Lecture Notes in Bioinformatics 3909

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

Alberto Apostolico Concettina Guerra
Sorin Istrail Pavel Pevzner
Michael Waterman (Eds.)

Research in Computational Molecular Biology

10th Annual International Conference, RECOMB 2006
Venice, Italy, April 2-5, 2006
Proceedings

Springer

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

Alberto Apostolico
Concettina Guerra
University of Padova, Department of Information Engineering
Via Gradenigo 6/a, 35131 Padova, Italy
E-mail: {axa, guerra}@dei.unipd.it

Sorin Istrail
Brown University, Center for Molecular Biology and Computer Science Department
115 Waterman St., Providence, RI 02912, USA
E-mail: sorin@cs.brown.edu

Pavel Pevzner
University of California at San Diego
Department of Computer Science and Engineering
La Jolla, CA 92093-0114, USA
E-mail: ppevzner@cs.ucsd.edu

Michael Waterman
University of Southern California
Department of Molecular and Computational Biology
1050 Childs Way, Los Angeles, CA 90089-2910, USA
E-mail: msw@usc.edu

Library of Congress Control Number: 2006922626

CR Subject Classification (1998): F.2.2, F.2, E.1, G.2, H.2.8, G.3, I.2, J.3

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743
ISBN-10 3-540-33295-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-33295-4 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 2006
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11732990 06/3142 5 4 3 2 1 0

Preface

This volume contains the papers presented at the 10th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2006), which was held in Venice, Italy, on April 2–5, 2006. The RECOMB conference series was started in 1997 by Sorin Istrail, Pavel Pevzner and Michael Waterman. The table on p. VIII summarizes the history of the meetings. RECOMB 2006 was hosted by the University of Padova at the Cinema Palace of the Venice Convention Center, Venice Lido, Italy. It was organized by a committee chaired by Concettina Guerra. A special 10th Anniversary Program Committee was formed, by including the members of the Steering Committee and inviting all Chairs of past editions. The Program Committee consisted of the 38 members whose names are listed on a separate page.

From 212 submissions of high quality, 40 papers were selected for presentation at the meeting, and they appear in these proceedings. The selection was based on reviews and evaluations produced by the Program Committee members as well as by external reviewers, and on a subsequent Web-based PC open forum. Following the decision made in 2005 by the Steering Committee, RECOMB Proceedings are published as a volume of *Lecture Notes in Bioinformatics* (LNBI), which is co-edited by the founders of RECOMB. Traditionally, the *Journal of Computational Biology* devotes a special issue to the publication of archival versions of selected conference papers.

RECOMB 2006 featured seven keynote addresses by as many invited speakers: Anne-Claude Gavin (EMBL, Heidelberg, Germany), David Haussler (University of California, Santa Cruz, USA), Ajay K. Royyuru (IBM T.J. Watson Research Center, USA), David Sankoff (University of Ottawa, Canada), Michael S. Waterman (University of Southern California, USA), Carl Zimmer (Science Writer, USA), Roman A. Zubarev (Uppsala University, Sweden). The Stanislaw Ulam Memorial Computational Biology Lecture was given by Michael S. Waterman. A special feature presentation was devoted to the 10th anniversary and is included in this volume.

Like in the past, an important ingredient for the success of the meeting was represented by a lively poster session.

RECOMB06 was made possible by the hard work and dedication of many, from the Steering to the Program and Organizing Committees, from the external reviewers, to Venice Convention, Venezia Congressi and the institutions and corporations who provided administrative, logistic and financial support for the conference. The latter include the Department of Information Engineering of the University of Padova, the Broad Institute of MIT and Harvard (USA), the College of Computing of Georgia Tech. (USA), the US Department of Energy, IBM Corporation (USA), the International Society for Computational Biology

(ISCB), the Italian Association for Informatics and Automatic Computation (AICA), the US National Science Foundation, and the University of Padova.

Special thanks are due to all those who submitted their papers and posters and who attended RECOMB 2006 with enthusiasm.

