Molecular Evolution: Nearly Neutral Theory

Tomoko Ohta, *National Institute of Genetics, Mishima, Japan*

Nearly neutral theory is an extension of the neutral theory and contends that the borderline mutations, whose effects lie between the selected and the neutral classes, are important at the molecular level.

Advanced article

Article Contents

doi: 10.1002/9780470015902.a0001801.pub3

Neutralism and the Nearly Neutral Theory

Comparative studies of deoxyribonucleic acid (DNA) and protein sequences have revealed two important characteristics of molecular evolution. One is the apparent uniformity of evolutionary rate; for example, if one compares human haemoglobin α with that of gorilla, horse and carp, one realizes that the difference almost linearly increases with the divergence time. Another characteristic is that the stronger the constraint of a molecule, the slower its evolution. This may be seen by the slowest rate of evolution of histone IV and the fastest rate of fibrinopeptides. Selective constraint comes from the requirements for proteins to retain their structure and function. For fibrinopeptides, the constraint is weak and almost any amino acid substitution is acceptable in evolution; however, for histone IV, the constraint is so strong that almost no amino acid substitution is accepted. The level of constraints of other proteins usually fits between these two extremes. DNA evolution also shows this pattern, such that noncoding regions are changing rapidly compared with coding regions. These characteristics were not thought to be in accord with the neo-Darwinian theory, but may be explained by the neutral theory, which predicts that rates of change are a simple function of the neutral mutation rate (Kimura, 1968, 1983; King and Jukes, 1969). **See also**: Functional Constraint and Molecular Evolution; Molecular Clocks; Molecular Evolution: Introduction; Molecular Evolution: Neutral Theory

To explain the relationship between the evolutionary rate and the constraint, the neutral theory assumes that a certain fraction of new mutations are free of constraint (selectively neutral), while the rest have deleterious effects and are eliminated from the population. However, natural selection cannot be so simple as to be all or nothing, and the nearly neutral theory contends that the borderline mutations, whose effects lie between the selected and the neutral classes, are important at the molecular level (Ohta, 1973). Their fate is influenced by both random genetic drift and selection. **Figure 1** compares the selection, neutral and nearly neutral theories with regard to how new mutations are classified. **See also**: Drift: Theoretical Aspects; Mutations and the Genetic Code; Mutations and New Variation: Overview; Mutation–Selection Balance

Influence of Population Size on the Basic Results of Neutralism

A critical quantity in the discussion of molecular evolution is the fixation probability of mutant genes in the population, because the evolutionary rate of a protein or a gene becomes the product, mutation rate times fixation probability. 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 gene. Then, in a population of N individuals, the total number of mutations appearing in the population is $2Nv_g$ per generation; a fraction u of them fix in the population, and the rate of substitution per generation becomes

$$k_g = 2Nv_g u \quad [1]$$

Here, u depends on the magnitude of natural selection. For selectively neutral mutants, u is equal to the initial frequency, $1/(2N)$, and we have

$$k_g = v_g \quad [2]$$

Let us examine the fixation probability of the nearly neutral mutations. The simplest case is the semidominant gene with selective advantage, s, that may be positive or negative. Fixation probability is a monotonically increasing function of the product, $2Ns$, given in **Figure 2**, and the effectiveness of selection is determined by this product. **See also**: Population Genetics: Overview

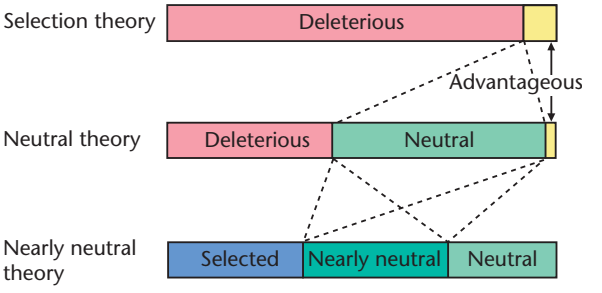

Figure 1 The classification of new mutants under the selection, neutral and nearly neutral theories. Note that while most selected mutants are deleterious, the group also includes advantageous mutants.

