Book Reviews


Recent advances in molecular biology have revolutionized our understanding of eucaryotic gene and genome structure, and revealed novel mechanisms of evolution at the molecular level. This has been the result of the rapid accumulation of DNA primary sequences and restriction maps. To population geneticists and molecular evolutionists, statistical analyses of these data are crucial, and the nine chapters in this book are a survey of these statistical methods.

Chapter 1, by Schaffer, deals with regression methods that predict DNA fragment lengths from mobilities on a gel. This problem is not difficult to solve statistically but it does have some practical importance. Gingeras, in chapter 2, discusses various computer algorithms for DNA sequence analysis but fails to tell us where the statistical problems do occur. He provides a useful list of available computer programs, but there is no discussion of the merits of each. However, he does mention the possibility of using artificial intelligence techniques for sequence analysis, and I, for one, would have liked to see a detailed discussion of this interesting area.

The next two chapters, by Ewens and Kaplan, respectively, are reviews of models and methods for studying evolutionary relatedness and genetic variability in populations based on restriction maps or DNA sequences. Ewens goes to great length to point out the several types of variability that one could consider and when each is appropriate; I found this discussion very useful. However, Kaplan does a more comprehensive job of studying the various estimators by computer simulations and data analysis.

Chapter 5, by Brown and Clegg, on the analysis of variation in related sequences, is perhaps the best in this collection. These authors provide new statistical methods and perform a comprehensive analysis of data on a repeated DNA sequence in maize knob heterochromatin. They repeatedly point out where the statistical problems lie, what they are, and what approaches may be taken to solve them. I was particularly interested in their method of inferring concerted evolution. In the next chapter, Felsenstein, a leader in the field of phylogenetic analysis, has described the statistical problems in this area very well. These two chapters are a must for statisticians interested in a new class of problems and are areas where they have much to contribute.

Errors in phylogenetic analysis that can occur through chance convergent evolution are studied by Templeton in chapter 7, and he suggests nonparametric methods as a remedy. I disagree with much of what he has to say. First, his probabilistics are all conditional on the type of convergent evolution—a fact we will not know. Second, the simulation studies of Kaplan (chap. 4) demonstrate that current methods work rather well, and it behooves Templeton to show the superiority of his methods by computer simulation.

Chapters 8 and 9 contain discussions of the use of DNA data in human genetics. Bishop et al. study the number of DNA polymorphisms necessary to map the human genome so that an unknown disease locus can be shown to be linked to such a marker. These studies are interesting, but it is not clear whether gene mapping will proceed using the strategy that these authors recommend. The last chapter, by Asmussen and Clegg, discusses the use of DNA markers in
prenatal diagnosis of genetic diseases. The results are correct, but I found their evolutionary analysis of the usefulness of marker genes irrelevant in the context of medical genetics.

This book has been printed attractively and contains few typographical errors. However, on page 53, formula 18 should be \( \theta / ((1 + \theta) \log n) \), and on page 97 the \( O^*(n) \) should be \( O^*(\eta) \). The editor has done an excellent job but he has slipped once—the reference to a human-rodent hybrid (line 9, p. 183)! I presume this is a human-rodent cell hybrid.

This is the first book that has attempted to gather statistical methods for analysis of DNA data, and I enjoyed it, but I also expected a great deal more. Considerable DNA sequence data exist, and it would have been useful to include more analyses than provided. This book will be useful to students of genetics and evolution, particularly for newcomers.

ARAVINDA CHAKRAVARTI
University of Pittsburgh


The early ancestors of all DNA-containing organisms had only a few genes. Today, the amount of DNA per genome encompasses four orders of magnitude from about \( 4 \times 10^6 \) nucleotide pairs (np) in bacteria and \( 4 \times 10^7 \) in fungi to more than \( 10^{10} \) in some primitive fishes, salamanders, and many plants. Polyploidy is one process by which genome size increases. Wen-Hsiung Li reviews other, more fundamental mechanisms that have been elucidated in recent years. One mechanism is gene elongation, an increase in gene size that may occur by tandem duplication of relatively short nucleotide sequences. An extreme example is a gene coding for collagen in chickens that consists of more than 50 exons each made up of five to a dozen tandem duplications of one basic sequence 9 np long. Complex genes evolve as well from simpler ones by the joining of small primordial genes having separate functions. The ancestral genes may be recognized in the different exons, each coding in the modern gene for a distinct protein domain, as in the gene encoding the constant region of the heavy chain of immunoglobulin \( \gamma \).

Complete genes also become duplicated in evolution. Some, like those coding for the ribosomal or transfer RNAs, exist in multiple copies that have remained identical with one another in structure and function. Others, such as the globin genes in vertebrates, diverge after duplication and acquire novel but related functions. And then, there are nucleotide sequences of unknown function that are multiplied many times in a genome; for example, the Alu sequences, of which there are about 300,000 in the human genome, each some 300 np long. The evolution of the DNA by duplication is a fascinating problem. I guess that we know only the tip of the iceberg and that exciting discoveries will be forthcoming.

