

The first step toward a definitive bioinformatics textbook

Bioinformatics: Sequence and Genome Analysis

By David W. Mount

Cold Spring Harbor Laboratory Press, \$95.00, ISBN 0-87969-597-8, 2001

Reviewed by A. Jamie Cuticchia

The Hospital for Sick Children, Toronto, Ontario, Canada

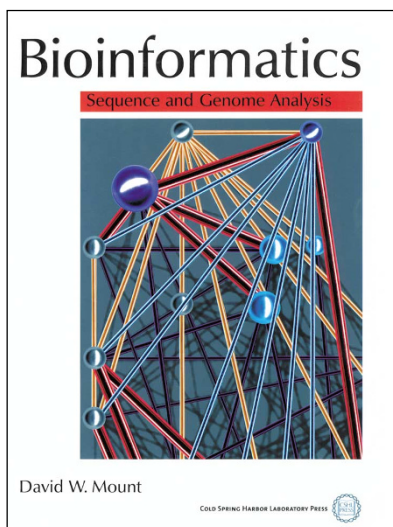
There are as many definitions of the word 'bioinformatics' as there are people who are willing to give one. Thus, to title a book *Bioinformatics*, even with the more specific subtitle, *Sequence and Genome Analysis*, begs for clashes and controversies over whether or not it lives up to its name. The book, written by University of Arizona professor David W. Mount, and published by Cold Spring Harbor Laboratory Press, helps by clearly stating in its glossary that bioinformatics is "an interdisciplinary field involving biology, computer science, mathematics, and statistics to analyze biological sequence data, genome content, and arrangement, and to predict the function and structure of macromolecules." So there.

In the past few years, there has been an explosion in the number of titles available on this topic. Authors like Waterman and Pevzner take a mathematical approach, whereas Bishop showcases a collection of tools and databases with a nearly exclusive European focus. Letovsky, on one hand, concentrates primarily on a single topic (databases), whereas Baxeianis and Oullette find it appropriate merely to collect a series of disjointed papers on topics they presumably find interesting, resulting in a publication that looks more like a journal than a book.

What has been missing from this field is a definitive text that sets the tone for future topics that might someday be included in an expanded definition of bioinformatics. Mount has provided such a text, and educators throughout the world who are struggling to create new courses in bioinformatics should rejoice. Will this book meet all the needs of all the people interested in this topic? Of course not. It may be a decade or more before a 'standard' text, such as Lehninger's for biochemistry, exists for this field. As the discipline of bioinformatics develops in what Jim Clark, founder of Netscape,

would call 'Internet time,' there will no doubt be many books on subspecialties of bioinformatics that, one hopes, will merge to form a definitive text.

To understand what this book is, one can start by addressing what the book is not. This is not a book for people who are unfamiliar with genetics. Although com-



puter scientists may well find this book an interesting overview of sequence analysis, the book concentrates far more on educating the reader on the principles behind the analysis, and not predominantly on the computer science involved. The book also is not for the researcher who just received an account on the departmental bioinformatics server and wants to know how to run the programs. The author, to his credit, goes out of his way to stick to the principles of the analysis (for example, how one computes a homology score) rather than providing a cookbook for using particular software packages.

Bioinformatics is for the biologist who wants to learn more about the fundamentals of DNA sequence analysis. An analogy

for the target audience would be readers who want to know the components of an automobile that make it go, rather than seeking information on how to drive from Tucson to Santa Fe, or on the physics of an internal combustion engine. This audience is probably the largest and, until now, the most neglected.

Totalling 564 pages divided into ten chapters, with an outstanding index and somewhat simplistic glossary, the book focuses on the following topics: sequence alignment, RNA secondary structure prediction, laboratory databases, gene and phylogenetic prediction, protein classification and structure prediction (very basic), and whole genome analysis. Strongest in the area of alignment and slightly weaker when the author attempts to discuss large scale analysis and proteins, the book contains no egregious statements or glaring omissions. The text is well formatted and easily read, with many figures and tables. Color is used both effectively and densely. The book is worth purchasing if only for the extensive bibliographies at the end of each chapter. Moreover, the quality and level of explanation in each chapter is generally consistent, something that cannot be said for 'compilation' texts. Arguably, the only section that is out of place is the "History of Bioinformatics," which reads more like a congratulatory testimony to friends of the author (as indicated by the book's dedication) than a definitive history of the development of the field. But at a time when publications such as *Genome Technology* exist as a genomics version of *People* magazine, this author's folly is forgivable.

Acknowledging the rapid evolution of bioinformatics, Mount highlights an accompanying web site to the book (<http://www.bioinformatics.org>), where he intends to make errata, updated information, and problem sets with solutions available to purchasers of the book. This is laudable, although the quality and sustainability of the site remains to be seen. That said, if the same care is taken in its maintenance as is evident in the production of the text, online readers will not be disappointed.

Bioinformatics is an excellent text for the biologist who wants to learn more about the field, and is well worth exploring by the instructor looking to tackle his or her first bioinformatics course. □