

Lecture Notes in Bioinformatics

6832

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

Osmar Norberto de Souza
Guilherme P. Telles Mathew J. Palakal (Eds.)

Advances in Bioinformatics and Computational Biology

6th Brazilian Symposium on Bioinformatics, BSB 2011
Brasilia, Brazil, August 10-12, 2011
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

Osmar Norberto de Souza
Faculdade de Informática - PUCRS
Avenida Ipiranga 6681, Prédio 32 - Sala 608, 90619-900, Porto Alegre-RS, Brazil
E-mail: osmar.norberto@pucrs.br

Guilherme P. Telles
Instituto de Computação - Unicamp
Avenida Albert Einstein, 1251, 13083-852, Campinas-SP, Brazil
E-mail: gpt@ic.unicamp.br

Mathew J. Palakal
Indiana University Purdue University Indianapolis
535 W. Michigan Str. IT 477, Indianapolis, IN 46202, USA
E-mail: mpalakal@iupui.edu

ISSN 0302-9743
ISBN 978-3-642-22824-7
DOI 10.1007/978-3-642-22825-4
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349
e-ISBN 978-3-642-22825-4

Library of Congress Control Number: 2011932980

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

LNCS Sublibrary: SL 8 – Bioinformatics

© Springer-Verlag Berlin Heidelberg 2011

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.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

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

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume contains the full papers and extended abstracts selected by peer review for presentation at the 6th Brazilian Symposium on Bioinformatics, BSB 2011, held in Brasília, Distrito Federal, Brazil, during August 10–12, 2011. The first three meetings of this series, which took place in 2002, 2003, and 2004, were called WOB (Workshop on Bioinformatics). In 2005, the conference name was changed from WOB to its current name BSB to reflect the increase both in the quality of the contributions and in the interest of the scientific community for the meeting. Since then the BSB proceedings have been published as a special issue of the subseries *Lecture Notes in Bioinformatics* (volumes 3594/2005, 4643/2007, 5167/2008, 5676/2009, and 6268/2010) of the *Lecture Notes in Computer Science* series by Springer.

BSB topics of interest embrace many areas of bioinformatics, ranging from theoretical aspects of problems in bioinformatics to applications in molecular biology, biochemistry, genetics, and associated subjects. The Program Committee Chairs are very thankful to the authors of all submitted papers, and especially to the Program Committee and additional reviewers for their careful work in helping to select the eight full papers and four extended abstracts to build up this proceedings volume. The members of BSB Organizing Committees also thank the keynote speakers, Mathew J. Palakal, João Setubal, Peter F. Stadler and Sergio Verjovski-Almeida, for their participation in the event.

The conference was sponsored by the Brazilian Computer Society (SBC), with further support from the Brazilian sponsoring agencies CNPq and CAPES. Special thanks go to Springer for their continued support by agreeing to publish the BSB 2011 proceedings volume and to the CEBioComp/SBC Steering Committee. Last but not least, our sincerest thanks to all who have supported the Brazilian Symposium on Bioinformatics along all these years. In 2011, we are in debt to the indispensable close help and assistance of Maria Emília Telles Walter. Mia, it did make a difference. Our gratitude also goes to Marcelo de Macedo Brígido and Maria Emília Telles Walter and their local team at UnB: Tainá Raiol, Daniel Saad Nunes, Halian Vilela, Felipe Lessa, Filipe Lima, Maria Beatriz W. Costa, Paulo Alvarez, and Tulio Conrado da Silva.

August 2011

Osmar Norberto de Souza
Guilherme P. Telles
Mathew J. Palakal

Organization

BSB 2011 was promoted by the Brazilian Computer Society (SBC) and was organized by the Institute of Biology and by the Institute of Exact Sciences of the University of Brasília (UnB), Brasília-DF, Brazil.

Conference Chairs

Marcelo de Macedo Brígido	Laboratory of Molecular Biology, Institute of Biology, UnB, Brazil
Maria Emília M.T. Walter	Department of Computer Science, Institute of Exact Sciences, UnB, Brazil

Program Chairs

Osmar Norberto de Souza	Faculty of Informatics–PUCRS, Brazil
Guilherme P. Telles	Institute of Computing–Unicamp, Brazil
Mathew J. Palakal	Indiana University Purdue University Indianapolis, USA

Steering Committee

Marcelo de Macedo Brígido	University of Brasília, Brazil
André C.P.L.F. de Carvalho	University of São Paulo–São Carlos, Brazil
Carlos E. Ferreira	University of São Paulo, Brazil
Katia Guimarães	Federal University of Pernambuco, Brazil
Sergio Lifschitz	Pontifical Catholic University of Rio de Janeiro, Brazil
Osmar Norberto de Souza	Pontifical Catholic University of Rio Grande do Sul, Brazil
Maria Emília M.T. Walter	University of Brasília, Brazil

Program Committee

Nalvo F. Almeida	Federal University of Mato Grosso do Sul, Brazil
Ana Lúcia C. Bazzan	Federal University of Rio Grande do Sul, Brazil
Marcelo Brígido	University of Brasília, Brazil
Mario Caccamo	TGAC-BBSRC, UK
André C.P.L.F. de Carvalho	University of São Paulo–São Carlos, Brazil
Ivan G. Costa	Federal University of Pernambuco, Brazil
Mark Craven	University of Wisconsin, USA