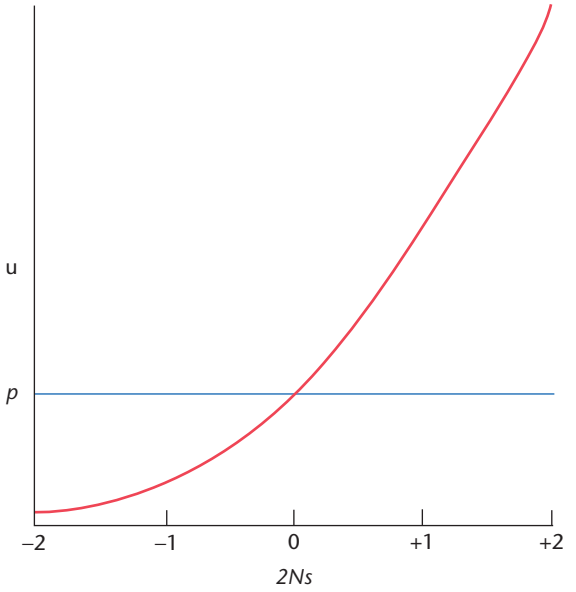

Figure 2 Fixation probability of a mutant in a finite population as a function of $2Ns$. p is the initial frequency of the mutant. The region of $2Ns < 0$ is that of slightly deleterious mutations.

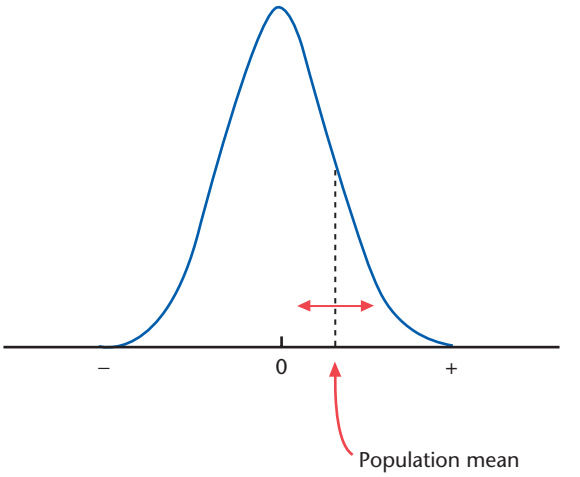

Figure 3 Hypothetical distribution of selection coefficient of new mutations with small effects. The population mean moves to the left or to the right by selection and drift. If selection is strong enough, the population mean moves to the right and most new mutants become deleterious. By drift, the population mean moves to the left and some mutants become advantageous (from Ohta, 1992).

In view of the importance of negative selection caused by constraints, it is likely that many nearly neutral mutants are very slightly deleterious, i.e. on the left side of $2Ns = 0$ in **Figure 2**. For such mutants, there is a negative correlation between the fixation probability and population size, provided that the selection coefficient is unchanged by population size (Ohta, 1973). Note that, due to the small absolute value of $2Ns$ in small populations, drift swamps the weak selection pressure against the nearly neutral allele, so that the allele is effectively neutral. However, in large populations, selection prevails over drift, and the fate of the allele will be determined by its deleterious effect on the phenotypes. Therefore, the chance of spreading by random drift is much higher in a small population than in a large population. The prediction may be approximately written as follows:

$$k_g \alpha v_g / N \qquad [3]$$

The nearly neutral theory has been much criticized on the ground that it predicts continuous deterioration. However note that, as Sella and Hirsh (2005) have clearly stated, when equilibrium is reached, and mutation, drift and selection balance each other, the number of slightly deleterious mutant substitutions is equal to that of slightly advantageous substitutions.

To be more quantitative, several models of nearly neutral mutations have been studied. It is convenient to assume the distribution of the mutants' effect around neutrality for the analyses. There are mainly two approaches: the shift model and the fixed model. The former is based on the assumption that the distribution remains the same when a mutant substitutes the previous allele, as the population mean shifts back to the original state. However, in the fixed model, the distribution is fixed irrespective of mutant substitutions, and the population mean is adjusted according to the effect of the fixed mutant. Here, the effect of each substitution remains and affects subsequent substitutions. Therefore, substitutions are interrelated in their effects on fitness. The best-studied case of the fixed model is to assume the normal distribution for the effects of nearly neutral mutations. **Figure 3** shows the distribution of the selection coefficients of new mutants around the population mean. By selection, the mean moves towards the right, whereas random drift brings about erratic movement. The effectiveness of selection is again determined by the product of population size and selection intensity that is measured by the standard deviation, σ_s, of the normal distribution. For nearly neutral mutations, $N\sigma_s$ is less than unity, where both random drift and selection affect the mutants' behaviour. Although the shift model and the fixed one are different, the evolutionary pattern becomes similar in both models, i.e. the negative correlation between the evolutionary rate and the population size, as in eqn [3], is predicted in both models. **See also**: Neutrality and Selection in Molecular Evolution: Statistical Tests

