

Lecture Notes in Bioinformatics

4175

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

Philipp Bächer Bernard M.E. Moret (Eds.)

Algorithms in Bioinformatics

6th International Workshop, WABI 2006
Zurich, Switzerland, September 11-13, 2006
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

Philipp Bücher

Ecole Polytechnique Fédérale de Lausanne, Switzerland

E-mail: Philipp.Bucher@isrec.unil.ch

Bernard M.E. Moret

Ecole Polytechnique Fédérale de Lausanne, Switzerland

E-mail: bernard.moret@epfl.ch

Library of Congress Control Number: 2006932026

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

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-540-39583-0 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-39583-6 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

springer.com

© Springer-Verlag Berlin Heidelberg 2006

Printed in Germany

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

Printed on acid-free paper SPIN: 11851561 06/3142 5 4 3 2 1 0

Preface

We are very pleased to present the proceedings of the *Sixth Workshop on Algorithms in Bioinformatics (WABI 2006)*, which took place in Zürich on September 11-13, 2006, under the auspices of the *International Society for Computational Biology (ISCB)*, the *European Association for Theoretical Computer Science (EATCS)*, and the *Eidgenössische Technische Hochschule Zürich (ETHZ)*.

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, that are founded on sound models, that are computationally efficient, and that 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. Specific topics of interest include, but are not limited to:

- Exact, approximate, and machine-learning algorithms for genomics, sequence analysis, gene and signal recognition, alignment, molecular evolution, population genetics and nucleotide polymorphism, structure determination or prediction, gene expression and gene networks, proteomics, functional genomics, and drug design.
- Methods, software and dataset repositories for the development and testing of such algorithms and their underlying models.
- High-performance approaches to computationally hard problems in bioinformatics, particularly optimization problems.

A major goal of the workshop is to bring together researchers spanning the range from abstract algorithm design to biological dataset analysis, so as to enable a dialogue between application specialists and algorithm designers, mediated by algorithm engineers and high-performance computing specialists. We believe that such a dialogue is necessary for the progress of computational biology, inasmuch as application specialists cannot analyze their datasets without fast and robust algorithms and, conversely, algorithm designers cannot produce useful algorithms without being conversant with the problems faced by biologists.

Part of this mix has been achieved for all six *WABI* events to date by collocating *WABI* with the *European Symposium on Algorithms (ESA)*, along with other occasional conferences or workshops, so as to form the interdisciplinary scientific meeting known as *ALGO*. This year, *ALGO 2006* comprised the *14th European Symposium on Algorithms (ESA 2006)*, the *6th Workshop on Algorithms in Bioinformatics (WABI 2006)*, the *4th Workshop on Approximation and Online Algorithms (WAOA 2006)*, the *2nd International Workshop on Parameterized and Exact Computation (IWPEC 2006)*, and the *6th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS 2006)*.

We received 100 submissions in response to our call for *WABI 2006* and were able to accept 36 of them, ranging from mathematical tools to experimental

studies of approximation algorithms and reports on significant computational analyses. Numerous biological problems are dealt with, including genetic mapping, sequence alignment and sequence analysis, phylogeny, comparative genomics, and protein structure. This year was the first in which *WABI* also called for machine-learning approaches along with combinatorial optimization, and we are delighted to feature five contributions from this area.

We would like to thank all authors for submitting their work to the workshop and all the presenters and attendees for their participation. We were particularly fortunate in enlisting the help of a very distinguished panel of researchers for our program committee, which undoubtedly accounts for the large number of submissions and the high quality of the presentations. Our heartfelt thanks go to all:

