PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

DNA and the Neutral Theory

M. Kimura

Phil. Trans. R. Soc. Lond. B 1986 **312**, 343-354
doi: 10.1098/rstb.1986.0012

Phil. Trans. R. Soc. Lond. B **312**, 343–354 (1986)
Printed in Great Britain

DNA and the neutral theory

By M. Kimura
National Institute of Genetics, Mishima, 411 Japan

The neutral theory claims that the great majority of evolutionary changes at the molecular (DNA) level are caused not by Darwinian selection but by random fixation of selectively neutral or nearly neutral mutants. The theory also asserts that the majority of protein and DNA polymorphisms are selectively neutral and that they are maintained in the species by mutational input balanced by random extinction. In conjunction with diffusion models (the stochastic theory) of gene frequencies in finite populations, it treats these phenomena in quantitative terms based on actual observations.

Although the theory has been strongly criticized by the 'selectionists', supporting evidence has accumulated over the years. Particularly, the recent outburst of DNA sequence data lends strong support to the theory both with respect to evolutionary base substitutions and DNA polymorphism, including rapid evolutionary base substitutions in pseudogenes. In addition, the observed pattern of synonymous codon choice can now be readily explained in the framework of this theory. I review these recent findings in the light of the neutral theory.

Introduction

When the neutral theory of molecular evolution was proposed 17 years ago (Kimura 1968*a*), DNA data did not exist. All that was available was a small number of protein sequences, mostly those of haemoglobins and cytochrome *c*. In addition, electrophoretic data on enzyme polymorphism started to become available for a few organisms including the human and a species of fruit fly, which stimulated population geneticists to ponder over their implications.

Since then, enormous amounts of data have accumulated. Especially, during the last few years, we have witnessed an outburst of DNA sequence data. We are now much better informed about the rate and pattern of evolution, and the amount of intraspecific variability at the molecular level. As a result, we can make a much more detailed examination of the neutral theory, particularly with respect to the issues on which the so-called 'neutralist–selectionist controversy' has centred.

The neutral theory

The neutral theory (or more precisely, the neutral-mutation–random-drift hypothesis) claims that the great majority of evolutionary changes at the molecular level are caused not by Darwinian selection acting on advantageous mutants, but by random fixation of selectively neutral or nearly neutral mutants. The theory does not deny the role of natural selection in determining the course of adaptive evolution, but it assumes that only a minute fraction of DNA changes are adaptive in nature.

The neutral theory also asserts that most of the intraspecific variability at the molecular level (including protein and DNA polymorphism) is essentially neutral, so that the majority of

polymorphic alleles are maintained in the species by the balance between mutational input and random extinction. It regards protein and DNA polymorphisms as a transient phase of molecular evolution and rejects the notion that the majority of such polymorphisms are adaptive and maintained in the species by some form of balancing selection.

As a scientific hypothesis, the neutral theory has the advantage that its underlying assumptions are sufficiently simple that its population genetical consequences can be worked out by using suitable mutational models. This enables us to examine the theory in quantitative terms by comparing theoretical predictions with observations. In this enterprise the diffusion equation method, or 'diffusion model' as it is usually called (Kimura 1964), which treats the behaviour of mutant alleles as a stochastic process, and which at one time was regarded as being too theoretical to be of actual use, has proved to be extremely useful or even indispensable. This contrasts with the traditional evolutionary theories which rely almost exclusively on verbal, qualitative arguments.

Before I proceed, I would like to present a few formulae which are useful in discussing molecular evolution and variation from the standpoint of the neutral theory (for more details, see Kimura 1983*a*). If we denote by k the rate of evolution in terms of mutant substitutions, and if we assume that this is caused by random fixation of selectively neutral mutants through random sampling drift, then

$$k = v_{\mathrm{T}} f_0, \tag{1}$$

where v_{T} is the total mutation rate and f_0 is the fraction of neutral mutants. In other words, under the neutral theory, the evolutionary rate is equal to the mutation rate to neutral alleles, provided that the same unit is used to measure both rates. Note that k represents the rate at which molecular mutants are substituted one after another in the course of evolution within the lineage. Each of these events takes a long time, that is, four times the effective population size, as shown by Kimura & Ohta (1969). Advantageous mutations may occur, but the theory assumes that they are so rare as to be negligible. In this formulation, $1 - f_0$ represents the fraction of definitely deleterious mutants which are eliminated from the population by negative selection and which do not contribute to evolution. Note that the neutral theory does not deny the occurrence of deleterious mutations. In fact, selective constraint due to such negative selection is a very important part of the neutralist explanation of some prominent features of molecular evolution, as I shall explain later.

