THE POPULATION DYNAMICS OF GENE SUBSTITUTION IN MOLECULAR EVOLUTION*

P. NARAIN
Indian Agricultural Statistics Research Institute, New Delhi 110 012, India.

ABSTRACT

The population dynamics of gene substitution at the molecular level is reviewed and discussed with particular reference to the stochastic approach in evolutionary studies. After a brief historical development of the theories on the mechanism of evolution, the diffusion process method is used to discuss the rate of gene substitution and the average time taken to achieve it. Some aspects of neutral theory of molecular evolution are reviewed. The rate of evolution at the molecular level is discussed with reference to haemoglobin in vertebrates. The level of heterozygosity in natural populations under infinite sites model is estimated by the conditional approach and compared with that obtained by the unconditional approach. Most of the treatment is based on the results obtained in relation to the neutral theory of molecular evolution and on stochastic problems and methods in population genetics.

INTRODUCTION

In a recent BBC Television series, shown on TV, one watched man's progress from the pre-historic times to the present day, in other words, man's own evolution. This series was by Jacob Bronowski a famous mathematician turned writer and humanist and was based on his book *The Ascent of Man*. More recently, a Japanese scientist discovered that man arose from an ancestral stock, common with the chimpanzee only 7 million years ago. On the evolutionary time scale, the evolution seems to have gone very much faster, particularly in the growth of our brain during the last several million years. The cranial capacity of the brain ba() doubled in 3 million years. By contrast, insects such as ants have remained quite unchanged for about 50 million years. Many such examples can be given. The fact that evolution occurs is well-recognized but what causes it to occur has been debated ever since the middle of the 19th century when, Charles Darwin ascribed it to 'Natural Selection'.

Since characters are transmitted from one generation to another by genetic factors, evolutionary changes are expressed in terms of the replacement of a given factor A by its counterpart a in a given population. In other words, it is the gene substitution at the population level which is reflected in evolution.

The process of gene substitution, on the evolutionary time scale, however, is made up of a sequence of events in which a rare mutant form which appeared, usually singly represented, in the population, finally spreads through the whole population reaching a frequency of 100% (fixation). For this, it takes, however, a considerable length of time. Usually a large number of mutants arise in each generation in a population. But most of them are lost by chance within a short interval of time without being used in evolution. The events of production of mutants and their fixation, however, go on incessantly for millions of

---

based on Prof. R. S. Verma Memorial Lecture delivered at the University of Delhi, Delhi, "A Date with Mathematicians, 1985", organized by the Delhi chapter of the Mathematical Association of India.

generations. There is then a statistical equilibrium, of a dynamic nature, in which mutational production of new alleles is counterbalanced by their random extinction, giving rise to a certain level of heterozygosity. To determine this, certain models of allelic mutation (such as infinite sites model of Kimura[1]) which are more realistic on the molecular level than the classic two-allele single locus model, are to be invoked.

For understanding the mechanism of evolution at the molecular level, the mathematical theory of the stochastic process of gene frequency change of mutant alleles in finite population has proved to be very effective. The most powerful method for treating such processes is the method of diffusion equations[2,3]. This method is used to develop the concepts of gene frequency distribution, probability of fixation of a mutant and the average length of time involved for each gene substitution. Some of the mathematical and statistical aspects of the process of gene substitution are reviewed in this paper. The evolutionary rate at the molecular level is discussed for the haemoglobin in vertebrates which has proved to be an excellent model system for such studies. Finally, the basic tenets of the neutral theory of molecular evolution are briefly summarized. However, before we take up these issues, a brief historical development of the various theories on the mechanism of evolution may not be out of place.

HISTORICAL DEVELOPMENT

Darwin[4], in his famous book *The Origin of Species* proposed that evolution occurred by natural selection. However, for adaptive evolution to be an inevitable process over time, it is necessary to have variation and to understand factors governing it. In this respect he failed to account for the origin and the precise nature of heritable variation. With the rediscovery of the laws of Mendel in 1900 this lacuna was cleared by establishing that mutation provides raw material for the evolution to act upon. There was, however, a controversy on the compatibility of Darwinism and Mendelism which was resolved with the development of population genetics, involving the mechanism of inheritance at the population level. The famous law of population genetics discovered simultaneously by Hardy[5], a mathematician, and Weinburg[6], a botanist, paved the way for the synthesis. Mendelism clearly indicated that the newer forms, better adapted to the environment and appearing in subsequent generations, can be understood in terms of mutations causing discontinuous variation. With respect to the simplest situation of a favourable and unfavourable character represented by gene *A* and *a* respectively, evolution is said to have occurred when *A* completely replaces *a*. In other words, it is gene substitution which is at the root of evolution. Further, since this process of gene substitution occurs in time in a progressive and continuous manner, such changes are accumulative within the species during evolution. It is only when one can quantitatively describe this process that the mechanism of evolution is better understood. Such studies involve mathematical and statistical approaches. Three personalities, viz., Fisher, Haldane and Sewall Wright contributed significantly in this connection. In particular, they determined the population consequences of Mendelian inheritance and applied them to evolution.

