher organisms evolves more rapidly than protein coding regions, i.e. those DNA regions that apparently do not carry genetic information in their primary structure are evolving rapidly. In mammals, the rate of nucleotide substitution in these regions is roughly 5×10^{-9} per site per year (e.g. see 49, 76, 63). The rate of synonymous substitutions in coding regions is slightly lower than this but is rather uniform among various genes, whereas the rate of amino acid replacement substitutions differs greatly from gene to gene. These values agree with the results of DNA hybridization studies (10, 104).

An as yet unsettled problem is whether the rate of DNA evolution depends upon generation length. An examination of amino acid sequences revealed little effect of generation time (55, 57). DNA hybridization studies, however, indicate that the longer the generation, the lower the DNA evolution rate (57, 55, 10, 104). This is still a controversial problem and is discussed later.

Another topic of interest is the isochore concept, the differentiation of chromosomal regions of warm blooded vertebrates into GC- and AT-rich segments (6, 41). Such segments are called "isochores" and have an average size well above 200 kilobases (6). A noteworthy fact is that most housekeeping genes locate in GC-rich isochores, whereas many tissue-specific genes are in AT-rich ones. In addition, codon usage is different in these two groups of genes (41). Thus, a codon bias in mammals appears to be largely determined by mutation pressure, but the bias in lower organisms with intermediate GC content such as in *E. coli* is influenced by selection because of the efficient selection in very large populations (102). Such selection must be nearly neutral, i.e. at the border between neutrality and selection.

Population Dynamics of Mutant Genes

A basic requirement for understanding the mechanisms of nucleotide or amino acid substitutions in evolution is to distinguish mutations from evolutionary substitutions. Numerous mutations appear in Mendelian populations in every generation, but the majority will be lost within a few generations. Thus, those mutations that contribute to evolution are a very small minority of all mutations. It is also necessary to understand the process of frequency increase of mutants in the population in the course of their substitution.

For neutral mutants, the process has been theoretically analyzed using the

diffusion equation method (51, 53). A neutral mutant, if it is ultimately fixed in the population, takes on the average 4N generations until this occurs.

$$t_1 = 4N, \qquad 1.$$

where N is the effective population size. If N is large, the time is very long. The rate of molecular evolution is measured by averaging the number of substitutions over very long period of time. It may be expressed as follows,

$$k = \lim_{T\to\infty} \frac{n(T)}{T} \qquad 2.$$

where T is the period, and n(T) is the number of mutant substitutions in this period. Obviously, for k to be measured accurately, $T>>N$.

Now consider a locus encoding a protein. Let the rate of occurrence of base substitutions in this DNA region be v_g per generation, and let u be the probability of fixation of a mutant. Then in a population of N individuals, the total number of mutations appearing in the population is $2Nv_g$ per generation, and a fraction u of them spread through the population, so the rate of substitution per generation becomes,

$$k_g = 2Nv_g u. \qquad 3.$$

Here u generally depends upon the magnitude of natural selection. It should be remembered that, at the molecular level, the number of nucleotide sites of a locus is so large that the probability of having identical mutations more than once is almost nil. Also, the probability of back mutation is negligibly small.

Let us now examine how natural selection influences the rate. If most substitutions are caused by Darwinian natural selection, and the average selective advantage of such substitutions is s with no dominance, the fixation probability is roughly twice the selective advantage (23, 35), and we have,

$$k_g = 4Nsv_g. \qquad 4.$$

Hence k_g depends on the product of three parameters, N, s and v_g. But when most substitutions are selectively neutral, the fixation probability is equal to the initial frequency, $1/2\ N$ (113, 46), and we have,

$$k_g = v_g. \qquad 5.$$

In other words, the evolutionary rate is simply equal to the mutation rate, and it is independent of population size (47).

The actual rate of molecular evolution seems to be roughly constant per year for each protein as reviewed in the previous section. In order to explain this fact by a selection model (Eq. 4), one has to assume that parameters like N and s have been nearly equal in various lineages. Such a situation could hardly hold in very different environments, particularly in the lines leading to

living fossils and in those with rapid phenotypic evolution. On the other hand, if the majority of substitutions are selectively neutral as was first proposed by Kimura (47), from formula (5), the observed pattern may be explained if the neutral mutation rate is constant per year. Here the possible problems are: (i) how the generation length affects the evolution rate, (ii) how fluctuation in the evolutionary rate is related to natural selection, and (iii) how selective constraints influence the rate. These problems are discussed later in relation to weak natural selection, i.e. the interaction of random genetic drift and selection.

Selective Constraints and Evolutionary Rate

The rate of evolution is different from protein to protein, and the difference reflects the degree of constraint as explained before. This constraint is directly connected to the function and structure of protein or RNA molecules, i.e. the more rigid their function and structure are, the lower is the rate. This point has been beautifully shown for fibrinopeptides, hemoglobin, cytochrome *c* and histone IV by Dickerson (19).

Such evolutionary features may be explained by the neutral mutation random drift hypothesis as follows. As in Kimura (49), v_T designates the rate of occurrence of new mutations in terms of nucleotide substitutions. Let f_0 be the fraction of such mutations that are selectively neutral, and the remaining $(1 - f_0)$ be the fraction of mutations that have a deleterious effect. Then the rate of neutral evolution becomes,

$$k = f_0 v_T, \qquad 6.$$

because deleterious mutations do not contribute to evolution. Favorable mutations are assumed to be too rare to have any statistical influence. Various degrees of selective constraint may be taken into account by f_0, e.g. for pseudogenes without constraint, $f_0 = 1$, whereas for amino acid replacement substitutions of histone IV (see Table 1), $f_0 = 0$. Here the question is whether or not mutations may be simply divided into neutral and deleterious classes. This problem is discussed later.

Polymorphisms at the Molecular Level

Elucidating the mechanism for maintaining genetic variability has been one of the most important problems of population genetics. Particularly at the molecular level, stimulated by the neutral mutation theory, numerous attempts were made around the 1970s to clarify the roles of selection and drift (59). According to the neutral theory, molecular polymorphisms are a phase of evolution (52), whereas to selectionists, they are actively maintained by balancing selection and independent of mutant substitutions.

The pioneering work on protein polymorphisms was published by Lewontin

& Hubby (60), who found that heterozygosity at enzyme loci measured by electrophoresis is about 12%, and that about one third of the loci are polymorphic in *D. melanogaster*. At about the same time, Harris (36) reported that heterozygosity by electrophoresis is about 6% in human. Since then, many investigators examined the level of protein polymorphisms of many species (5, 8, 100, 77).

