

Lecture Notes in Bioinformatics

3678

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

Aoife McLysaght Daniel H. Huson (Eds.)

Comparative Genomics

RECOMB 2005 International Workshop, RCG 2005
Dublin, Ireland, September 18-20, 2005
Proceedings



Springer

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

Aoife McLysaght
University of Dublin, Smurfit Institute of Genetics, Trinity College, Ireland
E-mail: mclysaga@tcd.ie

Daniel H. Huson
Tübingen University, Center for Bioinformatics (ZBIT)
72076 Tübingen, Germany
E-mail: huson@informatik.uni-tuebingen.de

Library of Congress Control Number: 2005932241

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

ISSN 0302-9743
ISBN-10 3-540-28932-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-28932-6 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

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11554714 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 models to Monte Carlo Markov Chain algorithms for Bayesian analysis of problems based on probabilistic models.

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 3rd RECOMB Comparative Genomics meeting, which was held in Dublin, Ireland, on September 18–20, 2005. The first two meetings of this series were held in Minneapolis, USA (2003) and Bertinoro, Italy (2004).

This year, 21 papers were submitted, of which the Program Committee selected 14 for presentation at the meeting and inclusion in this proceedings. Each submission was refereed by at least three members of the Program Committee. After completion of the referees' reports, an extensive Web-based discussion took place for making decisions. The RECOMB Comparative Genomics 2005 Program Committee consisted of the following 27 members: Vineet Bafna, Anne Bergeron, Mathieu Blanchette, Avril Coghlan, Dannie Durand, Nadia El-Mabrouk, Niklas Eriksen, Aaron Halpern, Rose Hoberman, Daniel Huson, Jens Lagergren, Giuseppe Lancia, Emmanuelle Lerat, Aoife McLysaght, Istvan Miklos, Bernard Moret, Pavel Pevzner, Ben Raphael, Marie-France Sagot, David Sankoff, Cathal Seoighe, Beth Shapiro, Igor Sharakhov, Mike Steel, Jens Stoye, Glenn Tesler and Louxin Zhan. We would like to thank the Program Committee members for their dedication and hard work.

RECOMB Comparative Genomics 2005 had several invited speakers, including: Anne Bergeron (Université du Québec à Montreal, Canada), Laurent Duret (Laboratoire de Biometrie et Biologie Evolutive, Université Claude Bernard, Lyon, France), Eddie Holmes (Department of Biology, Pennsylvania State University, USA), Jeffrey Lawrence (Department of Biological Sciences, University of Pittsburgh, USA), Stephan Schuster (Department of Biochemistry and Molecular Biology, Pennsylvania State University, USA), Ken Wolfe (Genetics Department, Trinity College Dublin, Ireland) and Sophia Yancopoulos (Institute for Medical Research, New York, USA).

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

RECOMB Comparative Genomics 2005 would like to thank Science Foundation Ireland (SFI) and Hewlett-Packard for providing financial support for the conference. We would like to thank the University of Dublin, Trinity College, for hosting the meeting. We would like to thank Nadia Browne for administrative support.

In closing, we would like to thank all the people who submitted papers and posters and those who attended RECOMB Comparative Genomics 2005 with enthusiasm.

September 2005

Aoife McLysaght and Daniel Huson

Table of Contents

Lower Bounds for Maximum Parsimony with Gene Order Data <i>Abraham Bachrach, Kevin Chen, Chris Harrelson, Radu Mihaescu, Satish Rao, Apurva Shah</i>	1
Genes Order and Phylogenetic Reconstruction: Application to γ -Proteobacteria <i>Guillaume Blin, Cedric Chauve, Guillaume Fertin</i>	11
Maximizing Synteny Blocks to Identify Ancestral Homologs <i>Guillaume Bourque, Yasmine Yacef, Nadia El-Mabrouk</i>	21
An Expectation-Maximization Algorithm for Analysis of Evolution of Exon-Intron Structure of Eukaryotic Genes <i>Liran Carmel, Igor B. Rogozin, Yuri I. Wolf, Eugene V. Koonin</i>	35
Likely Scenarios of Intron Evolution <i>Miklós Csűrös</i>	47
OMA, a Comprehensive, Automated Project for the Identification of Orthologs from Complete Genome Data: Introduction and First Achievements <i>Christophe Dessimoz, Gina Cannarozzi, Manuel Gil, Daniel Margadant, Alexander Roth, Adrian Schneider, Gaston H. Gonnet</i>	61
The Incompatible Desiderata of Gene Cluster Properties <i>Rose Hoberman, Dannie Durand</i>	73
The String Barcoding Problem is NP-Hard <i>Marcello Dalpasso, Giuseppe Lancia, Romeo Rizzi</i>	88
A Partial Solution to the C-Value Paradox <i>Jeffrey M. Marcus</i>	97
Individual Gene Cluster Statistics in Noisy Maps <i>Narayanan Raghupathy, Dannie Durand</i>	106
Power Boosts for Cluster Tests <i>David Sankoff, Lani Haque</i>	121
Reversals of Fortune <i>David Sankoff, Chungfang Zheng, Aleksander Lenert</i>	131

VIII Table of Contents

Very Low Power to Detect Asymmetric Divergence of Duplicated Genes <i>Cathal Seoighe, Konrad Scheffler</i>	142
A Framework for Orthology Assignment from Gene Rearrangement Data <i>Kristen M. Swenson, Nicholas D. Pattengale, Bernard M.E. Moret</i> ...	153
Author Index	167