The contributions of Fisher[7] in this field are embodied in his book *The Genetical Theory of Natural Selection*. He synthesized Darwinism, Mendelism and biometry and held the view that the rate and direction of evolution are almost exclusively determined by natural selection with mutation, migration and random drift playing subsidiary roles. The contributions of Haldane[8] are summarised in his book *The Causes of Evolution*. He is credited with developing the concept of the cost of natural selection which implies the proportion of genetic deaths due to substitution of one gene by another. Haldane[9] gave a quantitative measure for the rate of evolutionary change at the morphological level as a percentage change

in the mean length of a certain structure over time. He invented the term *darwin* implying the rate of change of 10^{-6} per year which means increase or decrease of size by a factor of e (the base of natural logarithm, i.e., 2.718) per million years. In terms of this unit, in the evolution of the horse, the height of certain teeth increased on the average by about 40 milidarwins, at a very slow rate. According to him, the rate of gene substitution is very low being one every 300 generations. Haldane[10] also realised the importance of mathematical methods in relation to the genetical theory of evolution. The work of Wright[11] dealing with the mathematical approach to the theory of evolution is published elsewhere. He emphasized the importance of random genetic drift due to finite population and the prevalence of non-additive gene interactions. Wright[12,13] developed a theory of evolution, which he later called 'the shifting balance theory'.

In the synthesis of Darwinism and Mendelism, besides the works of Fisher, Haldane and Wright, mention need be made of the contribution of Muller[14] who held the view that the very basis of adaptive evolution by natural selection is the fundamental nature of gene mutation. Not only is each gene self-reproducing but also mutated forms of a gene are again self-reproducing. The gene is the very basis of life whose criterion is the potentiality of evolution by natural selection.

Apart from the mathematical approaches, several workers[15–19] have provided evidences from fossil records (palaeontology), experimental natural populations, ecology, etc. to give rise to neo-Darwinism or a synthetic theory of evolution, in which ideas of Darwin and Mendel as well as methods of biometry were integrated with paramount importance to natural selection and negligible attention to random genetic drift due to finite population size. Subsequently, with the advent of molecular genetics giving gene a physical entity as a segment of deoxyribonucleic acid (DNA), attempts of Zuckerkandl and Pauling[20] led to estimation of evolutionary rates of nucleotide substitutions inside the genes. They were able to estimate the evolutionary rates of the amino acid substitutions by comparing amino acid sequences of proteins in related organisms—as for instance, comparison of haemoglobin among vertebrates—and by using palaeontological data. The contributions of Harris[21] as well as those of Lewontin and Hubby[22], using electrophoretic techniques, revealed considerable enzyme variability among individuals in several organisms. The mathematical approach of the population genetics using diffusion models[2] was fruitfully applied subsequently to determine how genes evolve at the molecular level. The evolutionary change at the molecular level indicated considerable incompatibility with the expectations of neo-Darwinism. Kimura[23] has made a comprehensive treatment of the subject and asserted that random genetic drift plays an important role in the context of evolution at the molecular level. In a sense it is non-Darwinian evolution since natural selection is not considered necessary for the evolution to occur. Naturally, it has given rise to a fierce debate between supporters of *selectionist* and *neutralist* interpretation of protein evolution. Kimura's approach is however, a stochastic one implying that evolution of a particular form occurs with a certain probability which need not always be unity.

THE DIFFUSION PROCESS

Mathematically, the stochastic process of gene frequency change can be approximated as a diffusion process with the random variable x representing the frequency of a gene A with time parameter changing continuously. At a given time, we have now a gene frequency distribution with a density function $f(x, t)$. It is possible to show, by methods often used in physics, that this density function satisfies the following Kolmogorov Forward Equation

$$\frac{\partial f(x,t)}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial x^2}[v(x)f(x,t)] - \frac{\partial}{\partial x}[m(x)f(x,t)] \qquad (1)$$

with $m(x)$ and $v(x)$ as given instantaneous drift and diffusion coefficients respectively. These represent, respectively, the expected mean change and the variance of change in gene frequency. By taking specific values of these various parameters and solving the resulting partial differential equations, it is possible to arrive at an analytical solution for the gene frequency distribution given by Kimura[24]. The distribution of unfixed classes, after becoming flat, decays at the rate of $1/2N$ per generation. In this stochastic process, the population size is usually expressed as an effective population size N_e which is usually smaller than the actual population size N since the progeny number does not follow a Poisson distribution. In the case of human beings, N_e is about half of N.

