THE ROYAL SOCIETY

doi 10.1098/rstb.2000.0724

Mechanisms of molecular evolution

Tomoko Ohta

National Institute of Genetics, Mishima, 411-8540, Japan (tohta@lab.nig.ac.jp)

Both drift and selection are important for nucleotide substitutions in evolution. The nearly neutral theory was developed to clarify the effects of these processes. In this article, the nearly neutral theory is presented with special reference to the nature of weak selection. The mean selection coefficient is negative, and the variance is dependent on the environmental diversity. Some facts relating to the theory are reviewed. As well as nucleotide substitutions, illegitimate recombination events such as duplications, deletions and gene conversions leave indelible marks on molecular evolution. Gene duplication and conversion are sources of the evolution of new gene functions. Positive selection is necessary for the evolution of novel functions. However, many examples of current gene families suggest that both drift and selection are at work on their evolution.

Keywords: weak selection; nearly neutral theory; evolution of new functions

1. INTRODUCTION

Studies of mechanisms of molecular evolution started more than 30 years ago, when Kimura (1968) and King & Jukes (1969) proposed that most mutant substitutions at the molecular level must be selectively neutral or nearly neutral. Neo-Darwinism was prevalent at that time, and most evolutionary biologists opposed the neutrality of molecular evolution. However, with the rapid accumulation of results, particularly at the DNA level, the neutral theory has expanded, and many evolutionists have recognized its significance. John Maynard Smith, a most influential evolutionist, was sceptical of the theory at the beginning, but has been quite ready to incorporate new findings of molecular genetics and related theories, as seen in his textbook (Maynard Smith 1989).

As to his scepticism, the neutral theory is not correct in the strict sense. Any changes in DNA can have some effects, even if very minute, and natural selection simply cannot detect mutations with very small effects. The nearly neutral theory started from the question, 'What are border line mutations between the selected and the neutral classes?' (Ohta & Kimura 1971; Ohta 1973). The theory has been evolving ever since, and its present status is summarized in this paper.

Another problem concerns the versatility of DNA. In fact, in eukaryote genomes, illegitimate recombination events of various kinds occur more frequently than previously thought. This versatile nature of the genetic material is important in providing useful genetic variability. Facts and theories on illegitimate recombination are reviewed here. In fact, this versatility is an important source of 'evolvability' for organismal evolution.

2. NUCLEOTIDE SUBSTITUTIONS

Nucleotide substitutions are the most common type of mutation at the molecular level. The debate on drift versus selection on these mutations has continued for more than 30 years, and we now have many good examples to which the nearly neutral theory is applicable. In such cases, both drift and selection operate; drift therefore predominates during periods of small population size, whereas selection operates mainly in large populations.

Weak selection pressure is thought not to be constant over space and time. It depends on environments as well as on genetic backgrounds. In this article I use the term environmental diversity to include both ecological factors and genetic backgrounds. Fluctuation of selective values should depend on interaction systems of various kinds, i.e. protein–DNA, protein–protein and protein–environment. For an exact analysis one requires a knowledge of numerous interaction systems, but it is impossible to know all these systems. My approach (Ohta 1972) is an attempt to evaluate roughly the overall effect of these systems on selection pressures on individual genes. In the 1970s I considered the meaning of near-neutrality when the selection pressure is not constant and depends on the environment. The main logic is as follows. Any new mutants can be advantageous under restricted conditions, but are generally disadvantageous in adapted systems. So, if the environment is diverse, it is almost impossible for a mutant to be advantageous under all conditions. In contrast, if the environment is uniform, a mutant will have a better chance of being advantageous. The probability of becoming advantageous for a mutant is larger in small populations than in large ones. In these circumstances it is very difficult to distinguish between advantageous, neutral and slightly deleterious mutant classes. This is the fixed model of molecular evolution (Ohta & Tachida 1990; Tachida 1991).

The main interest of this model is the prediction that the rate of mutant substitution is negatively correlated with the species population size. This prediction has an important bearing on the rate of molecular evolution. In general, large organisms with long generation times tend

to have small population sizes, and vice versa (Chao & Carr 1993). The mutation rate is thought to depend on the number of cell generations, and hence on the generation length of organisms. This generation-time effect on mutation rate is expected to partly cancel the population-size effect on the fixation probability of nearly neutral mutations (Ohta 1972, 1973).

