

# Lecture Notes in Bioinformatics

5541

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

Serafim Batzoglou (Ed.)

# Research in Computational Molecular Biology

13th Annual International Conference, RECOMB 2009  
Tucson, AZ, USA, May 18-21, 2009  
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 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

ISBN-13 978-3-642-02007-0 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<br>(Program Chair)  | Stanford University, USA                    |
| 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<br>(Conference Chair) | University of Arizona, USA                  |
| Manolis Kellis                        | Massachusetts Institute of Technology, USA  |
| Jens Lagergren                        | Stokholm 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<br>San Diego, USA |
| Terence P. Speed  | University of California Berkeley, USA                           |

### **Local Organization at the University of Arizona**

|                                       |  |
|---------------------------------------|--|
| John Kececioglu<br>(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

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

## 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  
 Fernandez-Baca, David  
 Fujita, Andre  
 Gat-Viks, Irit  
 Gerlach, Wolfgang  
 Goff, Loyal  
 Grãapl, Clemens  
 Gusfield, Dan  
 Haas, Stefan  
 Habib, Naomi  
 Haim, Wolfson, Efrat 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  
 Klau, Gunnar  
 Kojima, Kaname  
 Kolaczyk, Eric  
 Kosiol, Carolin  
 Kriete, Andres  
 Kuang, Rui  
 Kuttykrishnan, Sooraj  
 Lasserre, Julia  
 Lee, Byoungkoo  
 Lee, Ki Young  
 Lee, KiYoung  
 Li, Yong  
 Lilien, Ryan  
 Lin, Michael  
 Lonardi, Stefano  
 Ma, Xiaotu  
 Martins, Andre  
 McIlwain, Sean  
 Medvedev, Paul  
 Meshi, Ofer  
 MilaniÄ, Martin  
 Misra, Navodit  
 Mitrofanova, Antonina  
 Morgan, Alex  
 Morozov, Alexandre  
 Moses, Alan  
 Mosig, Axel  
 Mil, Mathias  
 Nakhleh, Luay  
 Navon, Roy  
 Ng, Julio  
 Novershtern, Noa  
 Obozinski, Guillaume  
 Paik, David  
 Pardo, Matteo  
 Pasaniuc, Bogdan  
 Perrier, Eric  
 Phillips, Michael  
 Pincus, Zach  
 Pop, Mihai  
 Qi, Yanjun  
 Raphael, Ben  
 Rasmussen, Matthew  
 Ray, Pradipta  
 Reynolds, Sheila  
 RingnÄr, Markus  
 Salari, Rahele  
 Sankararaman, Sriram  
 Satulovsky, Javier  
 Schelhorn, Sven-Eric  
 Schuster, Meromit  
 Sealfon, Rachel  
 Segal, Eran  
 Sennblad, Bengt



Shen-Orr, Shai  
Shimamura, Teppei  
Shlomi, Tomer  
Shmoish, Michael  
Shringarpure, Suyash  
Sievers, Fabian  
Sjöstrand, Joel  
Sohn, Kyung-Ah  
Song, Le  
Sridhar, Srinath  
Srivasa, Rohith  
Steel, Mike  
Steffen, Martin  
Steinfeld, Israel  
Steinhoff, Christine  
Subramanian, Ayshwarya  
Suthram, Silpa

Taliwal, Vikas  
Tamada, Yoshinori  
Tamura, Takeyuki  
Wittler, Roland  
Wu, Xuebing  
Xia, Charles Li  
Xu, Wei  
Yamaguchi, Rui  
Yanovsky, Vladimir  
Ye, Chun  
Yosef, Nir  
Yuan, Yuan  
Zaitlen, Noah  
Zerck, Alexandra  
Zhang, Kui  
Zheng, Chunfang