Usually, the rate of evolution (k) is expressed by taking one year as the unit length of time, while mutation rate is often measured per generation. Accordingly, (1) may be modified so that

$$k = (v_{\mathrm{T}}/g) f_0, \tag{2}$$

where g is the generation span. In other words, v_{T}/g is the total mutation rate per year. Next, let us consider intraspecific variability, and assume that, at a particular locus (or site), there are n possible allelic states which are selectively equivalent (that is, neutral), and that mutation occurs with a equal rate in all directions. Then, at equilibrium in which the mutational input and random extinction of alleles balance each other, the average heterozygosity at this locus (or site) is

$$\bar{H}_{\mathrm{e}} = 4N_{\mathrm{e}} v_0 / \{4N_{\mathrm{e}} v_0 [n/(n-1)] + 1\}, \tag{3}$$

where N_{e} is the effective population size and v_0 is the mutation rate for neutral alleles (Kimura 1968*b*). For an individual nucleotide site, there are four DNA bases so that $n = 4$. On the other

hand, for a gene as a whole, n is so large that we may put $n = \infty$. This leads to the infinite allele model which was proposed by Kimura & Crow (1964), and which is suitable for treating enzyme polymorphisms.

At an individual nucleotide site, one of the four DNA bases is fixed for most of the time. This ensures the existence of species-specific or 'consensus' DNA sequences. Let us call a population 'monomorphic' with respect to a particular site if the total frequency of 'variant' or less frequent alleles is q or less, where q is a small positive number such as $q = 0.01$. Then the probability of polymorphism per site is

$$P_{\text{poly}} = 1 - 4C_1 q^{\alpha}, \tag{4}$$

where $\alpha = 4N_e v_{\text{nuc}}^{(0)}$, $C_1 = \Gamma(\alpha+\beta)/\{F(\alpha+1)\,\Gamma(\beta)\}$ and $\beta = \alpha/3$, in which $v_{\text{nuc}}^{(0)}$ is the neutral mutation rate per site and $\Gamma(\cdot)$ stands for the gamma function (see p. 197 of Kimura (1983*a*) for details). Since α is small, $P_{\text{poly}} = 1 - q^{\alpha}$ is sufficiently accurate for our purpose.

By comparing (3) and (4) with (1) or (2), we note that the heterozygosity and also the probability of polymorphism tend to be high at a site in which the rate of evolution is high, although the relationship involved is not a simple, linear one.

Some misunderstandings

The neutral theory has been the target of a number of criticisms based on misunderstandings, so I shall try to discuss some of them.

First of all, genes involved in neutral evolution are not necessarily functionless as mistakenly suggested by some authors. By 'neutral evolution' I mean the cumulative genetic change caused by random drift under mutational pressure. What the neutral theory assumes is that mutant forms of each gene participating in neutral evolution are *selectively nearly equivalent*, namely, they can do the job equally well in terms of survival and reproduction of individuals. The fact that the protein and RNA molecules can tolerate many component substitutions without loss of their essential function, coupled with physiological homeostasis of organisms, is important in this context. Sometimes, neutral changes are called evolutionary 'noise', but I think this is a misnomer. Just as synonyms are not noise in language, it is not proper to regard the substitution of neutral alleles simply as noise or loss of genetic information. If the variants represent amino acid changes in a protein, this means that such changes are equally acceptable for the working of the protein in the body. Furthermore, this equality need not be exact; all that is required is that the resulting difference in fitness be small, say, for example, less than $1/(2N_e)$.

It is possible, and indeed likely, that the latitude for such interchangeability will increase as functional importance of a molecule or a part of one molecule decreases, and vice versa.

Sometimes, it is remarked that neutral alleles are by definition not relevant to adaptation, and therefore not biologically very important. I think that this is too short-sighted a view. Even if the so-called neutral alleles are selectively equivalent under a prevailing set of environmental conditions of a species, it is possible that some of them, when a new environmental condition is imposed, will become selected. Experiments suggesting this possibility have been reported by Dykhuizen & Hartl (1980) who called attention to the possibility that neutral alleles have a 'latent potential for selection'. I concur with them and believe that 'neutral mutations' can be the raw material for adaptive evolution.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Synonymous and other silent substitutions

The first strong evidence for the neutral theory which emerged with the advent of DNA (or RNA) sequence data was the preponderance of synonymous change, namely, the observation that nucleotide substitutions within codons that do not cause amino acid changes occur at a much higher rate than amino-acid-altering substitutions.