Synonymous substitutions are expected to be influenced by drift more than non-synonymous substitutions. Then, non-synonymous substitutions are expected to be more dependent on population size and the cancellation effect is stronger for them than for synonymous ones. I tested this prediction by comparing mammalian gene sequences (Ohta 1995). It was found that the effect of generation time on substitution rates is stronger for synonymous substitutions than for non-synonymous ones. In other words, the generation time effect is partially cancelled by the population size effect for non-synonymous substitutions, whereas the cancellation is very small for synonymous ones. Zhang (2000) extended the analyses to separate conservative and radical amino-acid substitutions further. He found that the ratio of radical rate to conservative rate is positively correlated with the ratio of non-synonymous rate to synonymous rate, and that the result is consistent with near-neutrality.

A very interesting case has been recently reported that fits the nearly neutral theory. Ludwig *et al.* (2000) found that *cis*-regulatory enhancers of the *even-skipped* stripe 2 of *Drosophila melanogaster* and its closely related species are undergoing sequence divergence, resulting in incompatible combinations of enhancers between species. They propose that stabilizing selection on the expression is maintaining phenotypic constancy but allowing mutant substitutions with weak effects, although some other forces might have affected divergence. They further argue that this system is ready to respond to environmental requirements.

Another notable characteristic of mammalian gene evolution is the large variance in the substitution rate (Gillespie 1991; Ohta 1995; Yang & Nielsen 1998). The variance is often four to six times larger than expected from the Poisson process under neutrality. The fixed model does not quite explain this large fluctuation. However, note that Cutler's analysis (Cutler 2000) indicates that this model might explain it if selection is mostly negative. Also, if one incorporates changes in population size, the variance becomes large for nearly neutral mutations (Araki & Tachida 1997).

I have emphasized the importance of population size for the behaviour of nearly neutral mutations. Here it should be remembered that, because of background selection and hitchhiking, tight linkage results in a decrease in the effective size of the population (reviewed in Charlesworth & Guttman 1996). In fact, examples are available that show such an effect when the recombination rate differs between regions of a chromosome (see, for example, Takano-Shimizu 1999). Linkage relationships therefore have a significant effect on the extent of near-neutrality.

3. ILLEGITIMATE RECOMBINATION

As well as nucleotide substitutions, illegitimate recombination events are important as sources of genetic variability at the molecular level. They include gene conversion and unequal crossing over, in which the length of the region ranges from a few to thousands of base pairs. Because of the large number of base pairs involved, their effects are likely to be larger on the average than those of nucleotide substitutions. In addition, the rate of occurrence might be locus specific. Therefore, assumptions about parameters for formulating evolution by illegitimate recombination become somewhat arbitrary.

A very common process relating to illegitimate recombination is evolution by gene duplication. In models of gene duplication it has been customarily assumed that, once redundant copies are available, useful mutations can accumulate in one of the copies, while another copy performs the original function (see, for example, Kimura 1983; Charlesworth 1985). In addition, a model of permanent heterozygosity has been put forward for evolution by gene duplication (Spofford 1969). I have tried to formulate the process in a slightly more general way (Ohta 1987, 1988); the results are summarized here.

My simulation results show how a gene family can evolve, together with the accumulation of non-functional (pseudo) genes. An important quantity is the ratio, R, of the rate of spread of useful mutations to that of deleterious ones,

$$R = u_+ v_+ / u_- v_-, \qquad (1)$$

where u_+ and v_+ are the fixation probability and the mutation rate of a beneficial mutant, and u_- and v_- are those of a deleterious one. This ratio is crucial for understanding the evolution of new functions. The formula tells us that u_+ should be fairly large, because v_+ is thought to be much smaller than v_-. For more discussions on various aspects of this problem, see Clark (1994), Walsh (1995) and Nowak *et al.* (1997).

It has often been noted that duplicate genes have been preserved for longer than expected from these simple models (see, for example, Nadeau & Sankoff 1997). Lynch & Force (2000) have shown that subfunctionalization of gene members might help to preserve duplicate genes. Because each subfunction is thought to be necessary, the deterioration of gene copies is prevented. This mode of evolution is caused mainly by the differentiation of regulatory elements. This would enhance the likelihood of evolution of new functions.