If protein polymorphisms are mostly neutral, they represent an intermediate stage of neutral gene substitution. Let us ask how much genetic variation is expected in a population when an equilibrium is reached between mutational input and random extinction in this process. A most convenient measure of genetic variation at a locus is a quantity called "virtual heterozygosity" (49), which is the probability that two randomly chosen alleles from the population will be different. In a randomly mating diploid population, heterozygosity and virtual heterozygosity are the same. The virtual heterozygosity at equilibrium has been shown to be,

$$\hat{H} = \frac{4Nv}{1 + 4Nv}, \qquad 7.$$

where v is the mutation rate (50, 69). This is remarkably simple in that only one product, Nv, comes into the formula. Nevertheless the formula is robust and applicable to many cases. As was pointed out before, the observed values of heterozygosity at enzyme loci are often around 12% in *Drosophila* species, and 5–6% in mammalian species. Then, Nv is predicted to be roughly 0.035 for *Drosophila,* and 0.015 in mammals, provided that most polymorphisms are selectively neutral and that the species are in equilibrium between random drift and mutation.

Of course, various species have greater or less heterozygosity. So far, *E. coli* is the species with highest value—47% (101). In several large animal species such as cheetah (78), polar bear (1), and elephant seal (8), the heterozygosity is practically nil. These species are thought to have had small population size in the past so that Nv is extremely small (49). However, if the neutral theory is correct for existing enzyme polymorphisms, the problem is still not settled. This is discussed further in the next section.

Recently, data on DNA sequence polymorphisms are becoming available. As expected from the neutral theory, polymorphisms in noncoding regions or at synonymous sites are much higher than at amino acid–replacing sites. The percentage of nucleotide differences among sampled alleles at the Adh locus in *D. melanogaster* was 0.13–1.22%, with an average value of 0.65% (56). Since an ordinary protein coding region consists of several hundred nucleotide sites, a heterozygosity of 0.65% per site is much higher than the 12% at the enzyme loci, even if the latter is measured by electrophoresis, which detects only a part of the total nucleotide change. In a way similar to the difference

of evolutionary rates between protein and noncoding DNA regions, polymorphisms are different between the two regions, i.e. the stronger the constraints, the fewer polymorphisms are observed.

A notable finding on DNA polymorphism is the contrasting patterns at the rosy region in *D. melanogaster* and *D. simulans* reported by Aquadro et al (3). These authors found that, unlike protein polymorphisms, DNA sequence variation in this region of *D. simulans* is estimated to be several times as great as that of *D. melanogaster*. Note that the average heterozygosity over many protein loci is almost the same in the two species but that geographic differentiation is more pronounced in *D. melanogaster* than in *D. simulans* (106). Aquadro et al (3) suggested that the effective population size of *D. simulans* is larger than that of *D. melanogaster*, and that the differences in species effective population size may be responsible for the pattern, i.e. weak selection would be effective on protein variation, but not on DNA variation. This fact has an important bearing on the nearly neutral theory and is discussed in the next section.

Meaning of Near Neutrality

In the previous sections, I have repeatedly emphasized that the rate of molecular evolution is strongly dependent upon selective constraints of proteins or nucleic acids. Under the neutral mutation–random drift theory, it is assumed that a certain fraction of new mutations are free of constraint or are selectively neutral, while the rest have deleterious effects and are selectively eliminated. An important question is how the two classes are distinguished by natural selection. Let us now examine theoretically the behavior of the mutants belonging to the borderline class. The critical quantity is the fixation probability of mutant genes. I examine the simplest case of a semidominant gene with selective advantage, s. Fixation probability in finite populations has been shown to become the function of the product, Ns, as follows (46, 69),

$$u = \frac{1 - e^{-4Nsp}}{1 - e^{-4Ns}}, \qquad 8.$$

where p is the initial frequency in the population, and is assumed to be much less than unity. Figure 1 shows the fixation probability of mutants as a function of Ns relative to the completely neutral ones ($Ns = 0$). As can be seen from the figure, the fixation probability is a continuous monotone function of Ns. Thus, when discussing molecular evolution by nearly neutral mutants, one has to consider all mutants around $Ns = 0$.

Let us examine the borderline mutations in some detail. Figure 2 shows classification of new mutations. The upper part shows the simple neutral model, and the lower part, the nearly neutral model. In the figure, neutral

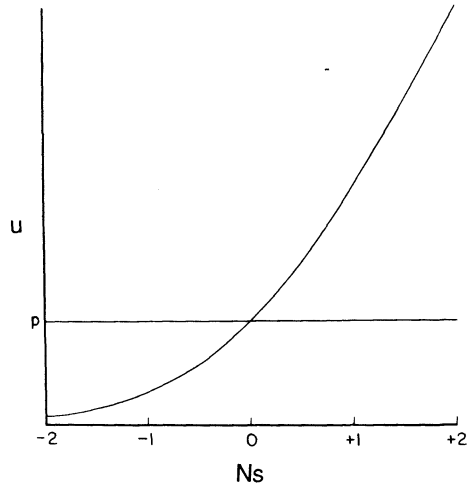

Figure 1 The fixation probability u of a mutant as a function of the product of population size and selection coefficient, Ns, relative to the completely neutral case. p is the initial frequency and the line, $u = p$, is the value for a neutral mutant.

means strictly neutral, deleterious means definitely deleterious, and nearly neutral means intermediate between neutral and selected. The behavior of mutations of the nearly neutral class is affected by both selection and drift.

As shown in Figure 1, the effectiveness of selection is determined by the product of the effective population size and the selection coefficient, Ns. Actual species have various population sizes from very small to very large, and therefore the effectiveness of selection will differ among species. In addition, physiological conditions may influence weak selection, e.g. a constraint on an enzyme function may differ between homeotherms and poikilotherms. The simple separation of new mutations into the deleterious and neutral classes will then not be satisfactory. There should be a substantial fraction of mutations that belong to the nearly neutral class.

Especially in view of the importance of negative selection caused by constraints, it is likely that many nearly neutral mutants are very slightly

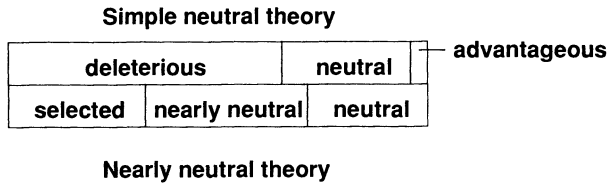

Figure 2 A schematic diagram shows the proportion of various classes of mutations. Deleterious mutants are definitely deleterious and neutral mutants are strictly neutral, while most selected mutants are deleterious, but the group also includes advantageous alleles. Nearly neutral mutants comprise a class intermediate between neutral and selected mutants.

deleterious, i.e. on the left side of $Ns = 0$ in Figure 1. My hypothesis of slightly deleterious mutants is based on this view and regards the near neutrality as the limit when the selection coefficient approaches zero (80–84). In other words, I propose that a substantial fraction of mutant substitutions at the molecular level are caused by random fixation of very slightly deleterious mutations.

