Lecture Notes in Bioinformatics 4205

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

Guillaume Bourque Nadia El-Mabrouk (Eds.)

Comparative Genomics

RECOMB 2006 International Workshop, RCG 2006
Montreal, Canada, September 24-26, 2006
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

Guillaume Bourque
Genome Institute of Singapore
Computational and Mathematical Biology
60 Biopolis Street, Singapore 138672
E-mail: bourque@gis.a-star.edu.sg

Nadia El-Mabrouk
Université de Montréal
Département d'Informatique et de recherche opérationnelle (DIRO)
Montréal, H3C 3J7, Canada
E-mail: mabrouk@iro.umontreal.ca

Library of Congress Control Number: 2006932402

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

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743
ISBN-10 3-540-44529-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-44529-6 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: 11864127 06/3142 5 4 3 2 1 0

Preface

The complexity of genome evolution poses many exciting challenges to developers of mathematical models and algorithms, who have recourse to a spectrum of algorithmic, statistical and mathematical techniques, ranging from exact, heuristic, fixed-parameter and approximation algorithms for problems based on parsimony to probabilistic models, requiring Monte Carlo Markov Chain algorithms for Bayesian analysis.

The annual RECOMB Satellite Workshop on Comparative Genomics (RECOMB Comparative Genomics) is a forum on all aspects and components of this field, ranging from new quantitative discoveries about genome structure and process to theorems on the complexity of computational problems inspired by genome comparison. The informal Steering Committee for this meeting consists of David Sankoff, Jens Lagergren and Aoife McLysaght.

This volume contains the papers presented at the 4th RECOMB Comparative Genomics meeting, which was held in Montreal, Canada, on September 24- 26, 2006. The first three meetings of this series were held in Minneapolis, USA (2003), Bertinoro, Italy (2004) and Dublin, Ireland (2005).

This year, 34 papers were submitted, of which the Program Committee selected 17 for presentation at the meeting and inclusion in these proceedings. Each submission was refereed by at least three members of the Program Committee.

RECOMB Comparative Genomics 2006 had several invited speakers, including: Lars Feuk (The Hospital for Sick Children, Toronto), Tao Jiang (University of California, Riverside), Fiona Brinkman (Simon Fraser University, Burnaby) Liqing Zhang (VirginiaTech, Blacksburg) Thomas J. Hudson (McGill University, Montreal).

In addition to the invited talks and the contributed talks, an important ingredient of the program was the lively poster session.

In closing, we would like to thank all the people who submitted papers and posters and those who attended RECOMB Comparative Genomics 2006 with enthusiasm.

September 2006 Guillaume Bourque and Nadia El-Mabrouk

Organization

RECOMB Comparative Genomics 2006 was organized by the DIRO (Department of Computer Science), the CRM (Centre de Recherches Mathématiques) and the Robert Cedergren Centre at the University of Montreal.

Steering Committee

Aoife McLysaght (Trinity College, Dublin)
Jens Lagergren (Stockholm Bioinformatics Centre & KTH)
David Sankoff (University of Ottawa)

Program Committee Chairs

Guillaume Bourque (Genome Institute of Singapore)
Nadia El-Mabrouk (DIRO, University of Montreal)

Program Committee

Lars Arvestad (School of Computer Science and Communication, Stockholm), Robert Beiko (The University of Queensland, Australia), Anne Bergeron (Université du Québec à Montréal, Canada), Mathieu Blanchette (McGill University), Michael Brudno (Department of Computer Science, University of Toronto), Cedric Chauve (Université du Québec à Montréal, Canada), Avril Coghlan (Trinity College, Dublin), Miklós Csürös (Computer Science Department, University of Montreal), Dannie Durand (Departments of Biological Sciences and Computer Science, Carnegie Mellon University), Niklas Eriksen (Mathematical Science, Chalmers University of Technology and Göteborg University, Sweden), Rose Hoberman (Computer Science Department, Carnegie Mellon University), Tao Jiang (Department of Computer Science and Engineering, University of California, Riverside), Aoife McLysaght (Trinity College, Dublin), Bernard Moret (Swiss Federal Institute of Technology, Lausanne, Switzerland), Laxmi Parida (IBM, Watson Research Center, Yorktown, USA), Yves van de Peer (Department of Plant Systems Biology, Ghent University, Belgium), Ben Raphael (Department of Computer Science & Engineering, University of California, USA), Cathal Seoighe (National Bioinformatics Network, University of Cape Town, South Africa), Eric Tannier (Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard, France), Glenn Tesler (Department of Mathematics, University of California, USA), Stacia Wyman (Department of Computer Science, Williams College, Williamstown).