There are now many examples of accelerated sequence divergence of duplicated genes (for a general review, see Li (1997), chapter 10). In fact, it seems that whenever new functions are acquired in evolution, an acceleration of amino-acid changes occurs, for example the emergence of foetal haemoglobin from embryonic haemoglobin in primates (Goodman *et al.* 1987), stomach lysozyme in ruminants (Irwin & Wilson 1990), visual pigments in mammals (Yokoyama & Yokoyama 1990), toxins in snails (Duda & Palumbi 1999) and α-mannosidases in diverse groups of eukaryotes (Gonzalez & Jordan 2000). However, in some cases it is difficult to determine whether or not positive selection worked to accelerate amino-acid substitutions. The effect might be caused simply by the relaxation of selective constraints (Li 1985). The comparison of polymorphism patterns with divergence patterns between closely related species would be

helpful in distinguishing positive selection from relaxation, as has been done for the single-locus case (McDonald & Kreitman 1991).

Gene conversion often contributes to the generation of useful genetic variability. In many multigene families with diverse functions, such as those of immunoglobulins, T-cell receptors and the major histocompatibility complex, gene conversion provides useful genetic variability (see, for example, Ohta 1991, 1999; Parham & Ohta 1996). Here again, positive selection seems to be efficient in maintaining this variability.

4. DISCUSSION

It has been shown repeatedly that analyses separately estimating the rates of synonymous and non-synonymous substitutions are very efficient for detecting natural selection. However, it should be noted that synonymous substitutions are not completely neutral. Akashi (1994, 1995), showed by population-genetic analyses that very weak selection on codon usage bias has been at work. Eyre-Walker (1999) has also demonstrated weak selection by examining the frequency distributions of synonymous variants within populations. The difference in substitution patterns between the synonymous and non-synonymous sites must therefore reflect their relative intensities of selection. Notable results have already been obtained on this subject. Akashi (1995) estimated that the selection intensity on codon bias, measured in terms of the product of population size and selection coefficient, is about -1 for *Drosophila simulans*, and that the value might be considerably smaller for *D. melanogaster* because of its small population size. In addition, Sawyer *et al*. (1987) estimated the average selection coefficient of naturally occurring amino-acid polymorphisms at the 6-phosphogluconate dehydrogenase locus of *Escherichia coli* to be -1.6×10^{-7}.

A popular approach for detecting selection is to compare the within-population divergence pattern with the between-population divergence pattern separately at synonymous and non-synonymous sites, as was first done by McDonald & Kreitman (1991). For mitochondrial genes an excess of non-synonymous within-population divergence is often found (see, for example, Nachman *et al*. 1994; Rand & Kann 1996). Such an excess is likely to be due to slightly deleterious effects of amino-acid substitutions. For nuclear genes, a similar excess, as well as the opposite case, i.e. an excess of non-synonymous between-species divergence, is found (see, for example, McDonald & Kreitman 1991; Long & Langley 1993). The latter case is probably caused by advantageous mutant substitutions (reviewed in Moriyama & Powell 1996).

It is highly desirable to connect molecular evolution with phenotypes. There are cases of positive selection that can be connected to phenotypes. For slightly deleterious mutations there is still a large gap between the observation at the molecular level and that at the phenotypic level. Mukai (1964) estimated the rate of occurrence of slightly deleterious mutations on viability to be 0.14 per second chromosome of *D. melanogaster* per generation. We do not know whether this class includes a significant fraction of the slightly deleterious (nearly neutral) mutations involved in molecular evolution. It is possible that a significant fraction of viability polygenes are products of illegitimate recombination or of transposable element insertions.

More recent work along this line has suggested that a large fraction of accumulated deleterious mutations in the *Caenorhabditis elegans* genome cannot be detected by fitness assays (Davies *et al*. 1999). The result would imply that these undetectable mutations include nearly neutral ones. Thatcher *et al*. (1998) also found that many genes found in the yeast genome do not have any discernible effect on fitness, and suggest that these genes have important functions in environments not tested in laboratories, or have very small effects not detectable under laboratory conditions. Mutations of such genes are nearly neutral.