April 2006

Alberto Apostolico
RECOMB 2006 Program Chair

Organization

Program Committee

Tatsuya Akutsu	(Kyoto University, Japan)
Alberto Apostolico	Chair (Accademia Nazionale Dei Lincei, Italy, and Georgia Tech., USA)
Gary Benson	(Boston University, USA)
Mathieu Blanchette	(McGill, Canada)
Philip E. Bourne	(University of California San Diego, USA)
Steve Bryant	(NCBI, USA)
Andrea Califano	(Columbia University, USA)
Andy Clark	(Cornell University, USA)
Gordon M. Crippen	(University of Michigan, USA)
Raffaele Giancarlo	(University of Palermo, Italy)
Concettina Guerra	(University of Padova, Italy, and Georgia Tech., USA)
Dan Gusfield	(University of California, Davis, USA)
Sridhar Hannenhalli	(University of Pennsylvania, USA)
Sorin Istrail	(Brown University, USA)
Inge Jonassen	(University of Bergen, Norway)
Richard M. Karp	(University of California, Berkeley, USA)
Simon Kasif	(Boston University, USA)
Manolis Kellis	(MIT, USA)
Giuseppe Lancia	(University of Udine, Italy)
Thomas Lengauer	(GMD Sant Augustin, Germany)
Michael Levitt	(Stanford, USA)
Michal Linial	(The Hebrew University in Jerusalem, Israel)
Jill Mesirov	(Broad Institute of MIT and Harvard, USA)
Satoru Miyano	(University of Tokyo, Japan)
Gene Myers	(HHMI, USA)
Laxmi Parida	(IBM T.J. Watson Research Center, USA)
Pavel A. Pevzner	(University of California San Diego, USA)
Marie-France Sagot	(INRIA Rhone-Alpes, France)
David Sankoff	(University of Ottawa, Canada)
Ron Shamir	(Tel Aviv University, Israel)
Roded Sharan	(Tel Aviv University, Israel)
Steve Skiena	(State University of New York at Stony Brook, USA)
Terry Speed	(University of California, Berkeley, USA)
Jens Stoye	(University of Bielefeld, Germany)
Esko Ukkonen	(University of Helsinki, Finland)
Martin Vingron	(Max Planck Institute for Molecular Genetics, Germany)
Michael Waterman	(University of Southern California, USA)
Haim J. Wolfson	(Tel Aviv University, Israel)

Steering Committee

Sorin Istrail	RECOMB General Vice-chair (Brown, USA)
Thomas Lengauer	(GMD Sant Augustin, Germany)
Michal Linial	(The Hebrew University of Jerusalem, Israel)
Pavel A. Pevzner	RECOMB General Chair (University of California, San Diego, USA)
Ron Shamir	(Tel Aviv University, Israel)
Terence P. Speed	(University of California, Berkeley, USA)
Michael Waterman	RECOMB General Chair (University of Southern California, USA)

Organizing Committee

Alberto Apostolico	(Accademia Nazionale dei Lincei, Italy, and Georgia Tech., USA)
Concettina Guerra	Conference Chair (University of Padova, Italy, and Georgia Tech., USA)
Eleazar Eskin	Chair, 10th Anniversary Committee (University of California, San Diego)
Matteo Comin	(University of Padova, Italy)
Raffaele Giancarlo	(University of Palermo, Italy)
Giuseppe Lancia	(University of Udine, Italy)
Cinzia Pizzi	(University of Padova, Italy, and Univ. of Helsinki, Finland)
Angela Visco	(University of Padova, Italy)
Nicola Vitacolonna	(University of Udine, Italy)

Previous RECOMB Meetings

Date/Location	Hosting Institution	Program Chair	Conference Chair
January 20-23, 1997 Santa Fe, NM, USA	Sandia National Lab	Michael Waterman	Sorin Istrail
March 22-25, 1998 New York, NY, USA	Mt. Sinai School of Medicine	Pavel Pevzner	Gary Benson
April 22-25, 1999 Lyon, France	INRIA	Sorin Istrail	Mireille Regnier
April 8-11, 2000 Tokyo, Japan	University of Tokyo	Ron Shamir	Satoru Miyano
April 22-25, 2001 Montréal, Canada	Université de Montréal	Thomas Lengauer	David Sankoff
April 18-21, 2002 Washington, DC, USA	Celera	Gene Myers	Sridhar Hannenhalli
April 10-13, 2003 Berlin, Germany	German Federal Ministry for Education & Research	Webb Miller	Martin Vingron
March 27-31, 2004 San Diego, USA	UC San Diego	Dan Gusfield	Philip E. Bourne
May 14-18, 2005 Boston, MA, USA	Broad Institute of MIT and Harvard	Satoru Miyano	Jill P. Mesirov and S. Kasif

The RECOMB06 Program Committee gratefully acknowledges the valuable input received from the following external reviewers:

Josep F. Abril
Mario Albrecht
Gabriela Alexe
Julien Allali
Miguel Andrade
Brian Anton
Sebastian Böcker
Vineet Bafna
Melanie Bahlo
Nilanjana Banerjee
Ali Bashir
Amir Ben-Dor
Asa Ben-Hur
Yoav Benjamini
Chris Benner
Gyan Bhanot
Trond Hellem Bø
Elhanan Borenstein
Guillaume Bourque
Frederic Boyer
Dan Brown
Trevor Bruen
Renato Bruni
David Bryant
Jeremy Buhler
Nello Cristianini
Jo-Lan Chung
Barry Cohen
Inbar Cohen-Gihon
Matteo Comin
Ana Teresa Freitas
Miklos Csuros
Andre Dabrowski
Alessandro Dal Palú
Sanjoy DasGupta
Gianluca Della Vedova
Greg Dewey
Zhihong Ding
Atsushi Doi
Dikla Dotan
Agostino Dovier
Oranit Dror
Bjarte Dysvik
Nadia El-Mabrouk
Rani Elkon
Sean Escola
Eleazar Eskin
Jean-Eudes Duchesne
Jay Faith
David Fernandez-Baca
Vladimir Filkov
Sarel Fleishman
Kristian Flikka
Menachem Former
Iddo Friedberg
Menachem Fruman
Irit Gat-Viks
Gad Getz
Apostol Gramada
Alex Gray
Steffen Grossmann
Jenny Gu
Roderic Guigo
Matthew Hahn
Yonit Halperin
Tzvika Hartman
Christoph Hartmann
Nurit Haspel
Greg Hather
Morihiro Hayashida
Trond Hellem Bø
D. Hermelin
Katsuhisa Horimoto
Moseley Hunter
Seiya Imoto
Yuval Inbar
Nathan Intrator
David Jaffe
Martin Jambon
Shane Jensen
Euna Jeong
Tao Jiang
Juha Kärkkäinen
Hans-Michael Kaltenbach
Simon Kasif
Klara Kedem
Alon Keinan
Wayne Kendal
Ilona Kifer
Gad Kimmel
Jyrki Kivinen
Mikko Koivisto
Rachel Kolodny
Vered Kunik
Vincent Lacroix
Quan Le
Soo Lee
Celine Lefebvre
Hadas Leonov
Jie Liang
Chaim Linhart
Zsuzsanna Lipták
Manway Liu
Aniv Loewenstein
Claudio Lottaz
Claus Lundegaard
Hannes Luz
Aaron Mackey
Ketil Malde
Kartik Mani
Thomas Manke
Yishay Mansour
Adam Margolin
Florian Markowetz
Setsuro Matsuda
Alice McHardy
Kevin Miranda
Leonid Mirny
Stefano Monti
Sayan Mukherjee
Iftach Nachman
Masao Nagasaki

Rei-ichiro Nakamichi
Tim Nattkemper
Ilya Nemenman
Sebastian Oehm
Arlindo Oliveira
Michal Ozery-Flato
Kimmo Palin
Kim Palmo
Paul Pavlidis
Anton Pervukhin
Pierre Peterlongo
Kjell Petersen
Nadia Pisanti
Gianluca Pollastri
Julia Ponomarenko
Elon Portugaly
John Rachlin
Sven Rahmann
Jörg Rahnenführer
Daniela Raijman
Ari Rantanen
Ramamoorthi Ravi
Marc Rehmsmeier
Vicente Reyes
Romeo Rizzi
Estela Maris Rodrigues
Ken Ross
Juho Rousu
Eytan Ruppin
Walter L. Ruzzo
Michael Sammeth
Oliver Sander
Zack Saul
Simone Scalabrin
Klaus-Bernd Schürmann
Michael Schaffer
Stefanie Scheid
Alexander Schliep
Dina Schneidman
Russell Schwartz
Paolo Serafini
Maxim Shatsky
Feng Shengzhong
Tetsuo Shibuya
Ilya Shindyalov
Tomer Shlomi
A. Shulman-Peleg
Abdur Sikder
Gordon Smyth
Yun Song
Rainer Spang
Mike Steel
Israel Steinfeld
Christine Steinhoff
Kristian Stevens
Aravind Subramanian
Fengzhu Sun
Christina Sunita Leslie
Edward Susko
Yoshinori Tamada
Amos Tanay
Haixu Tang
Eric Tannier
Elisabeth Tillier
Wiebke Timm
Aristotelis Tsirigos
Nobuhisa Ueda
Igor Ulitsky
Sandor Vajda
Roy Varshavsky
Balaji Venkatachalam
Stella Veretnik
Dennis Vitkup
Yoshiko Wakabayashi
Jianyong Wang
Junwen Wang
Kai Wang
Li-San Wang
Lusheng Wang
Tandy Warnow
Arieh Warshel
David Wild
Virgil Woods
Terrence Wu
Yufeng Wu
Lei Xie
Chen Xin
Eric Xing
Zohar Yakhini
Nir Yosef
Ryo Yoshida
John Zhang
Louxin Zhang
Degui Zhi
Xianghong J. Zhou
Joseph Ziv Bar
Michal Ziv-Ukelson

