

Preface

The idea that all life on earth traces back to a common origin dates back at least to Charles Darwin's *Origin of Species*. Ever since, biologists have tried to piece together parts of this "tree of life" based on what we can observe today: fossils, and the evolutionary signal that is present in the genomes and phenotypes of different organisms. Mathematics has played a key role in helping transform genetic data into phylogenetic (evolutionary) trees and networks. In this book, I will explain some of the central concepts and basic results in phylogenetics, which benefit from several branches of mathematics, including combinatorics, probability, and algebra.

As well as providing an overview of this field, I have also tried to highlight many of the advances in this field that have taken place since my earlier book with Charles Semple (*Phylogenetics*, Oxford University Press, 2003). It is quite amazing how much this field has developed in the past dozen years. It soon became clear that one of the hardest tasks would be deciding what to leave out in order to complete this book on schedule. Rather than attempting a comprehensive survey of the current state of the art, this book provides an updated summary of the main theory, supplemented by a selection of topics that have mathematical appeal and either proven or potential biological relevance. In confining the scope to mathematical topics, there is comparatively little detail about the many impressive advances that have been made by the theoretical computer science community in developing more efficient algorithms or resolving outstanding computational complexity questions.

The chapters roughly follow the outline of the series of 10 lectures I gave at the NSF/CBMS Conference on Mathematical Phylogeny (Winthrop University) in June 2014, and is based around my November 2014 survey paper on phylogenetics in a special issue of the *American Mathematical Monthly* devoted to mathematical biology.

Where some material overlaps with the earlier book, *Phylogenetics*, I have tried to ensure that the presentation here is briefer, presented differently, or updated. For example, there is more emphasis here on stochastic models, particularly in the new material in Chapters 3, 8, and 9. Topics such as phylogenetic diversity are now described, and there is a more detailed treatment in Chapter 9 of how species trees and gene trees evolve. Chapter 10 provides an introduction to the very active research area of phylogenetic networks (which complements the excellent 2010 book [203] on this topic). Several of the specialized topics in the earlier book are also omitted here.

In this book, phylogenies (rather than the more abstract X -tree notion) are central, hierarchies rather than split systems take precedence, and rooted trees are drawn growing upwards like biological trees (however, in the final chapter, I bow to convention and orient phylogenetic networks downwards). Keeping the mathematical requirements modest has precluded descriptions of some of the more technical results in detail. However, I hope this makes the text a bit more accessible for readers from other disciplines, particularly biologists, who wish to better understand the mathematical foundations of phylogenetics.

Regarding the structure of this book, most chapters rely on earlier ones in some way, though some chapters can be read without having covered all the earlier material; Fig. 1 illustrates the main flow of concepts between the chapters.

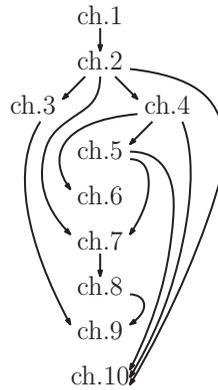


Figure 1. *The main flow of concepts between the chapters of this book.*

A number of exercises (and some examples) are highlighted in boxes throughout the chapters. The exercises marked with + are slightly more challenging than the others.