Lecture Notes in Bioinformatics 5724

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

Steven L. Salzberg Tandy Warnow (Eds.)

Algorithms in Bioinformatics

9th International Workshop, WABI 2009
Philadelphia, PA, USA, September 12-13, 2009
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

Steven L. Salzberg
Center for Bioinformatics and Computational Biology
and Department of Computer Science
University of Maryland
College Park, MD, USA
E-mail: salzberg@umiacs.umd.edu

Tandy Warnow
Department of Computer Sciences
The University of Texas at Austin
Austin, TX, USA
E-mail: tandy@cs.utexas.edu

Library of Congress Control Number: 2009933693

CR Subject Classification (1998): J.3, I.5, F.2.2, E.1, G.2, F.2

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743
ISBN-10 3-642-04240-6 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-04240-9 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 2009
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12749041 06/3180 5 4 3 2 1 0

Preface

These proceedings contain papers from the 2009 Workshop on Algorithms in Bioinformatics (WABI), held at the University of Pennsylvania in Philadelphia, Pennsylvania during September 12–13, 2009. WABI 2009 was the ninth annual conference in this series, which focuses on novel algorithms that address important problems in genomics, molecular biology, and evolution. The conference emphasizes research that describes computationally efficient algorithms and data structures that have been implemented and tested in simulations and on real data. WABI is sponsored by the European Association for Theoretical Computer Science (EATCS) and the International Society for Computational Biology (ISCB). WABI 2009 was supported by the Penn Genome Frontiers Institute and the Penn Center for Bioinformatics at the University of Pennsylvania.

For the 2009 conference, 90 full papers were submitted for review by the Program Committee, and from this strong field of submissions, 34 papers were chosen for presentation at the conference and publication in the proceedings. The final program covered a wide range of topics including gene interaction networks, molecular phylogeny, RNA and protein structure, and genome evolution. Extended versions of selected papers will appear in a special issue of the journal *Algorithms for Molecular Biology*, published by BiomedCentral. We thank all the members of the Program Committee for their detailed and timely reviews, and we also thank the authors of all the papers, whose contributions made the workshop possible. Thanks also to the 2009 Keynote speakers, Shelley Berger of the Wistar Institute (Philadelphia, Pennsylvania) and Elchanan Mossel of the Weizmann Institute of Science (Rehovot, Israel). Above all, we want to extend our appreciation to the Organizing Committee at the University of Pennsylvania – Junhyong Kim (Chair), Stephen Fisher, and Li-San Wang – who handled all of the local arrangements, the conference website, and many other details that made the conference a success.

September 2009

Steven L. Salzberg
Tandy Warnow

Organization

Program Chairs

Steven Salzberg	University of Maryland, College Park, USA
Tandy Warnow	University of Texas, Austin, USA

Program Committee

Elizabeth Allman	University of Alaska, Fairbanks, USA
Tanya Y. Berger-Wolf	University of Illinois, Chicago, USA
Paola Bonizzoni	UniversitáDegli Studi di Milano-Bicocca, Italy
Dan Brown	University of Waterloo, Canada
Rita Casadio	Universitá Degli Studi di Milano-Bicocca, Italy
Mark Clement	Brigham Young University, USA
Maxime Crochemore	Université de Marne-la-Vallé, Carte, France
Miklós Csűrös	Université de Montréal, Canada
Aaron Darling	University of California, Davis, USA
Nadia El-Mabrouk	Université de Montréal, Canada
Eleazar Eskin	University of California, Los Angeles, USA
Liliana Florea	University of Maryland, College Park, USA
Ganesh Ganapathy	Duke University, Durham, USA
Raffaele Giancarlo	Universitá degli Studi di Palermo, Italy
Concettina Guerra	Universitá di Padova, Italy
Roderic Guigó	University of Barcelona, Spain
Sridhar Hannenhalli	University of Pennsylania, Philadelphia, USA
Daniel Huson	University of Tübingen, Germany
Shane T. Jensen	University of Pennsylvania, Philadelphia, USA
Carl Kingsford	University of Maryland, College Park, USA
Ming Li	University of Waterloo, Canada
C. Randal Linder	University of Texas, Austin, USA
Stefano Lonardi	University of California, Riverside, USA
Lusheng Lusheng-Wang	City University of Hong Kong, Hong Kong
Jian Ma	University of California, Santa Cruz, USA
Erick Matsen	University of California, Berkeley, USA
István Miklós	Rényi Institute, Budapest, Hungary
Bernard Moret	Swiss Federal Institute of Technology, Lausanne, Switzerland
Burkhard Morgenstern	University of Göttingen, Germany
Elchanan Mossel	University of California, Berkeley, USA