RELATIVE CONTRIBUTION OF A MUTANT TO SUBSTITUTION

More important than the gene frequency distribution is the probability of fixation of a mutant being favoured during evolution[25,26]. For this purpose, we consider the Kolmogorov Backward Equation, the adjoint of (1), given by

$$\frac{\partial f(q, x; t)}{\partial t} = \frac{v(q)}{2}\frac{\partial^2 f(q, x; t)}{\partial q^2} + m(q)\frac{\partial f(q, x; t)}{\partial q}, \quad (2)$$

where q, the initial gene frequency at time $t=0$, is now a random variable with frequency x as fixed and the process is considered retrospectively by reversing the time sequence. When $x=1$, $f(q, 1; t)$ is denoted by $u(q, t)$, the probability that the gene whose initial frequency was q becomes fixed in the population by the tth generation.

A general formula for the ultimate probability of fixation defined by

$$u(q) = \lim_{t\to\infty} u(q, t), \quad (3)$$

is given by

$$u(q) = \int_0^q G(x)\,dx \Big/ \int_0^1 G(x)\,dx, \quad (4)$$

where

$$G(x) = \exp\left[-2\int \frac{m(x)}{v(x)}\,dx\right]. \quad (5)$$

For additive selection and binomial sampling,

$$m(x) = sx(1-x),\ v(x) = x(1-x)/2N_e, \quad (6)$$

where s is the selection coefficient in favour of gene A. These, when substituted in (4), give the formula

$$u(q) = \frac{1-\exp(-4N_e sq)}{1-\exp(-4N_e s)}. \quad (7)$$

This shows that the probability of fixation depends on the initial gene frequency q and the parameter $S = 4N_e s$. The probability of fixation (u) of an individual mutant can then be obtained by taking $q = 1/2N$ in (7) and approximating the expression for the probability of fixation. For the strictly neutral mutant in which $s \to 0$, the probability of fixation (u_0) is just the initial gene frequency $1/2N$. We can then consider the ratio $u:u_0$. This ratio gives the contribution of a mutant having selective advantage of s to gene substitution relative to that of a strictly neutral mutant and shows how the process of gene substitution is affected by the selective advantage or disadvantage of the gene. If s is less than $1/2N_e$, the mutant gene could be taken as almost neutral. We can then determine what happens to the gene substitution in the neighbourhood of the neutrality.

RATE OF GENE SUBSTITUTION AND TIME INVOLVED THEREIN

The process of accumulation of new mutants within the species during evolution can be quantified by studying the population dynamics of gene substitution. If v is the mutation rate per gamete per generation then, in a population of N individuals, $2Nv$ new distinct mutants would be introduced into the population in each generation. When the process is continued over a very long time, as in evolution, the rate

per generation of mutant substitution which is synonymous with the rate of evolution would be $2Nvu$. Of all the $2Nv$ new mutants appearing in each generation, only a fraction u eventually reach fixation. In the case of neutral mutations, we have seen that the probability of fixation is equal to the initial frequency, viz., $1/2N$ so that the rate of evolution is just equal to the mutation rate and independent of the population size. This is a very important result which led Kimura[27] to one of the premises of neutral theory of evolution in terms of the constancy of the evolutionary rate at the molecular level. In contrast, if we take advantageous mutants and assume that S is far greater than 1, then the probability of the fixation of the mutant is approximately $2sN_e/N$. Substituting this in the formula for rate of evolution, we find now that it depends on S as well as v. One should expect, therefore, in this case that the rate of evolution would depend strongly on the environment, being high for a species offered a new ecological opportunity but low for that kept in a stable environment. However, it is unlikely that the product Sv remains the same for diverse lineages of vertebrate evolution. Irrespective of whether the evolution at the phenotypic level is very rapid as in the line leading to man or has practically stopped as in the line leading to carp, the observed data at the molecular level do not support such a contingency.

In considering the population dynamics of mutant substitution, we need, in addition to the probability of gene fixation, the average length of time involved for each gene substitution. For getting quantitative estimates for the time to fixation, a general theory based on diffusion approximation was developed[3]. This involves conditioning the diffusion process for the contingency of eventual fixation, leading to conditioned forward diffusion equation given by

$$\frac{\partial f_{c1}(x;\,t)}{\partial t}=\frac{1}{2}\frac{\partial^2}{\partial x^2}[v(x)f_{c1}(x;\,t)] - \frac{\partial}{\partial x}[m_1^*(x)f_{c1}(x;\,t)], \tag{8}$$

where

$$m_1^*(x)=m(x)+v(x)\frac{\mathrm{d}}{\mathrm{d}x}[\log u(x)], \tag{9}$$

showing thereby that density function f_{c1} also satisfies (1) but with modified drift coefficient $m_1^*(x)$. The diffusion coefficient $v(x)$, however, remains the same in the conditional process. The adjoint of (8), the conditioned backward diffusion equation is then given by

$$-\frac{\partial u_{c1}(q;\,t)}{\partial t}=L_{c1}u_{c1}(q;\,t) \tag{10}$$

subject to boundary conditions

$$\lim_{x\to 0} u_{c1}(x;\,t)=\text{finite}$$

$$u_{c1}(1;\,t)=1, \tag{11}$$

where the operator L_{c1} is given by

$$L_{c1}=\frac{v(q)}{2}\frac{\partial^2}{\partial x^2}+m_1^*(q)\frac{\partial}{\partial q}. \tag{12}$$

