Preface

Necessity of Evolutionary Studies

Organisms on the earth are rich in diversity. Each organism also contains its own genome with many genes. This complex genetic system has been generated and constantly modified through eons of evolution since the origin of life. Evolutionary study is thus indispensable for gaining the unified view of life. Because even a single-cell bacterium is so complex, we have to study its genetic entity, that is, its genome, to acquire a comprehensive view of the organism. I will discuss evolution of organisms from the viewpoint of temporal changes of genomes and methods for their study in this book, titled *Introduction to Evolutionary Genomics*.

Evolutionary changes already started even before the origin of life, known as chemical evolution. Therefore, we need to dig down to the molecular level, starting from nucleotides and amino acids. In this sense, it is logically straightforward to study evolution of life at the molecular level, that is, molecular evolution. Molecular evolutionary study as a discipline was established only after biochemistry and genetics became the center of biology in the middle of the twentieth century. Because of this late start, there are still some molecular biologists who consider the study of evolution as carried out by specialized researchers, while there are some old-fashioned evolutionists who do not appreciate molecular-level studies. It would be my great pleasure if such people change their minds after reading this book. But, of course, I hope that the majority of readers of this book are young students and researchers.

What Is Evolution?

Evolution is the temporal process of life. Originally, the word "evolution" meant the development of embryo from egg. Charles Lyell was probably the first person to use this word in the modern meaning in his *Principles of Geology* published in the 1830s. Thanks to the pioneering works of Lamarck, Wallace, and Darwin, evolution as a biological phenomenon was gradually accepted during the last 200 years (e.g., [14]). Evolution, however, does not contain a predetermined pathway, unlike developmental processes. As the time arrow moves from past to present, life forms change. Therefore, any temporal change of organisms is evolution. Nowadays, the concept of evolution is sometimes extended to nonlife, such as evolution of the universe or evolution of human society.

What Is a Genome?

The word "genome" was coined by Hans Winkler, botanist, in 1920 [15]. Genes were already localized in chromosomes in the cell nucleus at that time, and Winkler joined two words, "gene" and "chromosome," to produce a new word "genome." Plants are often polyploid, and there was a need to designate a certain unit of chromosome sets. Later, Hitoshi Kihara defined genome as a minimum set of genes that are necessary for that organism [6]. This function-oriented definition is still used today, but a structure-oriented definition needs to be invented. Thus, I redefined genome as "a maximum unit of self-replicating body" [12]. A "self-replicating body" includes not only usual organisms but also organella and virus that need some help from organisms for their replication.

Kihara and his group conducted genome analysis on various wheat species in the early twentieth century, and he coined this famous couplet:

The history of the earth is recorded in the layers of its crust;
The history of all organisms is inscribed in the chromosomes.

This couplet was originally mentioned in his book written in Japanese [7] and is increasingly become evident as we now study evolution at the nucleotide sequence level.

The word genome also implies completeness. It is important to grasp all genetic information contained in a single genome, because this gene set mostly determines life patterns of its organism. However, all genes in one genome are not sufficient for that organism to exist. This insufficiency is clear for parasites. For example, leprosy-causing bacteria, *Mycobacterium leprae*, has many pseudogenes [1]. These bacteria probably lost their functional genes through a long parasite life due to dependence on host genomes. We should remember that all organisms on earth are interacting with each other. These kind of known host-parasite relationships are the only prominent examples. Even our own human genome gives a good example of dependency on nonhuman genomes.

Vitamins, by definition, cannot be synthesized inside human body, and we need to obtain them through various foods. For example, deficiency of vitamin C, or ascorbic acid, causes scurvy. Many nonhuman organisms do produce ascorbic acid, as they have its chemical pathway. A gene for enzyme L-gulonolactone oxidase (E.C. no. 1.1.3.8) became a pseudogene (nonfunctional) in the common ancestor of human and Old World monkeys, and we are no longer producing ascorbic acid [9]. In any case, an organism cannot survive alone. We have to always consider the environment surrounding an organism.

