Book Review

Bioinformatics for Biologists.

It is impossible to imagine modern high-throughput biology without bioinformatics. A modern biologist uses bioinformatics daily: designing primers for a PCR, interpreting microarray data or browsing a genome database. Bioinformatics has become an independent discipline such as genetics, molecular biology or immunology. Hence, it is clear that training in biology must adjust accordingly to include bioinformatics courses both for professional bioinformaticians and biologists who will use bioinformatics in their research. This book, Bioinformatics for Biologists, takes a first, correct, step by providing biologists with an introduction to the methods and algorithms of bioinformatics, while assuming minimal knowledge in computer science. The book’s key approach focuses on presenting biological problems and the bioinformatics tools available to address these problems along with their expected results and interpretation.

The book consists of five parts, each of which includes several chapters. Part I, Genomes, covers genome sequence analysis and applications of genetic variants such as single-nucleotide polymorphisms (SNPs). Chapter 1 describes how to find the genetic basis of disease and what precautions one has to keep in mind when performing such analysis. Chapter 2 addresses the task of selecting a tag SNP, which identifies, or tags, a haplotype block. Using only tag SNPs, rather than all SNPs in an association study, reduces genotyping cost. The author shows how the tag SNP selection problem can be interpreted as a set-covering problem and an integer-programming problem. The following chapter, Genome reconstruction: a puzzle with a billion pieces, describes the theory behind DNA sequence assembly. Here, authors cover the history and basis of Eulerian cycle, Hamiltonian cycle and De Bruijn graph and how they are implemented in genome assembly. Chapter 4 describes the dynamic programming algorithm and how it is applied to sequence alignment and gene prediction. In the next chapter, the author discusses the ways of testing genetic inheritance on the paternity inference example. The chapter explains the basic principles of statistical inference using Bayes’ Law and covers principles of probability statement formulation.

Part II, Gene Transcription and Regulation, includes three chapters. Chapter 6 describes how GC skew, G and C nucleotides strand asymmetry, is used in prediction of replication origin and terminus. The author discusses a theory that links genome changes with GC skew. The following chapter reviews transcription factor DNA-binding site modeling. It covers approaches for implementing position dependencies and methods of finding motifs in DNA sequences. Chapter 8 describes an algorithm for determining polysaccharide–protein-binding motifs using viral hemagglutinin protein and host glycans as an example.

In Part III, Evolution, chapters describe bioinformatics tools being used to unravel details of natural selection and evolution. Chapter 9 summarizes basic principles of genome rearrangements and algorithms that have been established for their analysis. The next chapter discusses a place for the Tree of Life and horizontal gene transfer in evolution theory. Upon conducting comparative analysis of phylogenetic trees, the authors conclude that the Tree of Life seems to be the major trend in evolution, despite the presence of a substantial amount of horizontal gene transfer events. Chapter 11 reviews large-scale genome rearrangements, and how these rearrangements are used in construction of phylogenetic trees.

Part IV, Phylogeny, has many interconnections with the previous chapters. Chapter 12 describes a cophylogeny reconstruction as a technique used to study coevolution. The next chapter covers approaches for constructing consensus phylogenetic trees, using the pantherine lineage of cats as an example. Also, the authors reviewed different evolutionary hypotheses and methodologies for
reconstructing phylogenetic trees of these big cats. Chapter 14 discusses phylogenetic construction using maximum parsimony and heuristics approaches, as well as computational difficulties that are involved in phylogeny estimation.

Part V, Regulatory Networks, includes two chapters that discuss how networks can be used to model many different biological processes. Chapter 15 provides a definition of networks and examples of networks of biological data. Also the author reviews the most commonly used network models, approaches for comparison and alignment of networks, and how network topology can be used to discover biological function. The last chapter of the book describes the inference of genetic regulatory networks from gene expression data. The author concentrates on development of a model for network inference, finding the best model and how the model can be applied to other biological data.

Each chapter is written by an expert(s) in the field and has its own introduction, hence every chapter is independent of others and the book can be read in any order. The book has many great, colorful illustrations and I enjoyed reading it. At the accompanying website, one can download presentations and other supporting material, which is a valuable resource for both students and teachers. Since the book was written for life science students, it is important to keep mathematics at an easy-to-digest level. This task was successfully accomplished in many chapters, though there are couple mathematically intense subjects that could use further simplification.

Overall, this book covers many bioinformatics topics. I believe several topics could be expanded in a future version of the text. First on that list are sequence alignment algorithms, since BLAST, one of most cited methods, has not been well covered in the book. In addition, it would be excellent if the next edition includes some coverage of approaches to explore the results of high-throughput technologies, in particular microarray and next-generation sequencing; protein and DNA structural analysis; gene finding methods and gene ontology enrichment analysis.

In conclusion, this book covers many topics in bioinformatics with a clear description of algorithms and is a good introductory textbook for biologists wishing to learn more about bioinformatics methods.

Artem Zykovich
Buck Institute for Research on Aging
8001 Redwood Boulevard, Novato
CA 94945, USA
E-mail: azykovich@buckinstitute.org