Book Review

Computational Genome Analysis: an Introduction
Richard C. Deonier, Simon Tavaré and Michael S. Waterman

I wish that I had had such a book 5 years ago, when I started to switch my career from being a control engineer to a computational biologist. Like many people following the same path, I was eager to prospect the biomedical fields in which my mathematical and computational expertise could be applied. Similarly, other researchers who were educated first as molecular biologists, will be pleased to see how they can apply computational approaches to solve the problems in their field. This book, Computational Genome Analysis: an Introduction, does a fine job in bridging these two fields together, by explaining the basics of a broad range of areas in computational genomics.

The book begins with a chapter called Biology in a Nutshell, which outlines key biological principles that are important in computational genomics. This section offers a concise overview of the biological concepts for audiences without biology-related backgrounds. Chapters 2 and 3 review various statistical tools that can be used to characterize biological words, i.e. short strings of letters in the genomic DNA and protein sequences. The topics include various probability distributions, introductions to Markov Chains, modeling the number of restriction sites in DNA and so on. Chapters 4 through 8 surround a central theme of the strategies and techniques used in completion of the Human Genome Project (HGP), from physical mapping of DNA to sequence alignments and assembly. This section clarifies the basic concepts and methodologies in describing the human genome, and elucidates the organization and the generation of the immense amount of information arising from the HGP. Chapter 9 highlights statistical frameworks in characterizing DNA signals, in particular, DNA sites where protein binding occurs. Chapters 10 and 11 summarize the techniques used in measuring and analyzing expression of genomic information. This section is potentially appealing for those who are interested in using DNA and protein microarrays or mass spectrometry. Chapters 12 and 13 review the methods for analyzing and describing genetic variations among (phylogeny) and within (population-based study) organisms. The book is concluded with a chapter, Comparative Genomics, by introducing how the variety of tools covered in the previous sections can be employed in the context of complete genomes.

As a textbook that was based on the authors’ lecture notes for a Computational Biology course for seniors and 1st-year graduate students at the University of Southern California, this book is organized in an incredibly friendly fashion. I especially like the section called ‘The Biological Problem’, at the beginning of each chapter, which helps the student focus on the biological topics while various mathematical formulations are discussed. Readers with only a basic programming background will love the fact that most analytical methods presented in the book are accompanied by implementations in the R language, a statistical environment that is fairly popular among statisticians and bioinformaticians.

Overall, this book covers many important concepts and areas in computational biology, and serves as a great guide for the researchers from a variety of backgrounds to start their path into this exciting interdisciplinary field.

Yunlong Liu
Division of Biostatistics,
Department of Medicine,
Center for Computational Biology
and Bioinformatics, Center for
Medical Genomics,
Indiana University School
of Medicine, Indianapolis, IN 46202 USA
E-mail: yunliu@iupui.edu

© The Author 2007. Published by Oxford University Press. For Permissions, please email: journals.permissions@oxfordjournals.org