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 Editor
Serafim Batzoglou
Computer Science Department
James H. Clark Center, 318 Campus Drive, RM S266
Stanford, CA 94305-5428, USA
E-mail: serafim@cs.stanford.edu

Library of Congress Control Number: Applied for

CR Subject Classification (1998): J.3, I.3.5, F.2.2, F.2, G.2.1

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743
ISBN-10 3-642-02007-0 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.com
© Springer-Verlag Berlin Heidelberg 2009
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12677146 06/3180 5 4 3 2 1 0
Preface

This volume contains the papers presented at RECOMB 2009: the 13th Annual International Conference on Research in Computational Molecular Biology held in Tucson, Arizona, USA, during May 18-21, 2009. The RECOMB conference series was started in 1997 by Sorin Istrail, Pavel Pevzner, and Michael Waterman. RECOMB 2009 was hosted by the University of Arizona, organized by a committee chaired by John Kececioglu, and took place at The Westin La Paloma Resort and Spa in Tucson, Arizona.

This year, 37 papers were accepted for presentation out of 166 submissions. The papers presented were selected by the Program Committee (PC) assisted by a number of external reviewers. Each paper was reviewed by three members of the PC, or by external reviewers acting as sub-reviewers to members of the PC. Following the initial reviews, there was an extensive Web-based discussion over a period of two weeks, leading to the final decisions. The RECOMB conference series is closely associated with the Journal of Computational Biology, which traditionally publishes special issues devoted to presenting full versions of selected conference papers.

RECOMB 2009 invited several distinguished speakers as keynotes and for a special session on “Personalized Genomics”. Invited speakers included Carlos D. Bustamante (Cornell University), Rade Drmanac (Complete Genomics), Mark Gerstein (Yale University), Eran Halperin (Navigenics), Michael Hammer (University of Arizona), Joanna Mountain (23andMe), Stephen Quake (Stanford University), Mostafa Ronaghi (Illumina), Pardis Sabeti (Harvard University), and Michael Snyder (Yale University).

RECOMB 2009 was only possible through the dedication and hard work of many individuals and organizations. Special thanks go to the PC and external reviewers for helping to form the conference program, and the organizers, chaired by John Kececioglu, for hosting the conference and providing the administrative, logistic, and financial support. Special thanks also go to our sponsors. The conference was overseen by the RECOMB Steering Committee. We thank Marina Sirota for help with editing the proceedings volume. Finally, we thank all the authors who contributed papers and posters, as well as the attendees of the conference for their enthusiastic participation.

March 2009

Serafim Batzoglou
Conference Organization

Program Committee

Tatsuya Akutsu Kyoto University, Japan
Serafim Batzoglou Stanford University, USA
  (Program Chair)
Gill Bejerano Stanford University, USA
Bonnie Berger Massachusetts Institute of Technology, USA
Michael Brent Washington University, USA
Mike Brudno University of Toronto, Canada
Jeremy Buhler Washington University, USA
Atul Butte Stanford University, USA
Rhiju Das Stanford University, USA
Colin Dewey University of Wisconsin, USA
Eleazar Eskin University of California Los Angeles, USA
Nir Friedman Hebrew University, Israel
James Galagan Broad Institute, USA
Eran Halperin UC Berkeley, USA
Alexander Hartemink Duke University, USA
Des Higgins University College Dublin, Ireland
Trey Ideker University of California San Diego, USA
Sorin Istrail Brown University, USA
Tao Jiang University of California Riverside, USA
Simon Kasif Boston University, USA
John Kececioglu University of Arizona, USA
  (Conference Chair)
Manolis Kellis Massachusetts Institute of Technology, USA
Jens Lagergren Stockholm University, Sweden
Thomas Lengauer Max Planck Institut-Informatik-für, Germany
Satoru Miyano Tokyo University, Japan
William Noble University of Washington, USA
Pavel Pevzner University of California San Diego, USA
Ron Pinter Technion, Israel
Aviv Regev Massachusetts Institute of Technology, USA
Knut Reinert Freie Universität Berlin, Germany
David Sankoff University of Ottawa, Canada
Russell Schwartz Carnegie Mellon University, USA
Eran Segal Weizmann Institute, Israel
Roded Sharan Tel Aviv University, Israel
VIII Organization