A similar observation came with the discovery that most eukaryotic genes contain intervening sequences or 'introns' which are removed when the mature messenger RNA is formed and which therefore do not participate in protein formation. It was found that evolutionary nucleotide substitutions in introns are also high, with a rate comparable to the synonymous substitutions or even higher.

Since natural selection acts on the phenotype of the organism in the determination of which the structure and function of proteins play a decisive role, one should expect that silent mutations which do not cause amino acid changes in proteins, other things being equal, would be much less subject to natural selection than those that cause amino acid changes. Yet, it is the silent substitutions that really accumulate at a higher rate per site in evolution.

These observations are quite consistent with a similar observation made previously on proteins. For example, when active insulin is formed from proinsulin, the middle segment C of proinsulin is removed and discarded. It was found that for this peptide C the rate of evolution in terms of amino acid substitutions is several times as fast as that of insulin (see p. 159 of Kimura 1983*a*).

A general rule that has emerged through these observations is that *molecular changes that are less likely to be subject to natural selection occur more rapidly in evolution*. This empirical rule can readily be understood from the standpoint of the neutral theory, because such molecular changes have a higher chance of being selectively neutral (that is, a larger f_0 in (1)) and therefore neutral evolution occurs at a higher rate (that is, a larger k in the same equation).

One interesting property of synonymous substitutions is that their rates are not only high in general but also are roughly equal to each other among different molecules.

Table 1 lists some estimates of the rate of evolutionary nucleotide substitutions for three molecules, presomatotropin (pregrowth hormone), β-globin and α-tubulin (for references to original data used for these estimations, see pp. 172–174 of Kimura (1983*a*).) In this table, k_2 stands for the rate of substitution per site per year at the second position of codons, taking 10^{-9} as the unit. Also, k_S denotes the rate of synonymous substitutions per site, which was obtained by dividing the synonymous component k'_S of the substitution rate at the third codon position by $\frac{2}{3}$. (For details, see p. 175 of Kimura (1983*a*); also readers may consult Kimura (1981*a*) and Kimura (1980) for details of the statistical methods used.) Since all the nucleotide changes at the second position and also the majority of changes at the first position of codons lead to amino acid changes, $2k_2$ roughly represents the rate of amino-acid-altering substitutions. It is interesting to note that in terms of amino acid substitutions, pregrowth hormone evolves more than a hundred times faster than α-tubulin, yet for synonymous changes, the former evolves only three or four times as fast as the latter.

Figure 1 illustrates the results of more extensive studies on this subject made by Miyata (1984); it is evident that the rates of synonymous substitutions are very close to each other even among proteins that differ widely in amino acid substitution rates. This suggests that the nature of the selective constraint is different for these two types of changes.

TABLE 1. ESTIMATES OF EVOLUTIONARY RATES

protein (comparison)	evolutionary rate† k_2	evolutionary rate† k_S
presomatotropins (human versus rat)	1.13±0.19	4.13±0.66
β-globin (human versus mouse and human versus rabbit)	0.59±0.12	2.48±0.48
α-tubulin (chicken versus rat)	0.008±0.005	1.18±0.13

† Units, 10^{-9} per site per year.

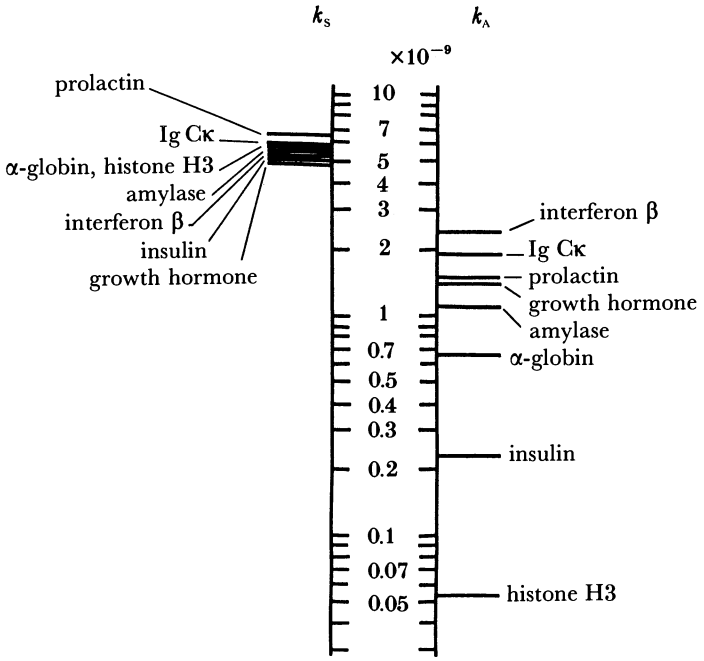

FIGURE 1. Distribution of synonymous substitution rates (k_S) contrasted with that of amino acid-altering substitution rates (k_A) for several protein genes (adapted from Miyata (1984) with minor modification).

