

Lecture Notes in Bioinformatics

5542

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

Ion Măndoiu Giri Narasimhan
Yanqing Zhang (Eds.)

Bioinformatics Research and Applications

5th International Symposium, ISBRA 2009
Fort Lauderdale, FL, USA, May 13-16, 2009
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

Ion Măndoiu

University of Connecticut

Computer Science & Engineering Department

371 Fairfield Way, Unit 2155, Storrs, CT 06269, USA

E-mail: ion@engr.uconn.edu

Giri Narasimhan

Florida International University

School of Computing and Information Sciences

Bioinformatics Research Group (BioRG)

11200 SW 8th Street, Room ECS254, University Park, Miami, FL 33199, USA

E-mail: giri@cs.fiu.edu

Yanqing Zhang

Georgia State University

Department of Computer Science

Atlanta, GA 30302-3994, USA

E-mail: yzhang@cs.gsu.edu

Library of Congress Control Number: Applied for

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

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-642-01550-6 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-01550-2 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: 12672264 06/3180 5 4 3 2 1 0

Preface

The 5th edition of the International Symposium on Bioinformatics Research and Applications (ISBRA 2009) was held during May 13–16, 2009 at Nova Southeastern University in Ft. Lauderdale, Florida. The symposium provides a forum for the exchange of ideas and results among researchers, developers, and practitioners working on all aspects of bioinformatics and computational biology and their applications.

The technical program of the symposium included 26 contributed papers, selected by the Program Committee from a number of 55 full submissions received in response to the call for papers. The technical program also included contributed papers and abstracts submitted to the Second Workshop on Computational Issues in Genetic Epidemiology (CIGE 2009), which was held in conjunction with ISBRA 2009. Additionally, the symposium included poster sessions and featured invited keynote talks by four distinguished speakers: Mikhail Gelfand from the Russian Academy of Sciences and Moscow State University spoke on evolution of regulatory systems in bacteria, Nicholas Tsinoremas from the Miller School of Medicine and the College of Arts and Sciences at the University of Miami spoke on bioinformatics challenges in translational research, Esko Ukkonen from the University of Helsinki spoke on motif construction from high-throughput SELEX data, and Shamil Sunyaev from Brigham and Women's Hospital and Harvard Medical School spoke on interpreting population sequencing data.

We would like to thank the Program Committee members and external reviewers for volunteering their time to review and discuss symposium papers. We would also like to thank the Chairs and the Program Committee of CIGE 2009 for enriching the technical program of the symposium with a workshop on an important and active area of bioinformatics research. We would like to extend special thanks to the Steering and General Chairs of the symposium for their leadership, and to the Finance, Publicity, Local Organization, Posters Chairs, and Web Master for their hard work in making ISBRA 2009 a successful event. Last but not least, we would like to thank all authors for presenting their work at the symposium.

May 2009

Ion Măndoiu
Giri Narasimhan
Yanqing Zhang

Organization

5th International Symposium on Bioinformatics Research and Applications (ISBRA 2009)

Steering Chairs

Dan Gusfield	University of California, Davis, USA
Yi Pan	Georgia State University, USA
Marie-France Sagot	INRIA, France

General Chairs

Matthew He	Nova Southeastern University, USA
Alexander Zelikovsky	Georgia State University, USA

Program Chairs

Ion Măndoiu	University of Connecticut, USA
Giri Narasimhan	Florida International University, USA
Yanqing Zhang	Georgia State University, USA

Publicity Chair

Raj Sunderraman	Georgia State University, USA
-----------------	-------------------------------

Finance Chair

Anu Bourgeois	Georgia State University, USA
---------------	-------------------------------

Poster Chairs

Yufeng Wu	University of Connecticut, USA
Craig E. Nelson	University of Connecticut, USA

Local Organization Chairs

Edward Keith	Nova Southeastern University, USA
Miguel A. Jimenez-Montano	Universidad Veracruzana, Mexico