Levels of Heterozygosity

If mutants are completely neutral, the level of polymorphism is in equilibrium between mutational input and loss by random drift in a stable population. The expected heterozygosity in equilibrium is

$$H = \frac{4Nv_g}{1 + 4Nv_g} \qquad [4]$$

(Kimura and Crow, 1964). The formula is robust and applicable to many cases, such as protein loci, DNA segments and single nucleotide sites.

In the 1970s, data on protein polymorphisms measured by electrophoresis had accumulated, and the narrow range of observed heterozygosity of various species was thought not to be in accord with the neutral prediction. For example, the heterozygosity is often around 12% in *Drosophila* species and 5–6% in mammalian species. Then, Nv_g is predicted to be 0.035 for *Drosophila* and 0.015 in mammals if protein polymorphisms are neutral. How is such a narrow range possible? It may be accounted for by the nearly neutral theory, as the deleterious effects of some variants prevent the increase of heterozygosity even when the population size becomes larger (Ohta, 1974).

Data on DNA sequence polymorphisms are now available. As expected from the neutral theory, polymorphisms in noncoding regions or at synonymous sites are higher than at amino acid replacing sites. In other words, similar to the difference of evolutionary rates, polymorphisms show less heterozygosity when the constraints are stronger. Detailed examination has, however, revealed some intriguing patterns; for example, in contrast to similar levels of heterozygosity of proteins between *Drosophila melanogaster* and *Drosophila simulans*, the level of DNA polymorphism is higher in the latter than in the former. In addition, the proportion of replacement polymorphisms in the coding region is higher in *D. melanogaster* than in *D. simulans*. These facts may reflect weak and context dependent selection as discussed later. Note that silent sites or noncoding regions may not be strictly neutral, as by the weak selection indicated on codon bias as a result of the relative abundance of different transfer ribonucleic acids (tRNAs) (Akashi, 1995). **See also**: Codon Usage in Molecular Evolution

Generation Time Effect

In the 1970s, the data suggested that the rate of protein evolution seemed to be constant per year, whereas genome divergence measured by DNA hybridization appeared to be constant per generation. If the neutral theory (eqn [2]) is applicable, such a generation time effect should also be observed for protein evolution. In the 1980s, comparative studies of DNA sequences revealed the generation time effect for synonymous substitutions but not for amino acid substitutions (Li *et al*., 1987).

This seemingly contradictory observation is explained by the nearly neutral theory, such that most substitutions in the genomic DNA of higher organisms are neutral but that amino acid substitutions are nearly neutral (Ohta, 1973). As explained before, there is a negative correlation between population size and the rate of mutant substitution for nearly neutral mutations. Equation [3] is now modified to measure the rate per year,

$$k\alpha v_g/(Ng) \qquad [5]$$

where k is the evolutionary rate per year and g is the generation time in years. In general, large organisms have a long generation time and small population size, and vice versa. Therefore, N and g in the denominator tend to cancel each other, resulting in rough constancy.

This prediction was tested by comparing patterns of synonymous and nonsynonymous substitutions in 49 gene sequences of three orders, primates, artiodactyls and rodents (Ohta, 1995). The results indicated that the generation time effect is more conspicuous for synonymous substitutions than for nonsynonymous substitutions. In other words, the rodent branch is much longer than the primate branch for synonymous substitutions, but the difference in the two branches is not so large for nonsynonymous substitutions. Primates generally have longer generation times than rodents, and the difference in the patterns of the two types of substitutions is consistent with the nearly neutral theory. The prediction has been verified by using a large dataset of primates and rodents (Chimpanzee Sequencing and Analyses Consortium, 2005). **See also**: Molecular Evolution: Patterns and Rates

