

Near-Neutrality, Robustness, and Epigenetics

Tomoko Ohta*

Department of Population Genetics, National Institute of Genetics, Mishima, Japan

*Corresponding author: E-mail: tohta@lab.nig.ac.jp.

Accepted: 14 February 2011

Abstract

The nearly neutral theory emphasizes the interaction of drift and weak selection in evolution. With progress of genome biology, the applicability of the nearly neutral theory has expanded. The genome-wide analyses of synonymous and nonsynonymous substitutions at protein-coding regions show prevalence of very weak selection. Many patterns of evolution of gene regulation are also in agreement with the nearly neutral prediction. Our consideration on near-neutrality expands in relation to the progress on molecular understanding of robustness and epigenetics. Both are bridges to link genotypes with phenotypes and important for understanding how weak selection and drift interact in the evolution of complex systems.

Key words: near-neutrality, robustness, epigenetics.

Introduction

The nearly neutral theory was put forward in 1973 as an extension of the neutral theory (Ohta 1973). Simply put, the nearly neutral theory posits that there exists a significant class of alleles that are sufficiently weakly selected such that the evolutionary dynamics are governed by both drift and selection. At the beginning, the theory was directed toward amino acid substitution in evolution. There are now various data from amino acid substitutions to genomic structural variations relating to the nearly neutral theory. In the following, I present progress in the applicability of the theory; near-neutrality is widely observed at the protein level, and divergence pattern of gene expression is mostly consistent with near-neutrality.

Another interesting development on near-neutrality is found in the concept of robustness of gene regulatory networks. The nearly neutral theory has become more significant in understanding morphological evolution because gene regulation is thought to be the basis of organismal development, and nearly neutral mutations would increase under robust systems as compared with sensitive systems. Intimately, related subjects are epigenetic phenomena of which our understanding is rapidly increasing. Its relation to near-neutrality is also discussed. Note that both robustness and epigenetics lie between genotype and phenotype.

Evolution of Protein-Coding Regions

The nearly neutral theory is directed toward understanding the interaction of drift and selection. Figure 1 shows the

comparison of the selection, the neutral and the nearly neutral theories on how new mutations are classified. An important prediction of the nearly neutral theory is the negative correlation between the evolutionary rate and the effective population size of species. This is because, in the nearly neutral class, slightly deleterious mutations are much more abundant than slightly advantageous ones (Ohta 1992). Also, if selection coefficients are dependent upon conditions determined by genetic backgrounds and environments, even slightly advantageous mutant substitutions may become rapid in small populations as compared with large populations (Ohta 1972).

Comparison of the pattern of synonymous and nonsynonymous substitutions at the protein-coding regions gives opportunities for verifying the population size effect. Note that mammalian lineages are characterized as follows: organisms with long generation time tend to have small population size and vice versa. So population size tends to be large for rodents, middle for artiodactyls, and small for primates. I analyzed the 49 gene sequences of three orders, primates, artiodactyls, and rodents (Ohta 1995). The results are in accord with the prediction, that is, the rodent branch is much longer than the primate branch for the synonymous tree, but the difference of the two lineages is not so large for nonsynonymous tree. This is because of the population size effect of nearly neutral amino acid substitutions in comparison with synonymous changes. Genome data show similar pattern, for example, the ratio of nonsynonymous to synonymous divergence is higher in human–chimpanzee

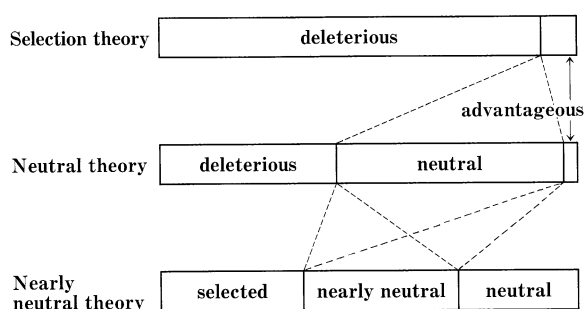


FIG. 1.—Diagram to show how new mutations are classified under the selection, the neutral, and the nearly neutral theories. The diagram of the nearly neutral theory is made by folding that of the neutral theory. Therefore, the selected class includes both advantageous and deleterious mutations and the nearly neutral class, both slightly advantageous and slightly deleterious one.

comparison than in mouse–rat comparison (The Chimpanzee Sequencing and Analysis Consortium 2005).

