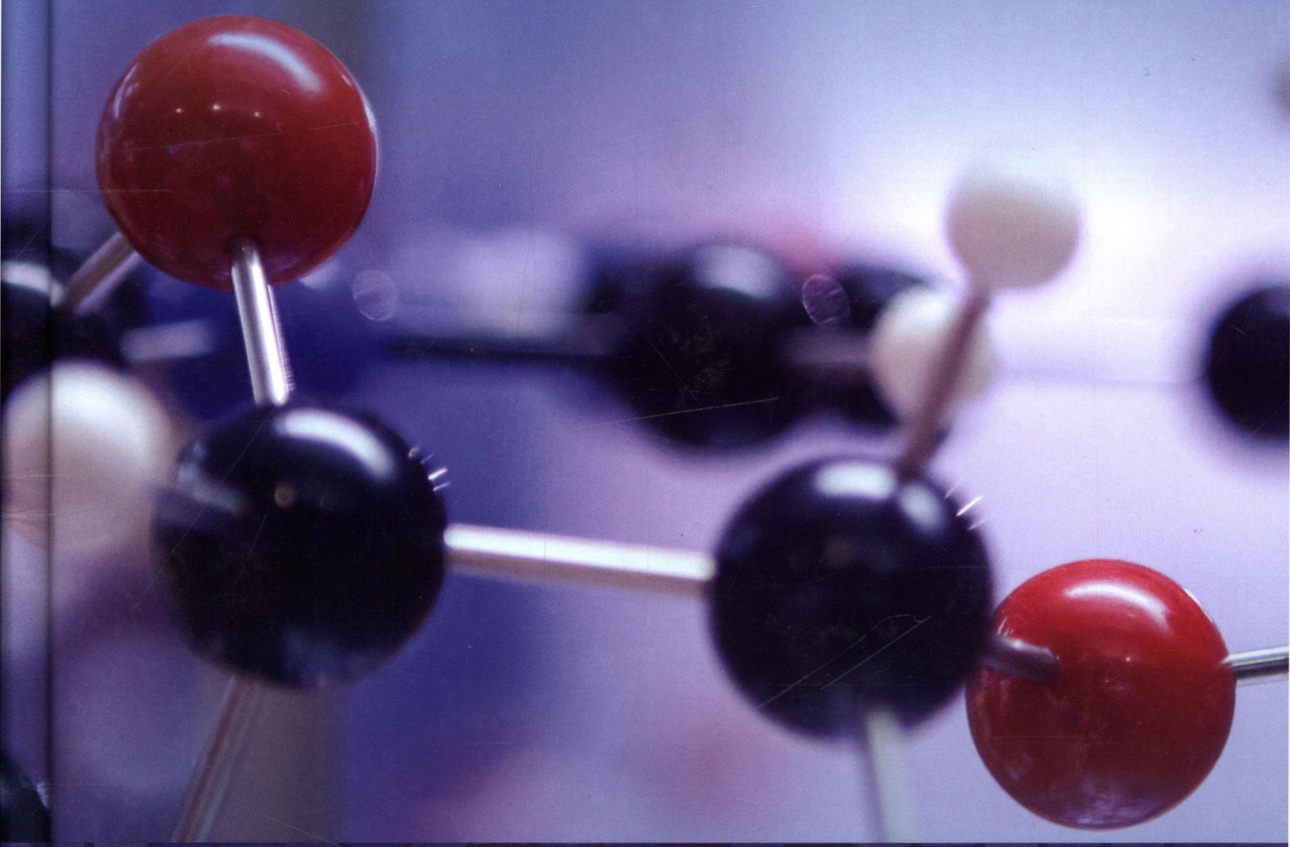
Basic Applied Bioinformatics



Chandra Sekhar Mukhopadhyay
Ratan Kumar Choudhary
Mir Asif Iquebal

WILEY Blackwell

An Accessible Guide that Introduces Students in All Areas of Life Sciences to Bioinformatics

Basic Applied Bioinformatics provides practical guidance in bioinformatics and helps students to optimize parameters for data analysis and then to draw accurate conclusions from the results. In addition to parameter optimization, the text will also familiarize students with relevant terminology. *Basic Applied Bioinformatics* is written as an accessible guide for graduate students studying bioinformatics, biotechnology, and other related sub-disciplines of the life sciences.