Vitalism Versus Mechanism

If we consider the history of biology, one viewpoint presents a controversy between vitalism and mechanism. Vitalism maintains that life has a unique law that does not exist in nonlife forms; thus, it is dualistic. Mechanism is monistic, for it states that life only follows physicochemical laws that govern inorganic matters. In other words, there is no specific difference between organism and inorganic matters according to the mechanistic view. The long history of biology may be considered a series of victories of mechanism against vitalism (e.g., [11]). For example, we easily recall the theory of heart as a pump for blood circulation by William Harvey in the seventeenth century and the discovery of enzyme function in a cell-free system by Eduard Buchner in the nineteenth century. Biochemistry and genetics are two main fields of biology where the mechanistic viewpoint is always emphasized. Molecular biology inherited this aspect from these two disciplines.

Some biologists, however, were strong proponents of vitalism. Hans Driesch, developmental biologist in the early twentieth century, examined the development of sea urchin and discovered that two- or four-cell-stage embryos can develop adult individuals even after they were separated. Because of this utterly mysterious process, he proposed the existence of "entelechy" only in organism [4]. It is true that the animal development is still not completely known. Yet, the modern developmental biology is clearly on the side of Wilhelm Roux, a contemporary of Driesch. Roux strictly followed the mechanistic view in his study.

There is still the remnant of a dualistic view similar to vitalism, to consider mind and body as totally separate entities. As vitalism tried to demarcate organism from inorganic matters, this dualistic view tries to demarcate mind from body. However, a logical consequence of the mechanistic view of life is of course to explain mind as some special organismic process, most probably a neuronal one, that is, mind exists in a body, and these two are inseparable. The mind-body dualism is illogical to begin with and scientifically wrong (e.g., [2]). It is unfortunate that some eminent neurologists such as Wilder Penfield [10] and John Eccles [5] maintained the dualistic view.

Everything Is History

Mechanism is about to declare its victory over vitalism, including the mind-body dualism. Yet, it still remains whether we can explain the whole life phenomenon completely only through mechanism. Life, with its eons of history, is a product of evolution, and there are so many chance effects. For example, spontaneous mutations appear randomly. Most of the mutations that last a long time in the history of life are selectively neutral, and they were chosen through the random genetic drift [8]. Furthermore, there are so many inorganic factors that drastically change the environment of the earth. Examples are volcanic activities, ice ages, continental drifts, and asteroid impacts. These seem to be all random from the organismic world. These historical processes where chance dominates are out of control of mechanism.

As mutations arise, some disappear while the others remain. This process is impossible to be fully explained through the logical cause and effect style of mechanism. This is not restricted to life. *Shinra bansho* (Japanese; "all matters and events in this universe," in English) themselves are transient and there always exists a history, as Hitoshi Kihara pointed out in his couplet. After all, everything is history. Therefore, the essence of natural science is to describe the history of the universe at various levels. Often it is claimed that the ultimate goal of natural science is to discover the laws of nature, and the description of nature is only one process to the eventual finding of laws. It fails to put first things first. So-called natural laws are mere tools for an effective description of natural phenomena. A phenomenon that can be described succinctly is relatively simple, while from a complex phenomenon, it is difficult to extract some laws. However, such difference comes from the phenomena themselves, and the objective of natural science should not be restricted to phenomena from which it is easy to find some laws. It should be noted, however, that giving a mere description of everything is not enough. Human ability to recognize the outer world is physically limited, and a structured description of the historical process is definitely necessary, depending on the content of each phenomenon. In this sense, the time axis, which is most important for organismal evolution, is obligatory for the description of nature itself. With the above argument in mind, I am quite confident that the very historical nature of genes with its self-replication mechanism has the key to overcome the mechanistic view of this universe.

Genome as a Republic of Genes

Another important feature of the genome is its completeness in the finite world. As biology experienced the mechanism versus vitalism controversy, it suffers another controversy on methodology, that is, reductionism versus holism. Organismal evolution is the summation of small genetic changes, and the whole process can be understood through the divide-and-conquer strategy. This reductionistic approach is very useful for the evolution of genes. Some people stress the importance of the holistic approach. However, I do not know of any profound discovery in natural science where a holistic approach was truly effective. Of course, I am not trying to say that the whole is a simple sum of parts. Reductionism is the only approach to understand any phenomenon. Then the "genome" comes in. Because of its completeness, the reductionistic approach in genome studies naturally brings us the whole world. In this sense, evolutionary genomics plays an important role as a unique test case of methodology in biology.