A preponderance of synonymous and other silent substitutions over amino-acid-altering ones appears to be a general rule in evolution; it holds even in the unusual situation described below, where the rate of evolution is speeded up about a million times.

Recently, Hayashida *et al.* (1985) investigated the evolution of influenza A virus genes (consisting of single-stranded RNA). These viruses are responsible for pandemic influenza, and they can undergo radical antigenic changes. The authors compared nucleotide sequences of homologous genes among strains that were isolated in different years, and found that the genes evolve at extremely high rates in clock-like fashion. The average rate of silent substitutions estimated was 1.1×10^{-2} per site per year, yet, compared with the corresponding rate of amino acid altering substitutions, it is only four or five times higher in this case. It is known that an outburst of pandemic influenza is caused by antigenic shift, but the number of amino acid sites responsible for surface antigens is quite limited. So, even in this case, neutral evolution predominates over adaptive evolution.

Gojobori & Yokoyama (1985) estimated the evolutionary rates of both the retroviral oncogene of Moloney murine sarcoma virus (*v-mos*^{MO} gene), and its cellular homologue (*c-mos*^{MO}

gene). They found that the former evolves nearly 0.8 million times faster than the latter. For the former (*v-mos*^{MO}), the rate of nucleotide substitutions at the first, second and third codon positions are respectively 1.31×10^{-3}, 0.56×10^{-3} and 2.06×10^{-3} per site per year. These show clearly that even in this case the rate of nucleotide substitution is highest at the third position, indicating the preponderance of synonymous changes.

These examples show that neutral evolution can occur in RNA viruses at extraordinarily high rates. The rapid evolution is caused by correspondingly high mutation rates due to a lack of the proofreading enzymes that ensure accurate replication, as pointed out by Gojobori & Yokoyama (1985), coupled with the very high replication rates of these viruses.

Evolutionary change in pseudogenes

The rapid evolutionary change observed in pseudogenes provided very strong support for the neutral theory in a previously unexpected way. Since I reviewed this topic extensively in my book (see pp. 178–183 of Kimura 1983*a*), I shall mention here only a few salient features. Generally speaking, a pseudogene is a region of DNA that shows definite homology with a known functional gene but has lost its ability to produce a functional product. It is sometimes called a 'dead gene'.

Comparison of pseudo globin genes with their normal counterparts revealed that base substitutions occurred at very high rates in the pseudogenes (Miyata & Yasunaga 1981; Li *et al.* 1981). What is really interesting is that the rates of substitutions are equally high in all three codon positions. According to Li *et al.* (1981), the estimated rate of substitutions in globin pseudogenes is $k_0 = 4.6 \times 10^{-9}$ per site per year. On the other hand, in normal globin genes, the estimated rates at the first, second and third codon positions are $k_1 = 0.71 \times 10^{-9}$, $k_2 = 0.62 \times 10^{-9}$ and $k_3 = 2.64 \times 10^{-9}$ respectively. In addition to base substitutions, pseudogenes accumulate deletions and additions also at very high rates in the course of evolution. It looks as if the pseudogenes have been liberated from the constraint of negative selection and are on the way to disintegration by accumulating various mutational changes (some of which must be highly damaging to normal globins) at the maximum speed allowable under mutational pressure and random drift. If this interpretation is correct, we may assume $f_0 = 1$ for pseudoglobins so that $k_0 = 4.6 \times 10^{-9}$ is equal to v_T/g in (2). Then, we can obtain a rough estimate of the fraction of neutral mutations among newly arisen amino-acid-altering mutations in normal haemoglobins by the ratio $(k_1 + k_2)/(2k_0)$, which turns out to be about 0.14. Similarly, we can estimate the fraction of neutral mutations among electrophoretically detectable changes, and this gives $f_{0(E)} \approx 0.14$ (Kimura 1983*b*). Since the amino acid substitution rate ($k_{aa} \approx 1.2 \times 10^{-9}$) of haemoglobins is very close to the median value of amino acid substitution rates of various proteins (Kimura 1974), we may regard these f_0 values as the representative values for the fractions of neutral mutations (P_{neut}) among coding loci in the mammalian genome.