Vincent Berry (U. Montpellier)
 Rita Casadio (U. di Bologna)
 Phoebe Chen (Deakin U.)
 Nadia El-Mabrouk (U. Montréal)
 Raffaele Giancarlo (U. di Palermo)
 David Gilbert (U. Glasgow)
 Roderic Guigo (U. Pompeu Fabra)
 Vasant Honavar (Iowa State U.)
 Daniel Huson (U. Tübingen)
 Jens Lagergren (KTH Stockholm)
 C. Randal Linder (U. Texas Austin)
 Joao Meidanis (U. Campinas)
 Satoru Miyano (Tokyo U.)
 Gene W. Myers (HHMI Janelia Farm)
 Luay Nakhleh (Rice U.)
 Cedric Notredame (CNRS Marseilles)
 Sven Rahmann (U. Bielefeld)
 Knut Reinert (Freie U. Berlin)
 Mikhail Roytberg (Russian Academy of Sciences)
 Marie-France Sagot (U. Claude Bernard)
 David Sankoff (U. Ottawa)
 Joao Setubal (U. Campinas)
 Adam Siepel (Cornell U.)
 Jijun Tang (U. South Carolina)
 Olga Troyanskaya (Princeton U.)
 Alfonso Valencia (CNB-CSIC)
 Jaak Vilo (Egeen Inc.)
 Tandy Warnow (U. Texas Austin)
 Lusheng Wang (City U. Hong Kong)
 Tiffani Williams (Texas A&M U.)
 Louxin Zhang (National U. Singapore)

We were fortunate to attract Ron Shamir, from Tel Aviv University, to address the joint conferences on topics in computational biomedicine, along with other distinguished speakers lecturing in more classical algorithmic areas: Erik Demaine (Massachusetts Institute of Technology), Lisa Fleischer (IBM T.J. Watson Research Labs), László Lovász (Eötvös Loránd University and Microsoft Research), and Kurt Mehlhorn (Max-Planck-Institute Saarbrücken).

Last but not least, we thank Michael Hoffman and his colleagues Angelika Steger, Emo Welzl, and Peter Widmayer, all at ETHZ, for doing a superb job of organizing the joint conferences.

We hope that you will consider contributing to future *WABI* events, through a submission or by participating in the workshop.

September 2006

Phillip Bücher and Bernard M.E. Moret
WABI'06 Program Co-Chairs

Table of Contents

Measures of Codon Bias in Yeast, the tRNA Pairing Index and Possible DNA Repair Mechanisms	1
<i>Markus T. Friberg, Pedro Gonnet, Yves Barral, Nicol N. Schraudolph, Gaston H. Gonnet</i>	
Decomposing Metabolomic Isotope Patterns	12
<i>Sebastian Böcker, Matthias C. Letzel, Zsuzsanna Lipták, Anton Pervukhin</i>	
A Method to Design Standard HMMs with Desired Length Distribution for Biological Sequence Analysis	24
<i>Hongmei Zhu, Jiaxin Wang, Zehong Yang, Yixu Song</i>	
Efficient Model-Based Clustering for LC-MS Data	32
<i>Marta Luksza, Bogusław Kluge, Jerzy Ostrowski, Jakub Karczmarski, Anna Gambin</i>	
A Bayesian Algorithm for Reconstructing Two-Component Signaling Networks	44
<i>Lukas Burger, Erik van Nimwegen</i>	
Linear-Time Haplotype Inference on Pedigrees Without Recombinations	56
<i>M.Y. Chan, Wun-Tat Chan, Francis Y.L. Chin, Stanley P.Y. Fung, Ming-Yang Kao</i>	
Phylogenetic Network Inferences Through Efficient Haplotyping	68
<i>Yinglei Song, Chunmei Liu, Russell L. Malmberg, Liming Cai</i>	
Beaches of Islands of Tractability: Algorithms for Parsimony and Minimum Perfect Phylogeny Haplotyping Problems	80
<i>Leo van Iersel, Judith Keijsper, Steven Kelk, Leen Stougie</i>	
On the Complexity of SNP Block Partitioning Under the Perfect Phylogeny Model	92
<i>Jens Gramm, Tzvika Hartman, Till Nierhoff, Roded Sharan, Till Tantau</i>	
How Many Transcripts Does It Take to Reconstruct the Splice Graph?...	103
<i>Paul Jenkins, Rune Lyngsø, Jotun Hein</i>	

