

## Subseries of Lecture Notes in Computer Science

### LNBI Series Editors

Sorin Istrail

*Brown University, Providence, RI, USA*

Pavel Pevzner

*University of California, San Diego, CA, USA*

Michael Waterman

*University of Southern California, Los Angeles, CA, USA*

### LNBI Editorial Board

Søren Brunak

*Technical University of Denmark, Kongens Lyngby, Denmark*

Mikhail S. Gelfand

*IITP, Research and Training Center on Bioinformatics, Moscow, Russia*

Thomas Lengauer

*Max Planck Institute for Informatics, Saarbrücken, Germany*

Satoru Miyano

*University of Tokyo, Tokyo, Japan*

Eugene Myers

*Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany*

Marie-France Sagot

*Université Lyon 1, Villeurbanne, France*

David Sankoff

*University of Ottawa, Ottawa, Canada*

Ron Shamir

*Tel Aviv University, Ramat Aviv, Tel Aviv, Israel*

Terry Speed

*Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia*

Martin Vingron

*Max Planck Institute for Molecular Genetics, Berlin, Germany*

W. Eric Wong

*University of Texas at Dallas, Richardson, TX, USA*

More information about this series at <http://www.springer.com/series/5381>

Daniel Figueiredo · Carlos Martín-Vide  
Diogo Pratas · Miguel A. Vega-Rodríguez (Eds.)

# Algorithms for Computational Biology

4th International Conference, AlCoB 2017  
Aveiro, Portugal, June 5–6, 2017  
Proceedings

*Editors*

Daniel Figueiredo  
University of Aveiro  
Aveiro  
Portugal

Carlos Martín-Vide  
Rovira i Virgili University  
Tarragona  
Spain

Diogo Pratas  
University of Aveiro  
Aveiro  
Portugal

Miguel A. Vega-Rodríguez  
University of Extremadura  
Caceres  
Spain

ISSN 0302-9743

Lecture Notes in Bioinformatics

ISBN 978-3-319-58162-0

DOI 10.1007/978-3-319-58163-7

ISSN 1611-3349 (electronic)

ISBN 978-3-319-58163-7 (eBook)

Library of Congress Control Number: 2017938562

LNCS Sublibrary: SL8 – Bioinformatics

© Springer International Publishing AG 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 International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

These proceedings contain the papers that were presented at the 4th International Conference on Algorithms for Computational Biology (AlCoB 2017), held in Aveiro, Portugal, during June 5–6, 2017.

The scope of AlCoB includes topics of either theoretical or applied interest, namely:

- Exact sequence analysis
- Approximate sequence analysis
- Pairwise sequence alignment
- Multiple sequence alignment
- Sequence assembly
- Genome rearrangement
- Regulatory motif finding
- Phylogeny reconstruction
- Phylogeny comparison
- Structure prediction
- Compressive genomics
- Proteomics: molecular pathways, interaction networks
- Transcriptomics: splicing variants, isoform inference and quantification, differential analysis
- Next-generation sequencing: population genomics, metagenomics, metatranscriptomics
- Microbiome analysis
- Systems biology

AlCoB 2017 received 24 submissions. Most papers were reviewed by three Program Committee members. There were also a few external reviewers consulted. After a thorough and vivid discussion phase, the committee decided to accept ten papers (which represents an acceptance rate of about 42%). The conference program included three invited talks and some poster presentations of work in progress.

The excellent facilities provided by the EasyChair conference management system allowed us to deal with the submissions successfully and to handle the preparation of these proceedings in time.

We would like to thank all invited speakers and authors for their contributions, the Program Committee and the external reviewers for their cooperation, and Springer for its very professional publishing work.

March 2017

Daniel Figueiredo  
Carlos Martín-Vide  
Diogo Pratas  
Miguel A. Vega-Rodríguez

# Organization

AlCoB 2017 was organized by the Center for Research and Development in Mathematics and Applications, CIDMA, University of Aveiro, Portugal, the Institute of Electronics and Informatics Engineering of Aveiro, IEETA, University of Aveiro, Portugal, and the Research Group on Mathematical Linguistics, GRLMC, Rovira i Virgili University, Tarragona, Spain.

## Program Committee

Can Alkan	Bilkent University, Turkey
Stephen Altschul	National Institutes of Health, USA
Yurii Aulchenko	PolyOmica, Netherlands
Timothy L. Bailey	University of Nevada, USA
Bonnie Berger	Massachusetts Institute of Technology, USA
Philipp Bucher	Swiss Federal Institute of Technology, Switzerland
Ken Chen	University of Texas MD Anderson Cancer Center, USA
Julio Collado-Vides	National Autonomous University of Mexico, Mexico
Eytan Domany	Weizmann Institute of Science, Israel
Dmitrij Frishman	Technical University of Munich, Germany
Terry Furey	University of North Carolina, USA
Olivier Gascuel	Pasteur Institute, France
Debashis Ghosh	University of Colorado, USA
Susumu Goto	Kyoto University, Japan
Osamu Gotoh	Institute of Advanced Industrial Science and Technology, Japan
Artemis Hatzigeorgiou	University of Thessaly, Greece
Javier Herrero	University College London, UK
Karsten Hokamp	Trinity College Dublin, Ireland
Fereydoun Hormozdiari	University of California, USA
Kazutaka Katoh	Osaka University, Japan
Lukasz Kurgan	Virginia Commonwealth University, USA
Gerton Lunter	University of Oxford, UK
Carlos Martín-Vide	Rovira i Virgili University, Spain, Chair
Zemin Ning	Wellcome Trust Sanger Institute, UK
William Stafford Noble	University of Washington, USA
Cedric Notredame	Center for Genomic Regulation, Spain
Christos Ouzounis	Centre for Research and Technology Hellas, Greece
Manuel C. Peitsch	Philip Morris International, Switzerland
Matteo Pellegrini	University of California, USA
Graziano Pesole	University of Bari, Italy
David Posada	University of Vigo, Spain

