THE EVOLUTION OF DNA SEQUENCES

MASATOSHI NEI
Center for Demographic and Population Genetics, The University of Texas, Health Science Center at Houston, P.O. Box 20334, Houston, TX 77225

Received April 6, 1988

DNA (RNA in some viruses) contains all the genetic information that is required for life. One can therefore argue that evolution of organisms should eventually be studied at the DNA level. This view is now widespread, and an increasing number of biologists are studying the evolutionary change of DNA in all kinds of organisms. While the gap between morphological evolution and DNA evolution remains largely unfilled, the study of DNA evolution has produced many interesting observations. In March 1985, the Royal Society of London organized a discussion meeting to evaluate the significance of these findings. This book represents the proceedings of the meeting. It contains 13 papers covering various aspects of molecular evolution, with emphasis on repeated DNA and transposons.

The book starts with a well-written paper by D. L. Hartl, M. Medhora, L. Green, and D. E. Dykhuizen on the evolution of DNA sequences in Escherichia coli. The authors present a comprehensive review of molecular population genetics of E. coli worked out mainly by B. R. Levin, R. Milkman, R. K. Selandier, and T. S. Whittam, as well as by the authors themselves. They show that E. coli is extremely variable and that polymorphic alleles at different loci are usually in strong linkage disequilibrium. This suggests that recombination is very rare in natural populations of this organism. Another topic of this article is the evolutionary dynamics of plasmids and transposons. Plasmids are self-replicating extrachromosomal DNA that can act as vectors to transfer a chromosomal gene from one organism to another. In natural isolates of E. coli, plasmids range in size from a few hundred to a few hundred thousand base pairs, but the distribution of sizes is bimodal. Transposons are DNA sequences that are capable of changing their location from one position to another in the chromosome. There are many different kinds of transposons in bacteria, but their evolutionary significance is largely unknown. Some biologists speculate that transposons are selfish DNAs that propagate in the genome like parasites; others believe that transposons bring about genomic rearrangement or mutations, some of which might increase the fitness of an individual. One thing is clear, however: the number of transposons cannot increase indefinitely, because individual fitness declines with increasing numbers of transposons.

The next two papers, written by W. R. Engels and J. F. Y. Brookfield, are concerned with the evolutionary biology of transposons in Drosophila. Engels discusses the P and I transposable elements. These elements have no obvious sequence homology but show similar genetic properties, including hybrid dysgenesis. A curious observation is that very old laboratory stocks of D. melanogaster (dating from the 1920's) lack in both elements, whereas most strains recently derived from natural populations contain the elements. M. G. Kidwell therefore proposed that the P and I elements invaded D. melanogaster during the period of 1930-1960. This hypothesis seems to be supported by the recent observation that the P element is harbored by species in the D. willistoni group but not by the sibling species of D. melanogaster. However, Engels points out several difficulties with this hypothesis and maintains that neither Kidwell's hypothesis nor his own "recent loss hypothesis" can properly explain the population dynamics of the P and I elements.

Brookfield discusses the mechanism of maintenance of transposable elements copia, 297, and 412 by fitting a mathematical model to data on the distribution of the number of transposons per chromosome. A neutral model based on a number of assumptions about the insertion and deletion of transposons fit available data reasonably well, but still there seem to be problems in assaying the number of transposons accurately.

The next three papers (by R. B. Flavell, H. C. MacGregor and S. K. Sessions, and C. J. Bostock) deal with repetitive DNA and chromosome evolution. According to Flavell, the number of base pairs required for the gene function of chromosomes in higher plants is about 10^7-10^8, but actual DNA content varies from 5 x 10^7 to 8 x 10^9 base pairs. Some of this variation is due to polyploidy, but most of the DNA in each haploid chromosome set appears to be in excess of the minimum required. The sequence of this "excess DNA" appears not to be highly conserved during evolution, and most of the sequences are repetitive DNA. The fraction of repetitive DNA is so high in many plant genomes (sometimes over 95%) that it plays a dominant role in determining chromosome size and structure. Repetitive DNA consists of various different families of sequences, but DNA belonging to the same family is often distributed both within and between chromosomes. The goal of Flavell's paper is to determine how these families of repetitive DNA are dispersed to different parts of the genome.

