

Epigenetics and Evolutionary Mechanisms

Tomoko Ohta*

National Institute of Genetics, Mishima, 411-8540, Japan

Abstract

In this article, an attempt to incorporate recent knowledge of epigenetics into the evolutionary theory is presented. As our interest is to clarify evolutionary mechanisms at the molecular level and to connect them to phenotype evolution, the interplay of drift and selection (near-neutrality) on molecular evolution is briefly reviewed. Epigenetic phenomena are partly controlled by genetic systems via chromatin structure, and special attention has been paid to the dynamic evolution of three gene families which encode chromatin components. These gene families are characterized by rapid birth and death of gene copy members, and weak diversity enhancing selection. Also the protein products contain disordered domain that provides flexible chromatin structure. The near-neutrality concept may be extended to their evolution. Here drift, selection and epigenetics become inseparable, and their interplay is thought to have been needed for the evolution of complex gene regulatory systems.

Keywords: Epigenetics; Chromatin structure; Chromatin component proteins; Near-neutrality; Interplay of drift and selection; Gene regulatory network; Evolution of complex systems

Introduction

Progress in genomics and epigenetics has prompted me to reconsider some of the basic models of evolution. The current main theory of evolution is Neo-Darwinism, which is based on population genetics. This field has developed in the last century by combining Mendelian genetics with Darwin's theory of natural selection.

In population genetics, the process of gene frequency change by selection has been formulated and such changes are thought to provide genetic mechanisms for evolution. However the Neo-Darwinian models have had little material bases for morphological evolution. Molecular biology has changed all fields of biology, and evolution should be no exception.

Brief History of Neutral and Nearly Neutral Theories

The first attack on Neo-Darwinism was the proposal of the neutral theory of molecular evolution by Kimura (1968) [1]. Here it was argued that random genetic drift rather than natural selection was the main force for evolutionary changes at the molecular level. It looked as if molecular evolution and phenotypic evolution were dichotomous, and material basis was still not available. The dichotomy was thought to come from the assumption that only a small minority of mutant substitutions is adaptive and phenotypic evolution is caused by these adaptive changes [1]. Here it was assumed that that majority of mutant substitutions are neutral and do not contribute to phenotypic changes. On the other hand, selectionists consider that any changes at the molecular level could not be neutral, and positive natural selection was the main cause. Very hot debate had been popular between the neutralists and the selectionists for decades.

I had been puzzled by the following three questions on the neutral theory, even if I belonged to the neutralist camp.

1. What are borderline mutations between the selected and the neutral classes of mutation?
2. Why the molecular clock on the rate of molecular evolution is year-dependent rather than generation-dependent?
3. Why does the heterozygosity on protein polymorphisms not much different among the species? Note that under the neutral theory, the

heterozygosity depends on population size, and species population size is much different among the species.

Many selectionists thought that the third one disproved the neutral theory. I recognized that, by bringing slightly deleterious mutations into the border between the selected and the neutral classes of mutation, the three problems could be explained Ohta [2].

Figure 1 gives a diagrammatic presentation on how new mutations are classified under the selection, the neutral and the nearly neutral theories Ohta [3]. A most notable difference on the prediction between the neutral and the nearly neutral theories is, the former predicts that the evolutionary rate is equal to the neutral mutation rate, whereas the latter predicts the negative correlation between the evolutionary rate and the species population size. This near-neutrality prediction has been verified by the genome-wide data (for a review see Ohta 2011) [4]. For a thorough review on weak selection of protein evolution, see Akashi et al. [5], in which population genetics analyses, as well as protein structure and function are focused.

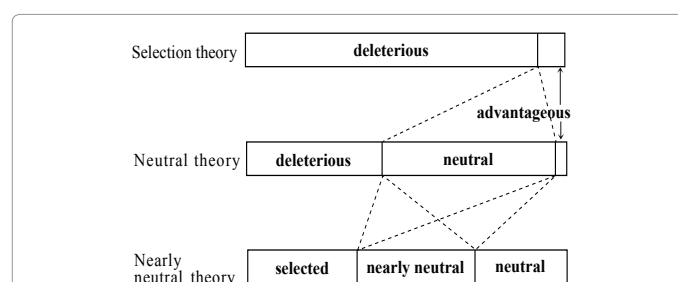


Figure 1: Diagram showing the proportion of various classes of new mutations for the selection, the neutral and the nearly neutral theories.

