

Lecture Notes in Bioinformatics

5267

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

Craig E. Nelson Stéphane Vialette (Eds.)

Comparative Genomics

International Workshop, RECOMB-CG 2008
Paris, France, October 13-15, 2008
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

Craig E. Nelson

University of Connecticut

Division of Genetics and Genomics, Molecular and Cell Biology

Beach Hall, 354 Mansfield Rd, Storrs, CT 06269, USA

E-mail: craig.nelson@uconn.edu

Stéphane Vialette

Université Paris-Est, IGM-LabInfo UMR 8049

5 Bd Descartes, Champs sur Marne, 77454 Marne la Vallée, France

E-mail: vialette@univ-mlv.fr

Library of Congress Control Number: 2008936578

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 3-540-87988-9 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-87988-6 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 2008

Printed in Germany

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

Printed on acid-free paper SPIN: 12533763 06/3180 5 4 3 2 1 0

Preface

Over the last decade great investments have been made in the acquisition of enormous amounts of gene sequence data from a diverse collection of organisms. Realizing the full potential of these investments will require the continued development of computational tools for comparative genomics and the intelligent application of these tools to address biologically relevant questions. The RECOMB Workshop on Comparative Genomics (RECOMB-CG) is devoted to bringing together scientists working on all aspects of comparative genomics, from the development of new computational approaches to genome sequence analysis and comparison, to the genome-wide application of computational tools to study the evolutionary dynamics of prokaryotic and eukaryotic genomes.

This volume contains the 19 papers presented at the 6th Annual RECOMB-CG workshop held during October 13–15 at the École Normale Supérieure, in Paris, France. The papers selected for presentation and published in these proceedings were selected from 48 submissions from scientists around the world. Each paper was reviewed by at least three members of the Program Committee in a stringent and thoughtful peer-review process.

The conference itself was enlivened by invited keynote presentations from Laurent Duret (Université Claude Bernard), Aviv Regev (Broad Institute), Chris Ponting (University of Oxford), Olga Troyanskaya (Princeton University), and Patricia Wittkopp (University of Michigan). These talks were supplemented by both presentation of the papers in this volume and a series of “late-breaking talks” selected from a wide-ranging and provocative poster session. Together, these talks and papers highlighted the acceleration of comparative genomics tools and applications. From the inference of evolution in genetic regulatory networks, to the divergent fates of gene and genome duplication events, to the importance of new computational approaches to unraveling the structural evolution of genomes, these presentations illustrate the crucial role of comparative genomics in understanding genome function.

RECOMB-CG 2008 would not have been possible without the participation of the many scientists who contributed their time and effort to making the conference a success. We thank the scientists who submitted their work for presentation at the conference, those members of the Program Committee who made every effort to ensure fair and balanced review of the many papers submitted for consideration at this year’s workshop, the members of the local Organizing Committee for arranging all the myriad details of the organizational aspects of the event, and the continued efforts of the Steering Committee for their ongoing dedication and guidance. RECOMB-CG 2008 is also deeply indebted to its sponsors including the Centre National de la Recherche Scientifique (CNRS), the GdR BiM, the Université Paris-Est Marne-la-Vallée, the Université de Nantes and the Institut National de Recherche en Informatique et Automatique (INRIA

Rhône-Alpes), and to the École Normale Supérieure Paris for hosting the conference.

It is the continued support and dedication of this community that allows RECOMB-CG to bring together comparative genomics researchers from across the globe to exchange ideas and information and focus the force of comparative genomics on improving our understanding genome evolution and function.

July 2008

Craig E. Nelson
Stéphane Vialette

Organization

Program Chairs

Craig E. Nelson University of Connecticut, USA
Stéphane Vialette Université Paris-Est, France

Program Committee

Lars Arvestad	KTH, Sweden
Véronique Barriel	Muséum National d'Histoire Naturelle, France
Anne Bergeron	Université du Québec Montréal, Canada
Guillaume Blin	Université Paris-Est, France
Guillaume Bourque	Genome Institute of Singapore, Singapore
Jeremy Buhler	Washington University in St. Louis, USA
Pierre Capy	Université Paris-Sud, France
Cédric Chauve	Simon Fraser University, Canada
Avril Coghlan	Sanger Institute, UK
Miklós Csuros	Université de Montréal, Canada
Aaron Darling	University of Queensland, Australia
Sankoff David	University of Ottawa, Canada
Bernard Dujon	Institut Pasteur, France
Dannie Durand	Carnegie Mellon University, USA
Nadia El-Mabrouk	Université de Montréal, Canada
Niklas Erikson	Göteborg University, Sweden
Guillaume Fertin	Université de Nantes, France
Olivier Gascuel	Université de Montpellier, France
Henri Grosjean	Université Paris-Sud, France
Matthew Hahn	Indiana University, USA
Tao Jiang	University of California - Riverside, USA
Jens Lagergren	Stockholm Bioinformatics Centre and KTH, Sweden
Emmanuelle Lerat	Université Claude Bernard, France
Aoife McLysaght	University of Dublin, Ireland
Bernard Moret	École Polytechnique Fédérale de Lausanne, Switzerland
Michal Ozery-Flato	University of Tel-Aviv, Israel
Pierre Pontarotti	Université de Provence, France
Eduardo Rocha	Université Paris 6 et Institut Pasteur, France
Hugues Roest Crollius	École Normale Supérieure, France
Antonis Rokas	Vanderbilt University, USA
Marie-France Sagot	INRIA Rhône-Alpes, France
Cathal Seoighe	University of Cape Town, South Africa

