

Lecture Notes in Bioinformatics

7534

Edited by S. Istrail, P. Pevzner, and M. Waterman

Editorial Board: A. Apostolico S. Brunak M. Gelfand

T. Lengauer S. Miyano G. Myers M.-F. Sagot D. Sankoff

R. Shamir T. Speed M. Vingron W. Wong

Subseries of Lecture Notes in Computer Science

Ben Raphael Jijun Tang (Eds.)

Algorithms in Bioinformatics

12th International Workshop, WABI 2012
Ljubljana, Slovenia, September 10-12, 2012
Proceedings



Springer

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

Volume Editors

Ben Raphael

Brown University

Department of Computer Science

P.O. Box 1910

Providence, RI 02912, USA

E-mail: braphael@cs.brown.edu

Jijun Tang

University of South Carolina

Department of Computer Science and Engineering

301 Main Street

Columbia, SC 29208, USA

E-mail: jtang@cse.sc.edu

ISSN 0302-9743

e-ISSN 1611-3349

ISBN 978-3-642-33121-3

e-ISBN 978-3-642-33122-0

DOI 10.1007/978-3-642-33122-0

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012945432

CR Subject Classification (1998): F.2, J.3, G.2.2, F.1, I.2.6, H.2.8, G.1.2

LNCS Sublibrary: SL 8 – Bioinformatics

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, 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.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume contains papers presented at the 12th Workshop on Algorithms in Bioinformatics (WABI), held September 10–12th, 2012, at University of Ljubljana, Slovenia. WABI 2012 was one of six workshops which, along with the European Symposium on Algorithms (ESA), constituted the ALGO annual meeting. WABI focuses on algorithmic advances in bioinformatics, computational biology, and systems biology with a particular emphasis on discrete algorithms and machine-learning methods that address important problems in molecular biology.

The 35 papers this volume include algorithms for a variety of biological problems including phylogeny, DNA and RNA sequencing and analysis, protein structure, and others. In addition to presentations of these papers, WABI 2012 featured a keynote address “Genome Rearrangement by Double Cut and Join” by Jens Stoye, Universität Bielefeld. The 35 papers presented here were selected from a total of 92 submissions. Each submission was reviewed by at least three members of the Program Committee. Following these initial reviews, submissions were further discussed by the Program Committee to arrive at the final list of selected papers. We sincerely thank the members of the Program Committee and their sub-referees for their hard work in reviewing the submissions. We also thank EasyChair for providing an outstanding online system for managing the review process.

We thank the Organizing Committee, Andrej Brodnik (Co-chair), Uroš Čibej, Gašper Fele-Žorž, Matevž Jekovec, Jurij Mihelič, Borut Robič (Co-chair), and Andrej Tolič, for their tireless efforts in managing all of the local arrangements required for a successful meeting. Finally, we offer special thanks Bernard Moret for his invitation to chair this year’s WABI meeting, his advice throughout the process, and his continuing commitment to WABI as a venue for discussing the latest algorithmic advances that impact molecular biology. We were honored to be part of this year’s WABI and look forward to the continued success of the meeting in the coming years.

