

## Shotgun coverage of human genome computing

Human Genome Computing,  
Second Edition

edited by **Martin J. Bishop**, Academic Press, 1998. US\$69.95 (xiv + 306 pages) 0 12 102051 7

My introduction to bioinformatics was annotating a genome sequence with paper and pen. My dog-eared printout of the 166-kb bacteriophage T4 genome sequence was about 80 pages in length, with every gene highlighted in colored marker. I don't expect to repeat that exercise again. A comparable printout of the human genome sequence would take about 1.5 million pages. The Human Genome Project would not be possible if our revolution were not proceeding hand in hand with a revolution in the Internet and in desktop computing. In particular, not just the generation but the distribution of the data is reliant on cutting-edge computational methods – from object-oriented databases to Java applets to the Web. Pushed to distribute efficiently one of the most complex data sets ever assembled, the genome project has been an early adopter of unstable but promising Web technologies, and the genetics community has surely never been so affected by proprietary-standards wars between major corporations such as Microsoft, Sun and Netscape.

Guide to Human Genome Computing, Second Edition is a multi-author collection of chapters on the various uses of computing in the Human Genome Project. Peculiarly absent from it is any sense of what audience it is intended for. Instead, the preface from the editor, Martin Bishop, is an overview of genome

computing that is almost bleak enough to make me change careers. Bishop speaks from the trenches about the enormity of the project, the numerous difficulties that must be surmounted, and the tedious nature of 'business as usual in factory scale operations' that is so common to the genome labs. OK, I knew all that (though I keep it repressed). But who will benefit from reading this book, and why, and what will they get out of it? Suppose I'm a bench molecular geneticist, and I'm wondering how I can access all this interesting genome data. Then the lucid introductory chapter from Lincoln Stein on accessing Web resources is a godsend. The tables in Michael Rhodes and Ramnath Elasarapu's chapter, on 'Biological Materials and Services', on how actually to get human and model-organism DNA clones shipped from genome labs are probably things I'd Xerox and stick to my office door. Numerous URLs throughout the book – for instance, from Guy Slater's overview of EST-sequence resources or Davidson *et al.*'s overview of gene expression databases – would get pasted into my Web browser's bookmark collection. But I'd look at a chapter such as Stephen Bryant's 'Managing Pedigree and Genotype Data', which has pages of example SQL database code, and head for the nearest pub to recover.

Suppose, however, that I'm a computer scientist who's been hired to do bioinformatics. Then Stein's chapter is old hat, but Bryant's chapter catches my eye, because it contains thoughtful observations on how to get geneticists to communicate complicated schemata in a pidgin language that can be converted to SQL. Other chapters cover theory and implementations in linkage analysis, comparative mapping, radiation-hybrid mapping, sequence-ready clone contig construction, and genome sequencing. These are generally useful overviews, but nobody will learn how to implement

working bioinformatics environments from these chapters, especially because the field's dirty secret is that genome informatics is woven together by an indescribable tangle of Perl scripts. (Larry Wall deserves either a medal or a beating for writing Perl – possibly both.) Simon Dear's chapter on genome-sequencing software is forthright about this, saying 'there can be no prescriptive solution because there is no single software package that addresses all the issues.'

Now suppose that I am a computational biologist, who is interested in algorithms for extracting biological meaning directly from DNA sequence information. There's even something for me here: two chapters on gene-prediction algorithms, which are written by Krogh, and by Milanese and Rogozin. Krogh's chapter includes a nice technical description of how hidden-Markov-model-based genefinders work, but, unfortunately, I can imagine all but a few statistically enlightened readers immediately turning the page.

Overall, I might compare this book to EST-sequence data. It is a grab bag of fragmentary topics, just like dbEST is a grab bag of fragmentary genes. The book provides a shotgun coverage of many disparate applications in genome computing, yielding information that is unquestionably useful, but scattered and incomplete. It is not a complete resource for any topic by itself, but it provides tags (references, URLs, and the phone numbers and email addresses of experts in the field) that, for anyone who has an interest in computational aspects, can lead to fuller understanding of the Human Genome Project.

### SEAN R. EDDY

Dept of Genetics, Washington University School of Medicine, 4566 Scott Ave, St Louis, MO 63110, USA.  
Email: eddy@genetics.wustl.edu

## On science and scientists

I Wish I'd Made You Angry Earlier.  
Essays on Science, Scientists and  
Humanity

by **Max F. Perutz**, Cold Spring Harbor Laboratory Press, 1998. US\$39.00 (354 pages) ISBN 0 87969 524 2

There are not many scientists who can write about science, scientists and humanity – and there are even fewer who can write in a manner that can be understood by scientists and by the general public. Max Perutz, however, can

do both. In this book, he presents 28 articles. Most of them appeared in the last ten years in periodicals such as the *New York Review of Books* – that is, publications that are not read by many scientists. Most of the articles are book reviews, and most of these books are autobiographies or biographies of great scientists. In his reviews, Perutz is able to condense these biographies to the essential. He has often met the scientists in question, or people who knew them, and so he can add revealing anecdotes. For example, in his article on Heisenberg, he mentions that, in 1947, Heisenberg said to a German refugee whose family had been murdered by the Nazis, 'The Nazis should have been left in power longer; then they would have become quite decent'.

Perutz condenses the biographies of Fritz Haber and his wife, Clara Immerwahr, down to 14 pages. What lives and deaths! He describes the facts and leaves judgement to the reader. From Haber he moves to Lise Meitner, Leo Szillard, Werner Heisenberg, Andrei Sakharov, François Jacob and, finally, to a particular enemy alien, Max Perutz. This autobiographical piece is actually the only text that has not been published before. In it, he describes his internment in England as an enemy alien. Together with other interned persons, he was to be shipped on a transatlantic liner away from England. The ship was hit by a torpedo. He states that he 'was always able to act with a certain coolness, that is to say reflecting before taking each