

## BOOK REVIEW

### Bioinformatics, a prodigious melting pot

#### Introduction to bioinformatics: a Theoretical and Practical Approach

SA Krawetz and DD Womble

Humana Press Inc., Totowa, New Jersey; 2003. 746 pp.  
\$89.50, paperback. ISBN 1-58829-241-X

*Heredity* (2004) 92, 594. doi:10.1038/sj.hdy.6800426

Reviewed by B Audit

Worldwide web resources like those available from the European Bioinformatics Institute and the National Center for Biotechnology Information form the highly visible front-end services offered by bioinformatics research. These resources allow numbers of bioinformatics experiments (eg sequence alignments) and a large amount of biological data (eg molecular databases) to be accessible by just a few mouse clicks. But, there is much more to bioinformatics research and computational biology than these online resources. Bioinformatics was founded by combining computational methods and experimental information to further our understanding of biological macromolecules in the context of living cells. Hence, this discipline has roots in biological sciences, of course, but also in computer sciences, statistics, physics and chemistry. It also implies that, to a certain extent, bioinformatics is an ill-defined terminology because it encompasses research interests based on a technical criterion (using a computer to study biological data) rather than on scientific questions. The ambitious aim of the book is to provide an introduction to the theoretical foundations as well as the practical aspects of the diverse facets of bioinformatics. Flipping through the pages of the book, the reader gets a good feeling of the broad range of expertise and scientific problems that make current bioinformatics research. However, there is a drawback to this, because bioinformatics refers to a *melting pot* of different scientific activities, the book lacks a unifying reading line.

The book is a collection of 36 articles written by 39 authors mostly from academia. Usefully, a *Glossary and Abbreviations* list and a *Suggested Reading* section accompany most contributions, a general index is proposed at the end of the book and many URLs pointing to relevant web sites are given along the text. All illustrations are printed in black and white, which is annoying when colour was used to convey information. In that case, the reader has to refer to the CD associated with the book that contains the complete set of figures and captions in PDF format. Compensating for this trouble, figures are also provided in JPEG format, which is very convenient to reuse this material in slide presentations. To complement the practical parts of the book, a selection of software applications and scripts discussed in the text are supplied on the CD.

The articles are grouped into four parts covering the main scientific issues underlying bioinformatics. Part I covers the biochemistry of nucleic acids and proteins, transcription/translation mechanisms and DNA replication, repair and recombination. Interestingly, this introduction to biology also presents some fundamentals about the cell structure and cell signalling, which constitute important knowledge to apprehend sequence information in the context of the cellular system. Part II is devoted to molecular genetics. The first three chapters are general and deal with different aspects of evolution and heredity: the evolution of coding sequences (gene evolution), the structure of noncoding DNA (repetitive DNA) and epigenetic mechanisms essential to the regulation of gene expression (chromatin structure, methylation patterns). The following three chapters are dedicated to genetic diseases, which are a very important consequence of sequence evolution. They cover the molecular basis and heredity of these diseases and provide a review of the online resources about clinical genetics. The last chapter introduces the related topic of population genetics. Part III is an introduction to Unix operating systems. Throughout six chapters, the reader gets an overview of what the Unix shell command line interface is, of how to install and administer a Unix OS as well as bioinformatics software applications and servers, and of how to use command line sequence analysis tools (example of the GCG package). Part IV is the longest spanning half of the book. It presents a number of computer applications and their theoretical foundation for four categories of problems: analysis of nucleic acid sequences from the assembly of sequencing fragments to statistical modelling, large-scale protein database and similarity searching, functional sequence analysis from motif discovery to 3D visualisation, and gene expression profiling with microarrays.

Overall, the book contains very valuable information on a wide range of topics related to bioinformatics research. Part IV forms the core of this work, alone it provides a very pragmatic approach to the main statistical methods and tools that are specific to bioinformatics. This part constitutes an operational overview of what can be achieved with the current state-of-the-art in bioinformatics research. In comparison, some of the topics considered in Part I, II and III are peripheral to bioinformatics. They are more completely introduced in specialised books where more space can be devoted to them. For example, the chapters of Part III, which present how to install, administer and use the Unix OS, are interesting to a curious reader familiar with other types of OS but are not an effective tutorial. Yet, the book forms an assorted patchwork where any reader full of questions about what happens behind the scenes in bioinformatics research will find material to nourish his/her curiosity.

B Audit

Laboratoire de Physique, Ecole Normale Supérieure de Lyon, 46,  
Allée d'Italie, 69364 Lyon Cedex 07, France  
E-mail: Benjamin.Audit@ens-lyon.fr