This hypothesis seems realistic when one realizes that the effect of an amino acid substitution in a protein often produces only a minor modification of a reaction coefficient (45). It is also in accord with a recent finding that many enzyme variants in *E. coli* are likely to be mildly deleterious based on statistical analyses of their frequency distribution (99). Molecular variants that disturb very slightly the secondary structure of molecules, e.g. by opening the stem region of a clover leaf structure of tRNA, may represent a mutant class with mild deleterious effects (82, 83). In this case, it is likely that a slightly deleterious base substitution is followed by a slightly advantageous (compensatory) substitution. This is supported by the observation that G.U or U.G pairs that represent intermediate steps account for about two thirds of all non-Watson-Crick pairs in the helical regions of analyzed tRNAs (38).

The model of very slightly deleterious mutation is related to the molecular clock in an important way. One problem with the molecular clock is that the observed substitution rate is constant per year rather than per generation. The generation-time effect is particularly evident when divergence is measured by DNA hybridization (57, 55, 10, 105). According to these studies, the rate of divergence of single-copy genomic DNA varies from 10^{-9} to 10^{-8} per base pair per year depending upon the taxon, and there appears to be a high negative correlation between the rate and generation length. Similar effects have been found when synonymous and other unimportant DNA divergences are examined (13, 34, 37, 66, 71, 109, 114). Thus, it seems that the rate of evolution of the majority of genomic DNA depends upon generation length, whereas the divergence rate of amino acid sequences is relatively insensitive to it.

In the early 1970s, I tried to explain this seemingly contradictory observation by using population genetics theory. The fundamental idea is that most genomic DNA of higher organisms can freely accumulate base substitutions (80), i.e. most new mutations are selectively neutral. On the other hand, amino acid substitutions are more likely to be influenced by natural selection, i.e. many of them may be regarded as nearly neutral, or very slightly deleterious. Let us now examine how slightly deleterious mutations are related to the generation-time effect. The left side of $Ns = 0$ in Figure 1 shows that there is a negative correlation between the fixation probability and population size, provided that the selection coefficient is unchanged by population size. In other words, the chance of spreading by random drift is much higher in a

small population than in a large population. To be more quantitative, several models of nearly neutral but slightly deleterious mutations have been studied by assuming some distribution of the selection coefficients for new mutations. I assumed an exponential function (85), Kimura (48), a gamma function, and Li (61), several classes of selection coefficients of new mutations. Again, all these models predict that the rate of mutant substitution is negatively correlated with the species population size, although the magnitude of the correlation depends on the model.

This prediction has an important bearing on the generation-time effect. In general, large organisms have a long generation length and small population size and vice versa. Thus there is a negative correlation between population size and generation length. On the other hand, the mutation rate per year will be lower in organisms with longer generation times as it is ordinarily believed (110, p. 254), and DNA divergence reflects such an effect. With selection this effect is expected to be partially cancelled by the negative correlation between the substitution rate and population size. Thus the amino acid substitution rate is expected to be relatively insensitive to generation length as compared to DNA divergence because amino acids are more likely to be subject to negative selection than are noncoding regions.

Actually, the small selection coefficients are not likely to remain unchanged in the long course of their substitutions (e.g. 25, 75). When varying selection intensity is examined, one notices that there are two quite different approaches. One is the so-called variable selection model in which selection coefficients fluctuate randomly around zero (e.g. see 26, 27, 30, 108). Here the selection coefficient varies from generation to generation or with some correlation between near generations, i.e. the variation is short term.

Another model considers long-term variation of the selection coefficient (81, 90, 70). In this model, the coefficients remain stable for a certain period of time or within a certain local area, but may vary in another period or area through a change of environmental conditions, which include both the internal genetic background and external ecological factors. Here the coefficient varies on a longer time scale than in the previous model. This model is difficult to formulate mathematically, and no precise analysis has yet been worked out.

We have performed extensive Monte Carlo simulations (97, 107). In the simulations, the selection coefficient of a new mutant in a local colony is assumed to follow the normal distribution with mean μ_s and variance σ_s^2. It is assumed that there are l habitats in the total population, and that each colony occupies a habitat. Let $s_{k,i}$ be the selection coefficient for the k-th mutant in the i-th colony and assume they are independent, so that the selection intensity fluctuates independently over space. This model is different from the landscape model of Gillespie (27) because we are concerned with the interaction effect of random drift and natural selection, whereas Gillespie treated the situation

of strong selection and weak mutation. Let N be the effective population size of a local colony, and $N_T = lN$ be the total size. The most important parameter is twice the product of the population size and the standard deviation of the selection coefficient, $2N\sigma_s$, because σ_s is the measure of selection intensity in this model.

It should be noted that the present model is different from those of Ohta (85), Kimura (48) and others, because here the distribution of selection coefficients is fixed irrespective of mutant substitutions and is the same as Kingman's (54) house-of-cards model. Let us call the present one the fixed model, and the previous one the shift model. Note that in the shift model, the population mean is reset to zero after each substitution. What do these models imply about the real process? Consider the structure of a protein. In the shift model, an amino acid substitution has no effect on the other amino acid substitutions in the protein. Therefore each substitution is independent. On the other hand, in the fixed model, the effect of each substitution remains and affects subsequent substitutions by changing the mean fitness. Therefore substitutions are interrelated in their effects on fitness. In terms of the higher order structure of a protein, substitutions that occur at many amino acid sites do not behave independently, and the fixed model I believe is a better description of nature. The shift model also implies, unrealistically, that proteins can improve or deteriorate indefinitely by successive substitutions, whereas this process may stop under the fixed model. In a study of hemoglobin genes, Goodman (32) suggested that evolution had accelerated after gene duplication but had slowed down in later phylogeny, and he argued that this reflects improvement of gene functions. The fixed model is consistent with this scenario.

Simulations were performed by choosing realistic values for the products of $2N$ and other parameters. The mean fitness of the population fluctuates with time because of random drift. Figure 3 shows the distribution of the selection coefficient of new mutations around the population mean. If selection is strong enough, the mean moves towards the right without fluctuation. For nearly neutral mutations, it moves erratically but tends to increase. When the mean is positive, the average selection coefficient of new mutations becomes negative, i.e. new mutations are slightly deleterious on the average.