# Table of Contents

|   |     |
|---|-----|
| Searching Protein 3-D Structures in Linear Time . . . . .   | 1   |
| <i>Tetsuo Shibuya</i>   |     |
| Optimization-Based Peptide Mass Fingerprinting for Protein Mixture<br>Identification . . . . .              | 16  |
| <i>Zengyou He, Chao Yang, Can Yang, Robert Z. Qi,<br/>Jason Po-Ming Tam, and Weichuan Yu</i>                |     |
| Boosting Protein Threading Accuracy . . . . .   | 31  |
| <i>Jian Peng and Jinbo Xu</i>   |     |
| New Perspectives on Gene Family Evolution: Losses in Reconciliation<br>and a Link with Supertrees . . . . . | 46  |
| <i>Cedric Chauve and Nadia El-Mabrouk</i>   |     |
| A Probabilistic Graphical Model for Ab Initio Folding . . . . .   | 59  |
| <i>Feng Zhao, Jian Peng, Joe DeBartolo, Karl F. Freed,<br/>Tobin R. Sosnick, and Jinbo Xu</i>               |     |
| Topology-Free Querying of Protein Interaction Networks . . . . .  | 74  |
| <i>Sharon Bruckner, Falk Hüffner, Richard M. Karp, Ron Shamir, and<br/>Roded Sharan</i>                     |     |
| Cross Species Expression Analysis of Innate Immune Response . . . . .                                       | 90  |
| <i>Yong Lu, Roni Rosenfeld, Gerard J. Nau, and Ziv Bar-Joseph</i>   |     |
| Haplotype Inference in Complex Pedigrees . . . . .  | 108 |
| <i>Bonnie Kirkpatrick, Javier Rosa, Eran Halperin, and<br/>Richard M. Karp</i>                              |     |
| Storage and Retrieval of Individual Genomes . . . . .   | 121 |
| <i>Veli Mäkinen, Gonzalo Navarro, Jouni Sirén, and Niko Välimäki</i>  |     |
| An Online Approach for Mining Collective Behaviors from Molecular<br>Dynamics Simulations . . . . .         | 138 |
| <i>Arvind Ramanathan, Pratul K. Agarwal, Maria Kurnikova, and<br/>Christopher J. Langmead</i>               |     |
| Parameter Synthesis in Nonlinear Dynamical Systems: Application to<br>Systems Biology . . . . .             | 155 |
| <i>Alexandre Donzé, Gilles Clermont, Axel Legay, and<br/>Christopher J. Langmead</i>                        |     |

|   |     |
|---|-----|
| Spatial Clustering of Multivariate Genomic and Epigenomic Information . . . . .   | 170 |
| <i>Rami Jaschek and Amos Tanay</i>  |     |
| How Many Bootstrap Replicates Are Necessary? . . . . .  | 184 |
| <i>Nicholas D. Pattengale, Masoud Alipour, Olaf R.P. Bininda-Emonds, Bernard M.E. Moret, and Alexandros Stamatakis</i>                          |     |
| A Robust Bayesian Two-Sample Test for Detecting Intervals of Differential Gene Expression in Microarray Time Series . . . . .                   | 201 |
| <i>Oliver Stegle, Katherine Denby, David L. Wild, Zoubin Ghahramani, and Karsten M. Borgwardt</i>   |     |
| Incorporating Nucleosomes into Thermodynamic Models of Transcription Regulation . . . . .   | 217 |
| <i>Tali Raveh-Sadka, Michal Levo, and Eran Segal</i>  |     |
| Combinatorial Algorithms for Structural Variation Detection in High Throughput Sequenced Genomes . . . . .                                      | 218 |
| <i>Fereydown Hormozdiari, Can Alkan, Evan E. Eichler, and S. Cenk Sahinalp</i>  |     |
| Optimizing PCR Assays for DNA Based Cancer Diagnostics . . . . .  | 220 |
| <i>Ali Bashir, Qing Lu, Dennis Carson, Benjamin Raphael, Yu-Tsueng Liu, and Vineet Bafna</i>  |     |
| The Multi-State Perfect Phylogeny Problem with Missing and Removable Data: Solutions via Integer-Programming and Chordal Graph Theory . . . . . | 236 |
| <i>Dan Gusfield</i>   |     |
| COE: A General Approach for Efficient Genome-Wide Two-Locus Epistasis Test in Disease Association Study . . . . .                               | 253 |
| <i>Xiang Zhang, Feng Pan, Yuying Xie, Fei Zou, and Wei Wang</i>   |     |
| Overlapping Pools for High Throughput Targeted Resequencing . . . . .   | 270 |
| <i>Snehit Prabhu and Itsik Pe'er</i>  |     |
| Deep Sequencing of a Genetically Heterogeneous Sample: Local Haplotype Reconstruction and Read Error Correction . . . . .                       | 271 |
| <i>Oswaldo Zagordi, Lukas Geyrhofer, Volker Roth, and Niko Beerenwinkel</i>   |     |
| Lifting Prediction to Alignment of RNA Pseudoknots . . . . .  | 285 |
| <i>Mathias Möhl, Sebastian Will, and Rolf Backofen</i>  |     |
| Detection of Locally Over-Represented GO Terms in Protein-Protein Interaction Networks . . . . .  | 302 |
| <i>Mathieu Lavallée-Adam, Benoit Coulombe, and Mathieu Blanchette</i>   |     |

