Lecture Notes in Bioinformatics 4751

Edited by S. Istrail, P. Pevzner, and M. Waterman

Subseries of Lecture Notes in Computer Science

Glenn Tesler Dannie Durand (Eds.)

Comparative Genomics

International Workshop, RECOMB-CG 2007
San Diego, CA, USA, September 16-18, 2007
Proceedings

Springer

Volume Editors

Glenn Tesler
University of California, San Diego
Department of Mathematics
9500 Gilman Drive, La Jolla California 92093-0112, USA
E-mail: gptesler@math.ucsd.edu

Dannie Durand
Carnegie Mellon University
Departments of Biological Sciences and Computer Science
Pittsburgh, PA 15213, USA
E-mail: durand@cmu.edu

Library of Congress Control Number: 2007934768

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 3-540-74959-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-74959-2 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12124237 06/3180 5 4 3 2 1 0

Preface

The wealth of genomic data available today is a potential goldmine for basic research and economic development in the biomedical sciences. Comparison of related genomes offers enormous inferential power, revealing a wealth of knowledge about genome evolution, genetic function, and cellular processes. Recognition of this fact has spurred efforts to sequence a range of closely related primate and mammalian genomes, as well as concerted efforts to sequence multiple genomes in the yeast, *Drosophila*, and nematode lineages. Computational strategies to interpret and exploit these data are essential in order to realize the full value of these growing scientific resource. The annual RECOMB Satellite Workshop on Comparative Genomics (RECOMB-CG) is an interdisciplinary forum on all aspects of genome comparison, ranging from quantitative discoveries about genome structure to algorithms for comparative inference to theorems on the complexity of computational problems required for genome comparison.

This volume contains the papers presented at the Fifth Annual RECOMB Satellite Workshop on Comparative Genomics held September 16–18, 2007 in La Jolla, at the University of California, San Diego. Eighteen papers were submitted, of which the Program Committee selected 14 for presentation at the meeting and inclusion in these proceedings. Each submission was reviewed by at least three members of the Program Committee. The program also included a lively poster session. A session of short talks presenting late-breaking results, selected a few weeks before the meeting from the submitted poster abstracts, provided an opportunity to hear about provocative results from works in progress. In addition to contributed presentations, we were honored by plenary talks given by the invited speakers: Francesca Ciccarelli (European Institute of Oncology), Michael B. Eisen (University of California, Berkeley), Matthew Hahn (Indiana University), Katherine S. Pollard (University of California, Davis), Oliver A. Ryder (Zoological Society of San Diego), and Ajit Varki (University of California, San Diego). This year the meeting was held jointly with the new RECOMB Satellite Workshop on Computational Cancer Biology (RECOMB-CCB), and included a joint keynote address by Barbara J. Trask (Fred Hutchinson Cancer Research Center), sponsored by the law firm of Morrison & Foerster, LLP.

RECOMB-CG presentations focus on emerging problems, data, and technologies. This year's invited talks gave particular attention to the evolution of primate genomes and the use of comparative methods for identifying genomic novelties that make us uniquely human. Whole-genome approaches to species tree reconstruction were a dominant theme among the contributed papers. In contrast to the frequently misleading practice of inferring a species tree from a single gene family, comparative genomics research is spawning approaches for deriving phylogenetic signal from entire genomes. Whole-genome methods discussed at RECOMB-CG 2007 included analysis of conserved intron positions

and gene order conservation. Another emerging theme was gene duplication. Some papers investigated the role of gene duplication in the evolution of genetic novelties. Other work viewed duplications as a form of uncertainty and proposed methods to address this source of noise in reconstruction of genome rearrangements. A third research thrust discussed at the meeting focussed on novel approaches to inferring ancestral character states, and genome rearrangement distances and phylogenies. Finally, presentations on gene family evolution, as well as the use of comparative methods in inferring regulatory motifs and networks, complemented results on large-scale, spatial genomics.