A second key prediction of the nearly neutral theory is abundance of low-frequency polymorphisms as compared with strict neutrality. Data on DNA polymorphisms reveal prevalence of nearly neutral amino acid substitutions. For example, Hughes et al. (2003) have analyzed human SNP data and found that amino acid changing SNPs show reduced diversity as compared with synonymous SNPs in agreement with the nearly neutral prediction.

The McDonald–Kreitman (M-K) test (McDonald and Kreitman 1991), that compares the relative numbers of synonymous and nonsynonymous substitutions within a population and those between closely related species, is popular for detecting selection. Several reports have been published on data analyses by using an extended M-K test. For example, Smith and Eyre-Walker (2002) analyzed data of *Drosophila simulans* and *D. yakuba* and emphasized that 45% of amino acid substitutions have been fixed by positive selection.

Sawyer et al. (2007) have performed a solid analysis of X chromosome data of *Drosophila melanogaster* and *D. simulans*, that is, polymorphisms in 91 genes in an African population of the former and the divergence between the two species. The result is interesting: about 70% of amino acid changing polymorphisms are slightly deleterious, but about 95% of fixed differences are positively selected. They also estimated that about 50% of amino acid changes are nearly neutral, if the near-neutrality is defined as $|Nes| < 1$. They conclude that weak selection of which substantial fraction is nearly neutral is efficient for adaptive evolution of proteins. Therefore, adaptive protein evolution and near-neutrality are nicely unified.

Functional approach on amino acid substitutions is becoming to have impact on near-neutrality. Lunzer et al. (2010) examined functional effects of single amino acid substitutions in isopropylmalate dehydrogenase (IMDH) of

Escherichia coli. They have found that the transition from *E. coli* IMDH to *Pseudomonas aeruginosa* IMDH cannot be explained without assuming slightly deleterious amino acid substitutions followed by compensatory substitutions. This nearly neutral process is thought to be caused by cryptic epistasis among amino acids within the protein molecule.

By using genome-wide polymorphism data of closely related species of yeast, Elyashiv et al. (2010) have performed detailed statistical analyses to estimate the effect of purifying selection. The yeast species used are characterized by having several completely isolated subpopulations with varying sizes. They have found negative correlation between heterozygosity and the ratio of nonsynonymous to synonymous polymorphisms on these subpopulations, whereas this ratio is larger for the within-population polymorphisms than on fixed differences among subpopulations. These results have verified the prevalence of weak purifying selection supporting the nearly neutral theory. They have further found that the shifts in intensity of selection across populations can be explained by the change of a single parameter, that is, population size.

Evolution of Gene Regulation

Evolution of gene regulation is under interplay of drift and selection. The best-studied case of gene regulation is the interaction of transcription factors and enhancers in the upstream of the transcription start site. Gene regulation is thought to be in a well-balanced state, and any mutations that disturb the balance are deleterious. Ludwig et al. (2000) have found that the enhancer regions are in constant turnover, that is, nucleotide substitutions at the enhancer region of the gene, even skipped, are occurring within the allowed latitude of stabilizing selection on this gene expression. In other words, the turnover is nearly neutral.

By measuring the level of mRNA in various organs of human and chimpanzee, Khaitovich et al. (2006) have clarified the relationship between diversity of mRNA levels among human individuals and the divergence of mRNA levels between human and chimpanzee. They have found that the results for various organs have been mostly consistent with the nearly neutral prediction. The exception was testis such that the human–chimpanzee divergence is too rapid as compared with diversity among human individuals. The authors interpret that this rapid divergence reflects positive selection on gene expression for testis.

By using quantitative genetics approach, Bedford and Hartl (2009) examined the observed variance of mRNA levels among seven *Drosophila* species. They have found that the variance initially increases with divergence time but eventually saturates. Based on this observation, they estimated the intensity of stabilizing selection on expression levels and have found that the selection is weak, may be regarded as mostly nearly neutral. However, they emphasize that

divergence among the species is mostly attributed to positive selection, and therefore, weak selection is effective.