It is reassurig that $f_{0(E)} \approx 0.14$ thus obtained agrees quite well with the corresponding value ($P_{neut} = 0.14 \pm 0.06$) obtained by using data on the distribution of rare variant alleles together with that of polymorphic alleles which were detected by the electrophoretic method (Kimura 1983*b*).

Compared with protein polymorphism, data on DNA sequence polymorphism of nuclear genes are still very scanty. However, the study of Kreitman (1983) who determined DNA sequences of 11 cloned alcohol dehydrogenase (*Adh*) genes in *D. melanogaster*, together with

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Bodmer & Ashburner's (1984) study on *Adh* sequences in *D. simulans* and a few other sibling species, show that synonymous (or silent) sites are much more polymorphic and evolve much faster than non-synonymous sites, in qualitative agreement with the neutral theory. A more recent and extensive study by Aquadro *et al.* (1985) on DNA sequence variation around the *Adh* gene region in natural populations of *D. melanogaster* revealed extensive variation due to base substitutions, insertions and deletions. Of particular interest is their finding that length variation due to unique sequence insertions or deletions and transposable element insertions are very common in this species.

Unequal usage of synonymous codons as one aspect of selective constraint

It is now well known that synonymous codons are used quite unequally or in 'non-random' fashion in many genes of various organisms (see, for example, Grantham 1980; Grantham *et al.* 1980). In fact, non-random usage is the rule rather than the exception, and often this has been (and still is) quoted as evidence against the neutrality of synonymous base substitutions in evolution. At the beginning, this seemed to be a very puzzling and troublesome problem for the neutralists. Fortunately, however, it has become clear, mainly due to a series of papers by Ikemura (1980, 1981*a*,*b*; see also Ikemura (1985) for review), that the unequal usage of synonymous codons is a result of selective constraint mainly caused by unequal availability of the cognate tRNA species in the cell at least in unicellular organisms such as *Escherichia coli* and yeast. He made the important discovery that among synonymous codons for an amino acid, the most frequently used codon invariably corresponds to the most abundant isoaccepting tRNA species. For example, there are six codons coding for leucine, but in *E. coli*, the codon CUG is used most frequently. This matches with the observation that among the cognate tRNA species for leucine in this organism, the one ($tRNA_1^{Leu}$) that recognizes this particular codon is the most abundant one.

On the other hand, in yeast, the codon UUG is used most frequently, and, at the same time the tRNA recognizing this codon is much more abundant than other tRNA species for leucine. The observation that the synonymous codon choice-pattern is similar among different genes within an organism also adds support to Ikemura's theory.

An additional rule that has emerged through recent studies by Ikemura and others is that more highly expressed genes tend to show stronger choice pattern among synonymous codons. In fact, Ikemura (1985) showed that a very strong correlation exists between the number of molecules per cell and the F_{op} value, namely the fraction by which the 'optimal' codon (corresponding to the most abundant cognate tRNA species) is used.

These observations can readily be understood on the neutral theory by noting that in evolution synonymous substitutions are constrained in such a way that they do not deviate much from the established pattern of relative availability of isoaccepting tRNA species, since otherwise translational efficiency would be reduced in the cell thereby causing loss of Darwinian fitness. Such a constraint (negative selection) must be stronger for more highly expressed genes which must be more directly related to the survival and reproduction of individuals (or cells) than less highly expressed genes. Also, it has been found that essentially the same codon choice pattern is shared by a very wide group of related organisms. For example, the coding 'dialect' of *E. coli* is similar to those of other Enterobacteriacae (Ikemura 1985).

I want to bring out one more observation that is highly pertinent to the neutralist–selectionist

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

controversy. As shown by Miyata (1982) and Ikemura (1985), and as exemplified in table 2, *stronger choice among synonymous codons tends to slow down rather than accelerate the synonymous base substitutions in evolution*. If synonymous substitutions were caused by positive Darwinian selection, one should expect that the stronger choice would accelerate evolution rather than retard it, but the actual observations are the other way around.

TABLE 2. CODON CHOICE PATTERN AND EVOLUTIONARY RATE

molecule (protein)	number of molecules†	F_{op}	evolutionary distance K_S	evolutionary distance K_{aa}
omp A	3×10^4	0.92	0.18	0.07
trp A	under 1000	0.61	1.34	0.15

† Number of protein molecules per genome for omp A has been measured in *E. coli* (Ikemura 1985), while that for trp A is an inferred value (T. Ikemura, personal communication).

