

Lecture Notes in Bioinformatics

5251

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

Keith A. Crandall Jens Lagergren (Eds.)

Algorithms in Bioinformatics

8th International Workshop, WABI 2008
Karlsruhe, Germany, September 15-19, 2008
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

Keith A. Crandall

Brigham Young University

Department of Biology

Provo, UT, USA

E-mail: keith_crandall@byu.edu

Jens Lagergren

KTH, Royal Institute of Technology

Computational Biology Department

Stockholm, Sweden

E-mail: jensl@kth.se

Library of Congress Control Number: 2008934600

CR Subject Classification (1998): F.1, F.2.2, E.1, G.1-3, J.3

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-540-87360-0 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-87360-0 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 is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2008

Printed in Germany

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

Printed on acid-free paper SPIN: 12515653 06/3180 5 4 3 2 1 0

Preface

We are very pleased to present the proceedings of the *8th Workshop on Algorithms in Bioinformatics (WABI 2008)*, held as part of the *ALGO 2008* meeting at the University of Karlsruhe, Germany on September 15–19, 2008. *WABI 2008* covered all research on algorithmic work in bioinformatics and systems biology with an emphasis on discrete algorithms and machine-learning methods that address important problems in molecular biology. Such algorithms are founded on sound models, are computationally efficient, and have been implemented and tested in simulations and on real datasets. The goal of these proceedings is to present recent research results, including significant work-in-progress, and to identify and explore directions for future research. Original research papers were solicited in all aspects of algorithms in bioinformatics, targeting the following areas in particular:

- Exact, approximate, and machine-learning algorithms for sequence analysis, gene and signal recognition, alignment and assembly, molecular evolution, structure determination or prediction, gene expression, pathways, gene networks, proteomics, functional genomics, and drug design;
- High-performance computing approaches to computationally hard learning and optimization problems in bioinformatics;
- Methods, software, and dataset repositories for development and testing of such algorithms and their underlying models.

The goal of the workshop is to bring together an interdisciplinary group of individuals with research interests in bioinformatic algorithms. For this reason, this year's workshop was held in conjunction with the *16th Annual European Symposium on Algorithms (ESA)*, the *6th Workshop on Approximation and Online Algorithms (WAOA)*, the *8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS)*, and the *ALGO 2008 Graduate Students Meeting*. We believe this interdisciplinary group, especially with the specific inclusion of graduate students and young scientists, is essential for moving the field of bioinformatics forward in exciting and novel ways. We hope this volume attests to that fact with the assortment of exceptional papers contributed to the volume.

We received a total of 81 submissions in response to our call for papers for *WABI 2008* from which 32 were selected to be contributions to this volume. The topics range in biological applicability from genome mapping, to sequence assembly, to microarray quality, to phylogenetic inference, to molecular modeling.

This volume was made possible by the excellent submissions to *WABI 2008*. We thank all the authors for their submissions and especially those of papers selected for inclusion in this volume for their help in revising their work according to peer-review comments. We also thank our distinguished Program Committee

for their great help in making difficult decisions on which papers to include in this volume. Anyone who has served on a program committee knows that this involves a great deal of concentrated effort over a short period of time. We are truly grateful for the excellent reviews and recommendations received by our Program Committee. These individuals are listed on the next page.

We were very fortunate to attract Eytan Ruppin of Tel Aviv University to present the Keynote Address for *WABI 2008* on the topic of large scale in *silico* studies of human metabolic diseases. We appreciate his willingness to deliver this address and establish the quality of the meetings. Finally, we would like to thank Peter Sanders and Dorothea Wagner and their local organizing committee (Veit Batz, Reinhard Bauer, Lilian Beckert, Anja Blancani, Daniel Delling, Dennis Luxen, Elke Sauer, and Dominik Schultes) for their excellent job in organizing these joint conferences.

We hope that you will consider contributing to future *WABI* events, through a submission or by participating in the workshops.