Multiple Structure Alignment and Consensus Identification for Proteins	115
<i>Jieping Ye, Iyaylo Ilinkin, Ravi Janardan, Adam Isom</i>	
Procrastination Leads to Efficient Filtration for Local Multiple Alignment	126
<i>Aaron E. Darling, Todd J. Treangen, Louxin Zhang, Carla Kuiken, Xavier Messeguer, Nicole T. Perna</i>	
Controlling Size When Aligning Multiple Genomic Sequences with Duplications	138
<i>Minmei Hou, Piotr Berman, Louxin Zhang, Webb Miller</i>	
Reducing Distortion in Phylogenetic Networks	150
<i>Daniel H. Huson, Mike A. Steel, Jim Whitfield</i>	
Imputing Supertrees and Supernetworks from Quartets	162
<i>Barbara Hollan, Glenn Conner, Katharina T. Huber, Vincent Moulton</i>	
A Unifying View of Genome Rearrangements	163
<i>Anne Bergeron, Julia Mixtacki, Jens Stoye</i>	
Efficient Sampling of Transpositions and Inverted Transpositions for Bayesian MCMC	174
<i>István Miklós, Timothy Brooks Paige, Péter Ligeti</i>	
Alignment with Non-overlapping Inversions in $O(n^3)$ -Time	186
<i>Augusto F. Vellozo, Carlos E.R. Alves, Alair Pereira do Lago</i>	
Accelerating Motif Discovery: Motif Matching on Parallel Hardware	197
<i>Geir Kjetil Sandve, Magnar Nedland, Øyvind Bø Syrstad, Lars Andreas Eidsheim, Osman Abul, Finn Drabløs</i>	
Segmenting Motifs in Protein-Protein Interface Surfaces	207
<i>Jeff M. Phillips, Johannes Rudolph, Pankaj K. Agarwal</i>	
Protein Side-Chain Placement Through MAP Estimation and Problem-Size Reduction	219
<i>Eun-Jong Hong, Tomás Lozano-Pérez</i>	
On the Complexity of the Crossing Contact Map Pattern Matching Problem	231
<i>Shuai Cheng Li, Ming Li</i>	

A Fuzzy Dynamic Programming Approach to Predict RNA Secondary Structure	242
<i>Dandan Song, Zhidong Deng</i>	
Landscape Analysis for Protein-Folding Simulation in the H-P Model	252
<i>Kathleen Steinhöfel, Alexandros Skaliotis, Andreas A. Albrecht</i>	
Rapid <i>ab initio</i> RNA Folding Including Pseudoknots Via Graph Tree Decomposition	262
<i>Jizhen Zhao, Russell L. Malmberg, Liming Cai</i>	
Flux-Based <i>vs.</i> Topology-Based Similarity of Metabolic Genes	274
<i>Oleg Rokhlenko, Tomer Shlomi, Roded Sharan, Eytan Ruppin, Ron Y. Pinter</i>	
Combinatorial Methods for Disease Association Search and Susceptibility Prediction	286
<i>Dumitru Brinza, Alexander Zelikovsky</i>	
Integer Linear Programs for Discovering Approximate Gene Clusters	298
<i>Sven Rahmann, Gunnar W. Klau</i>	
Approximation Algorithms for Bi-clustering Problems	310
<i>Lusheng Wang, Yu Lin, Xiaowen Liu</i>	
Improving the Layout of Oligonucleotide Microarrays: Pivot Partitioning	321
<i>Sérgio A. de Carvalho Jr., Sven Rahmann</i>	
Accelerating the Computation of Elementary Modes Using Pattern Trees	333
<i>Marco Terzer, Jörg Stelling</i>	
A Linear-Time Algorithm for Studying Genetic Variation	344
<i>Nikola Stojanovic, Piotr Berman</i>	
New Constructive Heuristics for DNA Sequencing by Hybridization	355
<i>Christian Blum, Mateu Yábar Vallès</i>	
Optimal Probing Patterns for Sequencing by Hybridization	366
<i>Dekel Tsur</i>	
Gapped Permutation Patterns for Comparative Genomics	376
<i>Laxmi Parida</i>	

Segmentation with an Isochore Distribution 388
 Miklós Csűrös, Ming-Te Cheng, Andreas Grimm,
 Amine Halawani, Perrine Landreau

Author Index 401