Abbreviations: omp A, outer membrane major protein; trp A, tryptophan synthetase α; F_{op}, relative frequency of use of optimal codons (data, from Ikemura 1985); K_S, number of synonymous substitutions per nucleotide site; K_{aa}, number of amino acid substitutions per codon.

Evolutionary distances are estimated by comparing DNA sequences between *Escherichia coli* and *Salmonella typhimurium*. Data for trp A are from Nichols & Yanofsky (1979), and those for omp A are taken from Ikemura (1985).

FEATURES OF MOLECULAR EVOLUTION CONTRASTED WITH THOSE OF PHENOTYPIC EVOLUTION

There are at least two features that distinguish molecular evolution from phenotypic evolution (table 3). Molecular evolution is characterized by (i) constancy in rate and (ii) conservatism in mode (see Kimura (1983*a*) for review), while phenotypic evolution exhibits (i) irregularity in rate and (ii) opportunism in mode (see, for example, Simpson 1949). These features can readily be understood if we assume that phenotypic evolution is largely controlled by positive Darwinian selection that brings about adaptation of organisms to their environment, while molecular evolution is mainly caused by random fixation of selectively neutral or nearly neutral mutants under mutation pressure.

TABLE 3. CONTRAST BETWEEN MOLECULAR AND PHENOTYIC EVOLUTION

type of evolution	rate	mode
molecular	constant	conservative
phenotypic	irregular	opportunistic

As to the first feature of molecular evolution, enough evidence has now accumulated indicating that the rate of evolution in terms of nucleotide substitutions is approximately constant among lines for a given type of gene or DNA region. I believe that the usefulness of the rate constancy hypothesis (or 'molecular clock' concept) in constructing phylogenetic trees attests its validity. The fact that the constancy is in terms of physical time (year) even among organisms with very different generation spans has often been cited as evidence against the neutral theory, because constancy of v_T/g in (2) appears to be inconsistent with observations. In fact, traditional studies of mutations on visible and viability traits (including lethals) strongly suggest that the spontaneous mutation rate per generation, but not per year, is roughly equal

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

among different animals (such as *Drosophila*, mouse and man) whose generation spans are very different. It now appears, however, that many of these 'mutations' are caused or controlled by transposons and insertion sequences (see, for example, Rubin 1983; Mukai & Yukuhiro 1983; Mackay 1984). On the other hand, no definite data are available at present to settle the issue whether the mutation rate for nucleotide substitutions (with which the neutral theory is concerned) is proportional to year or generation. Experimental studies on this subject are much needed.

The conservative nature of molecular evolution has now been well established; those mutant substitutions that are less disruptive to the existing structure and function of a molecule occur more frequently in evolution than more disruptive ones. This is easy to understand from the neutral theory, because the more conservative the mutational change, the more likely it is to turn out to be selectively neutral.

Some years back, we (Kimura & Ohta 1974) enumerated five principles which govern molecular evolution, including constancy in rate and conservatism in mode. One of the principles states that functionally less important molecules or parts of a molecule evolve (in terms of mutant substitutions) faster than more important ones. This too can readily be understood by the neutral theory, because the fraction of neutral mutations must be higher in less important molecules or parts of one molecule, that is, f_0 in (1) becomes larger in them.

When this principle, accompanied by its neutralist explanation, was first proposed, much opposition was voiced by 'selectionists', but I am glad to say that it has become a part of common sense among molecular biologists. It is now a routine practice among them, to search for various signals by comparing a relevant region of homologous DNA sequences of diverse organisms and to pick out a constant (consensus) pattern, but disregard variable ones as unimportant.

Concluding remarks

Recent data on DNA sequences have strongly vindicated the neutral theory: I believe that evidence is now overwhelming that, at the molecular level, neutral evolution predominate over Darwinian evolution. Then, the question that immediately arises is why random fixation of selectively neutral or nearly neutral mutants prevails at the molecular level, even if Darwinian evolution by positive natural selection appears to be so important at the phenotypic level.

The answer to this question, I think, comes from the fact that the most common type of natural selection at the phenotypic level is stabilizing selection, to use Mather's (1953) teminology. Here it is important to note that natural selection acts directly on phenotype, but only secondarily on the molecular constitution of genes.