The results of our simulations show the properties of the model in detail. A panmictic population is easiest to treat, so consider the case of a single colony. Tachida (107) examined various quantities of biological interest. When $4N\sigma_s$ is larger than 3, several advantageous mutants are quickly fixed, and thereafter almost all new mutations are deleterious. Then the fitness of the population changes very slowly, and substitution almost stops. However, the system takes a very long time to reach true equilibrium. When $4N\sigma_s$ is in

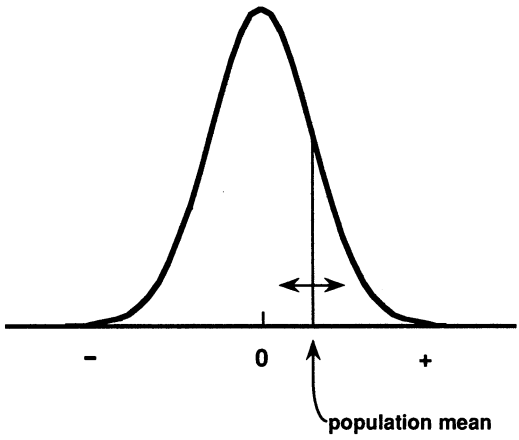

Figure 3 Distribution of the selection coefficient of new mutations. The population mean moves to the left or to the right by selection and drift.

the range 0.2 to 3, both random drift and selection affect the population fitness, which very slowly increases and reaches an equilibrium value in a period of the order of number of generations of the reciprocal of the mutation rate. Even after reaching equilibrium, substitution does not stop but is less frequent than in the completely neutral case. When $4N\sigma_s$ is less than 0.2, the mutants' behavior is almost the same as a completely neutral one. Tachida (107) called these three classes of mutations, selected ($4N\sigma_s > 3$), nearly neutral ($3 \geq 4N\sigma_s \leq 0.2$) and effectively neutral ($4N\sigma_s < 0.2$).

When the population is subdivided, and the selection coefficient fluctuates spatially, the situation is not so simple. Firstly, the selection intensity of the total population becomes the average of selection coefficients of the local colonies. If there are l habitats (colonies) and the selection coefficient of a colony is independently distributed as we assumed, the selection coefficient of the total population is normally distributed with the same mean, μ_s but with a smaller variance, σ_s^2/l, as expected. If the migration rate is high enough, the mutants' behavior becomes similar to that in panmictic populations with the variance of selection coefficient, σ_s^2/l.

Let us now examine the results for the cases where migration is less. Wright's island model was used, incorporating local extinction and recolonization, in the belief that natural populations often pass through bottlenecks, so that current populations are descended from a small number of founders (11). A most interesting question is: How does the total population size affect evolution by nearly neutral mutations? Recall that one problem of the nearly neutral theory is its prediction of a reduction of the evolutionary rate when population size becomes large. As mentioned before, this is related to the

constancy of the evolutionary rate, particularly the effect of generation length which is thought to be partially cancelled by the population-size effect. In general, the nearly neutral model predicts that the stronger the selection is, the larger the population-size effect becomes. Under the shift model with constant selection coefficients, the transition from the neutral to the selected class is quite steep as the population size increases (see Figure 1). Under the fixed model, the transition becomes more gentle, as Figure 4 shows. The transition from the neutral to the selected class of mutations is quite gentle especially when the migration rate is low.

The above study is a quantitative evaluation of my intuitive argument on fluctuating weak selection (81). The idea is that, in a stable environment, a random mutant needs to be beneficial only under restricted circumstances, whereas in a more variable environment, a mutant must be beneficial in many circumstances. Usually, the smaller the population, the more restricted the environmental variability. This idea is somewhat similar to Mani's (70) model, but the selection coefficient is assumed to be large in his study. In any case, the nearly neutral model now incorporates very slightly advantageous mutations for the region $s > 0$, and the concept of near neutrality approaches that of selectionists (e.g. 117).

What kind of evidence do we have to support the nearly neutral theory? Recent studies on DNA polymorphism in natural populations of *Drosophila*

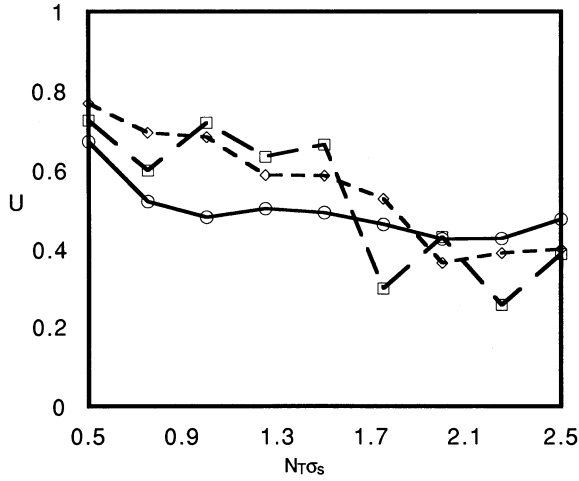

Figure 4 Results of simulations on fixation probability of mutants as functions of the product of total population size and selection intensity, $N_T\sigma_s$, relative to the completely neutral case. Each point is obtained by a simulation continued for $800/v$ generations, where v is the mutation rate. Three lines are for different levels of migration rate (m), (solid line) : $2Nm = 0.1$;(bold broken line) : $2Nm = 0.5$; (broken line): $2Nm = 4.0$.

reveal several facts that are better explained by the nearly neutral theory than by the strictly neutral one. As mentioned before, the pattern of DNA polymorphism at the rosy region in *Drosophila melanogaster* and *D. simulans* is quite different, i.e. unlike protein polymorphisms, DNA sequence heterozygosity in this region of *D. simulans* is estimated to be several times as great as that of *D. melanogaster* (3). Note that the average heterozygosity over many protein loci is almost the same in the two species but that geographic differentiation is more pronounced in *D. melanogaster* than in *D. simulans* (15). Aquadro et al (3) suggested that the differences in the effective population sizes of species may be responsible for the pattern, such that slightly deleterious mutations are common in proteins but not in DNA. Aquadro (2) studied more loci and found a similar pattern in all three loci examined; *per* in the X chromosome, Adh in the second chromosome, in addition to *rosy* in the third chromosome. By simulation studies, I have shown that the data fit the present model not only for heterozygosity but also for a proportion of polymorphism and the fixation index (95).

Another important subject related to the nearly neutral model is the rate of molecular evolution. For nearly neutral mutations, the effect of total population size on the substitution rate becomes larger as the mutant's effect gets stronger. Then the cancellation between generation length effect and negative selection (see above) would be more pronounced for amino acid substitutions than for synonymous changes. The data of Li et al (66) suggest that the generation-time effect is stronger on synonymous substitutions than on amino acid replacement substitutions, agreeing with the above prediction. By using the data of Li et al (66), Gillespie (31) obtained weighting factors. The factors were obtained such that the average is one, and the relative values remain the same. These are the ratios of the numbers of substitutions of the three branches for each of the mammalian orders, artiodactyls, rodents, and primates. Let w_a, w_r, and w_p be the weights for the above three orders respectively. For replacement substitutions, they turned out to be, $w_a = 0.885$, $w_r = 1.279$, $w_p = 0.836$. For synonymous substitutions, $w_a = 0.762$, $w_r = 1.611$, $w_p = 0.627$. Thus, amino acid replacement substitutions are less dependent on generation length. As I have repeatedly emphasized, the generation time effect is coupled with the population size effect. An interesting example is the differentiation of mitochondrial genome, among Hawaiian *Drosophila*(18). Mitochondrial evolution in two closely related lineages of Hawaiian *Drosophila* that have different histories shows that the rate is three times higher in lineages with repeated founder events than in lineages without bottlenecks.