Knut Reinert	Free University of Berlin, Germany
Peter Robinson	The Jackson Laboratory, USA
Julio Rozas	University of Barcelona, Spain
David Sankoff	University of Ottawa, Canada
Alejandro Schäffer	National Institutes of Health, USA
Xinghua Shi	University of North Carolina, USA
Nicholas D. Socci	Memorial Sloan Kettering Cancer Center, USA
Alexandros Stamatakis	Heidelberg Institute for Theoretical Studies, Germany
Granger Sutton	J. Craig Venter Institute, USA
Kristel Van Steen	University of Liège, Belgium
Arndt Von Haeseler	Center for Integrative Bioinformatics Vienna, Austria
Kai Wang	Columbia University, USA
Ioannis Xenarios	Swiss Institute of Bioinformatics, Switzerland
Jinn-Moon Yang	National Chiao Tung University, Taiwan
Shibu Yooseph	University of Central Florida, USA
Mohammed J. Zaki	Rensselaer Polytechnic Institute, USA
Daniel Zerbino	European Bioinformatics Institute, UK
Weixiong Zhang	Washington University in St. Louis, USA
Zhongming Zhao	University of Texas Health Science Center at Houston, USA

## Additional Reviewers

Xian Fan  
 Nam S. Vo  
 Chunfang Zheng

## Organizing Committee

Diana Costa	CIDMA, Aveiro
Daniel Figueiredo	CIDMA, Aveiro (Co-chair)
Carlos Martín-Vide	Tarragona (Co-chair)
Manuel A. Martins	CIDMA, Aveiro
Manuel Jesús Parra Royón	Granada
Armando J. Pinho	IEETA, Aveiro
Diogo Pratas	IEETA, Aveiro (Co-chair)
David Silva	London
Miguel A. Vega-Rodríguez	Cáceres

# Contents

## Invited Talks

Biomedical Applications of Prototype Based Classifiers and Relevance Learning. . . . .	3
<i>Michael Biehl</i>	
Describing the Local Structure of Sequence Graphs. . . . .	24
<i>Yohei Rosen, Jordan Eizenga, and Benedict Paten</i>	

## Graph Algorithms for Computational Biology

Mapping RNA-seq Data to a Transcript Graph via Approximate Pattern Matching to a Hypertext . . . . .	49
<i>Stefano Beretta, Paola Bonizzoni, Luca Denti, Marco Previtali, and Raffaella Rizzi</i>	
A Fast Algorithm for Large Common Connected Induced Subgraphs. . . . .	62
<i>Alessio Conte, Roberto Grossi, Andrea Marino, Lorenzo Tattini, and Luca Versari</i>	
Ranking Vertices for Active Module Recovery Problem. . . . .	75
<i>Javlon E. Isomurodov, Alexander A. Loboda, and Alexey A. Sergushichev</i>	
Computational Processes that Appear to Model Human Memory. . . . .	85
<i>John L. Pfaltz</i>	

## Phylogenetics

Inferring the Distribution of Fitness Effects (DFE) of Newly-Arising Mutations Using Samples Taken from Evolving Populations in Real Time . . .	103
<i>Philip J. Gerrish and Nick Hengartner</i>	
An Efficient Algorithm for the Rooted Triplet Distance Between Galled Trees . . . . .	115
<i>Jesper Jansson, Ramesh Rajaby, and Wing-Kin Sung</i>	
Clustering the Space of Maximum Parsimony Reconciliations in the Duplication-Transfer-Loss Model . . . . .	127
<i>Alex Ozdemir, Michael Sheely, Daniel Bork, Ricson Cheng, Reyna Hulett, Jean Sung, Jincheng Wang, and Ran Libeskind-Hadas</i>	



**Sequence Analysis and Other Biological Processes**

CSA-X: Modularized Constrained Multiple Sequence Alignment. . . . .	143
<i>T.M. Rezwatul Islam and Ian McQuillan</i>	
Quantifying Information Flow in Chemical Reaction Networks. . . . .	155
<i>Ozan Kahramanoğulları</i>	
Parallel Biological Sequence Comparison in Linear Space with Multiple Adjustable Bands . . . . .	167
<i>Gabriel H.G. Silva, Edans F.O. Sandes, George Teodoro, and Alba C.M.A. Melo</i>	
<b>Author Index</b> . . . . .	181