Adam Siepel Cornell University, USA
Mona Singh Princeton University, USA
Peter Stadler Universität Leipzig, Germany
Jens Stoye Universität Bielefeld, Germany
Josh Stuart University of California Santa Cruz, USA
Fengzhu Sun University of Southern California, USA
Olga Troyanskaya Princeton University, USA
Martin Vingron Max Planck Institute, Germany
Tandy Warnow University of Texas Austin, USA
Eric Xing Carnegie Mellon University, USA
Zohar Yakhini Agilent Laboratories

Steering Committee

Serafim Batzoglou Stanford University, USA
Sorin Istrail RECOMB Vice-Chair, Brown University, USA
Thomas Lengauer Max Planck Institute, Germany
Michal Linial Hebrew University, Israel
Pavel A. Pevzner RECOMB General Chair, University of California San Diego, USA
Terence P. Speed University of California Berkeley, USA

Local Organization at the University of Arizona

John Kececioglu (Conference Chair) University of Arizona
Maura Grohan BIO5 Institute and University of Arizona
Daphne Gillman BIO5 Institute and University of Arizona
Deborah Daun BIO5 Institute and University of Arizona
Steve Dix BIO5 Institute and University of Arizona
Previous RECOMB Meetings

<table>
<thead>
<tr>
<th>Dates</th>
<th>Hosting Institution</th>
<th>Program Chair</th>
<th>Conference Chair</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 20-23, 1997</td>
<td>Sandia National Lab</td>
<td>Michael Waterman</td>
<td>Sorin Istrail</td>
</tr>
<tr>
<td>Santa Fe, NM, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 22-25, 1998</td>
<td>Mt. Sinai School of Medicine</td>
<td>Pavel Pevzner</td>
<td>Gary Benson</td>
</tr>
<tr>
<td>New York, NY, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 22-25, 1999</td>
<td>INRIA</td>
<td>Sorin Istrail</td>
<td>Mireille Regnier</td>
</tr>
<tr>
<td>Lyon, France</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 8-11, 2000</td>
<td>University of Tokyo</td>
<td>Ron Shamir</td>
<td>Satoru Miyano</td>
</tr>
<tr>
<td>Tokyo, Japan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 22-25, 2001</td>
<td>Universite de Montreal</td>
<td>Thomas Lengauer</td>
<td>David Sankoff</td>
</tr>
<tr>
<td>Montreal, Canada</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 18-21, 2002</td>
<td>Celera</td>
<td>Gene Myers</td>
<td>Sridhar Hannenhalli</td>
</tr>
<tr>
<td>Washington, DC, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 10-13, 2003</td>
<td>German Federal Ministry for Education and Research</td>
<td>Webb Miller</td>
<td>Martin Vingron</td>
</tr>
<tr>
<td>Berlin, Germany</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 27-31, 2004</td>
<td>UC San Diego</td>
<td>Dan Gusfield</td>
<td>Philip E. Bourne</td>
</tr>
<tr>
<td>San Diego, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 14-18, 2005</td>
<td>Broad Institute of MIT and Harvard</td>
<td>Satoru Miyano</td>
<td>Jill P. Mesirov</td>
</tr>
<tr>
<td>Boston, MA, USA</td>
<td></td>
<td></td>
<td>and Simon Kasif</td>
</tr>
<tr>
<td>April 2-5, 2006</td>
<td>University of Padova</td>
<td>Alberto Apostolico</td>
<td>Concettina Guerra</td>
</tr>
<tr>
<td>Venice, Italy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 21-25, 2007</td>
<td>QB3</td>
<td>Terry Speed</td>
<td>Sandrine Dudoit</td>
</tr>
<tr>
<td>San Francisco, CA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 30 - April 2 2008</td>
<td>National University of Singapore</td>
<td>Martin Vingron</td>
<td>Limsoon Wong</td>
</tr>
<tr>
<td>Singapore, Singapore</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