Jens Stoye	Bielefeld University, Germany
Chuan-Yi Tang	National Tsing Hua University, Taiwan
Eric Tannier	INRIA Rhône-Alpes, France
Glenn Tesler	University of California - San Diego, USA
Louxin Zhang	National University of Singapore, Singapore

External Reviewers

Max Alekseyev	Charles Hébert	Irena Rusu
Sébastien Angibaud	Jian Ma	Maureen Stolzer
David Enard	Matthieu Muffato	Sara Vieira-Silva
Claire Herrbach	Aida Ouangraoua	

Steering Committee

Jens Lagergren	Stockholm Bioinformatics Centre and KTH, Sweden,
Aoife McLysaght	Trinity College Dublin, Ireland
David Sankoff	University of Ottawa, Canada

Sponsoring Institutions

CNRS	http://www.cnrs.fr/index.html
Université Paris-Est	http://www.univ-paris-est.fr
Université de Nantes	http://www.univ-nantes.fr
GDR Bioinformatique Moléculaire	http://www.gdr-bim.u-psud.fr
INRIA Rhône-Alpes	http://www.inrialpes.fr

Local Organization

Sèverine Bérard	Université de Montpellier, France
Guillaume Blin	Université Paris-Est France
Maxime Crochemore	Université Paris-Est France
Guillaume Fertin	Université de Nantes, France
Hugues Roest Crollius	École Normale Supérieure, France
Éric Tannier	INRIA Rhône-Alpes, France
Jean-Stéphane Varré	INRIA Futurs - Université de Lille, France
Stéphane Vialette	Université Paris-Est France

RECOMB CG 2008 website <http://igm.univ-mlv.fr/RCG08/>

Table of Contents

Algorithms for Exploring the Space of Gene Tree/Species Tree Reconciliations	1
<i>Jean-Philippe Doyon, Cedric Chauve, and Sylvie Hamel</i>	
Limitations of Pseudogenes in Identifying Gene Losses	14
<i>James C. Costello, Mira V. Han, and Matthew W. Hahn</i>	
Duplication Mechanism and Disruptions in Flanking Regions Influence the Fate of Mammalian Gene Duplicates	26
<i>Paul Rykin, Jin Jun, Edward Hemphill, and Craig Nelson</i>	
Estimating the Relative Contributions of New Genes from Retrotransposition and Segmental Duplication Events during Mammalian Evolution	40
<i>Jin Jun, Paul Rykin, Edward Hemphill, Ion Măndoiu, and Craig Nelson</i>	
Discovering Local Patterns of Co-evolution	55
<i>Yifat Felder and Tamir Tuller</i>	
Ancestral Reconstruction by Asymmetric Wagner Parsimony over Continuous Characters and Squared Parsimony over Distributions	72
<i>Miklós Csűrös</i>	
An Alignment-Free Distance Measure for Closely Related Genomes	87
<i>Bernhard Haubold, Mirjana Domazet-Lošo, and Thomas Wiehe</i>	
Gene Team Tree: A Compact Representation of All Gene Teams	100
<i>Melvin Zhang and Hon Wai Leong</i>	
Integrating Sequence and Topology for Efficient and Accurate Detection of Horizontal Gene Transfer	113
<i>Cuong Than, Guohua Jin, and Luay Nakhleh</i>	
An Evolutionary Study of the Human Papillomavirus Genomes	128
<i>Dunarel Badescu, Abdoulaye Baniré Diallo, Mathieu Blanchette, and Vladimir Makarenkov</i>	
An Algorithm for Inferring Mitogenome Rearrangements in a Phylogenetic Tree	143
<i>Matthias Bernt, Daniel Merkle, and Martin Middendorf</i>	
Perfect DCJ Rearrangement	158
<i>S�everine B�erard, Annie Chateau, Cedric Chauve, Christophe Paul, and Eric Tannier</i>	

Sorting Genomes with Insertions, Deletions and Duplications by DCJ . . .	170
<i>Sophia Yancopoulos and Richard Friedberg</i>	
A Fast and Exact Algorithm for the Median of Three Problem—A Graph Decomposition Approach	184
<i>Andrew Wei Xu</i>	
A Phylogenetic Approach to Genetic Map Refinement	198
<i>Denis Bertrand, Mathieu Blanchette, and Nadia El-Mabrouk</i>	
Sorting Cancer Karyotypes by Elementary Operations	211
<i>Michal Ozery-Flato and Ron Shamir</i>	
On Computing the Breakpoint Reuse Rate in Rearrangement Scenarios	226
<i>Anne Bergeron, Julia Mixtacki, and Jens Stoye</i>	
Hurdles Hardly Have to Be Heeded	241
<i>Krister M. Swenson, Yu Lin, Vaibhav Rajan, and Bernard M.E. Moret</i>	
Internal Validation of Ancestral Gene Order Reconstruction in Angiosperm Phylogeny	252
<i>David Sankoff, Chunfang Zheng, P. Kerr Wall, Claude dePamphilis, Jim Leebens-Mack, and Victor A. Albert</i>	
Author Index	265