

Lecture Notes in Bioinformatics

5462

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

Sanguthevar Rajasekaran (Ed.)

Bioinformatics and Computational Biology

First International Conference, BICoB 2009
New Orleans, LA, USA, April 8-10, 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 Editor

Sanguthevar Rajasekaran

University of Connecticut, Department of Computer Science and Engineering

257 ITE Building, 371 Fairfield Way, Storrs, CT 06269-2155, USA

E-mail: rajasek@engr.uconn.edu

Library of Congress Control Number: Applied for

CR Subject Classification (1998): H.2.8, J.3, I.5, I.2, H.3, F.1-2

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-642-00726-0 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-00726-2 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: 12633326 06/3180 5 4 3 2 1 0

Preface

This volume presents the proceedings of the First International Conference on Bioinformatics and Computational Biology (BICoB 2009). This conference was supported by the International Society for Computers and Applications (ISCA) and Springer.

Computational techniques have already enabled unprecedented advances in modern biology and medicine. This continues to be a vibrant research area with broadening of computational techniques and new emerging challenges. The Bioinformatics and Computational Biology (BICoB) conference has the goal of promoting the advancement of computing techniques and their application to life sciences. The topics of interest include (and are not limited to):

- Genome analysis: genome assembly; genome and chromosome annotation, gene finding; alternative splicing; EST analysis and comparative genomics
- Sequence analysis: multiple sequence alignment; sequence search and clustering; function prediction, motif discovery, functional site recognition in protein, RNA and DNA sequences
- Phylogenetics: phylogeny estimation; models of evolution; comparative biological methods; population genetics
- Structural Bioinformatics: structure matching, prediction, analysis and comparison; methods and tools for docking; protein design
- Analysis of high-throughput biological data: microarrays (nucleic acid, protein, array CGH, genome tiling, and other arrays); EST; SAGE; MPSS; proteomics; mass spectrometry
- Genetics and population analysis: linkage analysis; association analysis; population simulation; haplotyping; marker discovery; genotype calling
- Systems biology: systems approaches to molecular biology; multiscale modeling; pathways; gene networks

BICoB is interested in all areas of computing with an impact on life sciences including (but not limited to) algorithms, databases, languages, systems, and high-performance computing. Examples include:

- Parallel and high-performance techniques
- Unifying computational techniques
- Data and image mining techniques
- Approximation and randomized algorithms and systems
- Computational biology on emerging architectures and hardware accelerators

In BICoB 2009 there were three keynote speeches, ten invited talks, and 30 contributed presentations (selected from a total of 72 submissions). We are grateful to the keynote speakers and the invited speakers who contributed tremendously to the success of the conference. We are also thankful to all the authors who submitted papers, especially those who gave presentations at the conference.

The Program Committee members as well as reviewers recruited by them deserve a special thanks for their meticulous work in selecting the contributed papers. We are also grateful to Sahar Al Seesi for finalizing the papers for the proceedings.

April 2009

Sanguthevar Rajasekaran
Srinivas Aluru
Limsoon Wang

Organization

The First International Conference on BioInformatics and Computational Biology (BICoB) was organized by the International Society of Computers and Applications (ISCA) in collaboration with Springer

Executive Committee

General Chair	Sanguthevar Rajasekaran (University of Connecticut)
Program Co-chairs	Srinivas Aluru (Iowa State University) Limsoon Wang (National University of Singapore)
Steering Committee	Reda Ammar (University of Connecticut) Tao Jiang (University of California, Riverside) Vipin Kumar (University of Minnesota) Ming Li (University of Waterloo) Sanguthevar Rajasekaran, Chair (University of Connecticut) John Reif (Duke University) Sartaj Sahni (University of Florida, Gainesville)
Publicity Co-chairs	Ian Greenshields (University of Connecticut) Chun-Hsi Huang (University of Connecticut)
Keynote Speakers	Richard Karp, University of California, Berkeley Ming Li, University of Waterloo Vineet Bafna, University of California, San Diego
Invited Speakers	Srinivas Aluru, Iowa State University Vladimir Filkov, University of California, Davis Vipin Kumar, University of Minnesota Bin Ma, University of Waterloo Ion Mandoiu, University of Connecticut Satoru Miyano, University of Tokyo T.M. Murali, Virginia Tech. Mona Singh, Princeton University Wing-Kin Sung, National University of Singapore Dong Xu, University of Missouri, Columbia
Proceedings Coordinator	Sahar Al Seesi, University of Connecticut

