

Lecture Notes in Bioinformatics

3240

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

Inge Jonassen Junhyong Kim (Eds.)

Algorithms in Bioinformatics

4th International Workshop, WABI 2004
Bergen, Norway, September 17-21, 2004
Proceedings

Series Editors

Sorin Istrail, Celera Genomics, Applied Biosystems, Rockville, MD, USA

Pavel Pevzner, University of California, San Diego, CA, USA

Michael Waterman, University of Southern California, Los Angeles, CA, USA

Volume Editors

Inge Jonassen

University of Bergen

Department of Informatics and Computational Biology Unit, HIB

5020 Bergen, Norway

E-mail: Inge.Jonassen@ii.uib.no

Junhyong Kim

University of Pennsylvania

Goddard Laboratories, Biology Dept., School of Arts and Sciences

415 S University Avenue, Philadelphia, PA 19104, USA

E-mail: junhyong@sas.upenn.edu

Library of Congress Control Number: 2004111461

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

ISSN 0302-9743

ISBN 3-540-23018-1 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 is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2004

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik

Printed on acid-free paper SPIN: 11319306 06/3142 5 4 3 2 1 0

Preface

It gives us great pleasure to present the proceedings of the 4th Workshop on Algorithms in Bioinformatics (WABI 2004) which took place in Bergen, Norway, September 17–21, 2004. The WABI 2004 workshop was part of a five-conference meeting, which in addition to WABI, included ESA, WAOA, IWPEC, and ATMOS, hosted by the University of Bergen, Norway. See <http://www.iu.uib.no/algo2004/> for more details.

The Workshop on Algorithms in Bioinformatics covers research on all aspects of algorithmic work in bioinformatics. The emphasis is on discrete algorithms that address important problems in molecular biology. These are founded on sound models, are computationally efficient, and have been implemented and tested in simulations and on real datasets. The goal is to present recent research results, including significant work in progress, and to identify and explore directions of future research.

Original research papers (including significant work in progress) or state-of-the-art surveys were solicited on all aspects of algorithms in bioinformatics, including, but not limited to: exact and approximate algorithms for genomics, genetics, sequence analysis, gene and signal recognition, alignment, molecular evolution, phylogenetics, structure determination or prediction, gene expression and gene networks, proteomics, functional genomics, and drug design.

We received 117 submissions in response to our call for papers, and were able to accept 39 of these. In addition, WABI hosted one invited distinguished lecture, given to the entire ALGO 2004 conference, by Dr. Marie France Sagot of the INRIA Rhône-Alpes laboratories in France.

We would like to sincerely thank all the authors of submitted papers, and the participants of the workshop. We also thank the program committee for their hard work in reviewing and selecting the papers for the workshop. We were fortunate to have on the program committee the following distinguished group of researchers:

Pankaj Kumar Agarwal (Duke University)
Amihoud Amir (Bar-Ilan University)
Alberto Apostolico (Purdue University)
Gary Benson (MSSN, New York)
Alvis Brazma (EMBL-EBI, UK)
Olivier Gascuel (LIRMM, Montpellier)
Raffaele Giancarlo (University of Palermo)
David Gilbert (University of Glasgow)
Jan Gorodkin (KVL, Denmark)
Roderic Guigo (University of Pompeu Fabra)
Jacques van Helden (Université Libre de Bruxelles)
Daniel Huson (University of Tübingen)
Gregory Kucherov (Loria, France)
Nadia El-Mabrouk (University of Montreal)
Inge Jonassen (University of Bergen)

Junhyong Kim (University of Pennsylvania)
 Jens Lagergren (KTH, Sweden)
 Gad M. Landau (University of Haifa)
 Thierry Lecroq (Université de Rouen)
 Bernard Moret (University of New Mexico)
 Vincent Moulton (University of Uppsala)
 Roderic D.M. Page (University of Glasgow)
 David Sankoff (University of Ottawa)
 Joao Carlos Setubal (Virginia Polytechnic Institute)
 Jens Stoye (University of Bielefeld)
 Esko Ukkonen (University of Helsinki)
 Lisa Vawter (Aventis Inc., USA)
 Jaak Vilo (Egeen Inc., Estonia)
 Alfonso Valencia (CNB-CSIC, Spain)
 Martin Vingron (Max Planck Inst., Berlin)
 Tandy Warnow (University of Texas)
 Peter S. White (University of Pennsylvania)
 Louxin Zhang (National University of Singapore)