Organizing Committee

Nadia El-Mabrouk (DIRO, University of Montreal)
Guillaume Bourque(Genome Institute of Singapore)
Mathieu Blanchette (McGill University)
Franz Lang (Biochemistry Department, University of Montreal)
Vincent Ferretti (McGill University)
Sylvie Hamel (DIRO, University of Montreal)
Miklós Csürös (DIRO, University de Montreal)
François Major (DIRO, University of Montreal)
Mathieu Lajoie (DIRO, University of Montreal)

External Reviewers

Max Alekseyev
Guy Baele
Malay K. Basu
Mark Chaisson
Matteo Comin
Daniel Dalevi
Melissa Davis
Martin Frith
Neil Jones
Jacob Joseph
Steven Maere
Istvan Miklos
Narayanan Raghupathy
Magnus A. Rosenblad
David Sankoff
Benjamin Vernot

Sponsoring Institutions

Robert Cedergren Centre, University of Montreal
CRM (Centre de Recherches Mathématiques), University of Montreal
"Faculté des arts et des sciences", University of Montreal

Table of Contents

Reconstructing Domain Compositions of Ancestral Multi-domain Proteins 1
Behshad Behzadi, Martin Vingron

Domain Architecture in Homolog Identification 11
Nan Song, Robert D. Sedgewick, Dannie Durand

Inferring Positional Homologs with Common Intervals of Sequences 24
Guillaume Blin, Annie Chateau, Cedric Chauve, Yannick Gingras

On Genome Evolution with Accumulated Change and Innovation 39
Damian Wójtowicz, Jerzy Tiuryn

Paths and Cycles in Breakpoint Graphs of Random Multichromosomal Genomes 51
Wei Xu, Chunfang Zheng, David Sankoff

Common Intervals and Symmetric Difference in a Model-Free Phylogenomics, with an Application to Streptophyte Evolution 63
Zaky Adam, Monique Turmel, Claude Lemieux, David Sankoff

How Pseudo-boolean Programming Can Help Genome Rearrangement Distance Computation 75
Sébastien Angibaud, Guillaume Fertin, Irena Rusu, Stéphane Vialette

Sorting by Translocations Via Reversals Theory 87
Michal Ozery-Flato, Ron Shamir

Inferring Gene Orders from Gene Maps Using the Breakpoint Distance 99
Guillaume Blin, Eric Blais, Pierre Guillon, Mathieu Blanchette, Nadia El-Mabrouk

Ordering Partially Assembled Genomes Using Gene Arrangements 113
Éric Gaul, Mathieu Blanchette

Evolution of Tandemly Repeated Sequences Through Duplication and Inversion 129
Denis Bertrand, Mathieu Lajoie, Nadia El-Mabrouk, Olivier Gascuel

A PQ Framework for Reconstructions of Common Ancestors and Phylogeny 141
Laxmi Parida

Intron Loss Dynamics in Mammals 156
Jasmin Coulombe-Huntington, Jacek Majewski

Finding Maximum Likelihood Indel Scenarios 171
Abdoulaye Baniré Diallo, Vladimir Makarenkov, Mathieu Blanchette

Conservation Patterns in *cis*-Elements Reveal Compensatory Mutations 186
Perry Evans, Greg Donahue, Sridhar Hannenhalli

Transcription Factor Centric Discovery of Regulatory Elements in Mammalian Genomes Using Alignment-Independent Conservation Maps 200
Nilanjana Banerjee, Andrea Califano

Identifiability Issues in Phylogeny-Based Detection of Horizontal Gene Transfer 215
Cuong Than, Derek Ruths, Hideki Innan, Luay Nakhleh

Author Index 231