

Lecture Notes in Bioinformatics

6293

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

Vincent Moulton Mona Singh (Eds.)

Algorithms in Bioinformatics

10th International Workshop, WABI 2010
Liverpool, UK, September 6-8, 2010
Proceedings

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

Vincent Moulton

University of East Anglia

School of Computing Sciences

Norwich, NR4 7TJ, UK

E-mail: vincent.moulton@cmp.uea.ac.uk

Mona Singh

Princeton University

Lewis-Sigler Institute for Integrative Genomics

Department of Computer Science

Princeton, NJ 08544, USA

E-mail: mona@cs.princeton.edu

Library of Congress Control Number: 2010932425

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

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-642-15293-7 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-15293-1 Springer Berlin Heidelberg New York

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.

springer.com

© Springer-Verlag Berlin Heidelberg 2010

Printed in Germany

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

Printed on acid-free paper 06/3180

Preface

We are pleased to present the proceedings of the 10th Workshop on Algorithms in Bioinformatics (WABI 2010) which took place in Liverpool, UK, September 6–8, 2010. The WABI 2010 workshop was part of the four ALGO 2010 conference meetings, which, in addition to WABI, included ESA, ATMOS, and WAOA. WABI 2010 was hosted by the University of Liverpool Department of Computer Science, and sponsored by the European Association for Theoretical Computer Science (EATCS) and the International Society for Computational Biology (ISCB). See <http://algo2010.csc.liv.ac.uk/wabi/> for more details.

The Workshop in Algorithms in Bioinformatics highlights research in algorithmic work for bioinformatics, computational biology and systems biology. The emphasis is mainly on discrete algorithms and machine-learning methods that address important problems in molecular biology, that are founded on sound models, that are computationally efficient, and that have been implemented and tested in simulations and on real datasets. The goal is to present recent research results, including significant work-in-progress, and to identify and explore directions of future research.

Original research papers (including significant work-in-progress) or state-of-the-art surveys were solicited for WABI 2010 in all aspects of algorithms in bioinformatics, computational biology and systems biology. In response to our call, we received 83 submissions for papers and 30 were accepted. In addition, WABI 2010 hosted distinguished lectures by Eran Halperin, of Tel Aviv University and ICSI, Berkeley, and, together with ESA, Paolo Ferragina of University of Pisa. We would like to sincerely thank the authors of all submitted papers and the conference participants. We also thank the Program Committee and their sub-referees for their hard work in reviewing and selecting papers for the workshop.

We would especially like to thank Bernard Moret and Tandy Warnow for all of their advice and support in carrying out the role of being Co-chairs.

Thanks once again to all who participated in making WABI's 10th anniversary such a success. For us it has been an exciting and rewarding experience.

June 2010

Vincent Moulton
Mona Singh

Organization

Program Committee

Tatsuya Akutsu	Kyoto University, Japan
Bonnie Berger	MIT, USA
Tanya Berger-Wolf	University of Illinois, USA
Mathieu Blanchette	McGill University, Canada
Sebastian Böcker	University of Jena, Germany
Magnus Bordewich	University of Durham, UK
Mike Brudno	University of Toronto, Canada
Philipp Bucher	EPFL, Switzerland
Benny Chor	Tel Aviv University, Israel
Anne Condon	University of British Columbia, Canada
Lenore Cowen	Tufts University, USA
Keith Crandall	Brigham Young University, USA
Bhaskar Das Gupta	University of Illinois, USA
Nadia El-Mabrouk	University of Montreal, Canada
Liliana Florea	University of Maryland, USA
Olivier Gascuel	University of Montpellier, France
Barbara Holland	Massey University, New Zealand
Katharina Huber	University of East Anglia, UK
Daniel Huson	University of Tuebingen, Germany
Lydia Kavradi	Rice University, USA
Junhyong Kim	University Penn, USA
Carl Kingsford	University of Maryland, USA
Mehmet Koyuturk	Case Western, USA
Jens Lagergren	KTH, Sweden
Chris Langmead	CMU, USA
Ryan Lilien	University of Toronto, Canada
Ion Mandiou	University of Connecticut, USA
Joao Meidanis	Campinas University, Brazil
Satoru Miyano	Tokyo University, Japan
Bernard M.E. Moret	Swiss Federal Institute of Technology, Switzerland
Burkhard Morgenstern	University of Göttingen, Germany
Vincent Moulton	University of East Anglia, UK, Co-chair
Gene W. Myers	Janelia Farms, USA
Mihai Pop	University of Maryland, USA
Teresa Przytycka	NIH, USA
Cenk Sahinalp	Simon Fraser University, USA
David Sankoff	University of Montréal, Canada

VIII Organization

Russell Schwartz	CMU, USA
Joao Setubal	Virginia Tech., USA
Mona Singh	Princeton University, USA, Co-chair
Jens Stoye	University of Bielefeld, Germany
Glenn Tesler	UCSD, USA
Olga Vitek	Purdue University, USA
Lusheng Wang	City University Hong Kong, Hong Kong
Tandy Warnow	University of Texas Austin, USA
Chris Workman	Technical University of Denmark, Denmark
Louxin Zhang	National University of Singapore, Singapore