Unlike the type of natural selection which Darwin had in mind when he tried to explain evolution through the accumulation of small beneficial changes, stabilizing selection eliminates phenotypically extreme individuals and preserves those near the population mean. It acts to keep the *status quo*, rather than to produce a directional change. It was shown by Wright (1935) and Robertson (1956) that if genes are additive with respect to a quantitative character which is subject to stabilizing selection and if the mean and the optimum coincide, then the alleles involved behave as if negatively overdominant. Pursuing this problem further, I have shown (Kimura 1981*b*) that extensive neutral evolution can occur under stabilizing selection if a large number of loci or sites are involved in a quantitative character. This applies, for example, to the situation in which each individual in a mammalian species is heterozygous on the average

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

for one million nucleotide sites and the total selection intensity per individual is 50%. I believe that the concept of 'random drift under stabilizing selection' will help the selectionist camp to appreciate the merit of the neutral theory in understanding the mechanism of evolution and intraspecific variability at the molecular level. A similar view was presented by Milkman (1982) under the designation 'a unified selection theory'. More recently, he gave an interesting personal account on 'the rational resolution of the selectionist–neutralist controversy' along the lines I have outlined above (see pp. 328–334 of Milkman 1983).

Note that the selection involved here is not 'balancing selection' as routinely invoked by the selectionists. In fact, careful, large scale experiments by Mukai and his associates, using *Drosophila melanogaster* have produced no evidence for the three types of balancing selection, that is, overdominance, frequency-dependent selection and diversifying selection at work in protein polymorphism (see Mukai *et al.* (1982) for review).

I must add here that there is a possibility that a certain fraction of nucleotide sites (presumably a large fraction) produce no phenotypic effects at all, and therefore are completely neutral with respect to natural selection. On the other hand, a certain fraction (probably a very small fraction) of nucleotide or amino acid substitutions are definitely advantageous for the species in adapting to new environments, and therefore they are subjected to straightforward positive natural selection.

Recently, Perutz (1983) has made a detailed stereochemical examination of amino acid substitutions among vertebrate haemoglobins in relation to species adaptation. He concluded that adaptations leading to response to new chemical stimuli have evolved by only a few (one to five) amino acid substitutions in key positions, while most of the amino acid replacements between species are functionally neutral. He says that the evidence supports my neutral theory.

For me, a really encouraging aspect of the neutral theory is that its position becomes stronger with increasing data.

I would like to close my talk by quoting from Haldane (1959) who, a quarter of a century ago, wrote as follows in discussing Darwin's theory of evolution by natural selection.

> 'The history of science makes it almost certain that facts will be discovered which show that the theory of natural selection is not fully adequate to account for evolution. But the same history makes it extremely improbable that these facts will be in any way related to the criticisms at present made of it. The physics of Newton and Galileo have proved inadequate in several respects, and are being replaced by relativistic and quantum mechanics. These, however, are even further from the medieval physical theories than were the theories of Galileo and Newton. They were discovered because when the consequences of Newtonian physics were fully worked out, certain facts disagreed with them. It was not possible in Newton's time to guess at these discrepancies,.... Darwinism will, I do not doubt, be modified. Like any other successful theory it will ultimately develop its own internal contradictions.'

Contribution number 1615 from the National Institute of Genetics, Mishima, Shizuoka-Ken, 411 Japan.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

References

Aquadro, C. F., Deese, S. F., Bland, M. M., Langley, C. H. & Laurie-Ahlberg, C. C. 1985 Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. *Genetics* (In the press.)

Bodmer, M. & Ashburner, M. 1984 Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. *Nature, Lond.* **309**, 425–430.

Dykhuizen, D. & Hartl, D. L. 1980 Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. *Genetics* **96**, 801–817.

Gojobori, T. & Yokoyama, S. 1985 Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4198–4201.

Grantham, R. 1980 Workings of the genetic code. *Trends biochem. Sci.* **5**, 327–331.

Grantham, R., Gautier, C. & Gouy, M. 1980 Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucl. Acids Res.* **8**, 1893–1912.

Haldane, J. B. S. 1959 Natural selection. In *Darwin's biological work* (ed. P. R. Bell), pp. 101–149. Cambridge University Press.

Hayashida, H., Toh, H., Kikuno, R. & Miyata, T. 1985 Evolution of influenza virus genes. *Molec. Biol. Evol.* **2**, 289–303.

Ikemura, T. 1980 The frequency of codon usage in *E. coli* genes: correlation with abundance of cognate tRNA. In *Genetics and evolution of RNA polymerase, tRNA and ribosomes* (ed. S. Osawa, H. Ozeki, H. Uchida & T. Yura), pp. 519–523. University of Tokyo Press.

Ikemura, T. 1981*a* Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. molec. Biol.* **146**, 1–21.

Ikemura, T. 1981*b* Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. molec. Biol.* **151**, 389–409.

Ikemura, T. 1985 Codon usage and tRNA content in unicellular and multicellular organisms. *Molec. Biol. Evol.* **2**, 13–34.