It can be verified that the probability

$$u_{c1}(q)=\lim_{t\to\infty} u_{c1}(q;\,t), \tag{13}$$

is unity as expected.

The various moments of the distribution of time until fixation of the mutant can be obtained by developing corresponding differential equations with the help of (10). For average time until fixation, we have

$$M_{c1}(q)=\int_0^\infty t\,u_{c1}(q;\,t)\,\mathrm{d}t. \tag{14}$$

Using (10), the differential equation for $M_{c1}(q)$ is obtained as

$$\frac{v(q)}{2}\frac{\mathrm{d}^2M_{c1}(q)}{\mathrm{d}q^2}+m_1^*(q)\frac{\mathrm{d}M_{c1}(q)}{\mathrm{d}q}+1=0, \tag{15}$$

subject to boundary conditions

$$\lim_{q\to 0} M_{c1}(q)=\text{finite}$$

$$M_{c1}(1)=0. \tag{16}$$

The solution of (15) is given by

$$M_{c1}(q)=\int_q^1 I(q)u(q)[1-u(q)]\,dq + \frac{1-u(q)}{u(q)}\int_0^q I(q)[u(q)]^2\,dq, \quad (17)$$

where

$$I(q)=2\int_0^1 G(q)\,dq, v(q)G(q). \quad (18)$$

In the case of pure random drift, $m(q)=0$, $v(q)=q(1-q)/2N_e$ and we get from (17),

$$M_{c1}(q)=-4N_e\left(\frac{1-q}{q}\right)\log_e(1-q). \quad (19)$$

This formula shows that the average time to fixation depends on the effective population size and the initial gene frequency. For the case of a mutant introduced singly in the population in the initial stages, this average time tends to a limit of $4N_e$. In other words, it takes, on the average, four times the effective population size for a selectively neutral mutant to reach fixation by random frequency drift. This can be compared with $1/v$ as the average time interval between the two consecutive fixations since the rate of gene substitution is simply equal to the mutation rate in the case of neutral alleles. Whenever $4N_e$ is greater than $1/v$, the population will show considerable transient polymorphism. But when $4N_e$ is smaller than $1/v$, the population is monomorphic most of the time. It may be noted that the rate of evolution in terms of the gene substitution is independent of the rate at which individual mutant alleles increase or decrease within the population. What matters is the average interval between two consecutive fixations. We may have two contrasting situations in both of which the rate of evolution may be identical but in one case individual mutant alleles increase much more rapidly than in the other case.

Apart from the mean length of time until fixation of a mutant, one can also work out the variance of the length of time to its fixation[28]. For a single mutant introduced into the population initially, the variance tends to a limiting value of $4.64\,N_e^2$ giving a coefficient of variation of about 54%.

SOME ASPECTS OF NEUTRAL THEORY OF MOLECULAR EVOLUTION

Perhaps the greatest contribution of the method of stochastic process to population genetic problems is the development of a stochastic theory of molecular evolution[27-29]. According to this theory, a majority of gene substitutions at structural loci in evolution are due to random fixation of neutral or nearly neutral mutations and further most of the protein polymorphisms observed in the present natural populations simply represent a transient state of gene substitution by random genetic drift. As already stated, this theory has generated a great deal of controversy among evolutionary biologists who sometimes call it non-Darwinian and hence not tenable.

The neutral theory of molecular evolution has given rise to debate on two things. Firstly, the rate of evolution in terms of mutant substitution is roughly constant per year for each protein among diverse lineage. Secondly, the observed level of heterozygosity is mostly found to lie in a narrow range of 0 to 20% among various species. We discuss these issues separately:

Rate of evolution

Comparative studies of amino acid sequences of proteins help us to obtain reliable estimates of the rates at which new mutant genes are incorporated into the species in evolution. We can bypass classical genetical analysis by directly comparing proteins coded by genes between, say, mice and whales even though they cannot be crossed with each other. During the past decade and a half, a large body of data have accumulated on protein sequences; the most important seems to be the globins, particularly among vertebrates. Such studies serve as the model system to discuss evolutionary rate at the molecular level. The globins are of two types, haemoglobin and myoglobin. The

haemoglobin molecule is a tetramer consisting of two identical α-chains and two identical β-chains.

