Lecture Notes in Bioinformatics 6268

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

Carlos E. Ferreira Satoru Miyano
Peter F. Stadler (Eds.)

Advances in Bioinformatics and Computational Biology

5th Brazilian Symposium on Bioinformatics, BSB 2010
Rio de Janeiro, Brazil, August 31-September 3, 2010
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

Carlos E. Ferreira
Universidade de São Paulo
IME-USP (MAC)
Rua do Matão, 1010, 05508-090 São Paulo, SP, Brazil
E-mail: cef@ime.usp.br

Satoru Miyano
University of Tokyo
Institute of Medical Science, Human Genome Center
4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan
E-mail: miyano@ims.u-tokyo.ac.jp

Peter F. Stadler
University of Leipzig
Department Computer Science and Interdisciplinary Center for Bioinformatics
Bioinformatics Group
Härtlestr. 16-18, 04107 Leipzig, Germany
E-mail: studla@bioinf.uni-leipzig.de

Library of Congress Control Number: 2010931861

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

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743
ISBN-10 3-642-15059-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-15059-3 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 2010
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

This volume contains the accepted full papers and extended abstracts of the 5th Brazilian Symposium on Bioinformatics held in Búzios, Rio de Janeiro, Brazil from August 31 to September 3, 2010.

The first three meetings of this series, which took place 2002, 2003, and 2004, were called WOB (Workshop on Bioinformatics). In 2005, the conference got its current name BSB and has since published its proceedings as a special issue of the series *Lecture Notes in Bioinformatics* (volumes 3594/2005, 4643/2007, 5167/2008, and 5676/2009).

Its topics of interest vary in many areas of bioinformatics, including sequence analysis, motifs, and pattern matching; biomedical text mining; biological databases, data management, integration; biological data mining; structural, comparative, and functional genomics; protein structure, modeling, and simulation; gene identification and regulation; gene expression analysis; gene and protein interaction and networks; molecular docking; molecular evolution and phylogenetics; computational systems biology; computational proteomics; statistical analysis of molecular sequences; algorithms for problems in computational biology; as well as applications in molecular biology, biochemistry, genetics, and associated subjects.

We would like to thank all referees and Program Committee members for their careful work in preparing this proceedings volume. Also, we want to acknowlegde the local organizers and their staff for making this meeting possible.

September 2010

Carlos E. Ferreira
Satoru Miyano
Peter F. Stadler

Organization

BSB 2010 was promoted by the Brazilian Computing Society (SBC) and was organized by the Informatics Department of the Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

Executive Committee

Conference Chair:	Sergio Lifschitz (PUC-Rio, Brazil)
Program Chairs:	Carlos E. Ferreira (University of São Paulo, Brazil)
	Satoru Miyano (University of Tokyo, Japan)
	Peter F. Stadler (University of Leipzig, Germany)
Steering Comittee:	André C. de Carvalho (University of São Paulo, Brazil)
	Carlos E. Ferreira (University of São Paulo, Brazil)
	Katia Guimarães (Federal University of Pernambuco, Brazil)
	Sergio Lifschitz (PUC-Rio, Brazil)
	Francisco M. Salzano (UFRGS, Brazil)
	João Carlos Setubal (Virginia Tech, USA)
	Osmar Norberto de Sousa (PUC-RS, Brazil)

Program Committee

Emil Alexov	Clemson University, USA
Nalvo F. Almeida Jr.	UFMS, Brazil
Masanori Arita	University of Tokyo, Japan
Junior Barrera	USP, Brazil
Ana Lucia C. Bazzan	UFRGS, Brazil
Igor Berezovsky	University of Bergen, Norway
Marcelo M. Brigido	UnB, Brazil
André C.P.L. Carvalho	USP, Brazil
Saikat Chakrabarti	NCBI, NIH, USA
Zanoni Dias	Unicamp, Brazil
Jessica Fong	NCBI, NIH, USA
Oxana Galzitskaya	Institute of Protein Science, Russia
Kátia S. Guimarães	UFPE, Brazil
Ronaldo F. Hashimoto	USP, Brazil
Paul Horton	AIST, Japan

Maricel Kann	UMBC, USA
Carl Kingsford	UMIACS, UMD, USA
Kengo Kinoshita	Tohoku University, Japan
Eugene Krissinel	EBI, UK
Ney Lemke	UNESP, Brazil
Thomas Madej	NIH, USA
Wojtek Makalowski	University of Münster, Germany
Ion Mandoiu	University of Connecticut, USA
Natalia F. Martin	Embrapa, Brazil
Wellington S. Martins	UFG, Brazil
Hideo Matsuda	Osaka University, Japan
José Carlos M. Mombach	UFSM, Brazil
Anna Panchenko	NIH, USA
Ron Pinter	Technion, Israel
Teresa Przytycka	NIH, USA
Marie-France Sagot	University Lyon, France
Cenk Sahinalp	Simon Fraser University, Canada
Francisco M. Salzano	UFRGS, Brazil
Hagit Shatkay	Queen's University, Canada
Alexander Schliep	Rutgers University, USA
Benjamin Shoemaker	NIH, USA
Marcilio M.C.P. Souto	UFRN, Brazil
Jurek Tiuryn	University of Warsaw, Poland
Maria Emilia T. Walter	UnB, Brazil
Ryo Yoshida	Institute of Statistical Math., Japan
Alex Zelikovsky	Georgia State University, Brazil
Jie Zheng	NIH, USA

Table of Contents

Full Papers

Evolution of the Long Non-coding RNAs MALAT1 and MENβ/ϵ 1
Peter F. Stadler

Granger Causality in Systems Biology: Modeling Gene Networks in Time Series Microarray Data Using Vector Autoregressive Models 13
André Fujita, Patricia Severino, João Ricardo Sato, and Satoru Miyano

Semi-supervised Approach for Finding Cancer Sub-classes on Gene Expression Data 25
Clerton Ribeiro, Francisco de Assis T. de Carvalho, and Ivan G. Costa

Bounds on the Transposition Distance for Lonely Permutations 35
Luis Antonio B. Kowada, Rodrigo de A. Hausen, and Celina M.H. de Figueiredo

Insights on Haplotype Inference on Large Genotype Datasets 47
Rogério S. Rosa and Katia S. Guimarães

Extended Abstracts

An SVM Model Based on Physicochemical Properties to Predict Antimicrobial Activity from Protein Sequences with Cysteine Knot Motifs 59
William F. Porto, Fabiano C. Fernandes, and Octávio L. Franco

Enabling Annotation Provenance in Bioinformatics Workflow Applications 63
Milene Pereira Guimarães and Maria Cláudia Cavalcanti

BAT: A New Biclustering Analysis Toolbox 67
Cristian Andrés Gallo, Julieta Sol Dussaut, Jessica Andrea Carballido, and Ignacio Ponzoni

Detection of Protein Domains in Eukaryotic Genome Sequences 71
Arli A. Parikesit, Peter F. Stadler, and Sonja J. Prohaska

Discretization of Flexible-Receptor Docking Data 75
K.S. Machado, A.T. Winck, D.D. Ruiz, and O. Norberto de Souza

Author Index 81