September 2008

Keith Crandall
Jens Lagergren

Conference Organization

Program Chairs

Keith A. Crandall, BYU, USA
Jens Lagergren, KTH, Sweden

Program Committee

Vineet Bafna, UCSD, USA
Dan Brown U. Waterloo, Canada
David Bryant, U. Auckland, New Zealand
Benny Chor, Tel Aviv, Israel
Mark Clement, BYU, USA
Keith A. Crandall (Co-chair), BYU, USA
Miklos Csuros, U. Montreal, Canada
Aaron Darling, U. Queensland, Australia
Matt Dyer, Virginia Tech, USA
Nadia El-Mabrouk, U. Montreal, Canada
Eleazar Eskin, UCLA, USA
Olivier Gascuel, LIRMM-CNRS, France
Roderico Guigo, U. Barcelona, Spain
Eran Halperin, Berkeley, USA
Arndt Von Haeseler, U. Vienna, Austria
Daniel Huson, U. Tübingen, Germany
John Kececioglu, U. Arizona, USA
Jens Lagergren (Co-chair), KTH, Sweden
Jim Leebens-Mack, U. Georgia, USA
Stefano Lonardi, UC Riverside, USA
Istvan Miklos, UPPS, Hungary
Bernard Moret, EPFL, Switzerland
Luay Nakhleh, Rice U., USA
Uwe Ohler, Duke U., USA
Laxmi Parida, IBM TJ Watson Research, USA
Ron Pinter, Technion, Israel
David Posada, U. Vigo, Spain
Ben Raphael, Brown, USA
Knut Reinert, Freie U. Berlin, Germany
Allen Rodrigo, U. Auckland, New Zealand
David Sankoff, U. Ottawa, Canada
Adam Siepel, Cornell U., USA
Mona Singh, Princeton U., USA

Saurabh Sinha, UIUC, USA
Jens Stoye, U. Bielefeld, Germany
Glenn Tesler, UCSD, USA
Esko Ukkonen, Helsinki, Finland
Steve Woolley, Washington U., USA
Christopher Workman, DTU, Denmark

Local Organization

Peter Sanders (Co-chair)
Dorothea Wagner (Co-chair)
Veit Batz
Reinhard Bauer
Lilian Beckert
Anja Blancani
Daniel Delling
Dennis Luxen
Elke Sauer
Dominik Schultes

External Reviewers

Max Alekseyev	Clemens Gröpl
José Augusto Amgarten Quitzau	Hugo Gutierrez de Teran
Sandro Andreotti	Eran Halperin
Alexander Auch	Elena Harris
Serdar Bozdag	Katharina Jahn
Robert Beiko	Crystal Kahn
Vincent Berry	Eun Yong Kang
Brona Brejova	Miriam Ruth Kantorovitz
Michael Brudno	Steven Kelk
Minh Bui Quang	Gad Kimmel
Robert Castelo	Dmitry Kondrashov
Mark Chaisson	Carolyn Kosiol
Cedric Chauve	Anne Kupczok
Benny Chor	Mathieu Lajoie
Matteo Comin	Guy Lapalme
Alexis Criscuolo	Sébastien Lemieux
Tahir Ejaz	Qiyuan Li
Olof Emanuelsson	William Majoros
N. Fernandez-Fuentes	Bob Mau
Philippe Gambette	Paul Medvedev
Jennifer Gardy	Martin Milanic
Claudio Garruti	Julia Mixtacki
Ilan Gronau	John Moriarty

Cedric Notredame
Sean O'Rourke
Sebastian Oehm
Stephan Ossowski
Simon Puglisi
Tobias Rausch
Anna Ritz
Regula Rupp
Michael Sammeth
Heiko Schmidt
Sagi Snir
Todd Treangen

Vladimir Vacic
Tomas Vinar
Wei-Bung Wang
David Weese
Sebastian Will
Roland Wittler
Haim Wolfson
Christopher Workman
Nir Yosef
Matthias Zytnicki
Yonghui Wu

Table of Contents

Multichromosomal Genome Median and Halving Problems	1
<i>Eric Tannier, Chunfang Zheng, and David Sankoff</i>	
A Branch-and-Bound Method for the Multichromosomal Reversal Median Problem	14
<i>Meng Zhang, William Arndt, and Jijun Tang</i>	
Decompositions of Multiple Breakpoint Graphs and Rapid Exact Solutions	25
<i>Andrew Wei Xu and David Sankoff</i>	
Read Mapping Algorithms for Single Molecule Sequencing Data	38
<i>Vladimir Yanovsky, Stephen M. Rumble, and Michael Brudno</i>	
Exact Transcriptome Reconstruction from Short Sequence Reads	50
<i>Vincent Lacroix, Michael Sammeth, Roderic Guigo, and Anne Bergeron</i>	
Post-Hybridization Quality Measures for Oligos in Genome-Wide Microarray Experiments	64
<i>Florian Battke, Carsten Müller-Tidow, Hubert Serve, and Kay Nieselt</i>	
NAPX: A Polynomial Time Approximation Scheme for the Noah’s Ark Problem	76
<i>Glenn Hickey, Paz Carmi, Anil Maheshwari, and Norbert Zeh</i>	
Minimum Common String Partition Parameterized	87
<i>Peter Damaschke</i>	
Hardness and Approximability of the Inverse Scope Problem	99
<i>Zoran Nikoloski, Sergio Grimbs, Joachim Selbig, and Oliver Ebenhöh</i>	
Rapid Neighbour-Joining	113
<i>Martin Simonsen, Thomas Mailund, and Christian N.S. Pedersen</i>	
Efficiently Computing Arbitrarily-Sized Robinson-Foulds Distance Matrices	123
<i>Seung-Jin Sul, Grant Brammer, and Tiffani L. Williams</i>	
Efficient Genome Wide Tagging by Reduction to SAT	135
<i>Arthur Choi, Noah Zaitlen, Buham Han, Knot Pipatsrisawat, Adnan Darwiche, and Eleazar Eskin</i>	