|  |     |
|--|-----|
| Protein Fragment Swapping: A Method for Asymmetric, Selective Site-Directed Recombination . . . . .                          | 321 |
| <i>Wei Zheng, Karl E. Griswold, and Chris Bailey-Kellogg</i>   |     |
| Simultaneous Alignment and Folding of Protein Sequences . . . . .  | 339 |
| <i>Jérôme Waldispühl, Charles W. O'Donnell, Sebastian Will, Srinivas Devadas, Rolf Backofen, and Bonnie Berger</i>           |     |
| Shared Peptides in Mass Spectrometry Based Protein Quantification . . .  | 356 |
| <i>Banu Dost, Nuno Bandeira, Xiangqian Li, Zhouxin Shen, Steve Briggs, and Vineet Bafna</i>                                  |     |
| Evaluating Between-Pathway Models with Expression Data . . . . .   | 372 |
| <i>Benjamin J. Hescott, Maek D.M. Leiserson, Lenore J. Cowen, and Donna K. Slonim</i>  |     |
| Sorting Signed Permutations by Inversions in $O(n \log n)$ Time . . . . .  | 386 |
| <i>Krister M. Swenson, Vaibhav Rajan, Yu Lin, and Bernard M.E. Moret</i>   |     |
| Finding Biologically Accurate Clusterings in Hierarchical Tree Decompositions Using the Variation of Information . . . . .   | 400 |
| <i>Saket Navlakha, James White, Niranjan Nagarajan, Mihai Pop, and Carl Kingsford</i>  |     |
| Identification and Frequency Estimation of Inversion Polymorphisms from Haplotype Data . . . . .                             | 418 |
| <i>Suzanne S. Sindi and Benjamin J. Raphael</i>  |     |
| On the Relationship between DNA Periodicity and Local Chromatin Structure . . . . .  | 434 |
| <i>Sheila M. Reynolds, Jeff A. Bilmes, and William Stafford Noble</i>  |     |
| Phylogenies without Branch Bounds: Contracting the Short, Pruning the Deep (Extended Abstract) . . . . .                     | 451 |
| <i>Constantinos Daskalakis, Elchanan Mossel, and Sebastien Roch</i>  |     |
| Detecting the Presence and Absence of Causal Relationships between Expression of Yeast Genes with Very Few Samples . . . . . | 466 |
| <i>Eun Yong Kang, Ilya Shpitser, Chun Ye, and Eleazar Eskin</i>  |     |
| An Adaptive and Memory Efficient Algorithm for Genotype Imputation . . . . .   | 482 |
| <i>Hyun Min Kang, Noah A. Zaitlen, Buhm Han, and Eleazar Eskin</i>   |     |
| A Statistical Framework for the Functional Analysis of Metagenomes . . .   | 496 |
| <i>Itai Sharon, Amrita Pati, Victor M. Markowitz, and Ron Y. Pinter</i>  |     |

|   |            |
|---|------------|
| Learning Models for Aligning Protein Sequences with Predicted<br>Secondary Structure..... | 512        |
| <i>Eagu Kim, Travis Wheeler, and John Kececioglu</i>                                      |            |
| <b>Author Index</b> .....   | <b>533</b> |