Luay K. Nakhleh	Rice University, Houston, USA
Macha Nikolski	Université Bordeaux, France
Laxmi Parida	IBM T.J. Watson Research Center, Yorktown Heights, USA
Kunsoo Park	Seoul National University, Korea
Ron Pinter	Israel Institute of Technology, Haifa, Israel
Cinzia Pizzi	University of Padova, Italy
David Posada	University of Vigo, Spain
Sven Rahmann	Technische Universität Dortmund, Germany
Ben Raphael	Brown University, Providence, USA
Knut Reinert	Freie Universität Berlin, Germany
Allen Rodrigo	University of Auckland, New Zealand
David Sankoff	University of Ottawa, Canada
Mona Singh	Princeton University, USA
Saurabh Sinha	University of Illinois, Urbana, USA
Steven Skiena	State University of New York, Stonybrook, USA
Peter F. Stadler	Universität Leipzig, Germany
Alexandros Stamatakis	Technische Universität München, Germany
Jens Stoye	Universität Bielefeld, Germany
Jijun Tang	University of South Carolina, Columbia, USA
Glenn Tesler	University of California, San Diego, USA
Olga Troyanskaya	Princeton University, USA
Alfonso Valencia	Spanish National Cancer Research Centre, Madrid, Spain
Chris Workman	Technical University of Denmark, Lyngby, Denmark

Local Organization

Junhyong Kim (Chair)	University of Pennsylvania, Philadelphia, USA
Stephen Fisher	University of Pennsylvania, Philadelphia, USA
Li-San Wang	University of Pennsylvania, Philadelphia, USA

Table of Contents

Minimum Factorization Agreement of Spliced ESTs 1
Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, Yuri Pirola, and Raffaella Rizzi

Annotating Fragmentation Patterns 13
Sebastian Böcker, Florian Rasche, and Tamara Steijger

biRNA: Fast RNA-RNA Binding Sites Prediction 25
Hamidreza Chitsaz, Rolf Backofen, and S. Cenk Sahinalp

Quantifying Systemic Evolutionary Changes by Color Coding Confidence-Scored PPI Networks 37
Phuong Dao, Alexander Schönhuth, Fereydoun Hormozdiari, Iman Hajirasouliha, S. Cenk Sahinalp, and Martin Ester

PMFastR: A New Approach to Multiple RNA Structure Alignment 49
Daniel DeBlasio, Jocelyne Bruand, and Shaojie Zhang

On the Upper Bound of the Prediction Accuracy of Residue Contacts in Proteins with Correlated Mutations: The Case Study of the Similarity Matrices 62
Pietro Di Lena, Piero Fariselli, Luciano Margara, Marco Vassura, and Rita Casadio

Constructing Majority-Rule Supertrees 73
Jianrong Dong, David Fernández-Baca, and F.R. McMorris

SCJ: A Variant of Breakpoint Distance for Which Sorting, Genome Median and Genome Halving Problems Are Easy 85
Pedro Feijão and João Meidanis

A Simple, Practical and Complete $O(\frac{n^3}{\log n})$-Time Algorithm for RNA Folding Using the *Four-Russians* Speedup 97
Yelena Frid and Dan Gusfield

