

Preface

The main objective of this book is to present and explain the statistical methods and computational algorithms developed in molecular evolution, phylogenetics, and phylogeography for the comparative analysis of genetic sequence data. Reconstruction of molecular phylogeny and inference of the molecular evolutionary process are considered problems of statistical inference, and likelihood and Bayesian methods are treated in depth as standard methods of data analysis. Heuristic and approximate methods are discussed from such a viewpoint as well and are often used to introduce the central concepts, because of their simplicity and intuitive appeal. However, the book does not dwell on proofs or mathematical niceties; it emphasizes care but not rigour.

Molecular Evolution: A Statistical Approach represents an expanded and updated treatment of my earlier research monograph *Computational Molecular Evolution*, published by Oxford University Press in 2006. The major change has been the far more comprehensive and extensive coverage of Bayesian methods, while the target audience has been expanded to include upper level undergraduate as well as graduate students. It can also be read by researchers working in such diverse fields as evolutionary biology, molecular systematics, population genetics, statistical phylogeography, bioinformatics and computational biology, computer science, and computational statistics. It is hoped that biologists who have used software programs to analyse their own data will find the book particularly useful in helping them understand the principles of the methods. For applied mathematicians, molecular studies of evolution are 'a source of novel statistical problems' (Neyman 1971), and this book will provide an accessible summary of the exciting and often unconventional inference problems in the field, some of which are yet unsolved.

Although this new book is written at a similar level of mathematical sophistication as my 2006 work, I have taken care to assist the biologist readers who may find the mathematical arguments challenging. First, every important mathematical result is followed by a verbal rendering, and it is reportedly possible to read the book while skipping the equations, at least at first reading. Second, I have included numerous examples of real data analysis and numerical calculations to illustrate the theory, in addition to the working problems at the end of each chapter. Many biologists find numerical calculations less intimidating than abstract formulae. Example datasets and small C and R programs that implement computational algorithms discussed in the book are posted on the web site for the book: <http://abacus.gene.ucl.ac.uk/MESA/>. Third, I have prepared a primer on probability and statistics, with an overview of mathematical results used in this book, for biologists who would like to grapple with the mathematical details in the book. This has been used as the pre-course reading material for an advanced workshop on Computational Molecular Evolution (CoME) that runs annually in Hinxton, Cambridge, and Heraklion, Crete, co-organized by Aidan Budd, Nick Goldman, Alexandros Stamatakis, and me. It is available at: <http://abacus.gene.ucl.ac.uk/PPS/PrimerProbabilityStatistics.pdf>.

The 2006 book was used as a textbook for graduate courses on bioinformatics and computational genomics in Peking University (2010) and in ETH Zurich (2011). I thank the students in those courses for their useful feedback. For instructors, I have found an early

coverage of the simulation chapter to be useful, as afterwards simulation projects can be assigned as homework when other chapters are taught.

I am grateful to a number of colleagues who read earlier drafts of chapters of this book and provided constructive comments and criticisms: Konstantinos Angelis, Mario dos Reis, Ed Susko, Chi Zhang, and Tianqi Zhu. The following colleagues read and commented on Chapter 9: Daniel Dalquen, Adam Leaché, Liang Liu, and Jim Mallet. Needless to say, all errors that remain are mine. (Please report errors and typos you discover to me at z.yang@ucl.ac.uk. Errata will be posted on the book's web site.) Thanks are also due to Helen Eaton, Lucy Nash, and Ian Sherman at Oxford University Press for their support and patience throughout the project.

Ziheng Yang

London

April 2014