Phenomena That May Explain Abundance of Weak Selection

In this section, let us consider why such a large fraction of mutations both at the protein level and at the level of gene regulation are subject to weak selection.

Robustness of Genetic Regulatory Systems

Although robustness of biological systems has been recognized since Waddington, molecular methods to tackle this subject have become available only recently. Our knowledge on this is now rapidly expanding and should be incorporated into discussions of evolution. Particularly, the concept seems to provide an important clue to the long-standing issue on how molecular changes and morphological evolution are linked. Also it is important to recognize that robustness increases the width of near-neutrality (Ohta 2002).

Wagner (2005) argues that pathways of gene expression are complex, and numerous networks are entangled and that the system becomes robust to any perturbations. He calls this kind of robustness, “distributed robustness.” Another type of robustness is found by the morphological anomaly induced by mutations of the Hsp90 (Rutherford and Lindquist 1998). Hsp90 is a molecular chaperone and helps to stabilize signal transduction systems, and its defects cause phenotypic anomalies. Participation of microRNAs to the stability of various genetic systems is also suggested (Hornstein and Shomron 2006). So there are various mechanisms to increase robustness.

Following Lehner (2010), let us survey various types of robustness, that is, mutational (genetic), environmental, and stochastic ones. For the analysis, data on large-scale synthetic lethal or sick screens of yeast and other model organisms are useful. By examining such data, Lehner (2010) has found that genetic robustness is often correlated with stochastic robustness and/or environmental robustness. In other words, when genetic perturbation by a mutation has minor effect but have large effects in combination with defects of other genes (i.e., synthetic lethal or sick and therefore robust), the gene shows resilience to environmental stress (environmental robustness) and/or stochastic change (stochastic robustness). This negative interaction is equivalent to robustness, and mutations will have initially small (or nearly neutral) effects. Lehner (2010) argues that, under variable environment, resilience to many environmental conditions is advantageous and selected for, and genetic robustness has evolved as a by-product of evolution of environmental robustness.

Yeast data also provide some important characteristics of quantitative genetic interactions relating to robustness. Beltrao et al. (2010) surveyed quantitative mapping of genetic

interactions of yeast data on colony size of mutant strains. By using the multiplicative model, they classified the interaction types into the following three, positive, neutral, and negative interactions. Positive interaction means that the fitness effect of the double mutants is less than that expected from the two single mutations. This type includes the suppressive and compensatory cases. The neutral interaction means that the two mutations are independent. Negative interaction means that effect of the double mutants is more than that expected from the two single mutations. This type includes robust cases. Beltrao et al. (2010) have focused on proteins involved in chromatin biology, such as chromosome segregation, transcription, and chromatin modeling/remodeling. They have found a module composed of three correlated protein clusters. Note that proteins form complex set consisting of functional modules, each of which is made of several clusters. Their significant finding is that, within each cluster, genetic interactions are mostly positive, whereas, among the clusters, they are mostly negative. Remember that the three clusters are involved in chromatin modeling/remodeling, and proteins in each cluster often physically associate. It has often been emphasized that genetic interactions are rapidly changing in evolution as compared with proteins (Roguev et al. 2008; Tischler et al. 2008). In the chromatin modeling/remodeling system above, Beltrao et al. (2010) have further noticed that positive interactions are more conserved than negative interactions. By assuming that interactive effects parallel with fitness values, negative interactions result in more robust and more nearly neutral systems than positive interactions. Above findings based on genome-wide data have significant implications for our considerations on evolution of complex networks.

Proteins are more conserved than genetic interaction systems as mentioned before. However, note that the previous example of slightly deleterious plus compensatory amino acid substitutions in evolution of IDMH (Lunzer et al. 2010) may be regarded as a robust system of amino acids that makes the protein.

Epigenetics

The meaning of epigenetics is broad and various processes involved in “the inheritance of variation above and beyond changes in the DNA sequence” are included (Bonasio et al. 2010). Here, let us define epigenetics as more than one phenotypes under the same genotype that are heritable for some generations. Recent progress on our understanding of this concept at the molecular level is remarkable. Epigenetics is intimately related to robustness as both lie between genotypes and phenotypes, and their incorporation into evolutionary problems is a most urgent task.