VIII Organization

Alberto Dávila	Fiocruz, Brazil
Zanoni Dias	University of Campinas, Brazil
Alan Durham	University of São Paulo, Brazil
Fazel Famili	Institute for Information Technology, Canada
Katia S. Guimarães	Federal University of Pernambuco, Brazil
Ronaldo Hashimoto	University of São Paulo, Brazil
Paul Horton	AIST, Japan
Maricel Kann	UMBC, USA
Paula Kuser-Falcão	Embrapa, Brazil
Alair Pereira do Lago	University of São Paulo, Brazil
Ney Lemke	IBB-Unesp, Brazil
Sergio Lifschitz	Pontifical Catholic University of Rio de Janeiro, Brazil
Ion Mandoiu	University of Connecticut, USA
Natalia F. Martins	Embrapa, Brazil
Wellington Martins	Federal University of Goiás, Brazil
Anna Panchenko	NIH, USA
Fábio Passeti	INCA, Brazil
Duncan Ruiz	Pontifical Catholic University of Rio Grande do Sul, Brazil
Alexander Schliep	Rutgers University, USA
João Setubal	Virginia Bioinformatics Institute, USA
Marcilio M.C.P. de Souto	Federal University of Pernambuco, Brazil
Rainer Spang	University of Regensburg, Germany
Peter F. Stadler	University of Leipzig, Germany
Jerzy Tiuryn	University of Warsaw, Poland
Maria Emilia M.T. Walter	University of Brasília, Brazil
Alex Zelikovsky	Georgia State University, USA

Additional Reviewers

Said Adi	Federal University of Mato Grosso do Sul, Brazil
Aleteia P.F. de Araujo	University of Brasília, Brazil
Francisco Eloi Araújo	Federal University of Mato Grosso do Sul, Brazil
Katti Faceli	Federal University of São Carlos, Brazil
Mileidy Gonzalez	NIH, USA
Carlos Higa	University of São Paulo, Brazil
Maristela T. de Holanda	University of Brasília, Brazil
Renato P. Ishii	Federal University of Mato Grosso do Sul, Brazil

David Martins-Jr.	Federal University of ABC, Brazil
Edson Matsubara	Federal University of Mato Grosso do Sul, Brazil
Mariana Mendoza	Federal University of Rio Grande do Sul, Brazil
Roberto Riga	Embrapa, Brazil
John Spouge	NIH, USA
David Swarbreck	The Genome Analysis Center, UK
Manoj Tyagi	NCBI/NIH, USA
Michel Yamagishi	Embrapa, Brazil

Sponsoring Institutions

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)
Brazilian Computer Society (SBC)
University of Brasília (UnB)

Table of Contents

Full Papers

MicroRNA or Not MicroRNA?	1
<i>David Langenberger, Sebastian Bartschat, Jana Hertel, Steve Hoffmann, Hakim Tafer, and Peter F. Stadler</i>	
Hierarchical Multilabel Protein Function Prediction Using Local Neural Networks	10
<i>Ricardo Cerri and André C.P.L.F. de Carvalho</i>	
Identifying Significant Features in HIV Sequence to Predict Patients’ Response to Therapies	18
<i>Samuel Evangelista de Lima Oliveira, Luiz Henrique de Campos Merschmann, and Leoneide Erica Maduro Bouillet</i>	
Gene Prediction by Multiple Spliced Alignment	26
<i>Rodrigo Mitsuo Kishi, Ronaldo Fiorilo dos Santos, and Said Sadique Adi</i>	
A New Algorithm for Sparse Suffix Trees	34
<i>Gustavo Akio T. Sacomoto and Alair Pereira do Lago</i>	
Analysis and Implementation of Sorting by Transpositions Using Permutation Trees	42
<i>Marcelo P. Lopes, Marília D.V. Braga, Celina M.H. de Figueiredo, Rodrigo de A. Hausen, and Luis Antonio B. Kowada</i>	
Improved Gene Expression Clustering with the Parameter-Free PKNNG Metric	50
<i>Ariel E. Bayá and Pablo M. Granitto</i>	
Efficiently Querying Protein Sequences with the Proteinus Index	58
<i>Felipe Alves da Louza, Ricardo Rodrigues Ciferri, and Cristina Dutra de Aguiar Ciferri</i>	

Extended Abstracts

SciPhy: A Cloud-Based Workflow for Phylogenetic Analysis of Drug Targets in Protozoan Genomes	66
<i>Kary A.C.S. Ocaña, Daniel de Oliveira, Eduardo Ogasawara, Alberto M.R. Dávila, Alexandre A.B. Lima, and Marta Mattoso</i>	

A Conceptual Model for Transcriptome High-Throughput Sequencing Pipeline	71
<i>Ruben Cruz Huacarpuma, Maristela Holanda, and Maria Emilia Walter</i>	
A Conceptual Many Tasks Computing Architecture to Execute Molecular Docking Simulations of a Fully-Flexible Receptor Model	75
<i>Renata De Paris, Fábio A. Frantz, Osmar Norberto de Souza, and Duncan D. Ruiz</i>	
<i>kGC</i> : Finding Groups of Homologous Genes across Multiple Genomes	79
<i>Guilherme P. Telles, Nalvo F. Almeida, Marcelo M. Brigido, Paulo Antonio Alvarez, and Maria Emilia Walter</i>	
Author Index	83