Program Committee

Richa Agarwala	National Institute of Health
Yutaka Akiyama	Tokyo Institute of Technology
Ziv Bar-Joseph	Carnegie Mellon University
Chiranjib Bhattacharyya	Indian Institute of Science
Kun-Mao Chao	National Taiwan University
Jake Chen	Indiana University
Francis Y.L. Chin	Hong Kong University
Bhaskar Dasgupta	University of Illinois, Chicago
Scott J. Emrich	University of Notre Dame
Oliver Eulenstein	Iowa State University
Vladimir Filkov	University of California, Davis
Dmitrij Frishman	Tech. University of Munich
Terry Gaasterland	University of California, San Diego
Gaston Gonnet	ETH, Zurich
Osamu Gotoh	Kyoto University
Wen-Lian Hsu	Academia Sinica
Ming-Yang Kao	Northwestern University
Marek Karpinski	University of Bonn
George Karypis	University of Minnesota
Danny Krizanc	Wesleyan University
Bin Ma	University of Waterloo
Ion Mandoiu	University of Connecticut
Shinichi Morishita	University of Tokyo
Giri Narasimhan	Florida International University
Enno Ohlebusch	University of Ulm
Yi Pan	Georgia State University
Srinivasan Parthasarathy	Ohio State University
Prabakaran Ponraj	Duke University
Mihai Pop	University of Maryland
David Posada	University of Vigo, Spain
Ben Raphael	Brown University
Knut Reinert	Freie University of Berlin
Isidore Rigoutsos	IBM TJ Watson Research Center
Marie-France Sagot	INRIA Rhone-Alpes
Mona Singh	Princeton University
Wing-Kin Sung	National University of Singapore
Sing-Hoi Sze	Texas A&M University
Jerzy Tiuryn	Warsaw University
Ugo Vaccaro	University of Salerno
Li-San Wang	University of Pennsylvania
Yufeng Wu	University of Connecticut
Dong Xu	University of Missouri, Columbia
Alex Zelikovsky	Georgia State University

Reviewers

M. Bader	J. He	U. Schöning
S. Balla	A. Kolinski	O. Schulztrieglaß
M. Bauer	T.-H. Lin	S. Soi
P. Biecek	Y. Lu	M. Song
K. Cao	R. Preisner	E. Szczurek
J. Davila	Y. Qi	D. Ucar
J. Ernst	S. Roy	B. Wilczynski
A. Fox	P. Ryvkin	

Sponsoring Institutions

University of Connecticut, Storrs

Booth Engineering Center for Advanced Technologies, Storrs

Table of Contents

Invited Talks

Association Analysis Techniques for Bioinformatics Problems	1
<i>Gowtham Atluri, Rohit Gupta, Gang Fang, Gaurav Pandey, Michael Steinbach, and Vipin Kumar</i>	
Analyzing and Interrogating Biological Networks (Abstract)	14
<i>Eric Banks, Elena Nabieva, Bernard Chazelle, Ryan Peterson, and Mona Singh</i>	
From Architecture to Function (and Back) in Bio-networks	16
<i>Vladimir Filkov</i>	
A New Machine Learning Approach for Protein Phosphorylation Site Prediction in Plants	18
<i>Jianjiong Gao, Ganesh Kumar Agrawal, Jay J. Thelen, Zoran Obradovic, A. Keith Dunker, and Dong Xu</i>	
Assembly of Large Genomes from Paired Short Reads	30
<i>Benjamin G. Jackson, Patrick S. Schnable, and Srinivas Aluru</i>	
Amino Acid Classification and Hash Seeds for Homology Search	44
<i>Weiming Li, Bin Ma, and Kaizhong Zhang</i>	
Genotype and Haplotype Reconstruction from Low-Coverage Short Sequencing Reads	52
<i>Ion Măndoiu</i>	
Gene Networks Viewed through Two Models	54
<i>Satoru Miyano, Rui Yamaguchi, Yoshinori Tamada, Masao Nagasaki, and Seiya Imoto</i>	
Identifying Evolutionarily Conserved Protein Interaction Modules Using GraphHopper	67
<i>Corban G. Rivera and T.M. Murali</i>	
The 2-Interval Pattern Matching Problems and Its Application to ncRNA Scanning	79
<i>Thomas K.F. Wong, S.M. Yiu, T.W. Lam, and Wing-Kin Sung</i>	

Refereed Papers

RNA Pseudoknot Folding through Inference and Identification Using TAG _{RNA}	90
<i>Sahar Al Seesi, Sanguthevar Rajasekaran, and Reda Ammar</i>	