This text outlines the basics of bioinformatics, including pertinent information such as downloading molecular sequences (nucleotide and protein) from databases; BLAST analyses; primer designing and its quality checking, multiple sequence alignment (global and local using freely available software); phylogenetic tree construction (using UPGMA, NJ, MP, ME, FM algorithm and MEGA7 suite), prediction of protein structures and genome annotation, RNASeq data analyses and identification of differentially expressed genes and similar advanced bioinformatics analyses. The authors, Chandra Sekhar Mukhopadhyay, Ratan Kumar Choudhary, and Mir Asif Iquebal are noted experts in the field and have come together to provide updated information on bioinformatics.

Salient features of this book include:

- Accessible and updated information on bioinformatics tools
- A practical step-by-step approach to molecular-data analyses
- Information pertinent to study a variety of disciplines including biotechnology, zoology, bioinformatics and other related fields
- Worked examples, glossary terms, problems and solutions.

Basic Applied Bioinformatics gives students studying bioinformatics, agricultural biotechnology, animal biotechnology, medical biotechnology, microbial biotechnology, and zoology an updated introduction to the growing field of bioinformatics.

About the Authors

Chandra Sekhar Mukhopadhyay is an Assistant Scientist (Senior Scale) at the School of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University (GADVASU) at Ludhiana, Punjab, India.

Ratan Kumar Choudhary is an Assistant Professor at the School of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University (GADVASU) at Ludhiana, Punjab, India.

Mir Asif Iquebal is a Scientist at the Centre for Agricultural Bioinformatics, Indian Council of Agricultural Research-Indian Agricultural Statistics and Research Institute (ICAR-IASRI) at Pusa, New Delhi, India.

Cover Design: Wiley
Cover Images: (top) @ mashuk/Gettyimages; (middle) @ alice-photo/Shutterstock; (bottom) © kentoh/Shutterstock

www.wiley.com/wiley-blackwell

WILEY Blackwell

Also available as an e-book

ISBN 978-1-119-24433-2

9 781119 244332

Mukhopadhyay
Choudhary
Iquebal

Basic Applied Bioinformatics

WILEY
Blackwell

Basic Applied Bioinformatics

Chandra Sekhar Mukhopadhyay
Ratan Kumar Choudhary
Mir Asif Iquebal

With contributions from

Ravi GVPPS Kumar, Sarika, Dinesh Kumar, Aditya Prasad Sahoo, Amit Kumar, Saurabh Jain, Surbhi Panwar, Ashwani Kumar, Harpreet Kaur Manku

WILEY Blackwell

This edition first published 2018
© 2018 John Wiley & Sons, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at http://www.wiley.com/go/permissions.

The right of Chandra Sekhar Mukhopadhyay, Ratan Kumar Choudhary, and Mir Asif Iquebal to be identified as the authors of this work has been asserted in accordance with law.

Registered Office(s)
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty
In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Mukhopadhyay, Chandra Sekhar, author. | Choudhary, Ratan Kumar, author. | Iquebal, Mir Asif, author.
Title: Basic applied bioinformatics / by Chandra Sekhar Mukhopadhyay, Ratan Kumar Choudhary, Mir Asif Iquebal.
Description: 1st edition. | Hoboken, NJ : Wiley, [2017] | Includes bibliographical references and index. |
Identifiers: LCCN 2017015387 (print) | LCCN 2017019742 (ebook) | ISBN 9781119244370 (pdf) | ISBN 9781119244417 (epub) | ISBN 9781119244332 (hardback)
Subjects: LCSH: Bioinformatics–Textbooks. | BISAC: MEDICAL / Biostatistics.
Classification: LCC QH324.2 (ebook) | LCC QH324.2 .M85 2017 (print) | DDC 572.80285–dc23
LC record available at https://lccn.loc.gov/2017015387

Cover Design: Wiley
Cover Images: (top) @ mashuk/Gettyimages; (middle) @ alice-photo/Shutterstock; (bottom) © kentoh/Shutterstock

Set in 10.5/13pt Times by SPi Global, Pondicherry, India

Printed in Singapore by C.O.S. Printers Pte Ltd

10 9 8 7 6 5 4 3 2 1

此为试读,需要完整PDF请访问: www.ertongbook.com

BASIC APPLIED BIOINFORMATICS

Dedicated to students, researchers and professionals

Preface

Bioinformatics, a discipline that attempts to make predictions about biological functions using data from molecular sequence (nucleotide and protein) analysis and involves application of information science to biology has, over the years, evolved exponentially in the genomics era. Today it has become an indispensable component of biological science, including its application in a number of applied areas.

There has long been a need among students and researchers for a primer book on the application of bioinformatics tools in various spheres of veterinary and agricultural sciences. This being the era of multi-tasking, research workers who do not possess a background in computer or bioinformatics often stumble over *in silico* analysis of molecular data. This book provides practical know-how for graduate students of bioinformatics, biotechnology and other streams of biological science, and also for those who need to learn bio-computational analyses of the large volume of molecular data that is being generated in thousands of laboratories throughout the world.

The topics considered in this book are the basic ones that a student or researcher of the fields above should know. In addition, this book covers the syllabi of the graduate or undergraduate course called "Introduction to Bioinformatics" (or course name similar to that) that is offered in several universities. Some of the chapters, covering areas such as genome annotation in prokaryotes and eukaryotes, an overview of microarray data analysis, use of MISA for microsatellite sequence identification and SNP mining, have also been introduced for out-of-the-box applications.

In general, the book serves as a reference book for those working in biocomputational research and studies. The contents of this book cover wider areas of bioinformatics. Several freely available software tools (online or offline) are available, and students and researchers can use them for *in silico* analysis. However, in some instances, students become stuck while optimizing parameters for data analysis and drawing appropriate inferences. Also, they are often not familiar with several terminologies. This book explains steps for parameter optimization of the tools being used, as well as the basic terminologies. The results obtained have been explained, to demonstrate how inferences are drawn.

This book can also serve as a practical manual for the elucidation of critical steps, with annotation and explanation. It begins with basic aspects of bioinformatics, including frequently used terminology, concept development, handling molecular sequences,

BLAST analyses, primer designing, phylogenetic tree construction, prediction of protein structures and genome annotation. In the last few chapters, some advanced topics of bioinformatics have been covered, such as analysis of transcriptome data, identification of differentially expressed genes and prediction of microRNA targets.

Each chapter demonstrates the steps with an example, which involves stepwise elucidation of the procedures and explanation of the obtained results. The practical methodology is depicted with screenshots of the software being used, along with legends to explain the screenshot view. New terminologies introduced in some chapters have been provided. Additionally, four or five questions are given at the end of each chapter, with any hints which are deemed to be required for some questions.

We believe that there could be some unintentional mistakes remaining in this book. We sincerely request the reader to apprise the editors for typographical or other errors, if found. It is very common that the version of molecular sequences in public repository is updated over time, or sometimes one or more sequence entries are deleted. The readers are requested to update the editors about such changes. Similarly, the uniform resource locators (URLs) of the websites containing bioinformatics tools or databases can change suddenly. We will be careful to update these changes in the next edition of the book. Readers are also requested to assist us in this regard.

It is hoped that this book will be a useful primer for beginners of this fast-expanding field.

此为试读,需要完整PDF请访问: www.ertongbook.com

Acknowledgements

We thank Ms. Mindy Okura-Marszycki and Mr. Vishnu Narayanan of Wiley for their timely help and encouragements. We extend a special note of thanks to Prof. G.S. Brah, Founder Director, and Prof. Ramneek Verma, Director, of the School of Animal Biotechnology, GADVASU, Ludhiana, for providing the conducive working environment and for inspiring us to contribute to the field of bioinformatics. They evaluated some of the chapters and raised critical questions for improving the quality. The authors thankfully appreciate Miss J.K. Dhanoa and Ms. H.K. Manku for thoroughly checking the syntax of the manuscript, helping in editing and framing the diagrams in the proper format. The chapters were also evaluated for lucidness and ease of understanding by the graduate students of the Iowa State University, Mrs. Supreet Kaur (MSc Biochemistry) and Shravanti Krishna (PhD Biochemistry). Sincere thanks are extended to Dr. Nikhlesh Singh (Assistant Professor, Physiology, the University of Tennessee Health Science Center (UTHSC), Memphis, USA), Dr Monson Melissa (Postdoc Research Associate, Animal Science, Iowa State University) and Dr. Sangita Singh (Post Doc., Department of Food Science and Human Nutrition, Iowa State University) for their constructive criticisms to improve the chapters. Dr. Shivani Sood, Assistant Prof. (Biotechnology, Mukand Lal National College, Yamuna Nagar, Haryana, India), deserves special mention for critically checking the chapters and giving constructive input. All the freely available software and databases covered in this book are duly acknowledged. We are obliged to all those who have directly or indirectly contributed to writing this book.

Our sources of inspiration have been our families, colleagues and students. Nevertheless, we bow before the Almighty and Mother Nature for giving us the potential to accomplish the task.

List of Abbreviations

AFLP	amplified fragment length polymorphism
ASCII	American Standard Code for Information Interchange
BAC	bacterial artificial chromosome
BAM	binary alignment/map
BIC	Bayesian information criterion
BLAST	basic local alignment search tool
BWA	Burrows–Wheeler algorithm
BWT	Burrows–Wheeler transformation
cDNA	complementary DNA
CINEMA	color interactive editor for multiple alignments
cRNA	complementary RNA
dbGaP	database of genotypes and phenotypes
dbVar	database of variation
DDBJ	DNA data bank of Japan
DEG	differentially expressed genes
DNA	deoxyribonucleic acid
DP	dynamic programming
EMBL	European Molecular Biology Laboratory
EST	expressed sequence tag
ExPASy	expert protein analysis system
F81 model	Felsenstein (1981) model
FASTA	fast all
FDQN	fully quantified domain name
FPKM	fragment per kilobase of exon per million mappable reads
GATK	genome analysis toolkit
gi or GI	gene identity
GO	gene ontology
GOR	Garnier, Osguthorpe, and Robson
GSS	genome survey sequence
GTF	gene transfer format
GTR	generalized time-reversible
GUI	graphical user interface

GWAS	genome-wide association studies
HKY85 model	Hasegawa, Kishino and Yano (1985) model
IBL	internal branch length
InDels	insertion and deletions
INSDC	International Nucleotide Sequence Database Collaboration
IUPAC	International Union of Pure and Applied Chemistry
JALVIEW	Java alignment viewer
JC69 model	Jukes and Cantor (1969) model
K80 model	Kimura (1980) model
MACAW	multiple alignment construction and analysis workbench
MAFFT	multiple alignment using fast Fourier transform
ME	minimum evolution
MEGA	molecular evolution and genetic analysis
MISA	microsatellite identification tool
ML	maximum likelihood
MP	maximum parsimony
MSA	multiple sequence alignment
MSRE	methylation sensitive restriction enzymes
MUSCLE	multiple sequence comparison by log-expectation
mYa	million years ago
NBRF	National Biomedical Research Foundation
NCBI	National Center for Biotechnology Information
NGS	next-generation sequencing
NJ	neighbor joining
NWA	Needleman–Wunsch algorithm
ORF	open reading frame
OTU	operational taxonomic unit
PDB	protein data bank
pI/MW	isoelectric point to molecular weight ratio
PIR	protein information resource
PSD	protein sequence database
PWMs	position weight matrices
RCSB	Research Collaboratory for Structural Bioinformatics
RE	restriction enzyme
RF	reading frame
RFLP	restriction fragment length polymorphism
RPKM	read per kilobase of exon per million mappable reads
rRNA	ribosomal RNA
SAM	sequence alignment/map
SCOP	structural classification of protein
SIB	Swiss Institute of Bioinformatics
SNPs	single nucleotide polymorphisms
SPR	subtree pruning regrafting
SSR	simple sequence repeats
STS	sequence-tagged site

SWA	Smith–Waterman algorithm
T92 model	Tamura (1992) model
TBR	tree bisection reconnection
T-Coffee	tree-based consistency objective function for alignment evaluation
TFBS	transcription factor binding sites
TFs	transcription factors
TIS	translation initiation sites
TN93 model	Tamura and Nei (1993) model
T-P	transversion-parsimony
TPA	third-party annotation
TPM	transcripts per million
TRANSFAC	transcription regulatory factors
tRNA	transfer RNA
TTS	triplex-forming oligonucleotide target sequences
uGDT	unnormalized global distance test
UniProt	universal protein resource
UPGMA	unweighted pair group method with arithmetic mean
URL	uniform resource locator
VCF	variant call format
VNTR	variable number tandem repeat
WGS	whole genome shotgun
wwPDB	worldwide protein data bank
YAC	yeast artificial chromosome

此为试读,需要完整PDF请访问: www.ertongbook.com