The evolutionary tree of the genes determining haemoglobin is given by Maynard Smith[30]. The primitive jawless vertebrates are believed to have had a single-chain globin molecule in their blood some 500 million years ago. This is because the lampreys, the survivors of this group, have a single-chain globin. Some time during the origin of the vertebrates with jaws, about 450 million years ago, the gene specifying globin is believed to have duplicated and the two copies so produced are believed to have diverged to become the genes for the α and the β chains. Subsequently, α and β chains diverged to those for mammals and carp. The data generated due to the sequencing of amino acids of the globin molecules can be used to estimate the rate of evolution at the protein level. For instance, by comparing the sequences of chains of carp and mouse, we can calculate the number of amino acid substitutions needed to convert one into the other. Dividing this number by twice the time to a common ancestor, (900 million years in this case), we can obtain an estimate of evolutionary rate in terms of amino acid substitutions per polypeptide chain per year. Dividing again by the number of amino acid sites (141 for α chain and 146 for β chain) in the chain gives the estimate of the number of substitutions per site per year (table 1).

It is clear that the rate of evolution at the protein level is uniform among diverse lineages of vertebrates. One can say that haemoglobin must have evolved at a rate of approximately one amino acid substitution per site per billion years.

Table 1 *Average rates of evolution of haemoglobins*

Comparison	Amino acid substitutions per site per year ($\times 10^{-9}$)
Human α vs Carp α	0.89
Human α vs other mammal α	0.88
Human α vs human β	0.89
Human β vs other mammal β	1.19
Human β vs Lamprey globin	1.28
Mouse β vs other mammal β	1.40

Kimura[23] explains that such uniformity in evolutionary rates at molecular level is also found in other cases like cytochrome *c*, fibrinopeptides, pancreatic ribonuclease, lysozyme, insulin and histone. He suggests that we use 10^{-9}/a.a./yr as a unit of evolutionary time, calling it the *pauling*. Then the rate for haemoglobin is about one pauling but that for cytochrome *c* is 30 centi-pauling. Such evidences of constancy of evolutionary rate compare well with the theoretical finding, stated earlier, that for selectively neutral mutants the rate of evolution in terms of mutant substitution is equal to the mutation rate per gamete. The observed constancy can thus be explained by assuming that the majority of gene substitutions are due to random fixation of neutral mutations.

Heterozygosity

In order to examine quantitatively some aspects of heterozygosity, population genetic models for mutation production in a finite population are invoked. One such model is that of infinite sites[1]. This model assumes that the number of nucleotides for mutation is sufficiently large while the mutation rate per site is so low that whenever a mutation occurs it represents a new site in which no mutant forms are segregating within the population. Such a model seems to be realistic if we take the case of a cistron coding for the α chain of the mammalian haemoglobin. This polypeptide, consisting of 141 amino acids, as already stated, corresponds to 423 nucleotide sites. This allows 4^{423} or 10^{254} allelic states through base replacements alone. Thus for any one of these alleles there are 3×423 or 1269 other alleles that can be reached by a single step base replacement. The probability of returning to the original allele from any one of the derived alleles by further single base replacement is only 1 in 1269, assuming that all base replacements occur with equal probability.

The diffusion approach used by Kimura[1] in using this model takes into account all the

sample paths of the process that lead either to fixation or loss of the mutant form from the population. But since the population is usually large and the selection is very weak, the event of fixation has a very small probability. As such, it is more appropriate to consider only those sample paths of the process that lead to the loss of the mutant form from the population disregarding those in which they are fixed. In other words, one has to consider a process conditioned for the loss of alleles from the population. If we assume that on the average, in each generation, mutant forms appear in a population at r_m sites, a balance between the continued production of new mutants over many generations and their loss from the population is established. We can then study conditional stable distribution of the mutant frequency in different segregating sites following Narain[31,32]. Denoting this distribution by $\phi_{c0}(q, x)$, we have

$$\phi_{c0}(q, x)=r_m \int_0^\infty f_{c0}(q, x; t)\,dt \quad (0<x<1) \tag{20}$$

so that $\phi_{c0}(q, x)\,dx$ gives the expected number of sites having mutant in the frequency range x to $(x+dx)$ and conditional to their loss from the population. The expectation of an arbitrary function $g(x)$ with respect to this distribution, denoted by $I^g_{c0}(q)$ then satisfies the ordinary differential equation

$$\frac{v(q)}{2}\frac{d^2 I^g_{c0}(q)}{dq^2}+m^*_0(q)\frac{dI^g_{c0}(q)}{dq} + v_m g(q)=0, \tag{21}$$