RECOMB Tenth Anniversary Venue: il Palazzo Del Cinema del Lido di Venezia

Sponsors

DEPARTMENT OF
INFORMATION
ENGINEERING
UNIVERSITY OF PADOVA

BROAD
INSTITUTE

Georgia
Tech

National Science Foundation
WHERE DISCOVERIES BEGIN

iSCB
INTERNATIONAL
SOCIETY FOR
COMPUTATIONAL
BIOLOGY

AICA
Associazione Italiana per l'Informatica
ed il Calcolo Automatico

IBM

Table of Contents

Integrated Protein Interaction Networks for 11 Microbes
Balaji S. Srinivasan, Antal F. Novak, Jason A. Flannick, Serafim Batzoglou, Harley H. McAdams 1

Hypergraph Model of Multi-residue Interactions in Proteins: Sequentially–Constrained Partitioning Algorithms for Optimization of Site-Directed Protein Recombination
Xiaoduan Ye, Alan M. Friedman, Chris Bailey-Kellogg 15

Biological Networks: Comparison, Conservation, and Evolutionary Trees
Benny Chor, Tamir Tuller 30

Assessing Significance of Connectivity and Conservation in Protein Interaction Networks
Mehmet Koyutürk, Ananth Grama, Wojciech Szpankowski 45

Clustering Short Gene Expression Profiles
Ling Wang, Marco Ramoni, Paola Sebastiani 60

A Patient-Gene Model for Temporal Expression Profiles in Clinical Studies
Naftali Kaminski, Ziv Bar-Joseph 69

Global Interaction Networks Probed by Mass Spectrometry (Keynote)
Anne-Claude Gavin 83

Statistical Evaluation of Genome Rearrangement (Keynote)
David Sankoff 84

An Improved Statistic for Detecting Over-Represented Gene Ontology Annotations in Gene Sets
Steffen Grossmann, Sebastian Bauer, Peter N. Robinson, Martin Vingron 85

Protein Function Annotation Based on Ortholog Clusters Extracted from Incomplete Genomes Using Combinatorial Optimization
Akshay Vashist, Casimir Kulikowski, Ilya Muchnik 99

Detecting MicroRNA Targets by Linking Sequence, MicroRNA and Gene Expression Data
Jim C. Huang, Quaid D. Morris, Brendan J. Frey 114

RNA Secondary Structure Prediction Via Energy Density Minimization
Can Alkan, Emre Karakoc, S. Cenk Sahinalp, Peter Unrau, H. Alexander Ebhardt, Kaizhong Zhang, Jeremy Buhler 130

Structural Alignment of Pseudoknotted RNA
Banu Dost, Buhm Han, Shaojie Zhang, Vineet Bafna 143

Stan Ulam and Computational Biology (Keynote)
Michael S. Waterman 159

CONTRAlign: Discriminative Training for Protein Sequence Alignment
Chuong B. Do, Samuel S. Gross, Serafim Batzoglou 160

Clustering Near-Identical Sequences for Fast Homology Search
Michael Cameron, Yaniv Bernstein, Hugh E. Williams 175

New Methods for Detecting Lineage-Specific Selection
Adam Siepel, Katherine S. Pollard, David Haussler 190

A Probabilistic Model for Gene Content Evolution with Duplication, Loss, and Horizontal Transfer
Miklós Csűrös, István Miklós 206

A Sublinear-Time Randomized Approximation Scheme for the Robinson-Foulds Metric
Nicholas D. Pattengale, Bernard M.E. Moret 221

Algorithms to Distinguish the Role of Gene-Conversion from Single-Crossover Recombination in the Derivation of SNP Sequences in Populations
Yun S. Song, Zhihong Ding, Dan Gusfield, Charles H. Langley, Yufeng Wu 231

Inferring Common Origins from mtDNA (Keynote)
Ajay K. Royyuru, Gabriela Alexe, Daniel Platt, Ravi Vijaya-Satya, Laxmi Parida, Saharon Rosset, Gyan Bhanot 246