Kimura, M. 1964 Diffusion models in population genetics. *J. appl. Prob.* **1**, 177–232.

Kimura, M. 1968*a* Evolutionary rate at the molecular level. *Nature, Lond.* **217**, 624–626.

Kimura, M. 1968*b* Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* **11**, 247–269.

Kimura, M. 1974 Gene pool of higher organisms as a product of evolution. *Cold Spring Harb. Symp. quant. Biol.* **38**, 515–524.

Kimura, M. 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. molec. Evol.* **16**, 111–120.

Kimura, M. 1981*a* Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. natn. Acad. Sci. U.S.A.* **78**, 454–458.

Kimura, M. 1981*b* Possibility of extensive neutral evolution under stabilizing selection with special reference to non-random usage of synonymous codons. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5773–5777.

Kimura, M. 1983*a* *The neutral theory of molecular evolution*. Cambridge University Press.

Kimura, M. 1983*b* Rare variant alleles in the light of the neutral theory. *Molec. Biol. Evol.* **1**, 84–93.

Kimura, M. & Crow, J. F. 1964 The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

Kimura, M. & Ohta, T. 1969 The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771.

Kimura, M. & Ohta, T. 1974 On some principles governing molecular evolution. *Proc. natn. Acad. Sci. U.S.A.* **71**, 2848–2852.

Kreitman, M. 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature, Lond.* **304**, 412–417.

Li, W.-H., Gojobori, T. & Nei, M. 1981 Pseudogenes as a paradigm of neutral evolution. *Nature, Lond.* **292**, 237–239.

Mackay, T. F. C. 1984 Jumping genes meet abdominal bristles: hybrid dysgenesis-induced quantitative variation in *Drosophila melanogaster*. *Genet. Res.* **44**, 231–237.

Mather, K. 1953 The genetical structure of populations. In *Symp. Soc. Exp. Biol, VII, Evolution*, pp. 66–95. Cambridge University Press.

Milkman, R. 1982 Toward a unified selection theory. In *Perspectives on evolution* (ed. R. Milkman), pp. 105-118. Sunderland, Massachusetts: Sinauer Associates.

Milkman, R. (ed.) 1983 *Experimental population genetics*, Benchmark Papers in Genetics, 13. Stroudsburg, Pennsylvania: Hutchinson Ross.

Miyata, T. 1982 Evolutionary changes and functional constraints in DNA sequences. In *Molecular evolution, protein polymorphism and the neutral theory* (ed. M. Kimura), pp. 233–266. Tokyo: Japan Scientific Societies Press. Berlin: Springer-Verlag.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Miyata, T. 1984 Evolution of DNA; dynamically evolving eukaryotic genes. In *Introduction to molecular evolutionary study* (ed. M. Kimura), pp. 56–90. Tokyo: Baifukan. [In Japanese.]

Miyata, T. & Yasunaga, T. 1981 Rapidly evolving mouse α-globin-related pseudogene and its evolutionary history. *Proc. natn. Acad. Sci. U.S.A.* **78**, 450–453.

Mukai, T., Yamaguchi, O., Kusakabe, S., Tachida, H., Matsuda, M., Ichinose, M. & Yoshimaru, H. 1982 Lack of balancing selection for protein polymorphisms. In *Molecular evolution, protein polymorphism and the neutral theory* (ed. M. Kimura), pp. 81–120. Tokyo: Japan Scientific Societies Press. Berlin: Springer-Verlag.

Mukai, T. & Yukuhiro, K. 1983 An extremely high rate of deleterious viability mutations in *Drosophila* possibly caused by transposons in non-coding regions. *Proc. Japan Acad.* B **59**, 316–319.

Nichols, B. P. & Yanofsky, C. 1979 Nucleotide sequences of *trp A* of *Salmonella typhimurium* and *Escherichia coli*: An evolutionary comparison. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5244–5248.

Perutz, M. F. 1983 Species adaptation in a protein molecule. *Molec. Biol. Evol.* **1**, 1–28.

Robertson, A. 1956 The effect of selection against extreme deviants based on deviation or on homozygosis. *J. Genet.* **54**, 236–248.

Rubin, G. M. 1983 Dispersed repetitive DNAs in *Drosophila*. In *Mobile genetic elements* (ed. J. A. Shapiro), pp. 329–361. New York: Academic Press.

Simpson, G. G. 1949 *The meaning of evolution*. Yale University Press.

Wright, S. 1935 The analysis of variance and the correlations between relatives with respect to deviations from an optimum. *J. Genet.* **30**, 243–256.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES