

Bioinformatics—A Student's Companion

Kalibulla Syed Ibrahim
Guruswami Gurusubramanian
Zothansanga · Ravi Prakash Yadav
Nachimuthu Senthil Kumar
Shunmugiah Karutha Pandian
Probodh Borah · Surender Mohan

Bioinformatics—A Student's Companion

Kalibulla Syed Ibrahim
Department of Biotechnology
Mizoram University
Aizawl, Mizoram
India

Nachimuthu Senthil Kumar
Department of Biotechnology
Mizoram University
Aizawl, Mizoram
India

Guruswami Gurusubramanian
Department of Zoology
Mizoram University
Aizawl, Mizoram
India

Shunmugiah Karutha Pandian
Department of Biotechnology
Alagappa University
Karaikudi, Tamil Nadu
India

Zothansanga
Department of Biotechnology
Mizoram University
Aizawl, Mizoram
India

Probodh Borah
College of Veterinary Science
Assam Agricultural University
Guwahati, Assam
India

Ravi Prakash Yadav
Department of Biotechnology
Mizoram University
Aizawl, Mizoram
India

Surender Mohan
Department of Biotechnology and
Bioinformatics
SRM University
Sonapat, Haryana
India

ISBN 978-981-10-1856-5

ISBN 978-981-10-1857-2 (eBook)

DOI 10.1007/978-981-10-1857-2

Library of Congress Control Number: 2017933667

© Springer Science+Business Media Singapore 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

Biological research has progressed tremendously in the recent years due in part to the increased automation in the process of generation of data. On the other hand, storing, managing and interpreting huge volumes of data have become one of the most challenging tasks for the scientists. These have ultimately necessitated the application of computers paving the way for emergence of a highly interdisciplinary field of bioinformatics. Being interdisciplinary, it attracts people from mathematics, statistics, chemistry, physics, etc., but when it comes to learning bioinformatics, people often get confused as to where to start with.

Internet has revolutionized bioinformatics by offering various tools for free. Despite the richness of bioinformatics resources and methods, teaching bioinformatics to the beginners has still remained a daunting task for the biology teachers. It is therefore essential, in my opinion, that learners should be exposed to the practical lessons in bioinformatics using an easily explicable and 'how-to-do' approach. I wish to emphasize in particular the effort made by the authors for undertaking this challenging and exciting concept of developing a manual on bioinformatics, keeping in mind the readers' requirements from diverse backgrounds and also incorporating into it the syllabi of UG and PG courses of various Indian universities. This could serve not only as a model for teaching the practical component of the syllabi, but also for offering trainings to the beginners.

This manual is intended to introduce the basic algorithmic ideas and step-by-step procedures, with screenshots, in order to make better use of available tools to be able to develop novel computational ideas during their own careers later. Though the tools discussed over here are freely available over the internet along with the online manuals and help pages, originality of this work is on presenting the content on a self-study basis making it possible for the students to perform the basic

analyses even in the absence of an instructor. This manual is carefully drafted with a blend of theory, practice and examples. I wish that the authors would come out with more publications of this kind in the future.

Dr. T. Madhan Mohan
Adviser—DBT, Government of India

Preface

Bioinformatics is playing a significant role in the current revolution in modern biology. It attempts to unfurl the mystery of the biologically meaningful patterns of a sequence string of nucleotides. Teaching bioinformatics is now becoming an integral part of the core curriculum at most universities. We believe that it is only by understanding and wisely using these resources we will be able to push the frontier forward.

In this process we have come up with the following features, which should make this edition particularly helpful to you offering essential, practical answers to the day-to-day problems encountered in a successful career in modern biology. The target audience for this book is the biologists, especially the beginners, who want to learn how to analyze DNA or protein sequences in a simple but meaningful fashion. Readers are not expected to possess special background in statistics, mathematics, or computer science, rather just a basic knowledge of molecular biology and internet is all required.

This manual is expected to serve as a resource guide to some of the web-based analysis packages, including some standalone computing resources that are available free for academics to address some of the basic tasks of bioinformatics. In this manual, we focus on few theoretical aspects and emphasize more on applied concepts on PCR primer designing, sequence retrieval, sequence-similarity searching using BLAST, multiple-sequence alignment, phylogenetic analysis, RNA structure analysis, and protein sequence and structure analysis. Molecular docking is also included in this manual. This could be seen as an enriched task-oriented manual that will direct the reader through step-by-step instructions, completing specific bioinformatics analyses. Every chapter and exercise begins with an overview, followed by specific examples on how to perform and interpret the results, and concludes with exercises to test and explore some of the additional concepts. The manual finally concludes with the authors' recommendations on how to keep current with the ever-increasing information on bioinformatics through the appendix, providing links to the resources discussed and additional resources.

Bioinformatics—A Student's Companion presents a broad overview of the computational resources to help orient those just beginning to use computational

tools to address biological problems. We strongly believe that even a novice can quickly tackle each computational problem and arrive at a satisfying result when guided by this unique step-by-step approach on the collection of software with illustrative examples. Most exercises could be used for practical courses in advanced undergraduate- or graduate-level courses in life sciences.

We would like to thank the funding agencies—Bioinformatics Infrastructural Facility under the Biotechnology Information System Network (BTISNeT) of Department of Biotechnology, New Delhi at Mizoram University, Alagappa University, College of Veterinary Science (CVSc), Assam Agricultural University and the Department of Science and Technology—Science and Engineering Research Board (SERB), New Delhi. The literature access facility provided by DBT—DeLCON and UGC—INFLIBNET have facilitated writing of the manual. The administrative support of Mizoram University, Alagappa University, CVSc—Assam Agricultural University and SRM University are acknowledged. Thanks are also due to the master’s degree students who have been helping in testing the protocols and providing suggestions to improve them. Hopefully, the outcome has been a set of useful exercises that will be of help to students, researchers and the academic community.

Finally, we are grateful for the support from the publishers, Springer Nature, particularly Ms. Sivajothi and Ms. Madhurima Kahali, who have been incredibly patient with us.

Aizawl, India
 Aizawl, India
 Aizawl, India
 Aizawl, India
 Aizawl, India
 Karaikudi, India
 Guwahati, India
 Sonapat, India

Dr. Kalibulla Syed Ibrahim
 Dr. Guruswami Gurusubramanian
 Dr. Zothansanga
 Mr. Ravi Prakash Yadav
 Dr. Nachimuthu Senthil Kumar
 Dr. Shunmugiah Karutha Pandian
 Dr. Probodh Borah
 Mr. Surender Mohan

Contents

1	Nucleotide Analysis	1
1.1	Sequence Retrieval	2
1.2	Primer Designing	6
1.3	Designing Degenerate Primers	11
1.4	Reading of Sequence Trace Files Using Finch TV	14
1.5	Troubleshooting DNA Sequencing Problems	18
1.6	Editing Sequence Data	24
1.7	Sequence Assembly—CAP3 Program	28
1.8	Checking for Vector Contamination	32
1.9	Restriction Mapping Using NEBcutter	35
1.10	Gene Prediction Using ORF Finder (Open Reading Frame Finder)	40
1.11	Gene Prediction Using FGENESB	47
1.12	Dot-Plot	49
1.13	Global Sequence Alignment	56
1.14	Local Sequence Alignment	61
1.15	Basic Local Alignment Search Tool (BLAST)—Nucleotide BLAST	64
1.16	Interpreting BLAST Result	70
1.17	Multiple Sequence Alignment: T-Coffee for Small Alignments	76
1.18	Multiple Sequence Alignment—MUSCLE for Medium Alignments	84
1.19	Multiple Sequence Alignment—MAFFT for Large Alignments	87
1.20	Multiple Sequence Alignment and Phylogenetic Analysis Using MEGA	92

2	DNA Marker Analysis	117
2.1	Genetic Analysis Using NTSYSpc (Numerical Taxonomy System)	117
2.2	Principal Coordinate Analysis (PCORDA) Using NTSYSpc	124
2.3	Population Genetic Analysis Using PowerMarker	129
2.4	Dissimilarity Analysis—DARwin5 (Dissimilarity Analysis and Representation for Windows)	134
3	RNA Analysis	141
3.1	Predicting RNA Secondary Structure	141
3.2	Finding Repeats	145
4	Protein Sequence Analysis	149
4.1	Protein Sequence Retrieval from UniProtKB	149
4.2	Visualization of Features in a Multiple Sequence Alignment	154
4.3	Predicting Signal Peptides in Proteins Using SignalP 4.1 Server	156
4.4	Predicting Transmembrane Segments and Signal Peptides in Proteins Using Phobius	162
4.5	Predicting Subcellular Location Using TargetP	166
4.6	Protein BLAST (blastp)	171
4.7	Position-Specific Iterated (PSI)-BLAST	176
4.8	Creating Pattern from Alignment	180
4.9	Pattern-Hit Initiated (PHI)-BLAST	182
4.10	Domain Enhanced Lookup Time Accelerated-BLAST (DELTA-BLAST)	185
5	Protein Structure Analysis	191
5.1	Protein Primary Structure Analysis—ProtParam	191
5.2	Protein Secondary Structure Prediction	196
5.2.1	Secondary Structure Prediction Using SOPMA	197
5.2.2	Secondary Structure Prediction Using PSIPRED	201
5.3	Protein Tertiary Structure Prediction by Homology Modelling	204
5.3.1	Homology Modelling Using SwissModel	208
5.3.2	Protein Tertiary Structure Prediction by Threading (Fold Recognition)	215
5.4	Protein Tertiary Structure Analysis	223
5.4.1	RAMPAGE	223
5.4.2	Protein Structure Analysis Using SAVeS	226
5.5	Protein Structure Visualization	230
5.5.1	RasMol	230
5.5.2	PyMol	238
5.6	Protein Structure Alignment/Superimpose Using SuperPose	246
5.7	Protein Cleft Analysis	252

6 Protein–Ligand Interactions	257
6.1 Protein–Ligand Docking Using AutoDock4.1 and MGLTools . . .	257
6.2 Protein–Protein Docking Using ClusPro2.0	268
Appendix: Online Resources on Bioinformatics	275
References	279

About the Authors

Kalibulla Syed Ibrahim is postdoctoral fellow at the Department of Biotechnology, Mizoram University. He received his Ph.D. in Biotechnology from Alagappa University, Karaikudi. His fields of interest include protein expression and purification, microbial ecology, high throughput sequencing and data analysis. Besides having publications in peer reviewed journals he is also a recipient of DST-Young Scientist Award (YSS/2014/000657) for working in Human Microbiome of gastric cancer patients of Mizoram, India.

Guruswami Gurusubramanian is Professor of Zoology at Mizoram University with more than 23 years of experience in teaching and research. He received his Ph.D. in Zoology from the University of Madras, Chennai and served as Post Doctoral Humboldt Fellow at the Institut für Biologischen Pflanzenschutz, Germany. His contributions have been recognized with awards like Fellow of International Society for Ecological Communication, Poland, the Business Excellence Award from the International Business Council, and the Life Time Achievement Award from the Indian Compendium Bureau, New Delhi. He has published over 109 articles in leading peer-reviewed International journals, has contributed to two books published by leading publishers and holds four patents. He has been successful in attracting funding from various bodies like the UGC, DBT, CSIR and Central Silk Board as a Principal Investigator.

Zothansanga is currently working as a postdoctoral researcher at the Department of Biotechnology, Mizoram University with 5 years of teaching experience. His fields of interest include microbial diversity and metagenomics, molecular phylogeny, bioinformatics and microbial control of pests. He is presently working on microbial community analysis using next-generation sequencing methods.

Ravi Prakash Yadav is Research Scholar at the Department of Biotechnology, Mizoram University. He received his M.Sc. Biotechnology from the Kalinga Institute of Industrial Technology, Bhubaneswar (2010–2012). His fields of interest include computational biology/bioinformatics and cancer genomics. Currently, he is working on the study of genomic alterations in cell cycle regulatory genes that result in tumorigenesis.

Nachimuthu Senthil Kumar is Professor of Biotechnology at Mizoram University with more than 12 years of experience in teaching and research. He completed his Ph.D. in Zoology at Bharathiar University, Coimbatore and worked as a postdoctoral fellow at Sun Yat-sen University, Guangzhou, China. His fields of interest include bioinformatics and molecular phylogeny, cancer genomics and human genetics. With over 50 publications in reputed international peer-reviewed journals and authoring two books in leading publications, his contributions have been recognized with various awards from the CSIR and DBT. As a Principal Investigator, he has successfully completed research projects funded by government funding bodies like the DBT, CSIR and DST.

Shunmugiah Karutha Pandian is Professor of Biotechnology at Alagappa University. He received his Ph.D. from the Forensic Sciences Department (Government of Tamil Nadu) affiliated with the University of Madras. After his Ph.D., he moved to the Centre for Cellular and Molecular Biology (CCMB), Hyderabad followed by the National Tsing Hua University, Taiwan for postdoctoral research. An established pioneer in drug discovery from natural products, he has published 127 articles in leading peer-reviewed international (SCI) journals and has contributed to five books and two laboratory manuals published by leading publishers like Springer. He has been successful in completing nine Major Research Projects (MRPs) funded by government bodies like the UGC, DBT, DCI, CSIR and ICMR as a Principal Investigator. In recognition to his contributions, he has been awarded with the ‘Certificate of Achievement’ under the ‘Performance Linked Incentive Scheme’ of the Alagappa University, the ‘Shri. P.K. Das Memorial Best Faculty Award in Biotechnology’ by the Nehru Group of Institutions, and the Tamil Nadu Scientist Award (TANSA) in Biological Sciences for the year 2012.

Probodh Borah is Professor of Microbiology/Animal Biotechnology at Assam Agricultural University with more than 29 years of experience in teaching and research. He completed his Ph.D. in Veterinary Science at Assam Agricultural University, Guwahati and served as postdoctoral fellow at the University of California, USA. He has over 63 publications in peer-reviewed journals and has authored six books, two laboratory manuals and contributed three book chapters. He is an active member of the DBT Committee for North Eastern States. As a Principal Investigator, he has received funding from government agencies including the UGC, DBT, CSIR, ICMR and DST. He is also the Coordinator, Bioinformatics Infrastructure Facility and State Biotech Hub (Assam).

Surender Mohan is Assistant Professor of Biotechnology and Bioinformatics at SRM University, Haryana, India with more than four years of experience in teaching and research activities. He is also pursuing his doctoral research at Jawaharlal Nehru University (JNU), New Delhi, India in specializing in Infectious diseases. He has been awarded International Students Fellowship by National University of Singapore (NUS). His contributions have been recognized with collaborates with premier institutes like International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India in research activities and publications in peer-reviewed journals.