

2006. Sympatric speciation in palms on an oceanic island. *Nature* 441:210–213.
- Thompson, J. N. 2005. *The geographic mosaic of coevolution*. University of Chicago Press, Chicago, Illinois.
- Webb, C. O., D. D. Ackerly, M. A. McPeck, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* 33:475–505.
- Wiens, J. J., and M. J. Donoghue. 2004. Historical biogeography, ecology and species richness. *Trends Ecol. Evol.* 19:639–644.
- Susanne S. Renner, *Department of Biology, Systematic Botany, Munich University, Menzingerstr. 67, D-80638, Munich, Germany; E-mail: renner@lrz.uni-muenchen.de*

Syst. Biol. 55(4):698–700, 2006
 Copyright © Society of Systematic Biologists
 ISSN: 1063-5157 print / 1076-836X online
 DOI: 10.1080/10635150600899780

Statistical Methods in Molecular Evolution.—Rasmus Nielsen (editor). 2005. Springer Verlag, New York. 508 pp. ISBN 0-387-22333-9. \$89.95 (£61.50) (hardback).

Molecular evolution and its sister disciplines of phylogenetics, molecular systematics, and comparative genomics are progressing rapidly. Of particular importance has been the creation of powerful statistical tools, which have enabled biologists to coax increasingly more information from biological data. *Statistical Methods in Molecular Evolution* seeks to provide an introduction to these tools and the mathematical theory that underpins them.

Before discussing the contents of the book, it is useful to linger for a moment on its definition of the phrase “molecular evolution”: only the evolution of biological sequences or things that can be simply inferred from them, such as gene order or microsatellite length, are covered in the book. There is little to no discussion about the evolution of protein structures, alternative splicing, domain shuffling, or the currently fashionable topic of network evolution. This is not a criticism—a book can only cover a limited number of topics—but potential readers should be aware of its limitations. The target audience includes scientists with mathematical or biological backgrounds, with the intention of being accessible to most numerically minded biology research students. This is a difficult task to manage and there are inevitably chapters that many biologists will struggle with, because some ideas are presented in a relatively formal statistical way, or both deal with particularly difficult conceptual issues. The majority of authors, however, do try and guide less technically minded readers through some quite difficult subject matter, and in general they succeed, making the book a pleasant and informative read.

In terms of structure, the book consists of a collection of peer-reviewed chapters from significant innovators in their respective fields, tied together by the editorial hand of Rasmus Nielsen. The individual chapters feel distinct, and each reads as a review in its own right, which leads to the book’s main strengths and weaknesses. The weaknesses are typical of books with a large number of authors. There is redundancy of material, with

some basic topics being covered multiple times, particularly regarding the definition of some popular evolutionary models. There is also a frustrating inconsistency with nomenclature, with different symbols being used for the same term in different chapters. A glaring example of this is the commonly used parameter κ in nucleotide models, which is defined in the usual way as the transition/transversion rate ratio in all but one chapter, where the authors have atypically defined it as the transversion/transition rate ratio. Of more concern is that some topics are only sketchily covered, and some escape being covered at all. This tends to be particularly true for the justification of modeling choices, which occasionally have gaps in the logical progression building from their biological foundations. This will be a particular problem for statisticians using the book as an entry point into molecular evolution. There is also occasional patchiness in quality both within and between chapters. Although the overall quality is high, with some notably concise and informative expositions of difficult subjects, some chapters feel in places that they are drifting.

Dwelling on the negatives too long would be an injustice and, to some extent, miss the point. By pulling together the thoughts and opinions of leading researchers, and letting them write in-depth about their preferred topics, this book has managed a rare thing: to become more than the sum of its parts. It provides an introduction to a lot of the most exciting research going on in molecular evolution, describing it from different perspectives and providing readers with a key to explore different aspects of the literature. This is illustrated by the discussion of models of protein evolution, where chapters introduce perspectives and opinions from structural biology, phylogenetic inference, and statistics. The eclectic mix of topics also invites the reader to think of molecular evolution at different temporal and spatial levels, and to spot difficulties with our current approaches. For example, one chapter discusses evolution over short timescales and how population structure affects natural selection, whereas the next describes recent attempts to draw inferences about selection from more distant evolutionary relationships. The comprehensive coverage of both topics serves to highlight the similarities between

established approaches to this problem and how they might be able to learn from one another. Similarly, several chapters discuss hidden Markov models (HMMs). Two chapters discuss phylogenetic HMMs, where the evolutionary properties of a column in a sequence alignment depend on those of the previous column; a further chapter introduces the HMMs used for sequence alignment that model insertion and deletion. In addition to demonstrating the increasing importance of HMMs to molecular evolution, it may tempt readers to spot the problem that gaps introduce when describing protein structure and context-dependent mutation events.

Proceeding on to the details of the book: it is split into four sections that guide readers first through an introduction to core methodology and then on to three broad topics. The introductory section consists of four chapters that progressively build up a reader's knowledge of the statistical methodology used in molecular evolution. Starting with a discussion of Markov models of evolution, it proceeds through maximum likelihood and Bayesian inference, and concludes with a population genetics perspective. These chapters are all very strong, providing clear and edifying accounts of topics that are not readily available in any other single reference. The final chapter is the only in-depth coverage of population genetics provided in the book and, as such, feels isolated. This is disappointing because one of the pressing problems in molecular evolution is how to combine the benefits of population genetics and phylogenetics.

The second section consists of four chapters covering practical approaches for data analysis, providing detailed descriptions of the practicalities of particular computer programs and the theory behind how they work. It contains what many readers of the book will be immediately interested in: methodology for detecting the selective pressures acting on proteins, the use of molecular clocks, and practical tips on Bayesian inference. To provide an introduction to software whilst simultaneously describing advanced methodology is a difficult task, and each group of authors tries to address this problem in a different way, with varying degrees of success. The two chapters about Bayesian inference (MrBayes) and adaptive evolution (Paml) are probably the most successful at achieving these twin aims. The chapter about HyPhy, which may develop into an important tool for those studying the evolutionary process, has an excessively long account of the inner workings of the program, with only sketchy outlines and justification for the methodology. The chapter on molecular clocks (Multidivtime) goes too far in the other direction, providing an excellent discussion of the relevant theory, which would fit well elsewhere in the book, but offers little reference to the implementation.

The remaining two sections of the book are essentially a collection of reviews. The first of these covers statistical models that describe the evolution of genomic features. I found this to be one of the more interesting and informative sections of the book, which is indicative both of my personal research interests and the neoteric nature of the models discussed. It starts with one of the more estab-

lished subjects in this area, with a chapter on the evolution of proteins. This contains a wide breadth of material presented in a concise and thoughtful manner, covering both the simple models frequently used to infer phylogenetic trees and more complex models that attempt to use evolutionary information to inform about structure and function. Next is a discourse on microsatellite evolution, detailing the progressively more complex ways in which the extension and contraction of these elements have been statistically modeled.

The third chapter introduces the rapidly developing field of genome rearrangements, where the precise base-pair content of the genome is ignored and only the order of genetic markers, such as genes, are considered. It offers a route into what can, conceptually, be quite a difficult subject to understand, but one that is likely to grow in importance. It is conceivable that in the future these models will form the basis of genome aligners, and eventually be combined with nucleotide level models to provide a more useful and comprehensive description of genome evolution. The final chapter discusses what have become known as phylogenetic HMMs. Originally used to describe protein evolution, these models have now found a new lease of life in comparative genomics, where they are routinely used to find coding exons and other highly conserved elements. The methodology is clearly and intuitively described, which will enable those without a mathematical background to appreciate the principles behind phylogenetic HMMs, before developing an understanding based on the equations. The penultimate part of the chapter covers graphical models, giving readers a hook to explore new research directions for themselves.

The final section contains five chapters about inferences on molecular evolution, covering modeling in other important areas of sequence evolution. These include three diverse commentaries on base composition variation (GC content), estimating substitution matrices, and Bayesian inference of ancestral states. These are all worthy contributions, occasionally providing fresh insight into what are for the most part established areas of research. I shall concentrate my discussion on two chapters covering topics that have for some time needed a clear exposition in the literature: statistical alignment and the comparison of phylogenetic trees.

Over 20 years ago Sankoff and Kruskal established that phylogenetic inference and sequence alignment are two sides of the same coin. There has been progress since then, but the overwhelming majority of research has concentrated on the former—just look at the contents of this book! The development of more accurate methodology for alignment is of critical importance to molecular evolution, but statistical approaches have proved difficult and are yet to find widespread usage, due to limitations of the models used and the excessive time they take relative to score-based forms of alignment. This chapter discusses the founding papers of the research area, before progressing on to more complex and intricate approaches to the problem. It can be heavy going in places, but readers will benefit from an up-to-date

account that covers most of the major developments in the field.

The final chapter on assessing the uncertainty in phylogenetic inference addresses the long-established problem of evaluating the confidence that one should have in tree estimates. This chapter provides a balanced and long-needed general review, including both the Bayesian and frequentist perspectives and their relationship to one another. It would benefit any researcher using phylogenetic trees to familiarise themselves with this chapter. The authors conclude by stating that it is too hot a topic at present to issue a single, simple recommendation, but the often technical journey on which the authors carefully take the readers will enable them to make informed decisions.

The final judgement of any book published at the moment on molecular evolution will, perhaps, be interpreted through its comparison to Joe Felsenstein's *Inferring Phylogenies*. As one might expect, this book

covers topics in a narrower and deeper fashion than *Inferring Phylogenies*. The majority of the chapters competently cover their material in an appropriate level of detail, with a number offering outstanding commentary on important but technically difficult topics. The overlap between chapters and the range of backgrounds of the authors benefits the reader by demonstrating the relevance of different topics to different areas of research. On the negative side, the multiple authorships lead to a slightly raw feeling, with repetition and inconsistency, whereas *Inferring Phylogenies* guides the reader through phylogenetics with a warm and precise narration. In summary, *Statistical Methods in Molecular Evolution* is a strong book, with only a few minor faults, and it should rank highly on *Systematic Biology* readers' "to buy" lists.

Simon Whelan, Faculty of Life Sciences, University of Manchester, M13 9PT, United Kingdom; E-mail: simon.whelan@manchester.ac.uk