Comparing Bacterial Genomes by Searching Their Common Intervals . . .	102
<i>Sébastien Angibaud, Damien Eveillard, Guillaume Fertin, and Irena Rusu</i>	
Generalized Binary Tanglegrams: Algorithms and Applications	114
<i>Mukul S. Bansal, Wen-Chieh Chang, Oliver Eulenstein, and David Fernández-Baca</i>	
Three-Dimensional Multimodality Modelling by Integration of High-Resolution Interindividual Atlases and Functional MALDI-IMS Data	126
<i>Felix Bollenbeck, Stephanie Kaspar, Hans-Peter Mock, Diana Weier, and Udo Seiffert</i>	
Detecting Motifs in a Large Data Set: Applying Probabilistic Insights to Motif Finding	139
<i>Christina Boucher and Daniel G. Brown</i>	
A Biclustering Method to Discover Co-regulated Genes Using Diverse Gene Expression Datasets	151
<i>Doruk Bozdağ, Jeffrey D. Parvin, and Umit V. Catalyurek</i>	
Computational Protocol for Screening GPI-anchored Proteins	164
<i>Wei Cao, Kazuya Sumikoshi, Tohru Terada, Shugo Nakamura, Katsuhiko Kitamoto, and Kentaro Shimizu</i>	
Towards Large-Scale Molecular Dynamics Simulations on Graphics Processors	176
<i>Joseph E. Davis, Adnan Ozsoy, Sandeep Patel, and Michela Taufer</i>	
An Agent-Based Model of Solid Tumor Progression	187
<i>Didier Dréau, Dimitre Stanimirov, Ted Carmichael, and Mirsad Hadzikadic</i>	
Deciphering Drug Action and Escape Pathways: An Example on Nasopharyngeal Carcinoma	199
<i>Difeng Dong, Chun-Ying Cui, Benjamin Mow, and Limsoon Wong</i>	
Selection of Multiple SNPs in Case-Control Association Study Using a Discretized Network Flow Approach	211
<i>Shantanu Dutt, Yang Dai, Huan Ren, and Joel Fontanarosa</i>	
Biclustering Expression Data Based on Expanding Localized Substructures	224
<i>Cesim Erten and Melih Sözdinler</i>	
Constrained Fisher Scores Derived from Interaction Profile Hidden Markov Models Improve Protein to Protein Interaction Prediction	236
<i>Alvaro J. González and Li Liao</i>	

Improving Protein Localization Prediction Using Amino Acid Group Based Physicochemical Encoding	248
<i>Jianjun Hu and Fan Zhang</i>	
The Impact of Gene Selection on Imbalanced Microarray Expression Data	259
<i>Abu H.M. Kamal, Xingquan Zhu, Abhijit S. Pandya, Sam Hsu, and Muhammad Shoaib</i>	
Spatial Information and Boolean Genetic Regulatory Networks	270
<i>Matthieu Manceny, Marc Aiguier, Pascale Le Gall, Joan H��risson, Ivan Junier, and Fran��ois K��p��s</i>	
Modeling of Genetic Regulatory Network in Stochastic π -Calculus.....	282
<i>Myl��ne Maurin, Morgan Magnin, and Olivier Roux</i>	
fMRI Activation Detection by MultiScale Hidden Markov Model	295
<i>Fangyuan Nan, Yaonan Wang, and Xiaoping Ma</i>	
cnF2freq: Efficient Determination of Genotype and Haplotype Probabilities in Outbred Populations Using Markov Models	307
<i>Carl Nettelblad, Sverker Holmgren, Lucy Crooks, and ��rjan Carlborg</i>	
A Comprehensive Analysis Workflow for Genome-Wide Screening Data from ChIP-Sequencing Experiments	320
<i>Hatice Gulcin Ozer, Doruk Bozda��, Terry Camerlengo, Jiejun Wu, Yi-Wen Huang, Tim Hartley, Jeffrey D. Parvin, Tim Huang, Umit V. Catalyurek, and Kun Huang</i>	
A Fitness Distance Correlation Measure for Evolutionary Trees	331
<i>Hyun Jung Park and Tiffani L. Williams</i>	
Alignment and Analysis of Closely Related Genomes	343
<i>Allison Regier, Michael Olson, and Scott J. Emrich</i>	
Computational Prediction of Genes Translationally Regulated by Cytoplasmic Polyadenylation Elements	353
<i>Eric C. Rouchka, Xiangping Wang, James H. Graham, and Nigel G.F. Cooper</i>	
Multiple Sequence Alignment System for Pyrosequencing Reads.....	362
<i>Fahad Saeed, Ashfaq Khokhar, Osvaldo Zagordi, and Niko Beerenwinkel</i>	
A Bayesian Approach to High-Throughput Biological Model Generation	376
<i>Xinghua Shi and Rick Stevens</i>	

Parallel Selection of Informative Genes for Classification	388
<i>Michael Slavik, Xingquan Zhu, Imad Mahgoub, and Muhammad Shoaib</i>	
Simulation Methods in Uncovering New Regulatory Mechanisms in Signaling Pathways.....	400
<i>Jarostaw Smieja</i>	
GridSPiM: A Framework for Simple Locality and Containment in the Stochastic π -Calculus.....	409
<i>Stephen Tyree, Rayus Kuplicki, Trevor Sarratt, Scott Fujan, and John Hale</i>	
Mutual Information Based Extrinsic Similarity for Microarray Analysis	424
<i>Duygu Ucar, Fatih Altiparmak, Hakan Ferhatosmanoglu, and Srinivasan Parthasarathy</i>	
Graph Spectral Approach for Identifying Protein Domains	437
<i>Hari Krishna Yalamanchili and Nita Parekh</i>	
Author Index	449