July 2012

Ben Raphael
Jijun Tang

Organization

Program Committee

Tatsuya Akutsu	Kyoto University, Japan
Max Alekseyev	University of South Carolina, USA
Mathieu Blanchette	McGill University, Canada
Guillaume Blin	Université Paris-Est Marne-la-Vallée, France
Paola Bonizzoni	Università di Milano-Bicocca, Italy
Daniel Brown	University of Waterloo, Canada
Philipp Bucher	Swiss Federal Institute of Technology (EPFL), Switzerland
Sebastian Böcker	Friedrich Schiller University Jena, Germany
Lenore Cowen	Tufts University, USA
Minghua Deng	Peking University, China
Pufeng Du	Tianjin University, China
Nadia El-Mabrouk	University of Montreal, Canada
Guillaume Fertin	LINA, University of Nantes, France
Liliana Florea	Johns Hopkins University, USA
Anna Gambin	Warsaw University, Poland
Olivier Gascuel	LIRMM, CNRS - Université Montpellier 2, France
Katharina Huber	University of East Anglia, UK
Daniel Huson	University of Tübingen, Germany
Carl Kingsford	University of Maryland, College Park, USA
Jim Leebens-Mack	University of Georgia, USA
Xinghua Lu	University of Pittsburgh, USA
Ion Mandoiu	University of Connecticut, USA
Satoru Miyano	University of Tokyo, Japan
Bernard M.E. Moret	Swiss Federal Institute of Technology (EPFL), Switzerland
Burkhard Morgenstern	Universität Göttingen, Germany
Vincent Moulton	University of East Anglia, UK
Luay Nakhleh	Rice University, USA
Macha Nikolski	LaBRI/CNRS, Université Bordeaux, France
Ron Pinter	Technion - Israel Institute of Technology
Nadia Pisanti	Università di Pisa, Italy
Mihai Pop	University of Maryland, USA
Teresa Przytycka	National Institutes of Health (NIH), USA
Tal Pupko	Tel Aviv University, Israel
Sven Rahmann	University of Duisburg-Essen, Germany
Ben Raphael (Co-chair)	Brown University, USA
Marie-France Sagot	INRIA Grenoble Rhône-Alpes and Université de Lyon 1, Villeurbanne, France

VIII Organization

David Sankoff	University of Ottawa, Canada
Thomas Schiex	INRA Toulouse, France
Russell Schwartz	Carnegie Mellon University, USA
Charles Semple	University of Canterbury, New Zealand
Roded Sharan	Tel Aviv University, Israel
Jingna Si	Zhejiang University, China
Jörg Stelling	Swiss Federal Institute of Technology Zurich, Switzerland
Leen Stougie	Free University Amsterdam, The Netherlands
Jens Stoye	Bielefeld University, Germany
Krister Swenson	Université de Montréal / McGill University, Canada
Jijun Tang (Co-chair)	University of South Carolina, USA
Glenn Tesler	University of California, San Diego, USA
Jerzy Tiuryn	Warsaw University, Poland
Li-San Wang	University of Pennsylvania, USA
Lusheng Wang	City University of Hong Kong, China
Christopher Workman	Technical University of Denmark, Denmark
Feng Yue	University of California San Diego, USA
Alex Zelikovsky	Georgia State University, USA
Louxin Zhang	National University of Singapore, Singapore
Meng Zhang	Jilin University, China
Xiuwei Zhang	Swiss Federal Institute of Technology (EPFL), Switzerland
Ziding Zhang	China Agricultural University, China

Additional Reviewers

Atias, Nir	Daniels, Noah
Bankevich, Anton	De Givry, Simon
Berry, Vincent	Della Vedova, Gianluca
Biecek, Przemyslaw	Dondi, Riccardo
Brunetti, Sara	Dutkowski, Janusz
Bulteau, Laurent	Dvorkin, Mikhail
Burstein, David	Dührkop, Kai
Caciula, Adrian	Faller, Beata
Canzar, Stefan	Faraut, Thomas
Catanzaro, Daniele	Fassler, Raffael
Chauve, Cedric	Gambin, Tomasz
Chikhi, Rayan	Garde, Christian
Christinat, Yann	Ghiurcuta, Cristina
Cohen, Ofir	Glebova, Olga
Crochemore, Maxime	Gorecki, Pawel
Csuros, Miklos	Gottlieb, Assaf
D'Addario, Marianna	Grossi, Roberto

Halldorsson, Bjarni
 Hao, Tong
 Hoinka, Jan
 Huang, Hailiang
 Huang, Yang
 Humphries, Peter
 Jakobi, Tobias
 Jankowski, Aleksander
 Jean, Geraldine
 Jiang, Minghui
 Jiang, Shuai
 Kluge, Boguslaw
 Kochanczyk, Marek
 Kopczynski, Dominik
 Koskas, Michel
 Kulikov, Alexander
 Köster, Johannes
 Labarre, Anthony
 Lacroix, Vincent
 Lavallée-Adam, Mathieu
 Lefebvre, Arnaud
 Li, Fan
 Li, Tingting
 Li, Wei
 Lin, Yu
 Linz, Simone
 Liu, Zheng
 Mahmoody, Ahmad
 Mancuso, Nicholas
 Mangul, Serghei
 Martin, Marcel
 Navlakha, Saket
 Nefedov, Alexey
 Nurk, Sergey
 Oberhardt, Matthew
 Oren, Yaara
 Pardi, Fabio
 Patro, Robert
 Peeri, Michael

