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)
VIII Organization

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

<table>
<thead>
<tr>
<th>Date/Location</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>New York, NY, USA</td>
<td>Sorin Istrail</td>
<td>Mireille Regnier</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>April 22-25, 1999</td>
<td>INRIA</td>
<td>Sorin Istrail</td>
<td>Satoru Miyano</td>
</tr>
<tr>
<td>Lyon, France</td>
<td>University of Tokyo</td>
<td>Ron Shamir</td>
<td>Satoru Miyano</td>
</tr>
<tr>
<td>April 8-11, 2000</td>
<td>Tokyo, Japan</td>
<td>Ron Shamir</td>
<td>Satoru Miyano</td>
</tr>
<tr>
<td>March 22-25, 2001</td>
<td>Université de Montréal</td>
<td>Thomas Lengauer</td>
<td>David Sankoff</td>
</tr>
<tr>
<td>Montréal, Canada</td>
<td></td>
<td>Sorin Istrail</td>
<td>Satoru Miyano</td>
</tr>
<tr>
<td>April 18-21, 2002</td>
<td>Celeria</td>
<td>Gene Myers</td>
<td>Sridhar Hannenhalli</td>
</tr>
<tr>
<td>Washington, DC, USA</td>
<td>German Federal Ministry for Education &amp; Research</td>
<td>Webb Miller</td>
<td>Martin Vingron</td>
</tr>
<tr>
<td>April 10-13, 2003</td>
<td>Berlin, Germany</td>
<td>UC San Diego</td>
<td>Dan Gusfield</td>
</tr>
<tr>
<td>March 27-31, 2004</td>
<td>San Diego, USA</td>
<td>UC San Diego</td>
<td>David Sankoff</td>
</tr>
<tr>
<td>May 14-18, 2005</td>
<td>Broad Institute of MIT and Harvard</td>
<td>UC San Diego</td>
<td>David Sankoff</td>
</tr>
<tr>
<td>Boston, MA, USA</td>
<td>RECOMB General Chair (Brown, USA)</td>
<td>UC San Diego</td>
<td>Jill P. Mesirov and S. Kasif</td>
</tr>
</tbody>
</table>
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 Palù
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
RECOMB Tenth Anniversary Venue: il Palazzo Del Cinema del Lido di Venezia
Sponsors
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
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural Alignment of Pseudoknotted RNA</td>
<td>Banu Dost, Buhm Han, Shaojie Zhang, Vineet Bafna</td>
<td>143</td>
</tr>
<tr>
<td>Stan Ulam and Computational Biology (Keynote)</td>
<td>Michael S. Waterman</td>
<td>159</td>
</tr>
<tr>
<td>CONTRAlign: Discriminative Training for Protein Sequence Alignment</td>
<td>Chuong B. Do, Samuel S. Gross, Serafim Batzoglou</td>
<td>160</td>
</tr>
<tr>
<td>Clustering Near-Identical Sequences for Fast Homology Search</td>
<td>Michael Cameron, Yaniv Bernstein, Hugh E. Williams</td>
<td>175</td>
</tr>
<tr>
<td>New Methods for Detecting Lineage-Specific Selection</td>
<td>Adam Siepel, Katherine S. Pollard, David Haussler</td>
<td>190</td>
</tr>
<tr>
<td>A Probabilistic Model for Gene Content Evolution with Duplication, Loss, and Horizontal Transfer</td>
<td>Miklós Csúrós, István Miklós</td>
<td>206</td>
</tr>
<tr>
<td>A Sublinear-Time Randomized Approximation Scheme for the Robinson-Foulds Metric</td>
<td>Nicholas D. Pattengale, Bernard M.E. Moret</td>
<td>221</td>
</tr>
<tr>
<td>Algorithms to Distinguish the Role of Gene-Conversion from Single-Crossover Recombination in the Derivation of SNP Sequences in Populations</td>
<td>Yun S. Song, Zihong Ding, Dan Gusfield, Charles H. Langley, Yufeng Wu</td>
<td>231</td>
</tr>
<tr>
<td>Inferring Common Origins from mtDNA (Keynote)</td>
<td>Ajay K. Royyuru, Gabriela Alexe, Daniel Platt, Ravi Vijaya-Satya, Laxmi Parida, Saharon Rosset, Gyan Bhanot</td>
<td>246</td>
</tr>
<tr>
<td>Efficient Enumeration of Phylogenetically Informative Substrings</td>
<td>Stanislav Angelov, Boulos Harb, Sampath Kannan, Sanjeev Khanna, Junhyong Kim</td>
<td>248</td>
</tr>
<tr>
<td>Phylogenetic Profiling of Insertions and Deletions in Vertebrate Genomes</td>
<td>Sagi Snir, Lior Pachter</td>
<td>265</td>
</tr>
</tbody>
</table>
Maximal Accurate Forests from Distance Matrices
   Constantinos Daskalakis, Cameron Hill, Alexander 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