The genome of one organism usually contains many genes, and they are interacting with each other in a complex way. In this sense, Saitou [12] characterized a genome as a "republic of genes." This implies that one particular gene or gene group is not controlling the other genes as the "master control gene." In other words, there is no fixed role for one gene as the master or slave. Any gene has the potential to control or influence functions of other genes.

It is true that in some genomes there exist systems in which a single "master control" gene plays a crucial role and only after this gene is expressed, other downstream genes are expressed. This kind of master control gene, however, has also the possibility of receiving some influence from other genes. Therefore, a top-down system, or the "empire of genes" viewpoint, may not apply to genes in every genome [13].

The "republic of genes" also implies another assertion that locations of genes in chromosomes are not so important. Since RNA and protein molecules are transcribed and translated, respectively, they are expected to show a "trans effect" – to influence genes on other chromosomes or those at a remotely related location of the same chromosome. In contrast, the "cis effect" is to influence the genes on the same chromosome at a close location. In fact, phenomena such as enhancers and chromatin remodeling recently received much attention. However, their effects may be restricted.

Synteny, or the gene order conservation between species, may also play only the passive role. If the gene order is important for the coordination of gene expression, as in bacterial operons, this order may be selectively conserved. However, most of the syntenic regions seem to be only results of chance effects, that is, a gene order happened to be void of disruption.

Genome sequence reporting papers often try to stress the importance of the cis effect, for a considerable part of any genome sequence information is the gene order. Yet, we have to first consider the trans effect when proteins or small RNA molecules are in question. In contrast, the cis effect should first be considered when a DNA sequence itself is influencing other genome regions, because this may involve neither transcription nor translation.

Structure of This Book

This book consists of three parts: Part I – Basic Processes of Genome Evolution; Part II – Evolving Genes and Genomes; and Part III – Methods for Evolutionary Genomics. Part I includes five chapters (Chaps. 1–5). I explain the basics of molecular biology in Chap. 1, while the following four chapters are more specific to evolution. Charles Darwin defined evolution as "descent with modification" in the *Origin of Species* [3]. DNA replication, explained in Chap. 1, is fundamental to "descent," or the connection from parents to offsprings, while "modification" in modern terms is mutation. Mutation is thus covered in Chap. 2, and phylogeny is described in Chap. 3 as the descriptor of evolution. Mutation is already a random process, but randomness also dominates throughout the default process of genome evolution, namely, neutral evolution, covered in Chap. 4. The description of natural selection follows in Chap. 5.

There are five chapters (Chaps. 6–10) in Part II. A brief history of life starting from the origin of life is discussed in Chap. 6, followed by an explanation of genomes of various organism groups; prokaryotes in Chap. 7; eukaryotes in Chap. 8; vertebrates in Chap. 9; and humans in Chap. 10. The evolution of many genomes is amply discussed in this part. Lineage-specific evolutionary problems are also covered in appropriate chapters.

We then move to methods in Part III (Chaps. 11–17). Genome sequencing, phenotype data collection, and databases are explained in Chaps. 11–13, followed by homology searches, multiple alignments, and evolutionary distances in Chaps. 14 and 15. Methods of tree and network building are discussed rather in detail in Chap. 16. Chapter 17 is devoted to population genomics.

I created a page on evolutionary genomics under my laboratory website (http://www.saitou-naruya-laboratory.org/Evolutionary_Genomics/), which contains a web appendix, detailed references, and a detailed index. The web appendix consists of four parts: basic statistics, worked-out examples of evolutionary genomics, and updates of each chapter.

It should be noted that I published a textbook on evolutionary genomics, written in Japanese, in 2007 [13]. This book is an extension of that book. If you have any questions, please contact me at saitou.naruya@gmail.com. I hope you will enjoy reading this book.

Following Walter Pater's maxim – "all art constantly aspires towards the condition of music" – I would like to conclude this preface with my epigram:

All biology aspires to evolution.