Peng, Jian
 Penn, Osnat
 Pirola, Yuri
 Popa, Alexandru
 Puglisi, Simon
 Rizzi, Romeo
 Rusu, Irena
 Ryvkin, Paul
 Sacomoto, Gustavo
 Salari, Rahele
 Scheubert, Kerstin
 Sciortino, Marinella
 Scornavacca, Celine
 Sczyrba, Alexander
 Si, Jingna
 Sikora, Florian
 Sirotkin, Alexander
 Skums, Pavel
 Soueidan, Hayssam
 Subramanian, Ayshwarya
 Sundermann, Linda
 Sykulski, Maciej
 Thévenin, Annelyse
 Tsai, Ming-Chi
 van Iersel, Leo
 Venturini, Rossano
 Wang, Xi
 Wang, Ying
 Wilczynski, Bartek
 Willing, Eyla
 Xu, Wei
 Yang, Xiao
 Zhang, Bo
 Zhang, Ting
 Zheng, Jim
 Zheng, Yu
 Zhou, Wanding
 Zytnicki, Matthias

Table of Contents

Preserving Inversion Phylogeny Reconstruction	1
<i>Matthias Bernt, Kun-Mao Chao, Jyun-Wei Kao, Martin Middendorf, and Eric Tannier</i>	
Fast Phylogenetic Tree Reconstruction Using Locality-Sensitive Hashing	14
<i>Daniel G. Brown and Jakub Trzaskowski</i>	
Efficient Computation of Popular Phylogenetic Tree Measures	30
<i>Constantinos Tsirogiannis, Brody Sandel, and Dimitris Cheliotis</i>	
SibJoin: A Fast Heuristic for Half-Sibling Reconstruction	44
<i>Daniel G. Brown and Daniel Dexter</i>	
Reconstructing the Evolution of Molecular Interaction Networks under the DMC and Link Dynamics Models	57
<i>Yun Zhu and Luay Nakhleh</i>	
Estimating Population Size via Line Graph Reconstruction	69
<i>Bjarni V. Halldórsson, Dima Blokh, and Roded Sharan</i>	
Extracting Conflict-Free Information from Multi-labeled Trees	81
<i>Akshay Deepak, David Fernández-Baca, and Michelle M. McMahon</i>	
Reducing Problems in Unrooted Tree Compatibility to Restricted Triangulations of Intersection Graphs	93
<i>Rob Gysel, Kristian Stevens, and Dan Gusfield</i>	
An Optimal Reconciliation Algorithm for Gene Trees with Polytomies	106
<i>Manuel Lafond, Krister M. Swenson, and Nadia El-Mabrouk</i>	
Accounting for Gene Tree Uncertainties Improves Gene Trees and Reconciliation Inference	123
<i>Thi Hau Nguyen, Jean-Philippe Doyon, Stéphanie Pointet, Anne-Muriel Arigon Chifolleau, Vincent Ranwez, and Vincent Berry</i>	
RNA Tree Comparisons via Unrooted Unordered Alignments	135
<i>Nimrod Milo, Shay Zakov, Erez Katzenelson, Eitan Bachmat, Yefim Dinitz, and Michal Ziv-Ukelson</i>	