*Corresponding author: Tomoko Ohta, National Institute of Genetics, Mishima, 411-8540, Japan, E-mail: tohta@nig.ac.jp

Received July 31, 2013; Accepted August 17, 2013; Published August 20, 2013

Citation: Ohta T (2013) Epigenetics and Evolutionary Mechanisms. Human Genet Embryol 3: 113. doi:[10.4172/2161-0436.1000113](https://doi.org/10.4172/2161-0436.1000113)

Copyright: © 2013 Ohta T. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Evolution of Complex Systems

Let me now consider the problem on how gene regulatory systems have evolved. If you look at the developmental biology or systems biology books, you would be struck by the facts how gene regulatory networks are complex. A most significant question is, "How such complex systems could have evolved?" In the next, I present my view on the problem. Recent progress on epigenetics has deep impact here. Before going into details of the problem, I review the evidence that suggest the interplay of drift and selection again.

Khaitovitz et al. [6] investigated the pattern of divergence of gene expression between human and chimpanzee. By examining various tissues, they have compared the expression divergence between the two species and the expression diversity among individuals within species. They have found that the ratio between the two is similar in various tissues except testis and brain. From this fact, they have argued that evolution of gene expression is mostly under drift and weak selection. They have also argued that positive selection worked on the exceptional cases.

More quantitative genetics approach has been performed by Bedford and Hart [7]. They have studied the pattern of gene expression divergence among *Drosophila* species by using quantitative genetics model, and by applying stochastic population genetics. They have found that the divergence initially increases linearly with time, but that it eventually reaches a plateau, which has been caused by stabilizing selection. They successfully estimated the intensity of the stabilizing selection, that is mostly weak, such that about half of new mutations fitting in the range of near-neutrality.

It now seems that evolution of gene expression is also under the nearly neutral process. A fundamental issue here is how genotype and phenotype are connected. Figure 2 is the diagram to show the relationship between the two. We know that the robustness exists in some cases, i.e., different genotypes may give the same phenotype. In other cases, the same genotype may result in different phenotypes. The former depends on robust regulatory systems and the latter, on different environments or even on chance. For understanding such flexible systems, epigenetics becomes most important.

Epigenetics is a rapidly expanding field of biology, and it means inheritance phenomena not caused by genotypes. Developmental processes of higher organisms or environmental responses of bacteria may be epigenetic. The present interest lies not in the definitive process, but in flexible paths responding to variable environments. Our knowledge on molecular mechanisms for epigenetics in higher organisms has greatly expanded recently, and it is now clear that chromatin structure and function are mainly responsible. However

genetic mechanisms at the whole genome level may influence chromatin structure and the connection between epigenetics and genetics become highly complicated. Next I consider this problem in relation to the big project on DNA elements in the human genome.

The Encyclopedia of DNA Elements (ENCODE) project started to map all functional elements in human genome (ENCODE Project Consortium 2007, 2012) [8,9]. Various methods including the DNAase I hypersensitivity (DHS) has been applied together with analyses on human diversity data. It has been estimated that 80% of the human genome may be assigned to have biochemical functions at some tissues or at some developmental times. By combining the ENCODE regions with human diversity data, it is concluded that the regions are under negative selection. Subsequent ENCODE report Thurman et al. [10] presented a study of the DHS regions in some detail, and found that ~2.9 million DHSs contain all cis-regulatory elements. Some of the interacting systems between enhancers and promoters have been shown to be traceable by their analyses. Such interaction systems depend on chromatin accessibility and other conditions of chromatin structure. DHSs are again estimated to be under negative selection. It has also been pointed out that a significant fraction of DHSs are in transposable elements such as retroposon.

How can we understand such abundant negative selection working in the human genome? Are all DHSs really involved in gene regulatory systems? A relevant study some years ago is reviewed here. Hahn et al. [11] have analyzed genomes of 52 species of Eubacteria and Archaea, to find out whether transcription binding sites, such as TATAAT, over- or under-represented compared with random expectation. They have found that the binding sites are often under-represented. They thought that the under-representation is caused by negative selection against spurious binding sites, and estimated the average intensity of selection, that is very weak, Nes (product of the effective population size, N_e , and the selection coefficient, s) is -0.12 for Eubacteria and -0.06 for Archaea genomes. There is no reason to suppose that spurious sites are excluded in the ENCODE analyses. Because of epigenetics via chromatin modification in Eukaryotes, the intensity of negative selection may be even smaller than these estimates.

Some unsolved problems on functionality of the ENCODE sites are given below.

What fraction of the sites had originated from retroposon and other transposable elements?

How much spurious binding sites are included?

Can non-coding RNAs be classified into the biologically functional and non-functional classes?

