The book was found
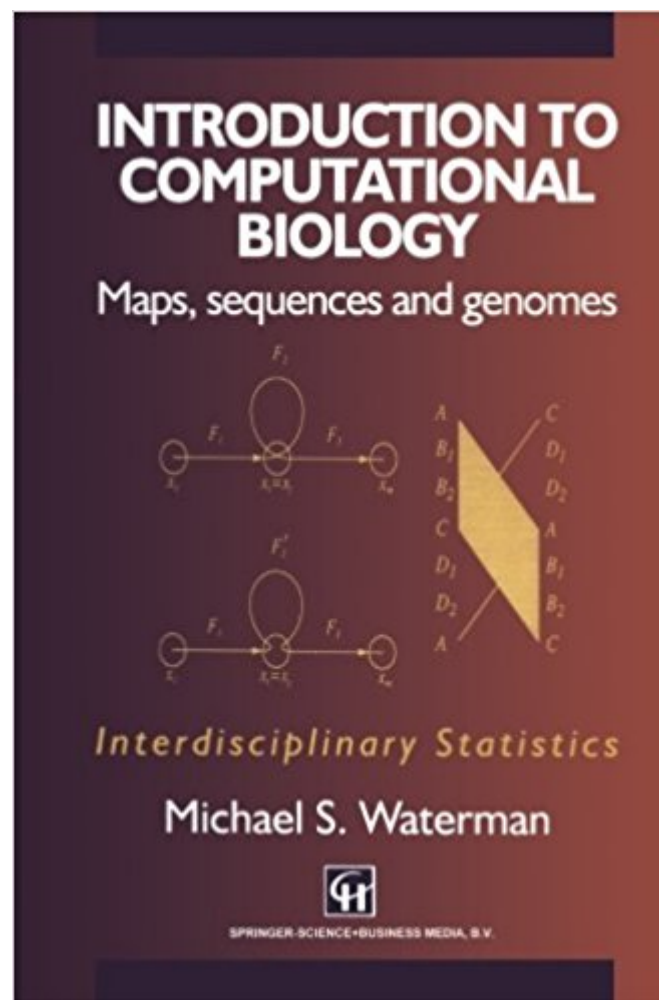
Introduction To Computational Biology: Maps, Sequences And Genomes (Chapman & Hall/CRC Interdisciplinary Statistics)

INTRODUCTION TO COMPUTATIONAL BIOLOGY

Maps, sequences and genomes

Interdisciplinary Statistics

Michael S. Waterman

SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.

Synopsis

Biology is in the midst of a era yielding many significant discoveries and promising many more. Unique to this era is the exponential growth in the size of information-packed databases. Inspired by a pressing need to analyze that data, Introduction to Computational Biology explores a new area of expertise that emerged from this fertile field- the combination of biological and information sciences. This introduction describes the mathematical structure of biological data, especially from sequences and chromosomes. After a brief survey of molecular biology, it studies restriction maps of DNA, rough landmark maps of the underlying sequences, and clones and clone maps. It examines problems associated with reading DNA sequences and comparing sequences to finding common patterns. The author then considers that statistics of pattern counts in sequences, RNA secondary structure, and the inference of evolutionary history of related sequences.Introduction to Computational Biology exposes the reader to the fascinating structure of biological data and explains how to treat related combinatorial and statistical problems. Written to describe mathematical formulation and development, this book helps set the stage for even more, truly interdisciplinary work in biology.

Book Information

Series: Chapman & Hall/CRC Interdisciplinary Statistics

Hardcover: 448 pages

Publisher: Chapman and Hall/CRC; 1 edition (June 1, 1995)

Language: English

ISBN-10: 0412993910

ISBN-13: 978-0412993916

Product Dimensions: 6.1 x 1 x 9.2 inches

Shipping Weight: 1.6 pounds (View shipping rates and policies)

Average Customer Review: 3.6 out of 5 stars   4 customer reviews

Best Sellers Rank: #1,093,387 in Books (See Top 100 in Books)   #43 inÃ Â Books > Science & Math > Mathematics > Applied > Biomathematics   #284 inÃ Â Books > Computers & Technology > Computer Science > Bioinformatics   #822 inÃ Â Books > Engineering & Transportation > Engineering > Bioengineering > Biotechnology

Customer Reviews

"I very much enjoyed the book, and was delighted to recommend itthe use of molecular biology to introduce and illustrate application of sophisticated mathematical theory was excellentas an

illustration of the challenges and rewards of collaborative work, it is ideal." -Statistics: Monash University