Table of Contents

Biomolecular Structure: RNA, Protein and Molecular Comparison

A Worst-Case and Practical Speedup for the RNA Co-folding Problem Using the <i>Four-Russians</i> Idea	1
<i>Yelena Frid and Dan Gusfield</i>	
Sparse Estimation for Structural Variability	13
<i>Raghavendra Hosur, Rohit Singh, and Bonnie Berger</i>	
Data Structures for Accelerating Tanimoto Queries on Real Valued Vectors	28
<i>Thomas G. Kristensen and Christian N.S. Pedersen</i>	
Sparsification of RNA Structure Prediction Including Pseudoknots	40
<i>Mathias Möhl, Raheleh Salari, Sebastian Will, Rolf Backofen, and S. Cenk Sahinalp</i>	
Prediction of RNA Secondary Structure Including Kissing Hairpin Motifs	52
<i>Corinna Theis, Stefan Janssen, and Robert Giegerich</i>	
Reducing the Worst Case Running Times of a Family of RNA and CFG Problems, Using Valiant's Approach	65
<i>Shay Zakov, Dekel Tsur, and Michal Ziv-Ukelson</i>	

Comparative Genomics

Reconstruction of Ancestral Genome Subject to Whole Genome Duplication, Speciation, Rearrangement and Loss	78
<i>Denis Bertrand, Yves Gagnon, Mathieu Blanchette, and Nadia El-Mabrouk</i>	
Genomic Distance with DCJ and Indels	90
<i>Marília D.V. Braga, Eyla Willing, and Jens Stoye</i>	
Listing All Sorting Reversals in Quadratic Time	102
<i>Krister M. Swenson, Ghada Badr, and David Sankoff</i>	

Haplotype and Genotype Analysis

Discovering Kinship through Small Subsets	111
<i>Daniel G. Brown and Tanya Berger-Wolf</i>	

Fixed-Parameter Algorithm for Haplotype Inferences on General Pedigrees with Small Number of Sites	124
<i>Duong D. Doan and Patricia A. Evans</i>	
Haplotypes versus Genotypes on Pedigrees	136
<i>Bonnie Kirkpatrick</i>	
Haplotype Inference on Pedigrees with Recombinations and Mutations	148
<i>Yuri Pirola, Paola Bonizzoni, and Tao Jiang</i>	
High-throughput Data Analysis: Next Generation Sequencing and Flow Cytometry	
Identifying Rare Cell Populations in Comparative Flow Cytometry	162
<i>Ariful Azad, Johannes Langguth, Youhan Fang, Alan Qi, and Alex Pothen</i>	
Fast Mapping and Precise Alignment of AB SOLiD Color Reads to Reference DNA	176
<i>Miklós Csűrös, Szilveszter Juhos, and Attila Bérces</i>	
Design of an Efficient Out-of-Core Read Alignment Algorithm	189
<i>Arun S. Konagurthu, Lloyd Allison, Thomas Conway, Bryan Beresford-Smith, and Justin Zobel</i>	
Estimation of Alternative Splicing Isoform Frequencies from RNA-Seq Data	202
<i>Marius Nicolae, Serghei Mangul, Ion Măndoiu, and Alex Zelikovsky</i>	
Networks	
Improved Orientations of Physical Networks	215
<i>Iftah Gamzu, Danny Segev, and Roded Sharan</i>	
Enumerating Chemical Organisations in Consistent Metabolic Networks: Complexity and Algorithms	226
<i>Paulo Vieira Milreu, Vicente Acuña, Etienne Birmelé, Pierluigi Crescenzi, Alberto Marchetti-Spaccamela, Marie-France Sagot, Leen Stougie, and Vincent Lacroix</i>	
Efficient Subgraph Frequency Estimation with G-Tries	238
<i>Pedro Ribeiro and Fernando Silva</i>	

Phylogenetics

Accuracy Guarantees for Phylogeny Reconstruction Algorithms Based on Balanced Minimum Evolution	250
<i>Magnus Bordewich and Radu Mihaescu</i>	
The Complexity of Inferring a Minimally Resolved Phylogenetic Supertree	262
<i>Jesper Jansson, Richard S. Lemence, and Andrzej Lingas</i>	
Reducing Multi-state to Binary Perfect Phylogeny with Applications to Missing, Removable, Inserted, and Deleted Data	274
<i>Kristian Stevens and Dan Gusfield</i>	
An Experimental Study of Quartets MaxCut and Other Supertree Methods	288
<i>M. Shel Swenson, Rahul Suri, C. Randal Linder, and Tandy Warnow</i>	
An Efficient Method for DNA-Based Species Assignment via Gene Tree and Species Tree Reconciliation	300
<i>Louxin Zhang and Yun Cui</i>	

Sequences, Strings and Motifs

Effective Algorithms for Fusion Gene Detection	312
<i>Dan He and Eleazar Eskin</i>	
Swiftly Computing Center Strings	325
<i>Franziska Hufsky, Léon Kuchenbecker, Katharina Jahn, Jens Stoye, and Sebastian Böcker</i>	
Speeding Up Exact Motif Discovery by Bounding the Expected Clump Size	337
<i>Tobias Marschall and Sven Rahmann</i>	
Pair HMM Based Gap Statistics for Re-evaluation of Indels in Alignments with Affine Gap Penalties	350
<i>Alexander Schönhuth, Raheleh Salari, and S. Cenk Sahinalp</i>	
Quantifying the Strength of Natural Selection of a Motif Sequence	362
<i>Chen-Hsiang Yeang</i>	
Author Index	375