Acknowledgments

I appreciate the kindness of many people who helped me to complete this book: Mrs. Mizuguchi Masako for her programming and drawing figures, Mrs. Kawamoto Tatsuko for her excellent secretarial work, and Mrs. Noaki Yoshimi for providing me some gel photos and chromatograms. I thank current Saitou Lab members (Dr. Sumiyama Kenta, Dr. Kirill Kryukov, Mr. Kanzawa Hideaki, Ms. Nilmini Hettiatachchi, and Mr. Isaac Babarinde), my former students (Dr. Matsunami Masatoshi, Dr. Timothy Jinam, Dr. Takahashi Mahoko, Dr. Suzuki Rumiko, Dr. Masuyama Waka, Dr. Kim Hyungcheol, Dr. Tomiki Takeshi, Dr. Kaneko Mika, Dr. Noda Reiko, Dr. Kitano Takashi, and Dr. OOta Satoshi), and my former postdocs (Dr. Sato Yukuto, Dr. Kawai Yosuke, Dr. Shimada Makoto, Dr. Takahashi Aya, Dr. Kim Chungon, and Dr. Liu Yuhua) for their various comments and for providing some parts of their works to the manuscripts of this book. I also thank Dr. Xie Lu for providing me some materials for Chap. 12. I thank Dr. Andreas Dress and Dr. Martin Vingnon, series editors of Computational Biology, to include this book into their series. Finally, I thank people of my good old Houston days (1982–1986), the most exciting period of my academic life: Dr. Lydia Aguilar-Bryan, Dr. Ranajit Chakraborty, Dr. Stephen P. Daiger, Dr. Robert Ferrell, Dr. Takashi Gojobori, Dr. Dan Graur, Dr. Craig Hanis, Dr. Wen-Hsiung Li, Dr. Chi-Cheng Luo, Dr. Partha Majumder, Dr. Pekka Pamilo, Dr. William Jack Schull, Dr. Bob Schwartz, Dr. Claiborne Stephens, Dr. Fumio Tajima, Dr. Naoyuki Takahata, Dr. Ken Weiss, Dr. Chung-I Wu, and, of course, Dr. Masatoshi Nei.

September, A.S. 0013 (2013 A.D.)

References

1. Cole, S. T., et al. (2001). Massive gene decay in the leprosy bacillus. *Nature, 409*, 1007–1011.
2. Damasio, A. (1994). *Descartes' error*. New York: Putnam.
3. Darwin, C. (1859). *On the origin of species*. London: John Murray.
4. Driesch, H. (1914). *The history and theory of vitalism*. London: MaCmillan.
5. Eccles, J. C. (1989). *Evolution of the brain: Creation of the self*. London/New York: Routledge.
6. Kihara, H. (1930). Genomanalyse bei *Triticum* und *Aegilops*. *Cytologia, 1*, 263–284.
7. Kihara, H. (1946). *Story on wheets (written in Japanese)*. Tokyo: Sogen-sha.
8. Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press.
9. Nishikimi, M., Fukuyama, R., Minoshima, S., Shimizu, N., & Yagi, K. (1994). Cloning and chromosomal mapping of the human nonfunctional gene for L-gulono-gamma-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. *Journal of Biological Chemistry, 269*, 13685–13688.
10. Penfield, W. (1975) The mystery of the mind: A critical study of consciousness and the human brain. Princeton: Princeton University Press.
11. Saitou, N. (1997). *Genes dream for 3.5 billion years (written in Japanese)*. Tokyo: Daiwa Shobo.
12. Saitou, N. (2004). *Genome and evolution (written in Japanese)*. Tokyo: Shinyo-sha.
13. Saitou, N. (2007). *Introduction to evolutionary genomics (written in Japanese).* Tokyo: Kyoritsu Shuppan.
14. Saitou, N. (2011). *Darwin: An introduction (written in Japanese)*. Tokyo: Chikuma Shobo.
15. Winkler, H. (1920). *Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche*. Jena: Fischer.

Springer

http://www.springer.com/978-1-4471-5303-0

Introduction to Evolutionary Genomics
Saitou, N.
2013, XXIII, 461 p. 227 illus., Hardcover
ISBN: 978-1-4471-5303-0