External Reviewers

Alber, Frank                      Candeias, Rogerio
Alekseyev, Max                    Chen, Rong
Andreotti, Sandro                 Chen, Xiaoyu
Anton, Brian                      Chen, Yang-ho
Arvestad, Lars                    Cho, Sungje
Aydin, Zafer                      Chor, Benny
Berry, Vincent                    Chuang, Han-Yu
Bandyopadhyay, Sourav             Chung, Ho-Ryun
Bar-Joseph, Ziv                    Cohen-Gihon, Inbar
Bauer, Markus                     Cowen, Lenore
Baumbach, Jan                     Csuros, Miklos
Bercovici, Sivan                  Dalca, Adrian
Bernhart, Stephan                 Diehl, Adam
Bielow, Chris                     Do, Chuong
Blom, Jochen                      Dudley, Joel
Bordewich, Magnus                 Dunbrack, Roland
Brejova, Bronislava               Elidan, Gal
Brown, Randall                    El-Hay, Tal
Bruckner, Sharon                  Eran, Ally
Ernst, Jason
Fernández-Baca, David
Fujita, Andre
Gat-Viks, Irit
Gerlach, Wolfgang
Goff, Loyal
Gräppl, Clemens
Gusfield, Dan
Haas, Stefan
Habib, Naomi
Haim, Wolfson, Efraïm Mashiach and
Han, Buhm
Hannum, Gregory
Hayashida, Morihiro
Haynes, Brian
Heinig, Matthias
Hildebrandt, Andreas
Hoffman, Michael
Hudek, Alexander
Hue, Martial
Husemann, Peter
Huson, Daniel
Imoto, Seiya
Inbar, Yuval
Jaimovich, Ariel
Jeong, Euna
Kaell, Lukas
Kamisetty, Hetunandan
Kamisetty, Hndetunandan
Kang, Hyun Min
Kato, Yuki
Kelley, Ryan
Kheradpour, Pouya
Kim, Seyoung
Kimmel, Gad
Kirkpatrick, Bonnie
Klaau, Gunnar
Kojima, Kaname
Kolaczyk, Eric
Kosiol, Carolin
Kriete, Andres
Kuang, Rui
Kuttykrishnan, Sooraj
Lasserre, Julia
Lee, Byoungkoo
Lee, Ki Young
Lee, Ki Young
Li, Yong
Lilien, Ryan
Lin, Michael
Lonardi, Stefano
Ma, Xiaotu
Martins, Andre
McIlwain, Sean
Medvedev, Paul
Meshi, Ofer
Milani, Martin
Misra, Navodit
Mitrofanova, Antonina
Morgan, Alex
Morozov, Alexandre
Moses, Alan
Mosig, Axel
Müll, Mathias
Nakhleh, Luay
Navon, Roy
Ng, Julio
Novershtern, Noa
Obozinski, Gaëlle
Paik, David
Pardo, Matteo
Pasaniuc, Bogdan
Perrier, Eric
Phillips, Michael
Pincus, Zach
Pop, Mihai
Qi, Yanjun
Raphael, Ben
Rasmussen, Matthew
Ray, Pradipta
Reynolds, Sheila
Ringnér, Markus
Safari, Rahele
Sankararaman, Sriram
Satulovsky, Javier
Schelhorn, Sven-Eric
Schuster, Meromit
Sealfon, Rachel
Segal, Eran
Sennblad, Bengt
# Table of Contents

Searching Protein 3-D Structures in Linear Time .......................... 1  
*Tetsuo Shibuya*

Optimization-Based Peptide Mass Fingerprinting for Protein Mixture Identification .................................................. 16  
Zengyou He, Chao Yang, Can Yang, Robert Z. Qi,  
Jason Po-Ming Tam, and Weichuan Yu

Boosting Protein Threading Accuracy ........................................ 31  
*Jian Peng and Jinbo Xu*

New Perspectives on Gene Family Evolution: Losses in Reconciliation and a Link with Supertrees ........................................ 46  
*Cedric Chauve and Nadia El-Mabrouk*

A Probabilistic Graphical Model for Ab Initio Folding ..................... 59  
Feng Zhao, Jian Peng, Joe DeBartolo, Karl F. Freed,  
Tobin R. Sosnick, and Jinbo Xu

Topology-Free Querying of Protein Interaction Networks ................ 74  
Sharon Bruckner, Falk Hüffner, Richard M. Karp, Ron Shamir, and  
Roded Sharan

