

BOOK REVIEW

Probability models for DNA sequence evolution

A probabilist's account of modern molecular population genetics

R Durrett

Springer-Verlag, New York; 2002. 240pp
€69.95, hardback. ISBN 0-387-95435-X

Heredity (2004) 92, 474. doi:10.1038/sj.hdy.6800419

Reviewed by Z Yang

This small book has five chapters, the first four of which cover theoretical population genetics. Chapter 1 is on the basic models. After a barebone introduction to genetics terminology, the author embarks on a formal probabilistic description of genetic drift and the coalescent, introducing the Wright–Fisher model, the Moran model, Ewens sampling formula, and the infinite-sites model. There is also a section on four-state nucleotide mutation/substitution models, but, apart from that, the book does not discuss models of DNA sequence evolution used in molecular phylogenetics, as the book title suggests. Compared with a standard population genetics text, the first chapter of the book has a very fast pace, with no historical perspective of the field, and the biological significance of the models discussed is taken for granted. I found that those features made the book refreshing to read. However, the reader does need some background knowledge of population genetics, as covered in an elementary textbook such as Hartl and Clark (1997).

Chapter 2 extends the basic coalescent model to account for deterministically changing population sizes, recombination, and population subdivision. Chapter 3 deals with natural selection. This covers the stochastic dynamics of allele frequencies under directional or balanced selection, re-deriving a number of results published by Kimura and others several decades ago using the diffusion approximation. The chapter also discusses evolution at a neutral locus linked to a selected locus, that is, background selection and genetic hitchhiking. Chapter 4 describes the popular tests of neutrality, including Tajima's D statistic, Fu and Li's D , the HKA test, and the McDonald–Kreitman test.

In those four chapters, the author did an admirable job in providing a concise and up-to-date summary of modern molecular population genetics. The details of derivations are often given, allowing the reader to follow the theory. In this regard, the author could have helped the reader by describing the assumptions of the models more explicitly; sometimes I had to study the proof to understand the model assumptions. An important area of study that is entirely missing is full likelihood-based coalescent methods, using computation-intensive methods such as importance sampling and Markov chain Monte Carlo. The book is limited to

models and methods analytically tractable, and estimation of parameters from read data examples used in the book typically takes a method-of-moments approach. However, for inference under parameter-rich models incorporating multiple processes such as population size change, recombination, and/or population structure, a method-of-moments approach based on simple statistics such as F_{st} does not seem workable, and methods that make more efficient use of information in the data become necessary. Intensive research in this area in the past 20 years has produced algorithms that are practically useful for real data analysis; see Donnelly and Tavaré (1997) and Balding *et al.* (2003).

Chapter 5 is on genome rearrangement and introduces models for describing chromosome sizes, inference of the number of inversions to transform one genome into another. It also covers translocations and gene duplications. This chapter is somewhat out of place, as it is not population genetics as the other four chapters are, and is not really probabilistic. Inference of rearrangement events is extremely important to comparative genomics. Current algorithms, however, are based on parsimony-style arguments, and attempt to locate the path that transforms one genome into another by a minimum number of rearrangement events. Even such a minimum-path approach already poses serious computational difficulties. To an evolutionary biologist, it would be interesting to estimate the rates at which genome-rearrangement events occur relative to nucleotide substitutions. The data generated from the various genome projects appear to contain sufficient information for such inference. However, a fully probabilistic approach to the problem does not seem tractable as the process of genome evolution over time is difficult to model. It is unclear whether a simulation algorithm might be developed, which might be combined with an approximate Bayes approach to enable computation through Markov chain Monte Carlo. This is a very active area of research and the next few years may well see exciting breakthroughs.

Overall, I found the book great fun to read. I recommend it to population geneticists who are interested in a concise modern summary of analytical results in the field, and to mathematicians who would like to work in this exciting area of research.

References

- Balding D, Bishop M, Cannings C (2003). *Handbook of Statistical Genetics*. 2nd edn. Wiley: New York.
- Donnelly P, Tavaré S (1997). *Progress in Population Genetics and Human Evolution*. Springer-Verlag: New York.
- Hartl DL, Clark AG (1997). *Principles of Population Genetics*. Sinauer Associates: Sunderland, MA.

Z Yang

Department of Biology, University College, London
E-mail: z.yang@ucl.ac.uk