where

$$m^*_0(q)=m(q)+v(q)\frac{d}{dq}[\log(1-u(q))], \tag{22}$$

subject to the boundary conditions

$$I^g_{c0}(0)=0, \quad \lim_{q\to 1} I^g_{c0}(q)=\text{finite}. \tag{23}$$

The various statistical properties of this distribution can be studied by assigning various forms to $g(q)$. In particular, by taking $g(q)=1$ and $q(1-q)$, we get respectively the total number of segregating sites, $I_c(q)$ and the average number of heterozygous nucleotide sites per individual, $H_c(q)$, conditional to loss of mutants. Taking $q=1/2N$, we get these properties for the case when the mutant form in each site is represented only once at the moment of its occurrence in a population of size N as given below:

$$\left.\begin{aligned} I_c(1/2N)&\approx 4N_e v \log_e(4N_e)\\ H_c(1/2N)&\approx \frac{8}{3}N_e v, \end{aligned}\right\} \tag{24}$$

where v is the total mutation rate and we assume $(N_e/N)=0.5$. The heterozygosity, i.e., the probability that a cistron (gene) would be heterozygous at one or more sites would then be

$$h_{c0}\approx 1-\exp[-H_c(1/2N)]. \tag{25}$$

A comparative picture of the total number of segregating sites and the heterozygosity for the conditional versus unconditional cases for varying N_e with mutation rate of 2×10^{-6} is shown in figure 1a, b.

If we assume that molecular neutral mutants occur at the rate of 2 per gamete per generation, with effective population size (N_e) as 10^5, we get the average number of heterozygous sites per individual as 5.3×10^5. This estimate is about two-thirds of that obtained by Kimura[1]. Similarly the heterozygosity is obtained as about 12% as against the value of 18% obtained by the unconditional approach. Nevo[33] reviewed genetic variation in natural populations of plants, animals and humans involving 243 species in which 14 or more loci were tested. Among 48 mammalian species, the average heterozygosity was found to be distributed almost uniformly between 0% and 7% although in two exceptional cases, these values were around 10%. This shows that the approach of Narain[31,32] giving an estimate of 12% is nearer the experimental values. The unconditional approach seems to be misleading as it

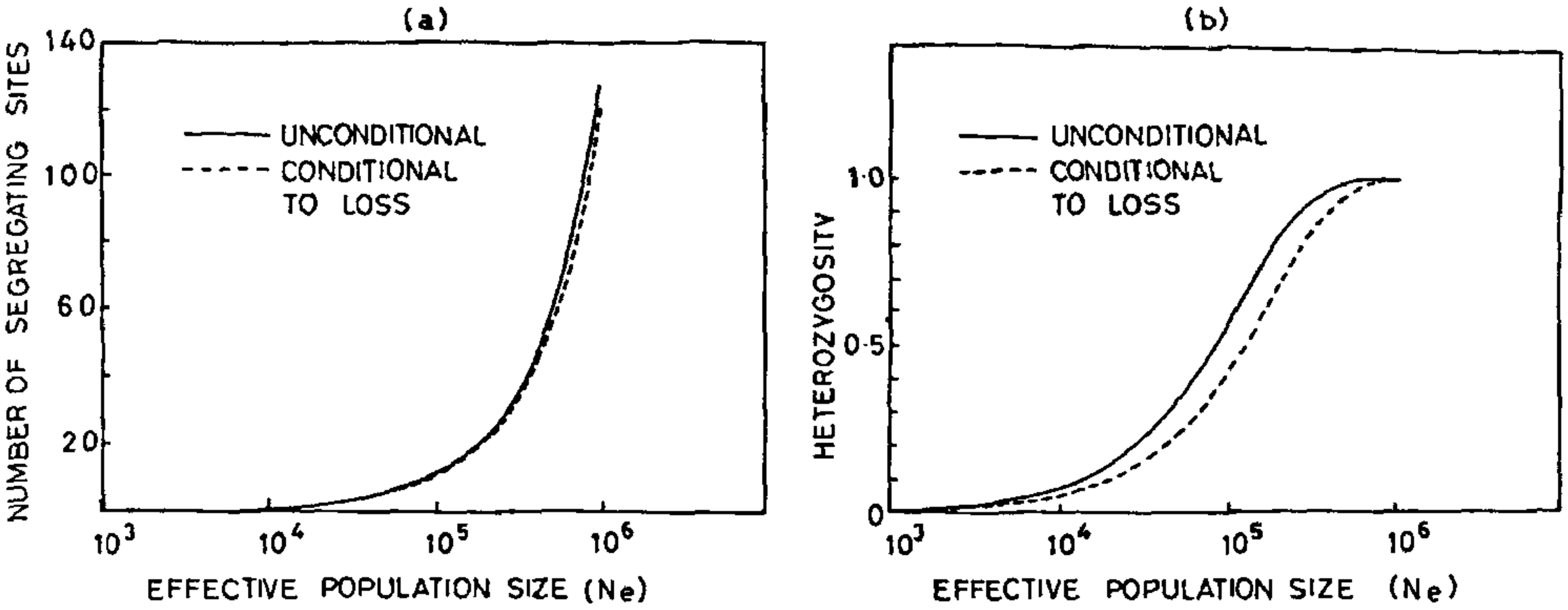

Figure 1a, b. Statistical properties of equilibrium distribution under steady flux of mutations for different values of N_e.

gives too heavy a weight to sample paths that rarely occur.