Comparing Variation Within and Between Species

DNA sequence data from samples within populations accumulated in the 1990s. Through comparative analyses of DNA sequences, it has become possible to estimate the number of nonsynonymous changes (supposedly selected) separately from that of synonymous changes (supposedly neutral). Starting with the work of McDonald and Kreitman (1991), many reports on synonymous and nonsynonymous polymorphisms have been published. A commonly used test compares the relative numbers of synonymous and nonsynonymous substitutions within a population with those between closely related species. Under the neutral theory, which assumes neutrality of both synonymous and nonsynonymous changes, the relative numbers of the two types of changes should be the same when measured within a population and when measured between closely related species. Departure from the neutral prediction was often observed. In some cases, excess of nonsynonymous differences was found for between-species comparisons, whereas the same excess was observed for within-species comparisons in other studies. It has usually been argued that advantageous mutant substitutions were responsible for the former and deleterious ones for the latter. **See also**: Molecular Evolution; Molecular Evolution: Rates; Molecular Evolution: Techniques; Variation, Within Species: Introduction

Sawyer *et al*. (2007) extended the analysis to X chromosome data of *D. melanogaster* and *D. simulans*. Polymorphisms in 91 genes in African populations of *D. melanogaster* and their divergence from *D. simulans* were analysed. Through sophisticated population genetic analyses with maximum likelihood, they have found that about 70% of amino acid polymorphisms are slightly deleterious, but about 95% of fixed differences are positively selected. They have also estimated that about half of new amino acid

changes have very small selection coefficient such that both drift and selection influence their behaviour, i.e. nearly neutral. Note that a large fraction of positively selected amino acid substitutions are also very weakly selected.

By using a similar data set of *Drosophila*, but with a very different model in which fluctuation of selection coefficient was assumed, the intensity of positive selection was estimated much larger than the earlier result (Mustonen and Lässig, 2007). The adequacy of the model needs to be examined.

It has been known that the variance of amino acid substitution rate is often larger than expected from the simple Poisson process. There exist numerous interactions among amino acids to fold a protein, and for it to function. If an amino acid substitution spreads in a population by drift and slightly disturbs this interacting system, compensatory substitutions would become slightly advantageous. Such substitution processes inflate the variance of evolutionary rates. Interacting systems at higher levels, such as among proteins and nucleic acids, may also influence the selective value of individual amino acid substitutions. In fact, some of the peculiar patterns observed may be caused by the shifting of such interactive systems. Further study is needed for an exact understanding of the nature of higher-order interactions.

It should also be recognized that genetic systems are robust at various levels, i.e. effects of mutations are often buffered or even silenced (Wagner, 2005). Some of the mechanisms of robustness have now been reported such as involvements of microRNA and heat shock protein (HSP) 90 in genetic pathways (Rutherford and Lindquist, 1998; Hornstein and Shomron, 2006). The range of near-neutrality may greatly increase by such mechanisms. These mechanisms would depend on genetic as well as on environmental conditions, and the selective force would be context dependent.

Is it possible to discriminate adaptive and nonadaptive changes from the observed patterns? It seems that once gene function attains a state sufficiently close to an optimum, genes are expected to evolve by nearly neutral mutant substitutions. Once in a while, shifting of interacting systems may cause a burst of mutant substitutions. Here, it is very difficult to find out whether environmental change or random genetic drift is really responsible for driving the shift. In some cases, however, adaptive changes may be observed, such as the emergence of a new function by gene duplication. There are several examples of duplicated genes that show accelerated evolution; for example, emergence of fetal haemoglobin from embryonic haemoglobin in primates, acquisition of stomach lysozyme of ruminants and evolution of genes responsible for visual pigments. In such cases, positive Darwinian selection must have worked.
See also: Adaptation and Natural Selection: Overview

Near-neutrality in Evolving Gene Regulation

Another important aspect of near-neutrality is evolution of gene regulation, i.e. it seems to be under interplay of drift and selection again. Usually, the regulation of gene expression is controlled by the interaction between transcription factors and regulatory elements that locate upstream of the coding regions, and is thought to be in a well-balanced state. Any mutations that disturb such a balance are deleterious; however, it has been found that the regulatory elements are in constant turnover. This is because binding sites of various transcription factors exist in multiples and some variations are allowed in their sequences. Ludwig *et al*. (2000) proposed that the stabilizing selection is working at a regulatory element, stripe2 element of the gene, *even-skipped*, of *D. melanogaster*, by showing that the chimaeric element between *D. melanogaster* and its closely related species does not function normally, even if the two native ones work perfectly. Thus, epistatic interaction exists among mutations and the constant turnover of this element is occurring within the allowed latitude by drift and stabilizing selection, i.e. turnover consists of slightly deleterious mutant substitutions and compensatory ones that subsequently occur.