Cross Species Expression Analysis of Innate Immune Response .......... 90  
Yong Lu, Roni Rosenfeld, Gerard J. Nau, and Ziv Bar-Joseph

Haplotype Inference in Complex Pedigrees ................................... 108  
Bonnie Kirkpatrick, Javier Rosa, Eran Halperin, and  
Richard M. Karp

Storage and Retrieval of Individual Genomes ................................ 121  
Veli Mäkinen, Gonzalo Navarro, Jouni Sirén, and Niko Välimäki

An Online Approach for Mining Collective Behaviors from Molecular Dynamics Simulations ........................................ 138  
Arvind Ramanathan, Pratul K. Agarwal, Maria Kurnikova, and  
Christopher J. Langmead

Parameter Synthesis in Nonlinear Dynamical Systems: Application to Systems Biology ........................................ 155  
Alexandre Donzé, Gilles Clermont, Axel Legay, and  
Christopher J. Langmead
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Clustering of Multivariate Genomic and Epigenomic Information</td>
<td>170</td>
</tr>
<tr>
<td><em>Rami Jaschek and Amos Tanay</em></td>
<td></td>
</tr>
<tr>
<td>How Many Bootstrap Replicates Are Necessary?</td>
<td>184</td>
</tr>
<tr>
<td><em>Nicholas D. Pattengale, Masoud Alipour, Olaf R.P. Bininda-Emonds, Bernard M.E. Moret, and Alexandros Stamatakis</em></td>
<td></td>
</tr>
<tr>
<td>A Robust Bayesian Two-Sample Test for Detecting Intervals of Differential Gene Expression in Microarray Time Series</td>
<td>201</td>
</tr>
<tr>
<td><em>Oliver Stegle, Katherine Denby, David L. Wild, Zoubin Ghahramani, and Karsten M. Borgwardt</em></td>
<td></td>
</tr>
<tr>
<td>Incorporating Nucleosomes into Thermodynamic Models of Transcription Regulation</td>
<td>217</td>
</tr>
<tr>
<td><em>Tali Raveh-Sadka, Michal Levo, and Eran Segal</em></td>
<td></td>
</tr>
<tr>
<td>Combinatorial Algorithms for Structural Variation Detection in High Throughput Sequenced Genomes</td>
<td>218</td>
</tr>
<tr>
<td><em>Fereydoun Hormozdiari, Can Alkan, Evan E. Eichler, and S. Cenk Sahinalp</em></td>
<td></td>
</tr>
<tr>
<td>Optimizing PCR Assays for DNA Based Cancer Diagnostics</td>
<td>220</td>
</tr>
<tr>
<td><em>Ali Bashir, Qing Lu, Dennis Carson, Benjamin Raphael, Yu-Tsung Liu, and Vineet Bafna</em></td>
<td></td>
</tr>
<tr>
<td>The Multi-State Perfect Phylogeny Problem with Missing and Removable Data: Solutions via Integer-Programming and Chordal Graph Theory</td>
<td>236</td>
</tr>
<tr>
<td><em>Dan Gusfield</em></td>
<td></td>
</tr>
<tr>
<td>COE: A General Approach for Efficient Genome-Wide Two-Locus Epistasis Test in Disease Association Study</td>
<td>253</td>
</tr>
<tr>
<td><em>Xiang Zhang, Feng Pan, Yuying Xie, Fei Zou, and Wei Wang</em></td>
<td></td>
</tr>
<tr>
<td>Overlapping Pools for High Throughput Targeted Resequencing</td>
<td>270</td>
</tr>
<tr>
<td><em>Snehit Prabhu and Itsik Pe’er</em></td>
<td></td>
</tr>
<tr>
<td>Deep Sequencing of a Genetically Heterogeneous Sample: Local Haplotype Reconstruction and Read Error Correction</td>
<td>271</td>
</tr>
<tr>
<td><em>Osvaldo Zagordi, Lukas Geyrhofer, Volker Roth, and Niko Beerenwinkel</em></td>
<td></td>
</tr>
<tr>
<td>Lifting Prediction to Alignment of RNA Pseudoknots</td>