Let us now turn our attention to the variance of the evolutionary rate. As discussed before, the variance is often larger than that under the simple Poisson process. Possible causes that inflate the variance are: (i) a difference of mutation rate among the lineages, (ii) weak selection that results in a

difference of intensity due to the change of population size, and environment, and (iii) other factors. The total variance would be determined by the sum of the several causes. The generation-time effect on mutation rate is partially cancelled by the population size effect, but the cancellation is measurable only when a large number of loci are examined, as the above weight analysis shows, and the variance is inflated by the difference of mutation rate. As to the second cause, the larger the mutants' effect becomes, the more inflation of the variance is expected. The comparison of the variance of the amino acid replacement substitution with that of synonymous substitution does not show clearly that the former variance is larger than the latter (28). One needs more data and analysis for clarification.

The final problem on near-neutrality concerns the assumption of fluctuating selection coefficients. Results from Dykhuizen & Hartl (21) suggest that many naturally occurring enzyme polymorphisms in *E. coli* are neutral or nearly neutral, but a latent potential for selection can be observed in the polymorphism at the 6-phosphogluconate dehydrogenase locus of *E. coli*. Sawyer et al (99) estimated that the average selection coefficient for that locus is approximately -1.6×10^{-7}. This estimate corresponds to our μ_s, and local values of the selection coefficient may be larger. Dean et al (17) examined the fitness effect of newly occurring amino acid substitutions at the β-galactosidase locus of *E. coli*, and again found that the majority of amino acid changes have minor effects on fitness. This would imply that the "Dykhuizen-Hartl" effect applies not only to naturally occurring polymorphic alleles but also to many newly arisen mutant alleles.

The model considered here incorporates spatial fluctuation of selection intensity but not a temporal fluctuation. It is therefore not quite realistic. However, for such large-scale fluctuation as is considered here, temporal changes of fitness would not usually occur while a mutant is on its way to fixation. The effect of temporal variation would be on changing the mean fitness of the population such that the value of μ_s is simply lowered on rare occasions. Then the results of the simulation studies would hold in a more general situation that incorporates temporal fluctuation of selection intensity.

Adaptation at the Molecular Level

So far we are mainly concerned with selection for keeping the gene function status quo. Once gene function attains a state sufficiently near an optimum, genes are expected to evolve by nearly neutral mutant substitutions. As long as fixed loci are considered, gene functions are usually kept as they are, and on rare occasions an environmental change causes a shift of gene function with the appropriate mutant substitutions. It may occur in protein coding regions as well as in regulatory regions such as a promoter or an enhancer. Such substitutions may constitute a major process of adaptation in bacteria.

Adaptation at the molecular level of higher organisms like mammals appears to include minor chromosomal changes such as duplication and illegitimate crossing-over. These changes would be acceptable under a genome structure with a large noninformational part. This is because noninformational part would provide flexibility on gene organization. The *Drosophila* genome seems to be the intermediate between mammalian and bacterial ones. It is interesting to find that the detailed analyses of DNA polymorphisms in *Drosophila* species reveal many different patterns among loci suggesting various ways of adaptation (2, 56). In many cases, the interaction of random drift and selection would be important.

In vertebrates, it is now known that multigene families are common. Organismal development is governed by spatially and temporally regulated expression of various gene families that are the products of hundreds of millions of years of evolution. Thus, the reorganization of genes by duplication or illegitimate recombination is very important for organismal evolution. Such a process is occurring much more frequently than was imagined by leading geneticists, Bridges (9), Muller (72) and Ohno (79), who thought of gene duplication as the major way to acquire new genes. Indeed, comparative studies of gene families show that the genetic material is more versatile than was previously thought, i.e. various illegitimate recombinational processes must have been rather common in evolution. In general, it has been thought that natural selection works to keep genes in status quo for those gene families that were established a long time ago. On the other hand, incipient gene families may be on the way to further progress, in the sense of acquiring more diverse function while positive Darwinian selection may be operating.

Population genetic models of established multigene families have been extensively analyzed by incorporating gene conversion (87, 88, 73, 74), unequal crossing-over (86), and duplicative transposition (14, 89). The results can satisfactorily explain the so-called concerted evolution of multigene families. In other words, each member of a gene family does not differentiate independently but evolves in concert with other members because of their functional interrelationship and because of the homogenizing effect of the above processes. The study of copy number regulation has just started (67). The origin of gene families with diverse function has also been studied by population genetics (91, 92). Starting from a single copy, my simulations have shown how beneficial genes may accumulate on the chromosomes under various conditions. With realistic values of deleterious mutation rate, positive selection is needed for acquiring gene families with desirable functions.

Is it possible to find incipient gene families in the process of acquiring new functions? There are now several examples of duplicated genes that show accelerated amino acid substitution relative to synonymous substitution as surveyed in (62, 93). The first example is the emergence of fetal hemoglobin

from embryonic hemoglobin in primates (33, 24). Higher primates have two duplicated hemoglobin γ genes, and other primates have only one. The γ genes of higher primates are turned on in fetal life, whereas the single gene of other primates is turned on in embryonic life and turned off at the beginning of fetal life. Examination of the DNA sequences of the γ gene family suggests that the amino acid substitution was accelerated in the duplicated genes in the period when the switch occurred from the embryonic gene to the fetal one phylogenetically (33, 24). The substitution rate is about 3.5 times the standard rate of hemoglobin genes. The next example is the β globin gene cluster of goat and sheep (64). However, in this case, it is difficult to judge whether the acceleration was caused by positive selection or simply by relaxation of selective constraint.

The evolution of the stomach lysozyme of ruminants is another interesting example. An amino acid sequence study has shown that the gene for stomach lysozyme arose by duplication of the gene for nonstomach lysozyme at the time when ruminants diverged from other mammals (42–44). It has been shown that, about 30 million years after divergence of ruminants from other mammals, the rate of amino acid substitution was three times as high as that of ordinary nonstomach lysozyme (42–44). Other examples surveyed in (93) include genes of the visual pigment (115), histocompatibility antigen (39, 40), immunoglobulin constant region (103), and protease inhibitor (16). For genes of the immune function, there seems to be no clear distinction between incipient and established gene families (94), and some multigene families belonging to the immunoglobulin superfamily are undergoing continuous reorganization via unequal crossing-over and gene conversion (98). Such examples reveal remarkable strategies to acquire enormous diversity of immune reaction, and positive selection must have operated for their origin, even if selective force may have been very weak at the level of an individual amino acid site (86).