The simplest form of repetitive DNA is tandem duplication of a sequence (usually a few hundred base pairs long) localized in a segment of chromosome. Evolution of this type of repetitive DNA can be explained

---

by unequal crossover, the rolling-circle model, the slippage replication model, or the aberrant in situ replication model. The last three models accommodate a rapid production of a large number of tandem repeats, but it is not yet clear whether these models really apply to germ-line cells. It should be noted that highly repetitive DNA is usually not transcribed and that its spreading through the genome seems to occur very rapidly, because closely related species often have it at different locations on chromosomes. The propagation of repetitive DNA over different chromosomes probably occurs through occasional interchromosomal translocation. Once repeated DNA is translocated from one chromosome to another, the homogeneity of repetitive DNA may be maintained by interchromosomal pairing. In addition to the above mechanisms, transposons also seem to play an important role of producing repetitive DNA, particularly defective transposable elements. MacGregor and Sessions examine the same problem as that addressed by Flavell using the European newts (Triturus), and their conclusion is essentially the same. So, the evolutionary change of highly repetitive DNA seems to be controlled by the same mechanism in both plants and animals.

Bostock has written one of the most interesting papers in this volume. The subject treated in this paper is DNA sequence amplification during development. A well-known example of this type of amplification is that of ribosomal-RNA (rRNA) genes in amphibians. Xenopus laevis has 400–600 copies of rRNA genes in the genome. In the process of oocyte formation, a small sample of these genes becomes amplified about 100,000 times, apparently by a process of rolling-circle replication to form the extrachromosomal nucleoli of the oocytes. Curiously, this type of amplification does not occur for SS RNA genes, of which the number in the genome is about 25,000. Why, then, did the gene-amplification system evolve only for rRNA genes? No answer to this question is available now. Gene amplification occurs in many other instances (e.g., in the process of formation of macronucleus from micronucleus in Tetrahymena and in enhancing resistance to the drug methotrexate in mammals). While the mechanism of this type of gene amplification is not fully understood, it may occasionally happen in germ-line genes. If this is the case, it will have a profound evolutionary implication. It seems to me that Britten and Kohne's (1968) original explanation of the evolution of repetitive DNA by saltatory replication should be seriously considered.

The article by G. A. Dover and D. Tautz and that by O. Smithies and P. A. Powers deal with evolution by gene conversion. Using the rRNA genes and the human globin genes \( \gamma \) and \( \beta \), both groups of authors show that gene conversion involving a small number of nucleotides apparently occurs more frequently than was previously thought. This may be true with the gene clusters examined here, because the sequence similarity between repeat genes is very high. Dover and Tautz, however, extrapolate their result even to the case of immunoglobulin heavy chain variable region (V) genes and criticize Gojobori and Nei's (1984) earlier conclusion, which was contrary. I would like to take the liberty to indicate that our unpublished statistical study did not show any significant partial gene conversion in V genes. It seems that their conclusion should be confined to repetitive genes, which show a high degree of similarity, not to multigene families with divergent repeat genes, such as the immunoglobulin and T-cell receptor gene families.

The major histocompatibility complex (MHC) plays an important role in the immune system and consists of two different classes of genes; class I and class II. W. F. Bodmer, J. Trowsdale, J. Young, and J. Bodmer discuss the evolution of class-II genes (HLA-D cluster) in humans. Unlike many other reviews on this subject, their paper is written with a background of population-genetics theory. They show that some polymorphic alleles at the DP locus have probably coexisted for about five million years in the same population. Unfortunately, the progress in this field is so rapid that this article is already somewhat outdated.

The last three papers are not directly related to repetitive DNA or transposons. W. M. Fitch invents a homomorphic method of studying the transition-transversion bias in nucleotide substitution and uses it for detecting concomitantly variable nucleotides (covariates). J. C. Avise presents a comprehensive review of the utility of mitochondrial DNA for the study of phylogeny construction and population structure. M. Kimura gives a good summary of his 1983 book on neutral theory, putting emphasis on DNA sequence data. Although these articles are well-written and very interesting, I shall not make any particular comments on them, because the results and views of these authors are already known to most molecular evolutionists.

The study of molecular evolution is wide-ranging, and it is difficult to cover all subjects in a small book like this one. The organizers chose their speakers carefully and produced a useful book on a timely subject. I recommend this book to anyone who is interested in evolution. It should, however, be noted that many efforts are now being made to fill the gap between morphological evolution and DNA evolution (e.g., Nathan et al., 1986). In the next decade or so we will probably know more about the molecular basis of evolution of morphological and physiological characters.

**Literature Cited**


Corresponding Editor: J. B. Mitton