It is necessary to answer these questions for understanding the meaning of negative selection the ENCODE project reported. Next, I consider how genetic and epigenetic phenomena are related.

Genetic Factors on Chromatin Function

Epigenetics depends on chromatin structure as explained above, and genetic mechanisms are responsible for a large fraction of remodeling of chromatin structure and function. I review here some interesting cases of dynamically evolving gene families encoding chromatin components.

Evolution of H1 Histone Gene Family

Histone proteins, H2A, H2B, H3 and H4 are core histones and make up nucleosomal core particles together with DNA. Histone H1

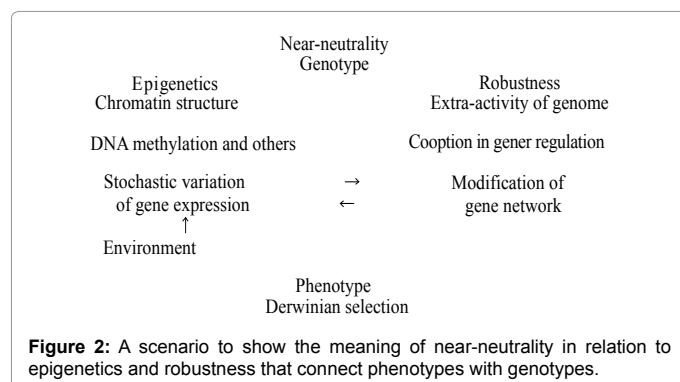


Figure 2: A scenario to show the meaning of near-neutrality in relation to epigenetics and robustness that connect phenotypes with genotypes.

members bind to nucleosome and linker DNA to help stabilization of chromatin structure. They also participate in gene regulation by remodeling of chromatin structure and function. I present here some characteristics of this gene family following the review article by Kowalski and Palyga [12].

H1 gene family usually consists of several gene members ranging from a single to more than ten copies. Each member of a family is differentially expressed in specialized cells and is called a subtype. Via differential expression, gene members perform different functions. An interesting property of the H1 histone is that it consists of the two main domains, the evolutionarily conserved globular domain (N-terminal), and the less constrained C-terminal domain. The latter is usually disordered. Thus this histone has very versatile structure and function.

Histone H1 globular domain has two DNA-binding sites. C-terminal (variable) domain (CtD) of H1 histone is relatively short and consists of a hydrophobic region and a basic segment. The CtD helps the globular domain to bind DNA and to provide versatile structure. The CtD is highly variable and often amino acid polymorphisms exist that have phenotypic effects in human. Because of the versatility, H1 histone is highly mobile and interacts with a number of non-histone-targets in a subtype-specific way, and provides diverse functions of chromatin. Therefore polymorphism in CtD may exhibit pleiotropic effects, some of which cause human disease.

Dynamic Evolution of Heterochromatin Protein 1 (HP1) Gene Family

Among other components of the heterochromatin, the gene family of for HP1 is interesting. Levine et al. [13] have presented the result of their phylogenetic analyses of *Drosophila* species. They have found structural diversity, lineage restriction and germ-line biased expression of *Drosophila* HP1 gene members. The HP1 proteins are characterized by the chromo-domain and the chromoshadow-domain. Some members contain both domains, and others, either one of them. For the former, the two domains are connected by the hinge region of which sequence and length are variable. Chromo-domain and the chromoshadow-domain have differentiated functions in chromatin remodeling. Because of such versatile structure of HP1 proteins, it is thought that they contribute to environmental responses of gene regulation.

Some of these interactions are being clarified for *Schizosaccharomyces pombe* HP1 proteins (Canzio et al. 2013) [14]. HP1 recognizes histone marks of chromatin, and drives a switch from an auto-inhibited state to a spreading-competent state of heterochromatin. In the former state, a histone-mimic sequence in HP1 inhibits the histone methyl mark recognition, and prevents spreading. Therefore heterochromatin dynamics depends on delicate balance among chromatin components. The chromo-domain, the hinge region and the chromoshadow-domain cooperate in the interplay via conformational change of HP1 proteins.

Rapid Evolution of the High-Mobility Group (HMG) Protein Family

The HMG proteins are abundant and highly mobile components of vertebrate chromatin. They are detected only in vertebrate and necessary for the specific interplay among chromatin proteins. Their gene family is dynamically evolving and, following Malicet et al. [15], the evolutionary pattern is reviewed here. HMG protein is made up of globular component that binds the nucleosome core particle, and of highly variable disordered region. There are three subfamilies

with differentiated function in human HMG proteins. Malicet et al. [15] examined the divergence pattern of one subfamily, HMGN, in detail. Evolution of this family is characterized by highly variable and disordered C-terminal region. Both length and amino acid sequence are rapidly changing in mammalian species. For example, the shared identity of amino acids between orthologous human and mouse HMGNs is less than half for the disordered region. The globular domain that contains nucleosome binding sites is conserved. Malicet et al. [15] have shown the difference in nucleosome interaction between the mouse and human HMGNs. Nevertheless, both induce chromatin decomposition, and the disordered region contributes to the specificity of chromatin function.