Several efforts were made[34] to develop appropriate statistical tests to justify the validity of the neutral theory. Gene frequency data on protein polymorphism from as many as 130 species were used to test the agreement of various statistical quantities with the values predicted from the neutral theory. For this purpose, the relationship between the mean and variance of heterozygosity in an equilibrium population could be used[35]. Under the neutral mutation hypothesis, the level of genetic variability in an equilibrium population is determined by mutation rate v and effective population size N_e. If we use the infinite allele model, in which new mutations are assumed to be always different from the pre-existing alleles in the population, the average heterozygosity per locus as shown by Kimura and Crow[36], is given by

$$H = M/(1-M), \tag{26}$$

where

$$M = 4N_e v. \tag{27}$$

Whether the observed heterozygosity agrees with the above theoretical value or not, is in practice, however very difficult to examine since the exact values of N_e and v in a population are hardly obtainable. However, the variance of single locus heterozygosity is given by[37]

$$V(h) = 2M/(1-M)^2(2-M)(3-M). \tag{28}$$

From this the neutral mutation hypothesis can be tested without knowing N_e and v separately. By estimating M from the observed value of average heterozygosity, we can compute the theoretical variance by using (28) and then compare it with the observed variance. Application of this technique to many organisms indicated good agreement between the theoretical and observed variances. This and several other tests showed[34] that most data on protein polymorphism can be accounted for by the neutral theory.

EVIDENCE AGAINST THE NEUTRAL THEORY

In the previous section on some aspects of the neutral theory of molecular evolution, we have given arguments which support the theory. However, not all of these arguments are necessarily convincing. The argument of the constancy of evolutionary rates, often termed *molecular clock*, is now losing ground. Fitch[38] showed that the region of a protein that evolves

is different in different lineages. Moreover, the variance of the rates appears to be two to three times larger than that expected under neutrality. More serious, it turns out that the rate of evolution is constant relative to absolute time and not to generation number as required by the neutral theory, although mutation rate is more nearly constant per generation than per year.

Recently, Gillespie[39] used a Cox process (Poisson process with a randomly varying rate) to model the molecular clock and concluded that the molecular evolution may be an *episodic process*; in other words, bursts of rapid evolution separated by long spells of very slow evolution. Such a property of molecular evolution is more easily accounted for by the action of natural selection than by the neutral allele model. Takahata[40], on the other hand, invoked a fluctuating neutral space model (fluctuating substitution rates through changes of selective constraints) to account for the enhanced variance of the number of substitutions. Using a time-dependent renewal process, the possibility that each gene substitution changes the degree of selective constraint is examined. Such an approach, however, does not exhibit any episodic nature of molecular evolution.