We would also like to thank the co-reviewers who assisted the program committee members in their work: Ali Al-Shahib, Lars Arvestad, Vineet Bafna, Vikas Bansal, Ugo Bastolla, Ann-Charlotte Berglund, Allister Bernard, Laurent Brehelin, Dave Bryant, Trond Hellem Bø, Sergio Carvalho Jr., Robert Castel, Sergi Castellano, Benny Chor, Richard Christen, Matteo Comin, Richard Coulson, Eivind Coward, Miklós Csűrös, Tobias Dezulian, Bjarte Dysvik, Ingvar Eidhammer, Isaac Elias, Eduardo Eyras, Pierre Flener, Kristian Flikka, Eva Freyhult, Ganesh Ganapathy, Clemens Groepl, Stefan Grunewald, Yann Guermeur, Michael Hallett, Sylvie Hamel, Chao He, Danny Hermelin, Pawel Herzyk, Matthias Höchsmann, Katharina Huber, Torgeir Hvidsten, Johan Kährström, Hans-Michael Kaltenbach, Michael Kaufmann, Carmel Kent, Mikko Koivisto, Tsvi Kopelowitz, Arnaud Lefebvre, Alice Lesser, Zsuzsanna Liptak, Olivier Martin, Gregor Obernosterer, Sebastian Oehm, Kimmo Palin, Kjell Petersen, Cinzia Pizzi, Mathieux Raffinot, Sven Rahmann, Pasi Rastas, Knut Reinert, Eric Rivals, Kim Roland Rasmussen, Mikhail Roytberg, Gabriella Rustici, Anastasia Samsonova, Erik Sandelin, Stefanie Scheid, Alexander Schliep, Beng Sennblad, Rileen Sinha, Örjan Svensson, Jinsong Tan, Gilleain Torrance, Aurora Torrente, Dekel Tsur, Juris Viksna, Alexey Vitreschak, Li-San Wang, Oren Weimann, R. Scott Winters, Peng Yin, Tomasz Zemojtel, Zhonglin Zhou, Michal Ziv-Ukelson.

We also would like to thank the WABI steering committee, Gary Benson, Olivier Gascuel, Raffaele Giancarlo, Roderic Guigo, Dan Gusfield, Bernard Moret, and Roderic Page, for inviting us to co-chair this program committee, and for their help in carrying out that task.

Finally we are grateful to Ole Arntzen at the Dept. of Informatics, University of Bergen, who helped with technical issues.

July 2004

Inge Jonassen and Junhyong Kim
 WABI 2004 Program Co-chairs

Table of Contents

Papers

Reversing Gene Erosion – Reconstructing Ancestral Bacterial Genomes from Gene-Content and Order Data	1
<i>Joel V. Earnest-DeYoung, Emmanuelle Lerat, and Bernard M.E. Moret</i>	
Reconstructing Ancestral Gene Orders Using Conserved Intervals	14
<i>Anne Bergeron, Mathieu Blanchette, Annie Chateau, and Cedric Chauve</i>	
Sorting by Reversals with Common Intervals	26
<i>Martin Figeac and Jean-Stéphane Varré</i>	
A Polynomial-Time Algorithm for the Matching of Crossing Contact-Map Patterns	38
<i>Jens Gramm</i>	
A 1.5-Approximation Algorithm for Sorting by Transpositions and Transreversals	50
<i>Tzvika Hartman and Roded Sharan</i>	
Algorithms for Finding Maximal-Scoring Segment Sets	62
<i>Miklós Csűrös</i>	
Gapped Local Similarity Search with Provable Guarantees	74
<i>Manikandan Narayanan and Richard M. Karp</i>	
Monotone Scoring of Patterns with Mismatches	87
<i>Alberto Apostolico and Cinzia Pizzi</i>	
Suboptimal Local Alignments Across Multiple Scoring Schemes	99
<i>Morris Michael, Christoph Dieterich, and Jens Stoye</i>	
A Faster Reliable Algorithm to Estimate the p-Value of the Multinomial llr Statistic	111
<i>Uri Keich and Niranjan Nagarajan</i>	
Adding Hidden Nodes to Gene Networks	123
<i>Benny Chor and Tamir Tuller</i>	
Joint Analysis of DNA Copy Numbers and Gene Expression Levels	135
<i>Doron Lipson, Amir Ben-Dor, Elinor Dehan, and Zohar Yakhini</i>	