Computing the Minimal Tiling Path from a Physical Map by Integer Linear Programming	148
<i>Serdar Bozdag, Timothy J. Close, and Stefano Lonardi</i>	
An Efficient Lagrangian Relaxation for the Contact Map Overlap Problem	162
<i>Rumen Andonov, Nicola Yanev, and Noël Malod-Dognin</i>	
A Faster Algorithm for RNA Co-folding	174
<i>Michal Ziv-Ukelson, Irit Gat-Viks, Ydo Wexler, and Ron Shamir</i>	
An Automated Combination of Kernels for Predicting Protein Subcellular Localization	186
<i>Cheng Soon Ong and Alexander Zien</i>	
Fast Target Set Reduction for Large-Scale Protein Function Prediction: A Multi-class Multi-label Machine Learning Approach	198
<i>Thomas Lingner and Peter Meinicke</i>	
Multiple Instance Learning Allows MHC Class II Epitope Predictions Across Alleles	210
<i>Nico Pfeifer and Oliver Kohlbacher</i>	
An Algorithm for Orienting Graphs Based on Cause-Effect Pairs and Its Applications to Orienting Protein Networks	222
<i>Alexander Medvedovsky, Vineet Bafna, Uri Zwick, and Roded Sharan</i>	
Enumerating Precursor Sets of Target Metabolites in a Metabolic Network	233
<i>Ludovic Cottret, Paulo Vieira Milreu, Vicente Acuña, Alberto Marchetti-Spaccamela, Fábio Viduani Martinez, Marie-France Sagot, and Leen Stougie</i>	
Boosting the Performance of Inference Algorithms for Transcriptional Regulatory Networks Using a Phylogenetic Approach	245
<i>Xiwei Zhang and Bernard M.E. Moret</i>	
Fast Bayesian Haplotype Inference Via Context Tree Weighting	259
<i>Pasi Rastas, Jussi Kollin, and Mikko Koivisto</i>	
Genotype Sequence Segmentation: Handling Constraints and Noise	271
<i>Qi Zhang, Wei Wang, Leonard McMillan, Jan Prins, Fernando Pardo-Manuel de Villena, and David Threadgill</i>	
Constructing Phylogenetic Supernetworks from Quartets	284
<i>Stefan Grünwald, Andreas Spillner, Kristoffer Forslund, and Vincent Moulton</i>	
Summarizing Multiple Gene Trees Using Cluster Networks	296
<i>Daniel H. Huson and Regula Rupp</i>	

Fast and Adaptive Variable Order Markov Chain Construction	306
<i>Marcel H. Schulz, David Weese, Tobias Rausch, Andreas Döring, Knut Reinert, and Martin Vingron</i>	
Computing Alignment Seed Sensitivity with Probabilistic Arithmetic Automata	318
<i>Inke Herms and Sven Rahmann</i>	
The Relation between Indel Length and Functional Divergence: A Formal Study	330
<i>Raheleh Salari, Alexander Schönhuth, Fereydown Hormozdiari, Artem Cherkasov, and S. Cenk Sahinalp</i>	
Detecting Repeat Families in Incompletely Sequenced Genomes	342
<i>José Augusto Amgarten Quitzau and Jens Stoye</i>	
Novel Phylogenetic Network Inference by Combining Maximum Likelihood and Hidden Markov Models (Extended Abstract)	354
<i>Sagi Snir and Tamir Tuller</i>	
A Local Move Set for Protein Folding in Triangular Lattice Models	369
<i>Hans-Joachim Böckenhauer, Abu Zafer M. Dayem Ullah, Leonidas Kapsokalivas, Kathleen Steinhöfel</i>	
Protein Decoy Generation Using Branch and Bound with Efficient Bounding	382
<i>Martin Paluszewski and Pawel Winter</i>	
Author Index	395