Khaitovich *et al*. (2006) have investigated this problem by measuring messenger RNA (mRNA) levels in various organs of human and chimpanzee. The pattern of diversity of mRNA levels among individuals of human was compared with that of divergence between human and chimpanzee. The results for various organs have been mostly consistent with the neutral and nearly neutral prediction provided that negative selection is common to keeping gene expression patterns. They also noted that testis was different from others, reflecting positive selection. So evolution of gene expression obeys in general the same principle as that of proteins.

It has been reported that many transcripts with unknown function and/or without constraint exist in human cells (The ENCODE Project Consortium, 2007). The finding would indicate more than necessary activities of the genome. Such activities are neutral and nearly neutral, and would provide opportunities for creating novel systems of gene regulation.

Remember that recruitment of regulatory proteins in developmental pathways is thought to be a fundamental mechanism for morphological evolution. With extra-transcripts, the recruitment would be more likely to occur than without such transcripts. Various trials and errors become possible under drift and selection, if extra-transcripts are available.

As compared with amino acid substitutions of protein evolution, selection on mutations in the regulatory elements would be strongly influenced by environmental factors. This is because the expression pattern of genes is directly related to morphological characters that are responsive to environmental changes. Hence, morphological evolution does not seem to be separable from the concept of near-neutrality.

References

Akashi H (1995) Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in *Drosophila* DNA. *Genetics* **139**: 1067–1076.

Chimpanzee Sequencing and Analyses Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69–87.

Hornstein E and Shomron N (2006) Canalization of development by microRNAs. *Nature Genetics* **38**: S20–S24.

Khaitovich P, Enard W, Lachmann M and Pääbo S (2006) Evolution of primate gene expression. *Nature Reviews Genetics* **7**: 693–702.

Kimura M (1968) Evolutionary rate at the molecular level. *Nature* **217**: 624–626.

Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge University Press.

Kimura M and Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725–738.

King JL and Jukes TH (1969) Non-Darwinian evolution: random fixation of selectively neutral mutations. *Science* **164**: 788–798.

Li W-H, Tanimura M and Sharp PM (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *Journal of Molecular Evolution* **25**: 330–342.

Ludwig MZ, Bergman C, Patel NH and Kreitman M (2000) Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**: 564–567.

McDonald JH and Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.

Mustonen V and Lässig M (2007) Adaptation to fluctuating selection in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **104**: 2277–2282.

Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* **246**: 96–98.

Ohta T (1974) Mutational pressure as the main cause of molecular evolution and polymorphisms. *Nature* **252**: 351–354.

Ohta T (1992) The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* **23**: 263–286.

Ohta T (1995) Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *Journal of Molecular Evolution* **40**: 56–63.

Rutherford SL and Lindquist S (1998) Hsp 90 as a capacitor for morphological evolution. *Nature* **396**: 336–342.

Sawyer SA, Parsch J, Zhang Z and Hartl DL (2007) Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **104**: 6504–6510.

Sella G and Hirsh AE (2005) The application of statistical physics to evolutionary biology. *Proceedings of the National Academy of Sciences of the USA* **102**: 9541–9546.

The ENCODE Project Consortium (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799–816.

Wagner A (2005) Distributed robustness versus redundancy as causes of mutational robustness. *BioEssays* **27**: 176–188.

Further Reading

Eyre-Walker A and Keightley PD (2007) The distribution of fitness effects of new mutations. *Nature Reviews Genetics* **8**: 610–618.

Hartl DL and Clark A (2007) *Principles of Population Genetics*, 4th edn. Sunderland, MA: Sinauer.

Lynch M and Cornery JS (2003) The origins of genome complexity. *Science* **302**: 1401–1404.

Ohta T (2002) Near neutrality in evolution of genes and gene regulation. *Proceedings of the National Academy of Sciences of the USA* **99**: 16134–16137.