An important characteristic of the above three cases of the rapidly evolving protein families is that they contain disordered variable region that provide diverse functions. Another notable property of the gene families is the rapid birth and death of gene copies. Here subtle balance among overlapping functions of gene copies is important. In addition, it has been proposed that chromatin activity may depend on the heat shock protein 90 Sawarkar and Paro [16]. Together with numerous chromatin component proteins, Hsp90 may regulate gene expression in specific ways.

Another significant subject on the robust gene regulation is the contribution of microRNA and other non-coding RNAs (Berezikov [17], Ebert and Sharp [18]). This subject covers a wide area of RNA biology, and is still in rapid expansion, and is not included in this article. However, it is noted that the majority of the evolution of non-coding RNAs may be under drift and selection, at least at the time of their origination. Later, some of RNAs may become indispensable, and be kept by negative selection.

For the evolution of complex systems, it is necessary to consider how all these systems are related to drift and selection. Furthermore “self-organization property” may be working together with the above systems (Bar-Yam et al.) [19]. Future study is needed here.

Conclusion

In this article, a most difficult problem, i.e., how enormous complexity in biological world could have evolved, has been discussed from various aspects. Significance of epigenetics via remodeling of chromatin structure is emphasized for understanding of this problem. It is recognized that the more we know about the mechanisms of gene regulation, the more complex systems are revealed. So one needs to know that evolutionary processes on gene regulation which is not as simple as can be described by Neo-Darwinian models. I also emphasize that our knowledge in all areas of biology, from molecular genetics and evolution, cell biology, population genetics, developmental biology, to systems biology need to be incorporated for the progress of science of evolution.

References

1. Kimura M (1968) Evolutionary rate at the molecular level. *Nature* 217: 624-626.
2. Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* 246: 96-98.
3. Ohta T (2002) Near-neutrality in evolution of genes and gene regulation. *Proc Natl Acad Sci U S A* 99: 16134-16137.
4. Ohta T (2011) Near-neutrality, robustness, and epigenetics. *Genome Biol Evol* 3: 1034-1038.
5. Akashi H, Osada N, Ohta T (2012) Weak selection and protein evolution. *Genetics* 192: 15-31.
6. Khaitovich P, Enard W, Lachmann M, Pääbo S (2006) Evolution of primate gene expression. *Nat Rev Genet* 7: 693-702.

7. Bedford T, Hartl DL (2009) Optimization of gene expression by natural selection. *Proc Natl Acad Sci U S A* 106: 1133-1138.
8. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447: 799-816.
9. ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57-74.
10. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, et al. (2012) The accessible chromatin landscape of the human genome. *Nature* 489: 75-82.
11. Hahn MW, Stajich JE, Wray GA (2003) The effects of selection against spurious transcription factor binding sites. *Mol Biol Evol* 20: 901-906.
12. Kowalski A, PaÅyga J (2012) Linker histone subtypes and their allelic variants. *Cell Biol Int* 36: 981-996.
13. Levine MT, McCoy C, Vermaak D, Lee YC, Hiatt MA, et al. (2012) Phylogenomic analysis reveals dynamic evolutionary history of the *Drosophila* heterochromatin protein 1 (HP1) gene family. *PLoS Genet* 8: e1002729.
14. Canzio D, Liao M, Naber N, Pate E, Larson A, et al. (2013) A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. *Nature* 496: 377-381.
15. Malicet C, Rochman M, Postnikov Y, Bustin M (2011) Distinct properties of human HMGN5 reveal a rapidly evolving but functionally conserved nucleosome binding protein. *Mol Cell Biol* 31: 2742-2755.
16. Sawarkar R, Paro R (2013) Hsp90@chromatin.nucleus: an emerging hub of a networker. *Trends Cell Biol* 23: 193-201.
17. Berezikov E (2011) Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet* 12: 846-860.
18. Ebert MS, Sharp PA (2012) Roles for microRNAs in conferring robustness to biological processes. *Cell* 149: 515-524.
19. Bar-Yam Y, Harmon D, de Bivort B (2009) Systems biology. Attractors and democratic dynamics. *Science* 323: 1016-1017.