9 November 1987; Revised 27 October 1988

1. Kimura, M., *Genetics*, 1969, **61**, 893.
2. Kimura, M., *J. Appl. Probab.*, 1964, **1**, 177.
3. Narain, P., *J. R. Stat. Soc. London*, 1974, **B36**, 258.
4. Darwin, C., *The origin of species by means of natural selection*, John Murray, London, 1859.
5. Hardy, G. H., *Science*, 1908, **28**, 49.
6. Weinberg, W., *Jahresh. Verein Vaterl. Naturk. Württem*, 1908, **64**, 368.
7. Fisher, R. A., *The genetical theory of natural selection*, University Press, Oxford, 1930, 2nd revised edition, Dover Publications, 1958.
8. Haldane, J. B. S., *The causes of evolution*, Harper and Row, New York, 1932.
9. Haldane, J. B. S., *Evolution*, 1949, **3**, 51.
10. Haldane, J. B. S., In: *Background to modern science*, (eds) J. Needham and W. Pagel. Cambridge University Press, 1938.
11. Wright, S., *Genetics*, 1931, **16**, 97.
12. Wright, S., *Proc. VI Int. Congr. Genet.*, 1932, **1**, 356.
13. Wright, S., In: *Mathematical topics in population genetics*, (ed.) K. Kojima, Springer–Verlag, Berlin, 1970.
14. Muller, H. J., In: *Proc. Int. Congr. Plant Sci.*, 1929, **1**, 897.
15. Simpson, G. G., *Tempo and mode in evolution*. Columbia University Press, 1944.
16. Dobzhansky, Th., *Genetics and the origin of species*, 1st edn, Columbia University Press. 1937.
17. Ford, E. B., *Ecological genetics*, 1st edn, Methuen, London, 1964, 2nd and 3rd edns. Chapman and Hall, London. 1965, 1971.
18. Mayr, E., *Animal species and evolution*, Harvard University Press, 1963.
19. Waddington, C. H., *The strategy of the genes*, George Allen and Unwin, London, 1957.
20. Zuckerkandl, E. and Pauling, L., In: *Evolving genes and proteins*, (ed.) V. Bryson and H. J. Vogel, Academic Press, New York, 1965.
21. Harris, H., *Proc. R. Soc. London*, 1966, **B164**, 298.
22. Lewontin, R. C. and Hubby, J. L., *Genetics*, 1966, **54**, 595.
23. Kimura, M., *The neutral theory of molecular evolution*, Cambridge University Press, 1983.
24. Kimura, M., *Proc. Natl. Acad. Sci. USA*, 1955, **41**, 144.
25. Kimura, M., *Ann. Math. Stat.*, 1957, **28**, 882.
26. Kimura, M., *Genetics*, 1962, **47**, 713.
27. Kimura, M., *Nature (London)*, 1968, **217**, 624.
28. Narain, P., *Genet. Res., Camb.*, 1970, **15**, 251.
29. King, J. L. and Jukes, T. H., *Science*, 1969, **164**, 788.
30. Maynard Smith, J., *The theory of evolution*, Penguin Books, 1958, 1975.
31. Narain, P., *Proc. Indian Natl. Sci. Acad.*, 1979, **B45**, 239.
32. Narain, P., *Proc. 70th Indian Sci. Congr.*, 1983, p. 1
33. Nevo, E., *Theor. Pop. Biol.*, 1978, **13**, 121.
34. Nei, M., In: *Vito Volterra Symp. on Mathematical Models in Biology*, (*Lecture Notes in Biomathematics*, **39**), C. Barigozzi, Springer-Verlag, Berlin, 1980.
35. Nei, M., Fuerst, P. A. and Chakraborty, R., *Nature (London)*, 1976, **262**, 491.

36. Kimura, M. and Crow, J. F., *Genetics*, 1964, **49**, 725.
37. Li, W. H. and Nei, M., *Genet. Res., Camb.*, 1975, **25**, 229.
38. Fitch, W. M., In: *Synthese, struktur, und Funktion des Hämoglobins*, (eds) H. Martin and L. Nowicki, J. F. Lehmanns Verlag, Munchen, 1971.
39. Gillespie, J. H., In: *Evolutionary processes and theory*, (eds) S. Karlin and E. Nevo, Academic Press, New York, 1986.
40. Takahata, N., *Genetics*, 1987, **113**, 169.

NEWS

TECHNICAL BULLETIN—A QUARTERLY HOUSE BULLETIN OF NATIONAL PHYSICAL LABORATORY

A special issue on the Birth Centenary of Professor C. V. Raman (7th November 1988) was released by Prof. D. S. Kothari, Chancellor, University of Delhi, Delhi, on 7th November 1988.

This special publication with colour photographs runs to fifty pages and contains a wealth of information for the young research workers and college students. Dr Krishan Lal, Deputy Director and Head, Materials Characterization Division, National Physical Laboratory, has written a preface 'a note on Professor C. V. Raman's Lecture on Fluorochromes'. He writes: "This lecture, obviously one of the last public lectures by Sir C. V. Raman, gives the flavour of the passion with which this great scientist pursued science all his life. It also gives a glimpse of some prominent facets of his great personality. He was deeply concerned with scientific education in the country, not only at higher levels but also at the grass-root level. His commitment to science, to truth, to the solid factural observations, is amply reflected all through. He extolled the fellow scientists to shed all lethargy and pursue science vigorously."

Individuals or Institutions interested in obtaining a copy (free of cost) may write to: Dr Krishan Lal, Deputy Director, National Physical Laboratory, New Delhi 110 012.

NOBEL PRIZES

Ten distinguished persons in the fields of Medicine, Economics, Physics, Chemistry and Literature were awarded the 1988 Nobel Prizes.

Three American professors, Leon Ledermann, Melvin Schwartz and Jack Steinberger shared the prize for Physics.

The prize for Chemistry went to three Federal German researchers, Johann Deisenhofer, Hartmut Michel and Robert Huber.

Sir James Black of Britain and Gertrude Elion of the United States won the prize for Medicine.

The prize for Economics was awarded to French economist Maurice Allais.

The winner of the literature prize is Nagib Mahfouz of Egypt.

The awards were presented by King Carl XVI Gustav of Sweden on Saturday the 10th December 1988.

The winners receive 2.5 million Swedish kroner each, diplomas, and gold medals bearing Nobel's image.