Tree Decomposition and Parameterized Algorithms for RNA Structure-Sequence Alignment Including Tertiary Interactions and Pseudoknots (Extended Abstract)	149
<i>Philippe Rinaudo, Yann Ponty, Dominique Barth, and Alain Denise</i>	
δ -TRIMAX: Extracting Triclusters and Analysing Coregulation in Time Series Gene Expression Data	165
<i>Anirban Bhar, Martin Haubrock, Anirban Mukhopadhyay, Ujjwal Maulik, Sanghamitra Bandyopadhyay, and Edgar Wingender</i>	
CLIIQ: Accurate Comparative Detection and Quantification of Expressed Isoforms in a Population	178
<i>Yen-Yi Lin, Phuong Dao, Faraz Hach, Marzieh Bakhshi, Fan Mo, Anna Lapuk, Colin Collins, and S. Cenk Sahinalp</i>	
Improved Lower Bounds on the Compatibility of Quartets, Triplets, and Multi-state Characters	190
<i>Brad Shuttters, Sudheer Vakati, and David Fernández-Baca</i>	
Succinct Multibit Tree: Compact Representation of Multibit Trees by Using Succinct Data Structures in Chemical Fingerprint Searches	201
<i>Yasuo Tabei</i>	
Comparing DNA Sequence Collections by Direct Comparison of Compressed Text Indexes	214
<i>Anthony J. Cox, Tobias Jakobi, Giovanna Rosone, and Ole B. Schulz-Trieglaff</i>	
Succinct de Bruijn Graphs	225
<i>Alexander Bowe, Taku Onodera, Kunihiko Sadakane, and Tetsuo Shibuya</i>	
Space-Efficient and Exact de Bruijn Graph Representation Based on a Bloom Filter	236
<i>Rayan Chikhi and Guillaume Rizk</i>	
From de Bruijn Graphs to Rectangle Graphs for Genome Assembly	249
<i>Nikolay Vyahhi, Alex Pyshkin, Son Pham, and Pavel A. Pevzner</i>	
MORPH-PRO: A Novel Algorithm and Web Server for Protein Morphing	262
<i>Natalie E. Castellana, Andrey Lushnikov, Piotr Rotkiewicz, Natasha Sefcovic, Pavel A. Pevzner, Adam Godzik, and Kira Vyatkina</i>	
How Accurately Can We Model Protein Structures with Dihedral Angles?	274
<i>Xuefeng Cui, Shuai Cheng Li, Dongbo Bu, Babak Alipanahi Ramandi, and Ming Li</i>	

Resolving Spatial Inconsistencies in Chromosome Conformation Data . . .	288
<i>Geet Duggal, Rob Patro, Emre Sefer, Hao Wang, Darya Filippova, Samir Khuller, and Carl Kingsford</i>	
MS-DPR: An Algorithm for Computing Statistical Significance of Spectral Identifications of Non-linear Peptides	301
<i>Hosein Mohimani, Sangtae Kim, and Pavel A. Pevzner</i>	
FinIS: Improved <i>in silico</i> Finishing Using an Exact Quadratic Programming Formulation	314
<i>Song Gao, Denis Bertrand, and Niranjan Nagarajan</i>	
Lightweight LCP Construction for Next-Generation Sequencing Datasets	326
<i>Markus J. Bauer, Anthony J. Cox, Giovanna Rosone, and Marinella Sciortino</i>	
Sign Assignment Problems on Protein Networks	338
<i>Shay Hourì and Roded Sharan</i>	
Sparse Learning Based Linear Coherent Bi-clustering	346
<i>Yi Shi, Xiaoping Liao, Xinhua Zhang, Guohui Lin, and Dale Schuurmans</i>	
A Simplified View of DCJ-Indel Distance	365
<i>Phillip E.C. Compeau</i>	
DCJ-indel Distance with Distinct Operation Costs	378
<i>Poly H. da Silva, Marília D.V. Braga, Raphael Machado, and Simone Dantas</i>	
Hidden Breakpoints in Genome Alignments	391
<i>Birte Kehr, Knut Reinert, and Aaron E. Darling</i>	
A Probabilistic Approach to Accurate Abundance-Based Binning of Metagenomic Reads	404
<i>Olga Tanaseichuk, James Borneman, and Tao Jiang</i>	
Tandem Halving Problems by DCJ	417
<i>Antoine Thomas, Aïda Ouangraoua, and Jean-Stéphane Varré</i>	
A Practical Approximation Algorithm for Solving Massive Instances of Hybridization Number	430
<i>Leo van Iersel, Steven Kelk, Nela Lekić, and Celine Scornavacca</i>	
Distributed String Mining for High-Throughput Sequencing Data	441
<i>Niko Välimäki and Simon J. Puglisi</i>	
Author Index	453