One puzzle is that some duplicated genes or nucleotide sequences remain identical for eons. This phenomenon is now called concerted evolution and may occur by gene conversion or unequal crossing-over. Norman Arnheim argues convincingly against a possible alternative to these processes, namely, natural selection acting separately but in parallel in each of the duplicated sequences. Particularly difficult to understand is the conservation of homology in sequences located in nonhomologous chromosomes, a problem explored by Arnheim using the human ribosomal genes as a model.

Classical genetics established that genes generally occupy fixed locations on chromosomes. The 1983 Nobel Prize in physiology or medicine was awarded to Barbara McClintock for demonstrating that some genetic elements move from
Allan Campbell distinguishes *episomes*, which can replicate independently of the chromosome, from *transposons*, which are found only in the inserted state and are of two types: true transposons, which include genes that determine phenotypic traits; and *insertion sequences*, which have no other known properties besides transposability and promoter or terminator activity. Transposons affect chromosome evolution because they induce rearrangements such as deletions and inversions. Modification of gene regulation may very well be the most lasting effect of transposons on evolution.

Restriction endonuclease digestion and DNA sequencing have laid open the exploration of organelle DNA. The large-scale evolution of animal mitochondrial DNA is reviewed by Wesley M. Brown, who argues for the stability of gene order for hundreds of millions of years, albeit differences exist between insects and vertebrates. These two large groups of organisms differ in DNA composition as well: G and C content amounts to somewhat more than 40% of the mitochondrial DNA of most vertebrates but to only 21% in *Drosophila melanogaster*. Nevertheless, as reviewed by John C. Avise and Robert A. Lansman, the mitochondrial DNA of animals exhibits high-sequence polymorphism within species—several times greater than for nuclear DNA. This mitochondrial polymorphism is mostly due to base substitutions rather than additions or deletions; and transitions prevail over transversions. The situation is altogether different in plants, where structural reorganization of the mitochondrial DNA appears to be common, but sequence homology is largely preserved.

Genetic polymorphism within species is further examined in other chapters. Robert K. Selander and Thomas S. Whittam review the most recent results concerning electrophoretically cryptic protein variation and present a valuable discussion of what we have learned about population structure from the distribution of protein polymorphisms in snails and in humans. Richard K. Koehn, Anthony J. Zera, and John G. Hall examine the contribution of natural selection to the maintenance of polymorphisms, whereas Masatoshi Nei emphasizes the significance of mutation from a neutralist point of view. Motoo Kimura restates with conviction his neutrality theory of molecular evolution in the light of the recent accumulation of DNA sequences.

*Evolution of Genes and Proteins* is a gem. It derives from a symposium held in June 1982 at the State University of New York in Stony Brook. Following strict editing rules, Nei and Koehn have gone as far as seems reasonable in order to integrate the separate contributions of many authors into a coherent book. The book, to the credit of editors and publisher, is well produced. Now may be high time to acknowledge Sinauer as an intelligent and innovative publisher that is contributing quite significantly to the current evolutionary literature.

Francisco J. Ayala
University of California, Davis


This volume, the first in (another) new series of costly monographs, is based on a symposium held in 1980 at the Third International Congress of Systematic and Evolutionary Biology convened in Vancouver, British Columbia. It consists of nine chapters divided into three major sections: I, Eutherian phylogeny and protein evolution (chaps. 1–5); II, Modeling the process of sequence divergence (chaps. 6, 7); and III, Prospects for investigating evolution through genomic DNA (chaps. 8, 9). Chapters 1, 8, and 9 were not part of the Vancouver symposium but were added subsequently to provide nonmolecular information on eutherian
phylogeny (chap. 1) and to describe the newest methods of dissecting and reconstructing genomic DNA evolution (chaps. 8, 9).

The five chapters constituting part I account for nearly 60% of the book. M. Novacek reviews background anatomical and fossil evidence bearing on higher eutherian phylogeny. Five major areas of instability in eutherian phylogeny are identified and discussed in light of evidence from comparative anatomy and palaeontology: (a) the Edentata, (b) the Carnivora, (c) the possible monophyly of the Rodentia, Lagomorpha, and related taxa, (d) the validity of the Archonta, and (e) the ungulate radiation. A composite cladogram for the major eutherian groups is presented and discussed. The chapter ends with a discussion of estimated divergence times for the eutherian orders and estimated rates of evolution among these taxa concluding "both molecular and taxonomic assessments of evolutionary rates in eutherians are limited critically by the imprecision of estimates of divergence dates from fossil and phylogenetic data" (p. 32).