Richard C. Deonier is Professor Emeritus in the Molecular and Computational Biology Section of the Department of Biological Sciences at the University of Southern California. Originally trained as a physical biochemist, His major research has been in areas of molecular genetics, with particular interests in physical methods for gene mapping, bacterial transposable elements, and conjugative plasmids. During 30 years of active teaching, he has taught chemistry, biology, and computational biology at both the undergraduate and graduate levels. Simon TavarA(c) holds the George and Louise Kawamoto Chair in Biological Sciences and is a Professor of Biological Sciences, Mathematics, and Preventive Medicine at the University of Southern California. Professor TavarA(c)'s research lies at the interface between statistics and biology, specifically focusing on problems arising in molecular biology, human genetics, population genetics, molecular evolution, and bioinformatics. His statistical interests focus on stochastic computation. Among the applications are linkage disequilibrium mapping, stem cell evolution, and inference in the fossil record. Dr. TavarA(c) is also a professor in the Department of Oncology at the University of Cambridge, England, where his group concentrates on cancer genomics. Michael S. Waterman is a University Professor, a USC Associates Chair in Natural Sciences, and Professor of Biological Sciences, Computer Science, and Mathematics at the University of Southern California. A member of the National Academy of Sciences and the American Academy of Arts and Sciences, Professor Waterman is Founding Editor and Co-Editor in Chief of the Journal of Computational Biology.His research has focused on computational analysis of molecular sequence data. His best-known work is the co-development of the local alignment Smith-Waterman algorithm, which has become the foundational tool for database search methods. His interests have also encompassed physical mapping, as exemplified by the Lander-Waterman formulas, and genome sequence assembly using an Eulerian path method.

This book gives a good survey of the different techniques employed by computational biologists. After a brief review of molecular biology in Chapter 1, the author treats the mathematical modeling of restriction maps in Chapter 2 using graph theory. His presentation is somewhat hurried, but he does give references and gives the reader three exercises at the end of the chapter. Multiple maps are treated in Chapter 3, wherein the author first makes use of probability theory, via the Kingman subadditive ergodic theorem. The proof is omitted but the author does a good job of explaining its

use in studying the double digest problem (DDP). The best part of this chapter is the author's explanation of the difficulties of using Kingman's results for solving the DDP, and goes on to discuss multiple solutions of the DDP. Graph theory is again used in the discussion. This sets up the discussion in Chapter 4, which outlines algorithms for the DDP. The author gives a very compact introduction to P- and NP-complete problems in the theory of computation, then proves that DDP is NP-complete. The author does a good job of discussing subsequent approximate methods used for the DDP, such as simulated annealing. Markov chains are introduced in the book here for the first time, but due to the shortness of the presentation, the reader should do outside reading as a back-up. The author does a great job of explaining the difficulties if measurement error is introduced in the DDP at the end of the chapter. Cloning is discussed in Chapter 5, with tools from probability theory used to deal with partial digest libraries. The chapter is really short though, and the working the problems at the end of the chapter is essential for the understanding the results of this chapter. The author switches gears in the next chapter, wherein physical maps are discussed. The discussion is fairly detailed and interesting. Sequencing is discussed in the next two chapters, and the treatment is very good. Hashing is introduced here, and psedocode is given throughout. The very important method of dynamic programming is outlined in Chapter 9, which is beautifully written, and again pseudocode abounds throughout. Genetic mapping is left out though, but the this, the longest chapter of the book, is a detailed introduction to this area. The results in this chapter are used to study multiple sequence alignment in Chapter 10, wherein hidden Markov models are introduced for the first time. The discussion of these models is very curt, but there are other books and notes available if the reader needs further guidance. The best chapter of the book follows, which discusses probability and statistics for sequence alignment. The theory of large deviations is brought in, and the author does an excellent job of discussing this important, and powerful theory. The reader's level of mathematical sophistication is assumed to be a lot greater than the rest of the book in this chapter. Knowledge of measure theory and martingales are assumed here. The author uses the very powerful tool of relative entropy, so indispensable in other applications of probability. The problem set at the end of the chapter is challenging but working them through is definitely worth the time involved. The next chapter also uses some heavy guns from probability theory to study sequence patterns. The author returns to matter of a more empirical nature in Chapter 13, which deals with RNA secondary structures. The reader with a background in simple combinatorial theory should find the reading straightforward and informative. Continuous-time Markov chains are introduced in the next chapter to study trees and sequences. The treatment here is rather hurried, so again the reader should work the exercises at the end of the chapter. The book ends with a

discussion of the literature and references. All in all a very nice book, worth the price, and worth spending time reading. The only minus might be the total omission of actual source code, but that really was not the intent of the book. Readers with a strong mathematical background will like the book, as well as anyone interested in going into the area of computational biology.

Advances in genetics and gene mapping has led to new challenges for statisticians. Dr. Carlson and "wired weird" have each given excellent reviews of the contents of this book. I need not repeat what they have said. Instead I choose to explain the importance of the material in the book so that you can appreciate the value of it.Microarray technology and the mapping of the human genome have enabled researchers to test genetic reactions to chemicals such as those that can be used in drug development. It is the hope of the pharmaceutical industry that drug products will be produced to target specific diseases and to use our newly acquired knowledge of human genetics to discover which patients will respond positively to a particular treatment and which patients will not. Such knowledge could improve the speed of drug development and save a great deal of the research money that is currently needed to take a drug through the long development process.Noisy data and multiplicity of testing create statistical problems familiar to statisticians but of a magnitude that requires new approaches. Also early efforts in microarray analysis were unsuccessful because of a lack of good experimental design. As statisticians become more and more involved in the design issues the methods are greatly improved and the chance to gain additional knowledge from the experiment is enhanced.This book provides the methods that researchers need to do the research needed in the exciting new discipline called "Computational Biology."