Local Organization Committee

Ahmed Albatineh	Nova Southeastern University, USA
Ricardo Carrera	Nova Southeastern University, USA
Josh Loomis	Nova Southeastern University, USA
Evan Haskell	Nova Southeastern University, USA
Saeed Rajput	Nova Southeastern University, USA
Reza Razeghifard	Nova Southeastern University, USA
Raisa Szabo	Nova Southeastern University, USA

Web Master

Zejin Jason Ding Georgia State University, USA

Program Committee

Srinivas Aluru Iowa State University, USA	Bhaskar Dasgupta University of Illinois at Chicago, USA
Danny Barash Ben-Gurion University, Israel	Colin Dewey University of Wisconsin-Madison, USA
Anne Bergeron Université du Québec à Montréal, Canada	Werner Dubitzky University of Ulster, UK
Tanya Berger-Wolf University of Illinois at Chicago, USA	Guillaume Fertin Université de Nantes, France
Daniel Berrar University of Ulster, UK	Liliana Florea George Washington University, USA
Olivier Bodenreider National Library of Medicine, NIH, USA	Jean Gao University of Texas at Arlington, USA
Paola Bonizzoni Univ. de Studi di Milano-Bicocca, Italy	Mikhail Gelfand IITP, Russia
Daniel Brown University of Waterloo, Canada	Michael Gribskov Purdue University, USA
Liming Cai University of Georgia, USA	Katia Guimarães Universidade Federal de Pernambuco, Brazil
Luonan Chen Osaka Sangyo University, Japan	Robert Harrison Georgia State University, USA

Jieyue He
Southeast University, China

Vasant Honavar
Iowa State University, USA

Lars Kaderali
University of Heidelberg, Germany

Ming-Yang Kao
Northwestern University, USA

George Karypis
University of Minnesota, USA

Yury Khudyakov
Centers for Disease Control and
Prevention, USA

Jing Li
Case Western Reserve University, USA

Yiming Li
National Chiao Tung University,
Taiwan

Guohui Lin
University of Alberta, Canada

Stefano Lonardi
University of California at
Riverside, USA

Jingchu Luo
Peking University, China

Osamu Maruyama
Kyushu University, Japan

Satoru Miyano
University of Tokyo, Japan

Bernard Moret
Ecole Poly. Fed. de Lausanne,
Switzerland

Craig Nelson
University of Connecticut, USA

Laxmi Parida
IBM T.J. Watson Research Center,
USA

Itsik Pe'er
Columbia University, USA

Mihai Pop
University of Maryland, USA

Teresa Przytycka
NCBI, USA

Sven Rahmann
Technical University of Dortmund,
Germany

Sanguthevar Rajasekaran
University of Connecticut, USA

Shoba Ranganathan
Macquarie University, Australia

Isidore Rigoutsos
IBM Research, USA

Cenk Sahinalp
Simon Fraser, Canada

David Sankoff
University of Ottawa, Canada

Russell Schwartz
Carnegie Mellon University, USA

João Carlos Setubal
Virginia Polytechnic Institute and
State University, USA

Mona Singh
Princeton University, USA

Steve Skiena
State University of New York
at Stony Brook, USA

Donna Slonim
Tufts University, USA

Ramanathan Sowdhamini
NCBS, India

Jens Stoye
Universität Bielefeld, Germany

Wing-Kin Sung
National University of
Singapore, Singapore

Sing-Hoi Sze
Texas A&M University, USA

Haixu Tang
Indiana University, USA

Gabriel Valiente
Technical University of
Catalonia, Spain

Jean-Philippe Vert
Ecole des Mines de Paris, France

Stéphane Vialette
Université Paris-Est
Marne-la-Vallée, France

Gwenn Volkert
Kent State University, USA

Li-San Wang
University of Pennsylvania, USA

Lusheng Wang
City University of Hong Kong, China

Carsten Wiuf
University of Aarhus, Denmark

Hongwei Wu
University of Georgia, USA

Yufeng Wu
University of Connecticut, USA

Dong Xu
University of Missouri-Columbia, USA

Kaizhong Zhang
University of West Ontario, Canada

Leming Zhou
University of Pittsburgh, USA

External Reviewers

Angibaud, Sébastien
Araujo, Flavia
Assareh, Amin
Astrovakaya, Irina
Bernauer, Julie
Blin, Guillaume
Chen, Shihyen
Comin, Matteo
DeRonne, Kevin
Della Vedova, Gianluca
Dewal, Ninad
Dondi, Riccardo
Ghods, MohammadReza
Guillemot, Sylvain
Harris, Elena
Husemann, Peter
Jahn, Katharina
Jin, Guangxu