An interesting quantitative character is the bristle number of *Drosophila*, which has been studied in detail. Mackay and co-workers (Mackay & Langley 1990; Mackay 1995) found that much existing variation in bristle number is attributable to alleles with large effects at a small number of loci. Residual variation is apparently caused by a large number of loci with small effects. These mutations are likely to be nearly neutral. Both types of mutation are thought to be important for the evolution of quantitative characters.

The evolution of gene families indicates that illegitimate recombination at the fine-structure level is very important. Gene conversion is effective in generating useful genetic variability. Unequal crossing over provides genetic redundancy by which new functions can be attained. It is reasonable to think that there might be evolutionary adjustment of the rates of their occurrences, just as for ordinary recombination. In fact, illegitimate recombination might be evolutionarily more ancient than ordinary recombination and might have contributed to the formation of new genes from the time of primitive organisms. The acquirement of novel gene functions by illegitimate recombination is thought to be essential to evolvability at the organismal level.

The evolution of recombination has long been of great interest (Maynard Smith 1978; Michod & Levin 1988). I should like to suggest that the mechanism of illegitimate recombination is also under evolutionary modification. The selective force on the rate of illegitimate recombination would be very weak, and the optimum rate would not often be attained. This is because the target of selection is variability generated by illegitimate recombination, and therefore the selective force to modify its rate is indirect.

I thank Professor Brian Charlesworth, Dr Tomoko Steen and an anonymous referee for their useful comments on the manuscript, and Ms Yuriko Ishii for her secretarial assistance. This is contribution no. 2376 from the National Institute of Genetics, Mishima, Japan.

REFERENCES

Akashi, H. 1994 Synonymous codon usage in *Drosophila melanogaster*: natural selection and translation accuracy. *Genetics* **136**, 927–935.

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

Araki, H. & Tachida, H. 1997 Bottleneck effect on evolutionary rate in the nearly neutral mutation model. *Genetics* **147**, 907–914.

Chao, L. & Carr, D. E. 1993 The molecular clock and the relationship between population size and generation time. *Evolution* **47**, 688–690.

Charlesworth, B. 1985 Recombination, genome size and chromosome number. In *The evolution of genome size* (ed. T. Cavalier-Smith), pp. 489–513. Chichester, UK: Wiley.

Charlesworth, B. & Guttman, D. S. 1996 Reductions in genetic variation in *Drosophila* and *E. coli* caused by selection at linked sites. *J. Genet.* **75**, 49–61.

Clark, A. G. 1994 Invasion and maintenance of a gene duplication. *Proc. Natl Acad. Sci. USA* **91**, 2950–2954.

Cutler, D. J. 2000 Understanding the overdispersed molecular clock. *Genetics* **154**, 1403–1417.

Davies, E. K., Peters, A. D. & Keightley, P. D. 1999 High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. *Science* **285**, 1748–1751.

Duda Jr, T. F. & Palumbi, S. R. 1999 Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc. Natl Acad. Sci. USA* **96**, 6820–6823.

Eyre-Walker, A. 1999 Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. *Genetics* **152**, 675–683.

Gillespie, J. H. 1991 *The causes of molecular evolution*. Oxford University Press.

Gonzalez, D. S. & Jordan, I. K. 2000 The α-mannosidases: phylogeny and adaptive diversification. *Mol. Biol. Evol.* **17**, 292–300.

Goodman, M., Czelusniak, J., Koop, B. F., Tagle, D. A. & Slightom, J. L. 1987 Globins: a case study in molecular phylogeny. *Cold Spring Harbor Symp. Quant. Biol.* **52**, 875–890.

Irwin, D. M. & Wilson, A. C. 1990 Concerted evolution of ruminant stomach lysozymes. *J. Biol. Chem.* **265**, 4944–4952.

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

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

King, J. L. & Jukes, T. H. 1969 Non-Darwinian evolution: random fixation of selectively neutral mutations. *Science* **164**, 788–798.

Li, W.-H. 1985 Accelerated evolution following gene duplication and its implication for the neutralist–selectionist controversy. In *Population genetics and molecular evolution* (ed. T. Ohta & K. Aoki), pp. 333–352. Tokyo and Berlin: Japan Scientific Society Press and Springer.

