Computational Biology

The Computational Biology series publishes the very latest, high-quality research devoted to specific issues in computer-assisted analysis of biological data. The main emphasis is on current scientific developments and innovative techniques in computational biology (bioinformatics), bringing to light methods from mathematics, statistics and computer science that directly address biological problems currently under investigation.

The series offers publications that present the state-of-the-art regarding the problems in question; show computational biology/bioinformatics methods at work; and finally discuss anticipated demands regarding developments in future methodology. Titles can range from focused monographs, to undergraduate and graduate textbooks, and professional text/reference works.

Author guidelines: springer.com > Authors > Author Guidelines

For other titles published in this series, go to
http://www.springer.com/series/5769

Jeremy J. Ramsden

Bioinformatics

An Introduction

Second edition

Springer

Jeremy J. Ramsden
Cranfield University
School of Applied Sciences
Bedfordshire, UK
j.ramsden@cranfield.ac.uk

Computational Biology Series ISSN 1568-2684
ISBN 978-1-84800-256-2 e-ISBN 978-1-84800-257-9
DOI 10.1007/978-1-84800-257-9

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2009920488

Printed on acid-free paper

Springer Science+Business Media
springer.com

Mi a tudvágyat szakhoz nem kötők,
Átpillantását vágyuk az egésznek.

IMRE MADÁCH

Preface to the Second Edition

Overview and aims. This book is intended as a self-contained guide to the entire field of bioinformatics, interpreted as the application of information science to biology. There is strong underlying belief that information is a profound concept underlying biology, and familiarity with the concepts of information should make it possible to gain many important new insights into biology. In other words, the vision underpinning this book goes beyond the narrow interpretation of bioinformatics sometimes encountered, which may confine itself to specific tasks such as the attempted identification of genes in a DNA sequence.

Organization and features. The chapters are grouped into three parts, covering the relevant fundamentals of information science, overviewing all of biology, and surveying applications. Thus Part I (fundamentals) carefully explains what information is, and discusses attributes such as value and quality, and its multiple significations of accuracy, meaning, and effect. The transmission of information through channels is described. Brief summaries of the necessary elements of set theory, combinatorics, probability, likelihood, clustering, and pattern recognition are given. Concepts such as randomness, complexity, systems and networks, needed for the understanding of biological organization, are also discussed. Part II (biology) covers both organismal (ontogeny and phylogeny, as well as genome structure) and molecular aspects. Part III (applications) is devoted to the most important practical applications of bioinformatics, notably gene identification, transcriptomics, proteomics, interactomics (dealing with networks of interactions), and metabolomics. These chapters start with a discussion of the experimental aspects (such as DNA sequencing in the genomics chapter), and then move on to a thorough discussion of how the data is analysed. Specifically medical applications are grouped in a separate chapter. A number of problems are suggested, many of which are open-ended and intended to stimulate further thinking. The bibliography points to specialized monographs and review articles expanding on material in the text, and includes guide references to very recently reported research not yet to be found in reviews.

Target audiences. This book is primarily intended as a textbook for undergraduates, for whom it aims to be a complete study companion. As such, it will also be useful to the beginning graduate student.

A secondary audience are physical scientists seeking a comprehensive but succinct guide to biology, and biological scientists wishing to better acquaint them-

selves with some of the physicochemical and mathematical aspects that underpin the applications.

It is hoped that all readers will find that even familiar material is presented with fresh insight, and will be inspired to new thoughts.

The author takes this opportunity to thank all those who gave him their comments on the first edition.

May 2008

Preface

This little book attempts to give a self-contained account of bioinformatics, so that the newcomer to the field may, whatever his point of departure, gain a rather complete overview. At the same time it makes no claim to be comprehensive: The field is already too vast—and let it be remembered that although its recognition as a distinct discipline (i.e., one after which departments and university chairs are named) is recent, its roots go back a long time.

Given that many of the newcomers arrive from either biology or informatics, it was an obvious consideration that for the book to achieve its aim of completeness, large portions would have to deal with matter already known to those with backgrounds in either of those two fields; that is, in the particular chapters dealing with them, the book would provide no information for them. Since such chapters could hardly be omitted, I have tried to consider such matter in the light of bioinformatics as a whole, so that even the student ostensibly familiar with it could benefit from a fresh viewpoint.

In one regard especially, this book cannot be comprehensive. The field is developing extraordinarily rapidly and it would have been artificial and arbitrary to take a snapshot of the details of contemporary research. Hence I have tried to focus on a thorough grounding of concepts, which will enable the student not only to understand contemporary work but should also serve as a springboard for his or her own discoveries. Much of the raw material of bioinformatics is open and accessible to all via the internet, powerful computing facilities are ubiquitous, and we may be confident that vast tracts of the field lie yet uncultivated. This accessibility extends to the literature: Research papers on any topic can usually be found rapidly by an internet search and, therefore, I have not aimed at providing a comprehensive bibliography.

In bioinformatics, so much is to be done, the raw material to hand is already so vast and vastly increasing, and the problems to be solved are so important (perhaps the most important of any science at present), we may be entering an era comparable to the great flowering of quantum mechanics in the first three decades of the twentieth century, during which there were periods when practically every doctoral thesis was a major breakthrough. If this book is able to inspire the student to take up some of the challenges, then it will have accomplished a large part of what it sets out to do.

Indeed, I would go further to remark that I believe that there are still comparatively simple things to be discovered and that many of the present directions of work in the field may turn out not to be right. Hence, at this stage in its development the most important thing is to illuminate that viewpoint that will facilitate new discoveries. This belief also underlies the somewhat more detailed coverage of the biological processes in which information processing in nature is embodied than might be considered customary.

A work of this nature depends on a long history of interactions, discussions, and correspondence with many present and erstwhile friends and colleagues, some of whom, sadly, are no longer alive. I have tried to reflect some of this debt in the citations. Furthermore, many scientific subjects and methods other than those mentioned in the text had to be explored before the ones best suited to the purpose of this work could be selected, and my thanks are due to all those who helped in these preliminary studies. I should like to add an especial word of thanks to Victoria Kechekhmadze for having so ably drawn the figures.

January 2004

Contents