Among various processes involved in epigenetics, DNA methylation is best studied. It is thought that DNA methylation is linked to chromatin structure and to participate in gene regulation. It has been known for some time that

the spectrum of genome methylation is very broad, that is, it is different among species and among loci within a genome and that its function is related to gene regulation. Following Suzuki and Bird (2008), let us briefly review progress on DNA methylation of higher organism genome. DNA methylation is found throughout the human genome except CpG islands. However, this global methylation pattern is limited to vertebrates, and mosaic patterns are observed in yeast and fruitfly. In plants, global pattern is seen, but the methylation system of transposons is different from that of animals. So there are similarities and differences among the species.

The importance of methylation is its impact on gene regulation. It has been known for some time that DNA methylation works to silence gene expression. In the following, I present summary of Suzuki and Bird (2008). DNA methylation pattern is different according to cell types and therefore may contribute to development. A notable progress on this topic is advancement of mapping of methylation of genomes. Comparison of such methylation maps provides basic data for considering the meaning of within-species (among individuals) and between-species differences of methylation patterns. Also active research is going on how DNA methylation map is different among various human tissues.

Next, I discuss an interesting idea on how DNA methylation may contribute to evolution of complex systems. Variation of DNA methylation at regulatory regions among individuals of the same genotype increases variation of gene expression and hence variability of phenotypes. Feinberg and Irizarry (2010) consider that the increase of variability is important for coping with environmental stresses and examined variation of methylation patterns in human and mouse tissues. They have identified the locations of variable regions of DNA methylation among individuals in mouse liver and termed them variably methylated regions (VMRs). They have extended the analysis to mouse brain, human liver, and human brain and found that VMRs are often associated with genes for development and morphogenesis. It has also been noted that the variation in methylation is likely to be caused by gain or loss of CpG in regions involved in gene regulation. Based on these findings, they argue that, if variable methylation at a regulatory region of a gene is advantageous under variable environments, its genetic basis, that is, loss or gain of CpG dinucleotides, is selected for, and provides plasticity in the developmental program. Note that variation of gene expression is intimately related to robustness because it is reciprocal of stochastic robustness (Kaneko 2007), and the latter is linked to genetic and environmental robustness (Lehner 2010). Both epigenetics and robustness are involved in processes that lie between genotype and phenotype, and their effects on weak selection must be significant.

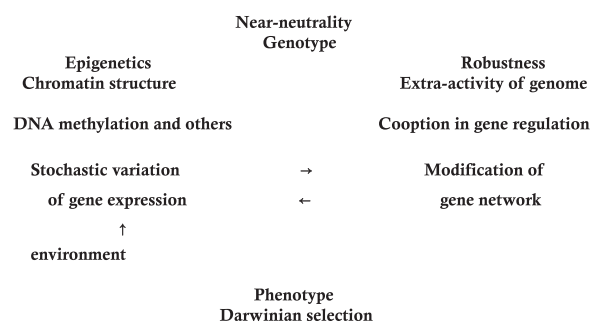


FIG. 2.—A scenario to show the meaning of near-neutrality in relation to epigenetics and robustness that connect phenotypes with genotypes.

Nearly Neutral Zone

Let us now consider the problem of drift and selection, that is, the nearly neutral zone of figure 1 in relation to robustness and epigenetics. Natural selection works on phenotypes, but its effects are transmitted by underlying genotypes, and the two are linked by epigenetics and robustness of organismal development. Figure 2 gives a scenario. Among epigenetic changes, only those that have minute or favorable effects would survive. As pointed out before, it is important to notice that epigenetic processes provide variable phenotypes that may be useful in changing environments because regulatory systems would readily respond to environmental stresses.

In relation to robustness, the human genome has extra-activity as shown by The ENCODE Project Consortium (2007), that is, transcriptional activity is more widespread in human genome than expected from protein-coding regions. Such activities are thought to be mainly nearly neutral, that is, they occur and may survive in allowed latitude under weak selection, drift, and robustness. Actual patterns on gene expression divergence analyzed by Bedford and Hartl (2009) and by Khaitovich et al. (2006), mentioned previously, may fit to the present scenario.