J. J. Beintema and J. A. Lenstra discuss the "Evolution of mammalian pancreatic ribonucleases" based on amino acid sequence information from 35 mammalian species. Phylogenetic trees for the enzymes as well as trees obtained by using other biological information are presented, compared, and discussed—all these trees differ only slightly. Rates of evolutionary change in the ribonucleases are estimated, and considerable variation in evolutionary rates is noted for different taxa. Some discussion is also given to the three-dimensional structure and enzyme activities of the ribonucleases.

W. W. De Jong contributes a summary of comparative studies of vertebrate eye lens proteins. Particular attention is paid to the alpha-crystallin chains, and a summary of amino acid sequence comparisons of lens proteins of 41 mammalian species is presented. The goal is to "establish the most probable pattern of relationships among the mammalian orders" (p. 96). De Jong found that many questions could not be resolved because the proteins studied had evolutionary rates too slow to provide sufficient data to address questions of branching order among the taxa of interest. This slow evolutionary rate is acknowledged as responsible for the highly irregular rate of substitution in the alpha-crystallin proteins.

The longest chapter, by M. Goodman, A. E. Romero-Herrera, H. Dene, J. Czelusniak, and R. E. Tashian, summarizes the contributions amino acid sequence information has made to our understanding of mammalian, especially primate, phylogeny. The logic and methodology of the maximum parsimony algorithm used by Goodman's laboratory for reconstructing phylogenies are reviewed. Both gene lineages and species lineages are reconstructed. Data from "553 polypeptide chains of 244 species" are analyzed and discussed. The recurrent theme throughout this analysis is the intrinsic variance observed in molecular evolution. It is clear from the data that some of the molecules studied exhibit a greater variance in their rates of evolution than others. Nevertheless, it is impressive that a correlation of .88 is obtained between the divergence times of mammalian lineages estimated from the fossil data, and from the overall molecular clock estimates (table 4). (Viewing molecular evolution as a statistical process, with a mean and a variance, seems to be a more powerful concept than viewing it as an absolute clock that sometimes runs fast and sometimes runs slowly.) Goodman et al.'s parsimony analysis identifies two periods of very rapid protein evolution which they suggest correlate with (1) the emergence of the tetrapods and (2) the origins of primates and the earliest eutherians. Much of this discussion has been presented before (references in bibliography), but it is all summarized here and will be of interest to new students of molecular evolution.

L. T. Hunt and the late M. O. Dayhoff cover "Evolution of chromosomal proteins" in the concluding chapter of section I. A brief review of eukaryote
chromosome structure and composition is presented, followed by a description of how computers can be used to detect homologous sequences within and between proteins. These data are then used to investigate phylogenetic relationships among proteins and protein families. The majority of this chapter focuses on the evolution of the histones, with evolutionary trees for histones H1–H5, as well as a phylogeny that includes all of the bovine nucleosome core histones. Trees are also described based on other chromosomal proteins including protamines and nonhistone chromosomal proteins.

Part II consists of a computer simulation study of the evolution of five proteins (M. Coates and S. Stone) and a study of gene and m-RNA structure in three protein families and attempts to define constraints on gene structure to be incorporated into their nonrandom REH evolutionary theory (R. Holmquist, D. Pearl, and T. H. Jukes). Both papers are concerned with modeling the process of molecular evolution to continue assessing whether significant amounts of evolution are indeed "neutral." A major conclusion of Coates and Stone's study is that each protein must be independently tested for its "clocklike" behavior—something that this reviewer had believed was common knowledge for the past decade! Perhaps a reason for continuing "molecular controversies" is the presentation of assertions that no molecule can be used as a clock because some molecules have an unacceptably high variance in their rates of molecular evolution.

The final section of the book was not part of the 1980 symposium but was added to present new methodologies available for the study of gene evolution. A. F. Scott and K. D. Smith describe restriction enzyme mapping of DNA as well as methods of studying all classes of repetitive DNA and single-copy DNA by hybridization analysis. A brief discussion of transposable elements and their role in evolution is also presented. In the concluding chapter D. Hewett-Emmett, P. J. Venta, and R. E. Tashian lucidly describe recombinant DNA technologies from cloning through nucleic acid sequencing and analysis. The evolutionary roles of gene structure (introns, exons, pseudogenes) and gene duplication are also ably presented.

Overall, this book is well presented, with few noticeable errors or typos. However, it is one of an ever-increasing proliferation of expensive volumes of symposia held years before the final publication date. Consequently, much of the information in this volume has already appeared in the reviewed literature. In contrast, other symposia dealing with molecular evolution have been published more rapidly and in paperback at much reduced prices. Such an approach to disseminating symposia papers seems highly preferable in being more timely and more readily available to graduate students and faculty alike.

LINDA R. MAXSON
University of Illinois at Urbana-Champaign