Kauffman, Chris
Kim, Dongchul
Kim, Yoo-Ah
Knapp, Bettina
Krishnan, Yamuna
Lara, James
Li, Weiming
Liu, Bo
Liu, Zhiping
Mangul, Serghei
Marschall, Tobias
Martin, Marcel
Mazur, Johanna
Monteiro, Carla
Offmann, Bernard
Palamara, Pierre
Podolyan, Yevgeniy
Pugalenthi, Ganesan

Radde, Nicole
Rizzi, Raffaella
Rosa, Rogerio
Rusu, Irena
Salari, Rahele
Schoenhuth, Alex
Sheikh, Saad
Stoffer, Deborah

Tripathi, Lokesh
Wittler, Roland
Wojtowicz, Damian
Wu, Lingyun
Zhao, Xingming
Zheng, Jie
Zola, Jaroslaw

Second Workshop on Computational Issues in Genetic Epidemiology (CIGE 2009)

Steering Committee

Andrew Allen	Duke University, USA
Ion Măndoiu	University of Connecticut, USA
Dan Nicolae	University of Chicago, USA
Yi Pan	Georgia State University, USA
Alex Zelikovsky	Georgia State University, USA

Program Chairs

Andrew Allen	Duke University, USA
Itsik Pe'er	Columbia University, USA

Program Committee

Dave Cutler	Emory University, USA
Frank Dudbridge	Cambridge University, UK
Eleazar Eskin	UCLA, USA
Eran Halperin	UC Berkeley/Tel Aviv University, USA/Israel
David Heckerman	Microsoft Research, USA
Chun Li	Vanderbilt University, USA
Eden Martin	Miami University, USA
Shaun Purcell	Harvard University, USA
Hongyu Zhao	Yale University, USA

Table of Contents

Evolution of Regulatory Systems in Bacteria (Invited Keynote Talk) . . .	1
<i>Mikhail S. Gelfand, Alexei E. Kazakov, Yuri D. Korostelev, Olga N. Laikova, Andrei A. Mironov, Alexandra B. Rakhmaninova, Dmitry A. Ravcheev, Dmitry A. Rodionov, and Alexei G. Vitreschak</i>	
Integrating Multiple-Platform Expression Data through Gene Set Features	5
<i>Matěj Holec, Filip Železný, Jiří Kléma, and Jakub Tolar</i>	
Practical Quality Assessment of Microarray Data by Simulation of Differential Gene Expression.	18
<i>Brian E. Howard, Beate Sick, and Steffen Heber</i>	
Mean Square Residue Biclustering with Missing Data and Row Inversions	28
<i>Stefan Gremalschi, Gulsah Altun, Irina Astrovskaya, and Alexander Zelikovskiy</i>	
Using Gene Expression Modeling to Determine Biological Relevance of Putative Regulatory Networks	40
<i>Peter Larsen and Yang Dai</i>	
Querying Protein-Protein Interaction Networks	52
<i>Guillaume Blin, Florian Sikora, and Stéphane Vialette</i>	
Integrative Approach for Combining TNF α -NF κ B Mathematical Model to a Protein Interaction Connectivity Map	63
<i>Mahesh Visvanathan, Bernhard Pfeifer, Christian Baumgartner, Bernhard Tilg, and Gerald Henry Lushington</i>	
Hierarchical Organization of Functional Modules in Weighted Protein Interaction Networks Using Clustering Coefficient	75
<i>Min Li, Jianxin Wang, Jianer Chen, and Yi Pan</i>	
Bioinformatics Challenges in Translational Research (Invited Keynote Talk)	87
<i>Nicholas F. Tsimoremas</i>	
Untangling Tanglegrams: Comparing Trees by Their Drawings	88
<i>Balaji Venkatachalam, Jim Apple, Katherine St. John, and Dan Gusfield</i>	
An Experimental Analysis of Consensus Tree Algorithms for Large-Scale Tree Collections	100
<i>Seung-Jin Sul and Tiffani L. Williams</i>	