Back-Translation for Discovering Distant Protein Homologies 108
Marta Gîrdea, Laurent Noé, and Gregory Kucherov

A Markov Classification Model for Metabolic Pathways 121
Timothy Hancock and Hiroshi Mamitsuka

Mimosa: Mixture Model of Co-expression to Detect Modulators of Regulatory Interaction 133
Matthew Hansen, Logan Everett, Larry Singh, and Sridhar Hannenhalli

Phylogenetic Comparative Assembly 145
Peter Husemann and Jens Stoye

K-Partite RNA Secondary Structures 157
Minghui Jiang, Pedro J. Tejada, Ramoni O. Lasisi, Shanhong Cheng, and D. Scott Fechser

Efficient Algorithms for Analyzing Segmental Duplications, Deletions, and Inversions in Genomes 169
Crystal L. Kahn, Shay Mozes, and Benjamin J. Raphael

Predicting Gene Structures from Multiple RT-PCR Tests 181
Jakub Kováč, Tomáš Vinař, and Broňa Brejová

A Tree Based Method for the Rapid Screening of Chemical Fingerprints 194
Thomas G. Kristensen, Jesper Nielsen, and Christian N.S. Pedersen

Generalizing the Four Gamete Condition and Splits Equivalence Theorem: Perfect Phylogeny on Three State Characters 206
Fumei Lam, Dan Gusfield, and Srinath Sridhar

Decoding Synteny Blocks and Large-Scale Duplications in Mammalian and Plant Genomes 220
Qian Peng, Max A. Alekseyev, Glenn Tesler, and Pavel A. Pevzner

A General Framework for Local Pairwise Alignment Statistics with Gaps 233
Pasi Rastas

MPSCAN: Fast Localisation of Multiple Reads in Genomes 246
Eric Rivals, Leena Salmela, Petteri Kiiskinen, Petri Kalsi, and Jorma Tarhio

Fast Prediction of RNA-RNA Interaction 261
Raheleh Salari, Rolf Backofen, and S. Cenk Sahinalp

FlexSnap: Flexible Non-sequential Protein Structure Alignment 273
Saeed Salem, Mohammed J. Zaki, and Chris Bystroff

A Non-parametric Bayesian Approach for Predicting RNA Secondary Structures 286
Kengo Sato, Michiaki Hamada, Toutai Mituyama, Kiyoshi Asai, and Yasubumi Sakakibara

Exact Score Distribution Computation for Similarity Searches in Ontologies 298
Marcel H. Schulz, Sebastian Köhler, Sebastian Bauer, Martin Vingron, and Peter N. Robinson

Linear-Time Protein 3-D Structure Searching with Insertions and Deletions 310
Tetsuo Shibuya, Jesper Jansson, and Kunihiko Sadakane

Visualizing Phylogenetic Treespace Using Cartographic Projections 321
Kenneth Sundberg, Mark Clement, and Quinn Snell

A Simulation Study Comparing Supertree and Combined Analysis Methods Using SMIDGen 333
M. Shel Swenson, François Barbançon, C. Randal Linder, and Tandy Warnow

Aligning Biomolecular Networks Using Modular Graph Kernels 345
Fadi Towfic, M. Heather West Greenlee, and Vasant Honavar

MADMX: A Novel Strategy for Maximal Dense Motif Extraction 362
Roberto Grossi, Andrea Pietracaprina, Nadia Pisanti, Geppino Pucci, Eli Upfal, and Fabio Vandin

Large-Scale Neighbor-Joining with NINJA 375
Travis J. Wheeler

A Unifying View on Approximation and FPT of Agreement Forests (Extended Abstract) 390
Chris Whidden and Norbert Zeh

Structural Alignment of RNA with Complex Pseudoknot Structure 403
Thomas K.F. Wong, T.W. Lam, Wing-Kin Sung, and S.M. Yiu

Improving Inference of Transcriptional Regulatory Networks Based on Network Evolutionary Models 415
Xiuwei Zhang and Bernard M.E. Moret

Author Index 429