Prospect

As explained in the previous section, recent studies of genetic variation of *Drosophila* species at the DNA level have revealed a number of unexpected properties such as the contrasting pattern of polymorphisms at the DNA and at the protein levels among the closely related species (2). The pattern now extends to the human population, as the nucleotide diversity is much less in human than in *Drosophila* populations in spite of a similar level of protein diversity (65). Also, it appears that each locus has its own characteristic pattern of polymorphism. In some loci, protein diversity appears to be very low as compared with DNA diversity within species, but not so in others (2, 56, 68). Thus the relative importance of drift and selection may differ from locus to locus, and such data provide excellent material for further study on

the mechanisms of maintaining polymorphism. For example, the population-size effect of the nearly neutral model can be used to discriminate it from the simple neutral model. Here the relative numbers of the amino acid altering and the synonymous or other unimportant substitutions are most convenient—a point discussed in the previous section. It may be difficult to discriminate the nearly neutral model from the selection model, in which most amino acid substitutions are caused by positive Darwinian selection (29, 30, 68). However, I would like to point out that the classical model of selection predicts that evolution is more rapid in large populations than in small ones (23; see page 102). Thus species with large population size accumulate more mutant substitutions under the selection model. This prediction is contrary to that of the nearly neutral model and will be useful for discriminating the two.

As was emphasized before, the genetic material is more versatile than previously thought, and such versatility has been used for organismal evolution. In particular, the complexity of higher organisms has been attained through numerous trials and errors of gene duplication and other illegitimate recombination of DNA. Various simulation studies have attempted to understand how Darwinian selection has been responsible for the origin of complicated genetic systems (91, 92). The results indicate that interaction between random genetic drift and selection is important, i.e. nearly neutral illegitimate recombination and nucleotide substitution are thought to be raw materials for organismal evolution.

My model of simulation studies on the origin of gene families incorporates gene duplication by unequal crossing-over and mutation (91, 92). The evolution of actual gene families seems to be more complicated than such a simple model, i.e. gene conversion occurs continuously in addition to unequal crossing-over and nucleotide substitution. For example, Irwin & Wilson (43) found that two different trees are obtained for the coding region and the 3′ untranslated region of stomach lysozyme genes of ruminant. Gene conversion had been more frequent in the coding region than in the untranslated region, because coding sequences are more conservative. There are other examples that show a similar pattern. Fitch et al (24) examined the emergence of fetal γ gene of the β gene cluster of primates, and again found that two γ genes show concerted evolution via frequent gene conversion. In this case, gene conversion had occurred not only in the coding region but also in the 5′ regulatory sequences, and the authors suggest that concerted evolution plus selection for favorable mutations would have been responsible for the emergence of the fetal gene. Another example is the evolution of the rat kallikrein gene family. Wine et al (111) suggest that gene conversion has played an important role in the evolution of the functional diversity of the duplicated genes after comparing gene sequences in detail. These three examples indicate that the frequent occurrence of gene conversion at the time

of evolution of a new function seems to be the rule rather than the exception. Basten & Ohta (5a) performed simulations to show that gene conversion is indeed effective in accelerating evolution by compensatory advantageous mutations. In the coming years, the combined study of experimental and theoretical analyses of various gene families will be a fascinating research project.

ACKNOWLEDGMENTS

I thank Dr. J. F. Crow, Dr. C. Aquadro, and Dr. C. Basten for going over the manuscript and offering numerous valuable comments. This is contribution No. 1915 from the National Institute of Genetics, Japan.

Literature Cited

1. Allendorf, F. W., Christiansen, F. B., Dobson, T., Eanes, W. F., Frydenberg, O. 1979. Electrophoretic variation in large mammals. I. The polar bear, *Thalarctos maritimus*. *Hereditas* 91: 19–22
2. Aquadro, C. F. 1990. Contrasting levels of DNA sequence variation in *Drosophila* a species revealed by "six-cutter" restriction map surveys. In *Molecular Evolution,* ed. M. Clegg, S. O'Brien, pp. 179–89. New York: Liss
3. Aquadro, C. F., Lado, K. M., Noon, W. A. 1988. The rosy region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA restriction map variation and divergence. *Genetics* 119: 875–88
4. Ayala, F. J., ed. 1976. *Molecular Evolution*. Sunderland, Mass: Sinauer
5. Ayala, F. J., Powell, J. R., Dobzhansky, Th. 1971. Polymorphisms in continental and island populations of *Drosophila willistoni*. *Proc. Nat. Acad. Sci. USA* 68:2480–83

5a. Basten, J. B., Ohta, T. 1992. Simulation study of a multigene family, with special reference to the evolution of compensatory advantageous mutations. *Genetics*. In press

6. Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., et al. 1985. The mosaic genome of warm-blooded vertebrates. *Science* 228:953–58
7. Bonnell, M. L., Selander, R. K. 1974. Elephant seals: genetic variation and near extinction. *Science* 184:908–9
8. Boyer, S. H., Crosby, E. F., Noyes, A. N., Fuller, G. F., Leslie, S. E., et al. 1971. Primate hemoglobins: Some sequences and some proposals concerning the character of evolution and mutation. *Biochem. Genet.* 5:405–48
9. Bridges, C. B. 1935. Salivary chromosome maps. *J. Hered.* 26:60–64
10. Britten, R. J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. *Science* 231:1393–98
11. Carson, H. L. 1976. The population flush and its genetic consequences. In *Population Biology and Evolution,* ed. R. C. Lewontin, pp. 123–37. Syracuse, NY: Syracuse Univ. Press
12. Cavalli-Sforza, L. L., Bodmer, W. F. 1971. *The Genetics of Human Populations*. San Francisco: W. H. Freeman
13. Chang, L-Y. E., Slightom, J. L. 1984. Isolation and nucleotide sequence analysis of the β-type globin pseudogene from human, gorilla and chimpanzee. *J. Mol. Biol.* 180:767–84
14. Charlesworth, B., Charlesworth, D. 1983. The population dynamics of transposable elements. *Genet. Res.* 42: 1–28
15. Choudhary, M., Singh, R. S. 1987. A comprehensive study of genetic variation in natural populations of *Drosophila melanogaster*. III. Variations in genetic structure and their causes between *Drosophila melanogaster* and its sibling species *Drosophila* simulans. *Genetics* 117:697–710
16. Creighton, T. E., Darby, N. J. 1989. Functional evolutionary divergence of