RECOMB-CG 2007 is indebted to the many individuals and organizations who contributed their support, dedication, and hard work. The Steering Committee supported us in all aspects of the meeting. The success of the meeting depends critically on the efforts of the Program Committee and their sub-reviewers. Their good judgment and constructive criticism engendered an exciting and high-quality scientific program. Paper and poster submission and selection were managed through the EasyChair Web site. We express our appreciation to Andrei Voronkov for providing this system. We are especially grateful to Anita McKee, Jennifer Zimmerman, and Doug Ramsey at UCSD and Annette McLeod at Carnegie Mellon University for administrative support. RECOMB-CG 2007 thanks the law firm Morrison & Foerster, LLP; the National Science Foundation; the University of California's Industry-University Cooperative Research Program; and the University of California, San Diego (UCSD) for financial support and UCSD and the California Institute for Telecommunications and Information Technology (Calit2) for hosting the conference.

Most important, we thank the invited speakers, the scientists who submitted papers and posters, the conference attendees, and the committee members and student volunteers who helped to make this meeting possible. It is the contribution of these individuals that makes RECOMB-CG an exciting scientific event.

September 2007

Glenn Tesler
Dannie Durand

Conference Organization

Program Committee Chairs

Glenn Tesler (University of California, San Diego, USA)
Dannie Durand (Carnegie Mellon University, USA)

Program Committee

Lars Arvestad (Kungliga Tekniska Högskolan, Sweden)
Vineet Bafna (University of California, San Diego, USA)
Serafim Batzoglou (Stanford University, USA)
Anne Bergeron (Université du Québec à Montréal, Canada)
Mathieu Blanchette (McGill University, Montreal, Canada)
Guillaume Bourque (Genome Institute of Singapore, Singapore)
David Bryant (The University of Auckland, New Zealand)
Jeremy Buhler (Washington University in St. Louis, USA)
Sourav Chatterji (University of California, Berkeley, USA)
Cedric Chauve (Université du Québec à Montréal, Canada)
Avril Coghlan (Wellcome Trust Sanger Institute, UK)
Miklos Csuros (University of Montreal, Canada)
Aaron Darling (University of Queensland, Australia)
Nadia El-Mabrouk (University of Montreal, Canada)
Niklas Eriksen (Chalmers University of Technology, Sweden)
Steffen Heber (North Carolina State University, USA)
Daniel Huson (Eberhard Karls Universität, Tübingen, Germany)
Tao Jiang (University of California, Riverside, USA)
Jens Lagergren (Kungliga Tekniska Högskolan, Sweden)
Aoife McLysaght (Trinity College, University of Dublin, Ireland)
Laxmi Parida (New York University and IBM, USA)
Marie-France Sagot (INRIA, France)
David Sankoff (University of Ottawa, Canada)
Marie Sémon (Université Claude Bernard Lyon 1, France)
Joao Setubal (Virginia Tech, USA)
Jens Stoye (Universität Bielefeld, Germany)
Haixu Tang (Indiana University, USA)
Eric Tannier (INRIA Rhône-Alpes, France)
Tiffani Williams (Texas A & M, USA)
Stacia Wyman (Fred Hutchinson Cancer Research Center, USA)
Liqing Zhang (Virginia Tech, USA)
Louxin Zhang (National University of Singapore, Singapore)
Yves van de Peer (Ghent University, Belgium)

External Reviewers

Laurent Gueguen (Université Claude Bernard Lyon 1, France)
Katharina Jahn (Universität Bielefeld, Germany)
Åsa Pérez-Bercoff (Trinity College, University of Dublin, Ireland)
Roland Wittler (Universität Bielefeld, Germany)
Chunfang Zheng (University of Ottawa, Canada)

Local Organizing Committee

Glenn Tesler (University of California, San Diego, USA)
Mark Chaisson (University of California, San Diego, USA)
Qian Peng (University of California, San Diego, USA)

Steering Committee

Jens Lagergren (Kungliga Tekniska Högskolan, Sweden)
Aoife McLysaght (Trinity College, University of Dublin, Ireland)
David Sankoff (University of Ottawa, Canada)

