2.7 Evolution of the Multigene Family: A Case of Dynamically Evolving Genes at Major Histocompatibility Complex

TOMOKO OHTA

Summary. It is now known that many multigene families exist in eukaryote genomes. Multigene families that are known to be evolving under continued occurrence of unequal crossing-over and gene conversion provide a different picture of evolution from that of the conventional population genetics. As an example of an evolving multigene family, evolution and variation at major histocompatibility complex loci are reviewed. Exceptionally high polymorphisms at the class I and class II loci of major histocompatibility complex (MHC) has been of great interest for many years. In addition, recent studies indicate that amino acid substitution at antigen recognition sites (ARS) is more rapid than synonymous substitution, contrary to the general pattern of nucleotide substitution in evolution. However, such acceleration seems to be limited to a certain period after gene duplication. In order to explain such an unusual pattern, a population genetic model of diversifying selection is constructed. To make the model fit the data, diversifying selection needs to work on enhancing diversity not only between alleles at the same locus, but also between genes at different loci belonging to the gene family. Gene conversion among genes is also incorporated by choosing parameter values that are thought to be realistic. Simulation studies reveal that very weak selection at individual amino acid sites can explain the unusual pattern of evolution and polymorphism at MHC loci under this model. The applicability of the present model is discussed by surveying other examples, such as evolution of protease inhibitors and immunogloblin variable regions that show a similar pattern of acceleration of amino acid substitutions.

Introduction

It is now established that genes of higher organisms are often split and that many repeated gene families exist. Comparative studies of these gene families show that genetic material is more versatile than previously thought, i.e., various “illegiti-

1 National Institute of Genetics, Mishima, 411 Japan
mate" recombinational processes are occurring, such as gene conversion or unequal crossing-over with substantial frequency. It is thought that such "versatility" has been of some advantage for acquiring new gene function, and hence for "progressive" evolution of higher organisms [1–3].

Versatility implies that the way genetic variability is supplied and the way natural selection works may be different from the previous theories based on Mendelian genes. Therefore, it is highly desirable to understand the origin and evolution of multigene families by means of population genetics. Here a population genetics approach is needed in order to understand how gene families have evolved, i.e., in order to know how natural selection and/or random genetic drift contributed to their origin and subsequent evolution, population genetic studies are required together with data analyses.

In general, it has been thought that natural selection works to keep genes in status quo for those gene families that were established tens of millions of years ago. The selection for keeping genes in status quo is called negative selection. On the other hand, incipient gene families may be in the way of further "progress" in the sense of acquiring more diverse function while positive natural selection may be operating. Population genetic models of the established multigene families have been extensively analyzed by incorporating gene conversion [4–7], unequal crossing-over [2] and duplicative transposition [8,9]. The results can satisfactorily explain the so-called concerted evolution of multigene families. In other words, each member of a gene family does not differentiate independently, but evolves in concert with other members because of their functional interrelationship and because of the homogenizing effect of the above processes.

The origin of gene families has also been studied by population genetics [3,10]. Starting from a single gene copy, my simulations have shown how beneficial genes may accumulate on the chromosome under various sets of parameter values on intensity of natural selection, and on rates of unequal crossing-over, of mutation and of random genetic drift. With realistic values of deteriorating mutation rate, positive selection is needed for acquiring duplicated genes with desirable functions.

In the real world, however, there is no clear-cut distinction between incipient and established gene families. Even ancient multigene families such as those of immunoglobulin genes are in continuous reorganization [11]. Remarkable examples are class I or class II genes at major histocompatibility complex of mammals. The number of loci and the extent of polymorphism of these genes are different from species to species, indicating continuous reorganization of gene arrangement. In this article, I review facts and theories on their evolution and variation. Discussion will be extended to other notable examples such as proteases and their inhibitors.

### Evolution and Variation at Major Histocompatibility Complex Loci

Genes of histocompatibility antigen are in the chromosomal region called major histocompatibility complex (MHC). For many years, these genes have attracted the attention of evolutionary and population geneticists because of their extraordi-