proteolytic enzymes and their inhibitors. *Trends Biochem. Sci.* 14:319–24
17. Dean, A. M., Dykhuizen, D. E., Hartl, D. L. 1988. Fitness effects of amino acid replacements in the β-galactosidase of *Escherichia coli*. *Mol. Biol. Evol.* 5:469–85
18. DeSalle, R.,Templeton, A. R. 1988. Founder effects and the rate of mitochondrial DNA evolution in Hawaiian *Drosophila*. *Evolution* 42:1076–84
19. Dickerson, R. E. 1971. The structure of cytochrome *c* and the rate of molecular evolution. *J. Mol. Evol.* 1:26–45
20. Dobzhansky, Th. 1937. *Genetics and the Origin of Species*. New York: Columbia Univ. Press. 1st ed.
21. Dykhuizen, D. E., Hartl, D. L. 1980. Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. *Genetics* 96:801–17
21a. Easteal, S. 1985. Generation time and the rate of molecular evolution. *Mol. Biol. Evol.* 2:450–53
22. Falconer, D. S. 1960. *Introduction to Quantitative Genetics*. London, New York: Longmans
23. Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. Oxford: Clarendon
24. Fitch, D. H. A., Bailey, W. J., Tagle, D. A., Goodman, M., Sieu, L., Slightom, J. L. 1991. Duplication of the γ-globin gene mediated by repetitive L1 LINE sequences in an early ancestor of simian primates. *Proc. Natl. Acad. Sci. USA* 88:7396–7400
25. Gillespie, J. H. 1974. Polymorphism in patchy environments. *Am. Nat.* 108: 145–51
26. Gillespie, J. H. 1978. A general model to account for enzyme variation in natural populations. V. The SAS-CFF model. *Theor. Popul. Biol.* 14:1–45
27. Gillespie, J. H. 1983. A simple stochastic gene substitution model. *Theor. Popul. Biol.* 23:202–15
28. Gillespie, J. H. 1986. Variability of evolutionary rates of DNA. *Genetics* 113:1077–91
29. Gillespie, J. H. 1986. Natural selection and the molecular clock. *Mol. Biol. Evol.* 3:138–55
30. Gillespie, J. H. 1987. Molecular evolution and the neutral allele theory. *Oxford Surveys Evol. Biol.* 4:10–37
31. Gillespie, J. H. 1989. Lineage effects and the index of dispersion of molecular evolution. *Mol. Biol. Evol.* 6:636–47
32. Goodman, M. 1976. Protein sequences in phylogeny. In *Molecular Evolution*, ed. F. J. Ayala, pp. 141–59. Sunderland, Mass: Sinauer
33. Goodman, M., Czelusniak, J., Koop, B. F., Tagle, D. A., Slightom, J. L. 1987. Globins: A case study in molecular phylogeny. *Proc. Cold Spring Harbor Symp. Quant. Biol.* 52:875–90
34. Goodman, M., Koop, B. F., Czelusniak, J., Weiss, M. L., Slightom, J. L. 1984. The η-globin gene; its long evolutionary history in the β-globin gene family of mammals. *J. Mol. Biol.* 180:803–23
35. Haldane, J. B. S. 1932. *The Causes of Evolution*. New York: Harper & Row
36. Harris, H. 1966. Enzyme polymorphisms in man. *Proc. R. Soc. London, Ser. B*, 164:298–310
37. Harris, S., Barrie, P. A., Weiss, M. L., Jeffreys, A. J. 1984. The primate ψβ1 gene; An ancient β-globin pseudogene. *J. Mol. Biol.* 180:785–801
38. Holmquist, R., Jukes, T. H., Pangburn, S. 1973. Evolution of transfer RMA. *J. Mol. Biol.* 78:91–116
39. Hughes, A. L., Nei, M. 1988. Pattern of nucleotide substitution at major histocompatibility complex loci reveals overdominant selection. *Nature* 335: 167–70
40. Hughes, A. L., Nei, M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proc. Natl. Acad. Sci. USA* 86:958–62
41. Ikemura, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2:13–34
42. Irwin, D. M., Wilson, A. C. 1989. Multiple cDNA sequences and the evolution of bovine stomach lysozyme. *J. Biol. Chem.* 264:11387–93
43. Irwin, D.M., Wilson, A. C. 1990. Concerted evolution of ruminant stomach lysozymes. *J. Biol. Chem.* 265:4944–52
44. Jollés, J., Pollés, P., Bowman, B. H., Prager, E. M., Stewart, C. -B., Wilson, A. C. 1989. Episodic evolution in the stomach lysozymes of ruminants. *J. Mol. Evol.* 28:528–35
45. Kacser, H., Burns, J. A. 1981. The molecular basis of dominance. *Genetics* 97:639–66
46. Kimura, M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713–19
47. Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* 217:624–26

48. Kimura, M. 1979. A model of effectively neutral mutations in which selective constraint is incorporated. *Proc. Natl. Acad. Sci. USA* 76:3440–44
49. Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge Univ. Press
50. Kimura, M., Crow, J. F. 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49:725–38
51. Kimura, M., Ohta, T. 1969. The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61:763–71
52. Kimura, M., Ohta, T. 1971. Protein polymorphism as a phase of molecular evolution. *Nature* 229:467–69
53. Kimura, M., Ohta, T. 1971. *Theoretical Aspects of Population Genetics*. Princeton: Princeton Univ. Press
53a. King, J. L., Jukes, T. H. 1969. Non-Darwinian evolution. *Science* 164:788–98
54. Kingman, J. F. C. 1978. A simple model for the balance between selection and mutation. *J. Appl. Probab.* 15:1–12
55. Kohne, D. E. 1970. Evolution of higher-organism DNA. *Q. Rev. Biophysics* 3(3):327–75
56. Kreitman, M. 1987. Molecular population genetics. *Oxford Surv. Evol. Biol.* 4:38–60
57. Laird, C. D., McConaughy, B. L., McCarthy B. J. 1969. Rate of fixation of nucleotide substitutions in evolution. *Nature* 224:149–54
58. Langley, C. H., Fitch, W. M. 1974. An examination of the constancy of the rate of molecular evolution. *J. Mol. Evol.* 3:161–77
59. Lewontin, R. C. 1974. *The Genetic Basis of Evolutionary Change*. New York, London: Columbia Univ. Press
60. Lewontin, R. C., Hubby, J. L. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54:595–609
61. Li, W-H. 1979. Maintenance of genetic variability under the pressure of neutral and deleterious mutations in a finite population. *Genetics* 92:647–67
62. Li, W.-H. 1985. Accelerated evolution following gene duplication and its implication for the neutralist-selectionist controversy. In *Population Genetics and Molecular Evolution*, ed. T. Ohta, K. Aoki, pp. 333–52. Tokyo: Jpn. Sci. Soc. Press
63. Li, W-H., Graur, D. 1991. *Fundamentals of Molecular Evolution*. Sunderland, Mass: Sinauer
64. Li, W-H., Gojobori, T. 1983. Rapid evolution of goat and sheep globin genes following gene duplication. *Mol. Biol. Evol.* 1:94–108
65. Li, W-H., Sadler, L. A. 1991. Low nucleotide diversity in man. *Genetics* 129:513–23
66. Li, W-H., Tanimura, M., Sharp, P. M. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* 25:330–42
67. Lyckegaard, E. M. S., Clark, A. G. 1991. Evolution of ribosomal RNA gene copy number on the sex chromosomes of *Drosophila melanogaster*. *Mol. Biol. Evol.* 8:458–74
68. McDonald, J. H., Kreitman, M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–54
69. Malécot, G. 1948. *Les mathématiques de l'hérédité*. Paris: Masson
70. Mani, G. S. 1984. A Darwinian theory of enzyme polymorphism. In *Evolutionary Dynamics of Genetic Diversity*, ed. G. S. Mani, pp. 242–98. Berlin: Springer
71. Moriyama, E. N. 1987. Higher rates of nucleotide substitution in *Drosophila* than in mammals. *Jpn. J. Genet.* 62: 139–47
72. Muller, H. J. 1936. Bar duplication. *Science* 83:528–30
73. Nagylaki, T. 1984. The evolution of multigene families under intrachromosomal gene conversion. *Genetics* 106: 529–48
74. Nagylaki, T. 1984. Evolution of multigene families under interchromosomal gene conversion. *Proc. Natl. Acad. Sci. USA* 81:3796–800
75. Nei, M. 1983. Genetic polymorphism and the role of mutation in evolution. In *Evolution of Genes and Proteins*, ed. M. Nei, R. K. Koehn, pp. 165–90. Sunderland, Mass: Sinauer
76. Nei, M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia Univ. Press
77. Nevo, E., Kim, Y. J., Shaw, C. R., Thaeler, C. S. Jr. 1974. Genetic variation, selection and speciation in *Thomomyo talpoides* pocket gophers. *Evolution* 28:1–23
78. O'Brien, S. J., Wildt, D. E., Bush, M.