<td>285</td>
</tr>
<tr>
<td><em>Mathias Möhl, Sebastian Will, and Rolf Backofen</em></td>
<td></td>
</tr>
<tr>
<td>Detection of Locally Over-Represented GO Terms in Protein-Protein Interaction Networks</td>
<td>302</td>
</tr>
<tr>
<td><em>Mathieu Lavallée-Adam, Benoit Coulombe, and Mathieu Blanchette</em></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Protein Fragment Swapping: A Method for Asymmetric, Selective Site-Directed Recombination</td>
<td>321</td>
</tr>
<tr>
<td>Wei Zheng, Karl E. Griswold, and Chris Bailey-Kellogg</td>
<td></td>
</tr>
<tr>
<td>Simultaneous Alignment and Folding of Protein Sequences</td>
<td>339</td>
</tr>
<tr>
<td>Jérôme Waldispühl, Charles W. O’Donnell, Sebastian Will, Srinivas Devadas, Rolf Backofen, and Bonnie Berger</td>
<td></td>
</tr>
<tr>
<td>Shared Peptides in Mass Spectrometry Based Protein Quantification</td>
<td>356</td>
</tr>
<tr>
<td>Banu Dost, Nuno Bandeira, Xiangqian Li, Zhouxin Shen, Steve Briggs, and Vineet Bafna</td>
<td></td>
</tr>
<tr>
<td>Evaluating Between-Pathway Models with Expression Data</td>
<td>372</td>
</tr>
<tr>
<td>Benjamin J. Hescott, Maek D.M. Leiserson, Lenore J. Cowen, and Donna K. Slonim</td>
<td></td>
</tr>
<tr>
<td>Sorting Signed Permutations by Inversions in $O(n \log n)$ Time</td>
<td>386</td>
</tr>
<tr>
<td>Krister M. Swenson, Vaibhav Rajan, Yu Lin, and Bernard M.E. Moret</td>
<td></td>
</tr>
<tr>
<td>Finding Biologically Accurate Clusterings in Hierarchical Tree</td>
<td>400</td>
</tr>
<tr>
<td>Saket Navlakha, James White, Niranjan Nagarajan, Mihai Pop, and Carl Kingsford</td>
<td></td>
</tr>
<tr>
<td>Identification and Frequency Estimation of Inversion Polymorphisms from Haplotype Data</td>
<td>418</td>
</tr>
<tr>
<td>Suzanne S. Sindi and Benjamin J. Raphael</td>
<td></td>
</tr>
<tr>
<td>On the Relationship between DNA Periodicity and Local Chromatin Structure</td>
<td>434</td>
</tr>
<tr>
<td>Sheila M. Reynolds, Jeff A. Bilmes, and William Stafford Noble</td>
<td></td>
</tr>
<tr>
<td>Phylogenies without Branch Bounds: Contracting the Short, Pruning the Deep (Extended Abstract)</td>
<td>451</td>
</tr>
<tr>
<td>Constantinos Daskalakis, Elchanan Mossel, and Sebastien Roch</td>
<td></td>
</tr>
<tr>
<td>Detecting the Presence and Absence of Causal Relationships between Expression of Yeast Genes with Very Few Samples</td>
<td>466</td>
</tr>
<tr>
<td>Eun Yong Kang, Ilya Shpitser, Chun Ye, and Eleazar Eskin</td>
<td></td>
</tr>
<tr>
<td>An Adaptive and Memory Efficient Algorithm for Genotype Imputation</td>
<td>482</td>
</tr>
<tr>
<td>Hyun Min Kang, Noah A. Zaitlen, Buhm Han, and Eleazar Eskin</td>
<td></td>
</tr>
<tr>
<td>A Statistical Framework for the Functional Analysis of Metagenomes</td>
<td>496</td>
</tr>
<tr>
<td>Itai Sharon, Amrita Pati, Victor M. Markowitz, and Ron Y. Pinter</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Learning Models for Aligning Protein Sequences with Predicted</td>
<td>512</td>
</tr>
<tr>
<td>Secondary Structure</td>
<td></td>
</tr>
<tr>
<td><em>Eagu Kim, Travis Wheeler, and John Kececioglu</em></td>
<td></td>
</tr>
<tr>
<td><strong>Author Index</strong></td>
<td>533</td>
</tr>
</tbody>
</table>