From the above scenario, it may be said that near-neutrality prevails at the genotype level but that natural selection may predominate at the phenotype level. Important points of this scenario are environmental contacts to gene expression and availability of enough room for modifying gene networks. Note also that, under the scenario, some proposals on epigenetic effects in evolution and variation (Jablonka and Lamb 2002; West-Eberhard 2003; Slatkin 2009) may become more understandable. As to “Chance and Necessity” by Jacques Monod, he might have considered natural selection at the phenotype level because molecular bases of epigenetics and robustness were not known 40 years ago. Here, I would like to emphasize again that this scenario may have significant implications for understanding evolution of enormously complicated systems at various levels in biological world.

Acknowledgments

I thank Dr D. Hartl for his valuable suggestions to improve the presentation. I also thank two anonymous referees for their useful comments.

Literature Cited

- Bedford T, Hartl DL. 2009. Optimization of gene expression by natural selection. *Proc Natl Acad Sci U S A*. 106:1133–1138.
- Beltrao P, Cagney G, Krogan NJ. 2010. Quantitative genetic interactions reveal biological modularity. *Cell* 141:739–745.
- Bonasio R, Tu S, Reinberg D. 2010. Molecular signals of epigenetic states. *Science*. 330:612–616.
- The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87.
- Elyashiv E, et al. 2010. Shifts in the intensity of purifying selection: an analysis of genome-wide polymorphism data from two closely related yeast species. *Genome Res*. doi: 10.1101/gr.108993.110.
- 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.
- Feinberg AP, Irizarry RA. 2010. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation and disease. *Proc Natl Acad Sci U S A*. 107:1757–1767.
- Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. *Nat Genet*. 38:S20–S24.
- Hughes AL, et al. 2003. Widespread purifying selection at polymorphic sites in human protein coding loci. *Proc Natl Acad Sci U S A*. 100:15754–15757.
- Jablonka E, Lamb MJ. 2002. The changing concept of epigenetics. *Ann New York Acad Sci*. 981:82–96.
- Kaneko K. 2007. Evolution of robustness to noise and mutation in gene expression dynamics. *PLoS One*. 2(5):e434.
- Khaitovich P, Enard W, Lachman M, Paabo S. 2006. Evolution of primate gene expression. *Nat Rev Genet*. 7:693–702.
- Lehner B. 2010. Genes confer similar robustness to environmental, stochastic and genetic perturbations in yeast. *PLoS One*. 5(2):e9035.
- Ludwig MZ, Bergman C, Patel NH, Kreitman M. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403:564–567.
- Lunzer M, Golding GB, Dean AM. 2010. Pervasive cryptic epistasis in molecular evolution. *PLoS Genet*. 6(10):e1001162.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.
- 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. 1992. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst*. 23:263–286.
- Ohta T. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J Mol Evol*. 40:56–63.
- Ohta T. 2002. Near-neutrality in evolution of genes and gene regulation. *Proc Natl Acad Sci U S A*. 99:16134–16137.
- Roguev A, et al. 2008. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*. 322:405–410.
- Rutherford SL, Lindquist S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* 396:336–342.
- Sawyer SA, Parsch J, Zhang Z, Hartl DL. 2007. Prevalence of positive selection among nearly neutral amino acid replacement in *Drosophila*. *Proc Natl Acad Sci U S A*. 104:6504–6510.
- Slatkin M. 2009. Epigenetic inheritance and the missing heritability problem. *Genetics* 182:845–850.
- Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415:1022–1024.
- Suzuki MM, Bird A. 2008. DNA methylation landscape: provocative insight from epigenomics. *Nat Rev Genet*. 9:465–476.
- Tischler J, Lerner B, Fraser AG. 2008. Evolutionary plasticity of genetic interaction networks. *Nat Genet*. 40:390–391.
- Wagner A. 2005. Distributed robustness versus redundancy as causes of mutational robustness. *BioEssays* 27:176–188.
- West-Eberhard MJ. 2003. Developmental plasticity and evolution. New York: Oxford University Press.

Associate editor: Bill Martin