1986. The cheetah in genetic peril. *Sci. Am.* 254(5):68–76
79. Ohno, S. 1970. *Evolution by Gene Duplication.* Berlin: Springer-Verlag
80. Ohta, T. 1972. Evolutionary rate of cistrons and DNA divergence. *J. Mol. Evol.* 1:150–57
81. Ohta, T. 1972. Population size and rate of evolution. *J. Mol. Evol.* 1:305–14
82. Ohta, T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 246:96–98
83. Ohta, T. 1974. Mutational pressure as the main cause of molecular evolution and polymorphisms. *Nature* 252:351–54
84. Ohta, T. 1976. *Role of very slightly deleterious mutations in molecular evolution and polymorphism. Theor. Popul. Biol.* 10:254–75
85. Ohta, T. 1977. Extension to the neutral mutation random drift hypothesis. In *Molecular Evolution and Polymorphism,* ed. M. Kimura, pp. 148–67. Mishima: Natl. Inst. Genet.
86. Ohta, T. 1980. *Evolution and Variation of Multigene Families.* New York: Springer-Verlag
87. Ohta, T. 1982. Allelic and non-allelic homology of a supergene family. *Proc. Natl. Acad. Sci. USA* 79:3251–54
88. Ohta, T. 1983. On the evolution of multigene families. *Theor. Popul. Biol.* 23:216–40
89. Ohta, T. 1986. Population genetics of an expanding family of mobile genetic elements. *Genetics* 113:145–59
90. Ohta, T. 1987. Very slightly deleterious mutations and the molecular clock. *J. Mol. Evol.* 26:1–6
91. Ohta, T. 1987. Simulating evolution by gene duplication. *Genetics* 115: 207–13
92. Ohta, T. 1988. Evolution by gene duplication and compensatory advantageous mutations. *Genetics* 120:841–47
93. Ohta, T. 1991. Multigene families and the evolution of complexity. *J. Mol. Evol.* 33:34–41
94. Ohta, T. 1991. Evolution of the multigene family: A case of dynamically evolving genes at major histocompatibility complex. In *Evolution of Life,* ed. S. Osawa, T. Honjo, pp. 145–59. Berlin: Springer
95. Ohta, T. 1992. Theoretical study of near neutrality. II. Effect of subdivided population structure with local extinction and recolonization. *Genetics* 130: 917–23
96. Ohta, T., Kimura, M. 1971. On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.* 1:18–25
97. Ohta, T., Tachida, H. 1990. Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. *Genetics* 126:219–29
98. Parham, P. 1989. Getting into the groove. *Nature* 342:617–18
98a. Romer, A. S. 1968. *The Procession of Life.* London: Weidenfeld & Nicolson
99. Sawyer, S. A., Dykhuizen, D E., Hartl, D. L. 1987. Confidence interval for the number of selectively neutral amino acid polymorphisms. *Proc Natl. Acad. Sci. USA* 84:6225–28
100. Selander, R. K. 1976. Genic variation in natural populations. In *Molecular Evolution,* ed. F. J. Ayala. Sunderland, Mass: Sinauer
101. Selander, R. K., Levin, B. R. 1980. Genetic diversity and structure in *Escherichia coli* populations. *Science* 210: 545–47
102. Sharp, P. M. 1989. Evolution at 'silent' sites in DNA. In *Evolution and Animal Breeding,* ed. W. G. Hill, T. F. C. Mackey, pp. 23–32. Wallingford: C. A. B. Int.
103. Sheppard, H. W., Gutman, G. A. 1981. Allelic forms of rat Kappa chain genes: evidence for strong selection at the level of nucleotide suquence. *Proc. Natl. Acad. Sci. USA* 78:7064–68
104. Sibley, C. G., Ahlquist, J. E. 1984. The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. *J. Mol. Evol.* 20:2–15
105. Sibley, C. G., Ahlquist, J. E. 1987. DNA hybridization evidence of hominoid phylogeny: Results from an expanded data set. *J. Mol. Evol.* 26: 99–121
106. Singh, R. S. 1989. Population genetics and evolution of species related to *Drosophila melanogaster. Annu. Rev. Genet.* 23:425–53
107. Tachida, H. 1991. A study on a nearly neutral mutation model in finite populations. *Genetics:* 128:183–92
108. Takahata, N., Kimura, M. 1979. Genetic variability maintained in a finite population under mutation and antocorrelated random fluctuation of selection intensity. *Proc. Natl. Acad. Sci. USA* 76:5813–17
109. Templeton, A. 1985. The phylogeny of the hominoid primates: A statistical analysis of the DNA-DNA hybridization data. *Mol. Biol. Evol.* 2:420–33
110. Watson, J. D. 1976. *Molecular Biology of the Gene.* Menlo Park, Calif: Benjamin. 3rd ed.

111. Wines, D. R., Brady, J. M., Southard, E. M., MacDonald, J. 1991. Evolution of the rat kallikrein gene family: gene conversion leads to functional diversity. *J. Mol. Evol.* 32:476–92
112. Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159
113. Wright, S. 1942. Statistical genetics and evolution. *Bull. Am. Math. Soc.* 48:223–46
114. Wu, C-I., Li, W-H. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA* 82:1741–45
115. Yokoyama, S., Yokoyama, R. 1990. Molecular evolution of visual pigment genes and other G-protein-coupled genes. In *Population Biology of Genes and Molecules*, ed. N. Takahata, J. F. Crow. Tokyo: Baifukan
116. Zuckerkandl, E., Pauling, L. 1965. Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins*, ed. V. Bryson, H. J. Vogel, pp. 97–166. New York: Academic Press
117. Zuckerkandl, E. 1987. On the molecular evolutionary clock. *J. Mol. Evol.* 26: 34–46