Counting Faces in Split Networks	112
<i>Lichen Bao and Sergey Bereg</i>	
Relationship between Amino Acids Sequences and Protein Structures: Folding Patterns and Sequence Patterns	124
<i>Alexander Kister</i>	
Improved Algorithms for Parsing ESLTAGs: A Grammatical Model Suitable for RNA Pseudoknots	135
<i>Sanguthevar Rajasekaran, Sahar Al Seesi, and Reda Ammar</i>	
Efficient Algorithms for Self Assembling Triangular and Other Nano Structures	148
<i>Vamsi Kundeti and Sanguthevar Rajasekaran</i>	
Motif Construction from High-Throughput SELEX Data (Invited Keynote Talk)	159
<i>Esko Ukkonen</i>	
Rearrangement Phylogeny of Genomes in Contig Form	160
<i>Adriana Muñoz and David Sankoff</i>	
Prediction of Contiguous Regions in the Amniote Ancestral Genome . . .	173
<i>Aïda Ouangraoua, Frédéric Boyer, Andrew McPherson, Éric Tannier, and Cedric Chauve</i>	
Pure Parsimony Xor Haplotyping	186
<i>Paola Bonizzoni, Gianluca Della Vedova, Riccardo Dondi, Yuri Pirola, and Romeo Rizzi</i>	
A Decomposition of the Pure Parsimony Haplotyping Problem	198
<i>Allen Holder and Thomas Langley</i>	
Exact Computation of Coalescent Likelihood under the Infinite Sites Model	209
<i>Yufeng Wu</i>	
Imputation-Based Local Ancestry Inference in Admixed Populations . . .	221
<i>Bogdan Paşaniuc, Justin Kennedy, and Ion Măndoiu</i>	
Interpreting Population Sequencing Data (Invited Keynote Talk)	234
<i>Shamil R. Sunyaev</i>	
Modeling and Visualizing Heterogeneity of Spatial Patterns of Protein-DNA Interaction from High-Density Chromatin Precipitation Mapping Data	236
<i>Juntao Li, Fajrian Yunus, Zhu Lei, Majid Eshaghi, Jianhua Liu, and R. Krishna Murthy Karuturi</i>	

A Linear-Time Algorithm for Analyzing Array CGH Data Using Log Ratio Triangulation	248
<i>Matthew Hayes and Jing Li</i>	
Mining of <i>cis</i> -Regulatory Motifs Associated with Tissue-Specific Alternative Splicing	260
<i>Jihye Kim, Sihui Zhao, Brian E. Howard, and Steffen Heber</i>	
Analysis of Cis-Regulatory Motifs in Cassette Exons by Incorporating Exon Skipping Rates	272
<i>Sihui Zhao, Jihye Kim, and Steffen Heber</i>	
A Class of Evolution-Based Kernels for Protein Homology Analysis: A Generalization of the PAM Model	284
<i>Valentina Sulimova, Vadim Mottl, Boris Mirkin, Ilya Muchnik, and Casimir Kulikowski</i>	
Irreplaceable Amino Acids and Reduced Alphabets in Short-Term and Directed Protein Evolution	297
<i>Miguel A. Jiménez-Montaño and Matthew He</i>	
A One-Class Classification Approach for Protein Sequences and Structures	310
<i>András Bánhalmi, Róbert Busa-Fekete, and Balázs Kégl</i>	
Prediction and Classification of Real and Pseudo MicroRNA Precursors via Data Fuzzification and Fuzzy Decision Trees	323
<i>Na'el Abu-halaweh and Robert Harrison</i>	
Author Index	335