Li, W.-H. 1997 *Molecular evolution*. Sunderland, MA: Sinauer.

Long, M. & Langley, C. H. 1993 Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* **260**, 91–95.

Ludwig, M. Z., Bergman, C., Patel, N. H. & Kreitman, M. 2000 Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**, 564–567.

Lynch, M. & Force, A. 2000 The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459–473.

McDonald, J. H. & Kreitman, M. 1991 Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652–654.

Mackay, T. F. C. 1995 The genetic basis of quantitative variation: numbers of sensory bristles of *Drosophila melanogaster* as a model system. *Trends Genet.* **11**, 464–470.

Mackay, T. F. C. & Langley, C. H. 1990 Molecular and phenotypic variation in the *achaete-scute* region of *Drosophila melanogaster*. *Nature* **348**, 64–66.

Maynard Smith, J. 1978 *The evolution of sex*. Cambridge University Press.

Maynard Smith, J. 1989 *Evolutionary genetics*. Oxford University Press.

Michod, R. E. & Levin, B. R. 1988 *The evolution of sex: an examination of current ideas*. Sunderland, MA: Sinauer.

Moriyama, E. N. & Powell, J. R. 1996 Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13**, 261–277.

Mukai, T. 1964 The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* **50**, 1–19.

Nachman, M. W., Boyer, S. N. & Aquadro, C. F. 1994 Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. *Proc. Natl Acad. Sci. USA* **91**, 6364–6368.

Nadeau, J. H. & Sankoff, D. 1997 Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. *Genetics* **147**, 1259–1266.

Nowak, M. A., Boerlijst, M. C., Cook, J. & Maynard Smith, J. 1997 Evolution of genetic redundancy. *Nature* **388**, 167–171.

Ohta, T. 1972 Population size and rate of evolution. *J. Mol. Evol.* **1**, 305–314.

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

Ohta, T. 1987 Simulating evolution by gene duplication. *Genetics* **115**, 207–213.

Ohta, T. 1988 Further simulation studies on evolution by gene duplication. *Evolution* **42**, 375–386.

Ohta, T. 1991 Multigene families and the evolution of complexity. *J. Mol. Evol.* **33**, 34–41.

Ohta, T. 1995 Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* **40**, 56–63.

Ohta, T. 1999 Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. *Immunol. Rev.* **167**, 319–325.

Ohta, T. & Kimura, M. 1971 On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.* **1**, 18–25.

Ohta, T. & Tachida, H. 1990 Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. *Genetics* **126**, 219–229.

Parham, P. & Ohta, T. 1996 Population biology of antigen presentation by MHC class I molecules. *Science* **272**, 67–74.

Rand, D. M. & Kann, L. M. 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* **13**, 735–748.

Sawyer, S. A., Dykhuizen, D. E. & Hartl, D. L. 1987 Confidence interval for the number of selectively neutral amino acid polymorphisms. *Proc. Natl Acad. Sci. USA* **84**, 6225–6228.

Spofford, J. B. 1969 Heterosis and the evolution of duplication. *Am. Nat.* **103**, 407–432.

Tachida, H. 1991 A study on a nearly neutral mutation model in finite populations. *Genetics* **128**, 183–192.

Takano-Shimizu, T. 1999 Local recombination and mutation effects on molecular evolution in *Drosophila*. *Genetics* **153**, 1285–1296.

Thatcher, J. W., Shaw, J. M. & Dickinson, W. J. 1998 Marginal fitness contributions of nonessential genes in yeast. *Proc. Natl Acad. Sci. USA* **95**, 253–257.

Walsh, J. B. 1995 How often do duplicated genes evolve new functions? *Genetics* **139**, 421–428.

Yang, Z. & Nielsen, R. 1998 Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **46**, 409–418.

Yokoyama, S. & Yokoyama, R. 1990 Molecular evolution of visual pigment genes and other G-protein-coupled genes. In *Population biology of genes and molecules* (ed. N. Takahata & J. F. Crow), pp. 307–322. Tokyo: Baifukan.

Zhang, J. 2000 Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J. Mol. Evol.* **50**, 56–68.