Sponsors

Morrison & Foerster, LLP (www.mofo.com)
National Science Foundation (www.nsf.gov)
Opportunity Award from the Industry-University Cooperative Research Program (www.ucdiscoverygrant.org)
Division of Physical Sciences, University of California, San Diego (physicalsciences.ucsd.edu)
Center for Algorithmic and Systems Biology, University of California, San Diego (casb.ucsd.edu)
California Institute for Telecommunications and Information Technologie (www.calitz.net)

Previous Meetings in This Series

1st RECOMB Satellite Workshop on Comparative Genomics
October 20–24, 2003
Institute for Mathematics and Its Applications (IMA), University of Minnesota, Minneapolis, USA
Program Chairs: Jens Lagergren (Stockholm Bioinformatics Centre, KTH, Sweden), Bernard Moret (University of New Mexico, USA), and David Sankoff (University of Ottawa, Canada)

2nd RECOMB Satellite Workshop on Comparative Genomics
October 16–19, 2004
Bertinoro International Center for Informatics, University of Bologna, Italy
Program Chairs: Jens Lagergren (Stockholm Bioinformatics Centre, KTH, Sweden), Aoife McLysaght (Trinity College, Ireland) and David Sankoff (University of Ottawa, Canada)

3rd RECOMB Satellite Workshop on Comparative Genomics
September 18–20, 2005
Trinity College, University of Dublin, Ireland
Program Chairs: Aoife McLysaght (Trinity College Dublin, Ireland) and Daniel Huson (Eberhard Karls Universität, Tübingen, Germany)

4th RECOMB Satellite Workshop on Comparative Genomics
September 24–26, 2006
University of Montreal, Quebec, Canada
Program Chairs: Nadia El-Mabrouk (University of Montreal, Canada) and Guillaume Bourque (Genome Institute of Singapore, Singapore)

Table of Contents

Multi-break Rearrangements: From Circular to Linear Genomes 1
Max A. Alekseyev

A Pseudo-boolean Programming Approach for Computing the Breakpoint Distance Between Two Genomes with Duplicate Genes 16
Sébastien Angibaud, Guillaume Fertin, Irena Rusu, Annelyse Thévenin, and Stéphane Vialette

Improving Inversion Median Computation Using Commuting Reversals and Cycle Information 30
William Arndt and Jijun Tang

Inferring a Duplication, Speciation and Loss History from a Gene Tree (Extended Abstract) 45
Cedric Chauve, Jean-Philippe Doyon, and Nadia El-Mabrouk

How to Achieve an Equivalent Simple Permutation in Linear Time 58
Simon Gog and Martin Bader

Baculovirus Phylogeny Based on Genome Rearrangements 69
Daniel Goodman, Noah Ollikainen, and Chris Sholley

Learning Gene Regulatory Networks via Globally Regularized Risk Minimization 83
Yuhong Guo and Dale Schuurmans

Evolution of Tandemly Arrayed Genes in Multiple Species 96
Mathieu Lajoie, Denis Bertrand, and Nadia El-Mabrouk

Selecting Genomes for Reconstruction of Ancestral Genomes 110
Guoliang Li, Jian Ma, and Louxin Zhang

A Heuristic Algorithm for Reconstructing Ancestral Gene Orders with Duplications 122
Jian Ma, Aakrosh Ratan, Louxin Zhang, Webb Miller, and David Haussler

Reconstructing an Inversion History in the *Anopheles Gambiae* Complex 136
Ai Xia, Maria V. Sharakhova, and Igor V. Sharakhov

Recovering True Rearrangement Events on Phylogenetic Trees 149
Hao Zhao and Guillaume Bourque

Parts of the Problem of Polyploids in Rearrangement Phylogeny 162
Chunfang Zheng, Qian Zhu, and David Sankoff

A Rigorous Analysis of the Pattern of Intron Conservation Supports the *Coelomata* Clade of Animals 177
Jie Zheng, Igor B. Rogozin, Eugene V. Koonin, and Teresa M. Przytycka

Author Index 193