The first name people learn in bioinformatics is the Smith-Waterman algorithm. Some people never learn anything else. This is by that Waterman. Although written in 1995, it still has some of the best discussion I've seen on the topics it addresses.The first few chapters deal with the "digest problem," reconstructing a DNA or protein sequence from the fragment sizes of enzyme digests. The technique is not used as much now as it was then, but it's always good to know the background of modern techniques.The digest problem doesn't stand alone, though. It introduces concepts - islands, anchors, etc. - that still matter. The problems in reconstructing molecules from digests yield the same kinds of intermediate results and the same ambiguities that arise in modern sequencing. As Waterman advances the discussion, shotgun sequencing appears as a logical extension, at least mathematically, of digest assembly.Sequence assembly involve end matching, perhaps in the presence of sequencing errors. That introduces the topic for which Waterman's name is famous,

approximate string matching. The next few chapter progress through dynamic programming and multiple alignments. The logical connections between the techniques shown are so tight that chapter boundaries are almost artificial. It was a real pleasure to see the computational and practical relationships laid out.The final topics, RNA structure and phylogenetic trees, lack the continuity that characterized the first dozen chapters. The RNA structure may be the weakest chapter in the book, but still a very competent introduction.Throughout, Waterman emphasizes mathematical rigor without insisting on uninformative theorems. Every topic is presented in rich detail, with special attention to scoring and background models. Perhaps there are newer discussions of some topics. I don't know of any clearer discussions, though. Best, I think, is how Waterman prepares the reader to ask all the right questions in any future discussion: what are the elements of the computation, how can elements be recombined, how good is a result, and how does the result stand out from the statistical background.The final chapter is what a bibliography should be. It doesn't just list authors, titles, and dates of publication. It actually discusses the contribution that each source made to this book. Rather than leave the reader to wander aimlessly among obscure titles, Waterman shows which sources are most informative on which topics. I wish more authors took the time for such commentary.This is a book worth having. It covers topics that I haven't seen elsewhere, and shows how many different topics relate to each other. It is rigorous without giving distracting detail. Most of all, it keeps the biology in sight of all calculations. Some authors seem to forget that anything exists but the arithmetic; Waterman puts the math clearly in the service of its subject. I enjoyed it immensely, and look forward to applying its content in my own research.

This text was really difficult to read; too many symbols, not enough examples. I didn't find it useful at all.

Download to continue reading...

Introduction to Computational Biology: Maps, Sequences and Genomes (Chapman & Hall/CRC Interdisciplinary Statistics) An Introduction to Systems Biology: Design Principles of Biological Circuits (Chapman & Hall/CRC Mathematical and Computational Biology) Statistics and Data Analysis for Microarrays Using R and Bioconductor, Second Edition (Chapman & Hall/CRC Mathematical and Computational Biology) Introduction to Proteins: Structure, Function, and Motion (Chapman & Hall/CRC Mathematical and Computational Biology) Algorithms in Bioinformatics: A Practical Introduction (Chapman & Hall/CRC Mathematical and Computational Biology) Computational Statistics Handbook with MATLAB, Third Edition (Chapman & Hall/CRC Computer

Science & Data Analysis) RNA-seq Data Analysis: A Practical Approach (Chapman & Hall/CRC Mathematical and Computational Biology) Environmental and Ecological Statistics with R, Second Edition (Chapman & Hall/CRC Applied Environmental Statistics) Introduction to High Performance Computing for Scientists and Engineers (Chapman & Hall/CRC Computational Science) Variational Methods in Image Processing (Chapman & Hall/CRC Mathematical and Computational Imaging Sciences Series) Generalized Linear Models, Second Edition (Chapman & Hall/CRC Monographs on Statistics & Applied Probability) Practical Statistics for Medical Research (Chapman & Hall/CRC Texts in Statistical Science) Dynamic Prediction in Clinical Survival Analysis (Chapman & Hall/CRC Monographs on Statistics & Applied Probability) Measure and Integral: An Introduction to Real Analysis, Second Edition (Chapman & Hall/CRC Pure and Applied Mathematics) Introduction to Set Theory, Third Edition, Revised and Expanded (Chapman & Hall/CRC Pure and Applied Mathematics) Introduction to Scientific Programming and Simulation Using R (Chapman & Hall/CRC The R Series) Introduction to Scientific Programming and Simulation Using R, Second Edition (Chapman & Hall/CRC The R Series) Introduction to Modern Cryptography, Second Edition (Chapman & Hall/CRC Cryptography and Network Security Series) Design of Experiments: An Introduction Based on Linear Models (Chapman & Hall/CRC Texts in Statistical Science) Introduction to Stochastic Processes (Chapman & Hall/CRC Probability Series)