

Preface

This book is intended to serve as an introduction to the rapidly developing area in comparative biology known as phylogenomics. Approximately 20 years ago, concurrent with the rapid and efficient sequencing of full genomes from living organisms, Jonathon Eisen coined the term "phylogenomics," a combination of phylogeny, which refers to the process whereby evolutionary trees are generated, and genomics, which represents the endeavor of obtaining genome-level data from organisms. Phylogenomics has developed into an important and compelling discipline in its own right over the past two decades. We developed this book in response to the students we have encountered over the last several years who are interested in applying genomics to comparative biology, specifically to phylogenetic, evolutionary, and population genetics problems.

Phylogenomics: A Primer is for advanced undergraduate students and graduate students in molecular biology, comparative biology, evolution, genomics, biodiversity, and bioinformatics. Depending on their educational training, students can focus on the topics in the book that are of the most interest to them. Students who do not have strong backgrounds in evolution or phylogenetics will find the chapters that discuss evolutionary principles and the manipulation of phylogenomic-level data particularly useful. Conversely, students who are adept in ecology, taxonomy, and biodiversity will have the opportunity to learn about the evolution of genes and populations at the phylogenomic level and become familiar with applying phylogenomics to their research.

We believe that there is no better way to understand the information that has been obtained about genes and genomes from life on this planet than to place it into context with the grand evolutionary experiment that has unfolded over the past 3.5 billion years. To this end, we have designed this book as a journey from the basic principles on which organic life has evolved, to the role of burgeoning databases in elucidating the function of proteins and organisms, and concluding with an interpretation of linear sequence information in the framework of organismal change.

Molecules are the currency of modern genomics and have an underlying linear arrangement of their component parts; that is, proteins and DNA can be broken down into linear sequences of amino acids and nucleotides, respectively. Chapters 1 and 2 present the essential principles underlying molecular biology and describe classical techniques used to analyze molecular sequences, including several high-throughput techniques that are known as "next-generation" approaches. Chapter 3 explores evolution at the population level and introduces phylogenetic tree building. As a convenience, we make a simple demarcation between the evolutionary studies that focus on populations (microevolution) and those that focus on species relationships at higher-level systematic relationships (macroevolution).

Chapters 4 through 9 discuss the storage and manipulation of genomics-level data to enable the generation of the data sets that are used in phylogenomics. These processes include accessing databases and Web programs such as PubMed, GenBank, and BLAST for downloading DNA and protein sequences, aligning

linear sequences, producing matrices for evolutionary analysis, and assembling and annotating genomes.

Chapters 10 through 17 focus on the construction of evolutionary trees via phylogenetic matrix construction. Various approaches to phylogenetic analysis are presented, including distance, likelihood, parsimony, resampling, and Bayesian inference. In addition, the phenomenon of incongruence in relation to tree building is described, as are the methods and programs by which phylogenomic trees are generated.

Chapters 18 through 21 focus on the application of modern phylogenomics at the gene and population levels. The transformation of population genetics by the use of DNA sequence information, the detection of natural selection on genes derived from genomic data, and the application of genome-level approaches to population genetics are essential to the understanding of natural populations in an evolutionary context.

The book concludes with a discussion of the basic applications of phylogenomics in the context of modern genome research. Chapter 22 demonstrates how phylogenetic matrices are constructed. Chapter 23 examines the use of genome content to understand evolution and phylogeny. The role of phylogenomics in biodiversity studies, specifically the construction of the tree of life, is explored in Chapter 24. Chapter 25 describes how genomics can be applied to environmental and ecological questions. The final chapter describes how phylogenomic data are used to study the function of genes and proteins.

Working through the applications described in this book does not require an extensive computer science background beyond basic skills such as using a terminal or Web browser. We present succinct summaries of each chapter and problem sets for the student to work through for each chapter. In addition, we include two chapters that describe the various programs that are used in phylogenomics analysis of species relationships (Chapter 17) and genome-level population genetics (Chapter 22). A few stand-alone programs are referenced as well, many of which can be downloaded and installed on either a Mac or PC.

Rob DeSalle
New York, New York

Michael Tessler
New York, New York

Jeffrey A. Rosenfeld
Newark, New Jersey