Searching for Regulatory Elements of Alternative Splicing Events Using Phylogenetic Footprinting	147
<i>Daichi Shigemizu and Osamu Maruyama</i>	
Supervised Learning-Aided Optimization of Expert-Driven Functional Protein Sequence Annotation	159
<i>Lev Soinov, Alexander Kanapin, and Misha Kapushesky</i>	
Multiple Vector Seeds for Protein Alignment	170
<i>Daniel G. Brown</i>	
Solving the Protein Threading Problem by Lagrangian Relaxation	182
<i>Stefan Balev</i>	
Protein-Protein Interfaces: Recognition of Similar Spatial and Chemical Organizations	194
<i>Alexandra Shulman-Peleg, Shira Mintz, Ruth Nussinov, and Haim J. Wolfson</i>	
ATDD: An Algorithmic Tool for Domain Discovery in Protein Sequences ..	206
<i>Stanislav Angelov, Sanjeev Khanna, Li Li, and Fernando Pereira</i>	
Local Search Heuristic for Rigid Protein Docking	218
<i>Vicky Choi, Pankaj K. Agarwal, Herbert Edelsbrunner, and Johannes Rudolph</i>	
Sequence Database Compression for Peptide Identification from Tandem Mass Spectra	230
<i>Nathan Edwards and Ross Lippert</i>	
Linear Reduction for Haplotype Inference	242
<i>Jingwu He and Alex Zelikovsky</i>	
A New Integer Programming Formulation for the Pure Parsimony Problem in Haplotype Analysis	254
<i>Daniel G. Brown and Ian M. Harrower</i>	
Fast Hare: A Fast Heuristic for Single Individual SNP Haplotype Reconstruction.....	266
<i>Alessandro Panconesi and Mauro Sozio</i>	
Approximation Algorithms for the Selection of Robust Tag SNPs	278
<i>Yao-Ting Huang, Kui Zhang, Ting Chen, and Kun-Mao Chao</i>	
The Minisatellite Transformation Problem Revisited: A Run Length Encoded Approach	290
<i>Behshad Behzadi and Jean-Marc Steyaert</i>	

A Faster and More Space-Efficient Algorithm for Inferring Arc-Annotations of RNA Sequences Through Alignment	302
<i>Jesper Jansson, See-Kiong Ng, Wing-Kin Sung, and Hugo Willy</i>	
New Algorithms for Multiple DNA Sequence Alignment	314
<i>Daniel G. Brown and Alexander K. Hudek</i>	
Chaining Algorithms for Alignment of Draft Sequence	326
<i>Mukund Sundararajan, Michael Brudno, Kerrin Small, Arend Sidow, and Serafim Batzoglou</i>	
Translation Initiation Sites Prediction with Mixture Gaussian Models	338
<i>Guoliang Li, Tze-Yun Leong, and Louxin Zhang</i>	
Online Consensus and Agreement of Phylogenetic Trees	350
<i>Tanya Y. Berger-Wolf</i>	
Relation of Residues in the Variable Region of 16S rDNA Sequences and Their Relevance to Genus-Specificity	362
<i>Maciej Liškiewicz, Hemant J. Purohit, and Dhananjay V. Raje</i>	
Topological Rearrangements and Local Search Method for Tandem Duplication Trees	374
<i>Denis Bertrand and Olivier Gascuel</i>	
Phylogenetic Super-networks from Partial Trees	388
<i>Daniel H. Huson, Tobias Dezulian, Tobias Klöpper, and Mike A. Steel</i>	
Genome Identification and Classification by Short Oligo Arrays	400
<i>Stanislav Angelov, Boulos Harb, Sampath Kannan, Sanjeev Khanna, Junhyong Kim, and Li-San Wang</i>	
Novel Tree Edit Operations for RNA Secondary Structure Comparison . . .	412
<i>Julien Allali and Marie-France Sagot</i>	
The Most Probable Labeling Problem in HMMs and Its Application to Bioinformatics	426
<i>Broňa Brejová, Daniel G. Brown, and Tomáš Vinař</i>	
Integrating Sample-Driven and Pattern-Driven Approaches in Motif Finding	438
<i>Sing-Hoi Sze, Songjian Lu, and Jianer Chen</i>	
Finding Optimal Pairs of Patterns	450
<i>Hideo Bannai, Heikki Hyvrö, Ayumi Shinohara, Masayuki Takeda, Kenta Nakai, and Satoru Miyano</i>	
Finding Missing Patterns	463
<i>Shunsuke Inenaga, Teemu Kivioja, and Veli Mäkinen</i>	
Author Index	475