Efficient Enumeration of Phylogenetically Informative Substrings
Stanislav Angelov, Boulos Harb, Sampath Kannan, Sanjeev Khanna, Junhyong Kim 248

Phylogenetic Profiling of Insertions and Deletions in Vertebrate Genomes
Sagi Snir, Lior Pachter 265

Maximal Accurate Forests from Distance Matrices
Constantinos Daskalakis, Cameron Hill, Alexandar Jaffe, Radu Mihaescu, Elehanan Mossel, Satish Rao 281

Leveraging Information Across HLA Alleles/Supertypes Improves Epitope Prediction
David Heckerman, Carl Kadie, Jennifer Listgarten 296

Improving Prediction of Zinc Binding Sites by Modeling the Linkage Between Residues Close in Sequence
Sauro Menchetti, Andrea Passerini, Paolo Frasconi, Claudia Andreini, Antonio Rosato 309

An Important Connection Between Network Motifs and Parsimony Models
Teresa M. Przytycka 321

Ultraconserved Elements, Living Fossil Transposons, and Rapid Bursts of Change: Reconstructing the Uneven Evolutionary History of the Human Genome (Keynote)
David Haussler 336

Permutation Filtering: A Novel Concept for Significance Analysis of Large-Scale Genomic Data
Stefanie Scheid, Rainer Spang 338

Genome-Wide Discovery of Modulators of Transcriptional Interactions in Human B Lymphocytes
Kai Wang, Ilya Nemenman, Nilanjana Banerjee, Adam A. Margolin, Andrea Califano 348

A New Approach to Protein Identification
Nuno Bandeira, Dekel Tsur, Ari Frank, Pavel Pevzner 363

Markov Methods for Hierarchical Coarse-Graining of Large Protein Dynamics
Chakra Chennubhotla, Ivet Bahar 379

Simulating Protein Motions with Rigidity Analysis
Shawna Thomas, Xinyu Tang, Lydia Tapia, Nancy M. Amato 394

Predicting Experimental Quantities in Protein Folding Kinetics Using Stochastic Roadmap Simulation
Tsung-Han Chiang, Mehmet Serkan Apaydin, Douglas L. Brutlag, David Hsu, Jean-Claude Latombe 410

An Outsider's View of the Genome (Keynote)
Carl Zimmer 425

Alignment Statistics for Long-Range Correlated Genomic Sequences
Philipp W. Messer, Ralf Bundschuh, Martin Vingron, Peter F. Arndt 426

Simple and Fast Inverse Alignment
John Kececioglu, Eagu Kim 441

Revealing the Proteome Complexity by Mass Spectrometry (Keynote)
Roman A. Zubarev 456

Motif Yggdrasil: Sampling from a Tree Mixture Model
Samuel A. Andersson, Jens Lagergren 458

A Study of Accessible Motifs and RNA Folding Complexity
Ydo Wexler, Chaya Zilberstein, Michal Ziv-Ukelson 473

A Parameterized Algorithm for Protein Structure Alignment
Jinbo Xu, Feng Jiao, Bonnie Berger 488

Geometric Sieving: Automated Distributed Optimization of 3D Motifs for Protein Function Prediction
Brian Y. Chen, Viacheslav Y. Fofanov, Drew H. Bryant, Bradley D. Dodson, David M. Kristensen, Andreas M. Lisewski, Marek Kimmel, Olivier Lichtarge, Lydia E. Kavraki 500

A Branch-and-Reduce Algorithm for the Contact Map Overlap Problem
Wei Xie, Nikolaos V. Sahinidis 516

A Novel Minimized Dead-End Elimination Criterion and Its Application to Protein Redesign in a Hybrid Scoring and Search Algorithm for Computing Partition Functions over Molecular Ensembles
Ivelin Georgiev, Ryan H. Lilien, Bruce R. Donald 530

10 Years of the International Conference on Research in Computational Molecular Biology (RECOMB) (Keynote)
Sarah J. Aerni, Eleazar Eskin 546

Sorting by Weighted Reversals, Transpositions, and Inverted Transpositions
Martin Bader, Enno Ohlebusch 563

A Parsimony Approach to Genome-Wide Ortholog Assignment
Zheng Fu, Xin Chen, Vladimir Vacic, Peng Nan, Yang Zhong, Tao Jiang 578

Detecting the Dependent Evolution of Biosequences
Jeremy Darot, Chen-Hsiang Yeang, David Haussler 595

Author Index 611