Preface

We are pleased to present the proceedings of the 10th Workshop on Algorithms in Bioinformatics (WABI 2010) which took place in Liverpool, UK, September 6–8, 2010. The WABI 2010 workshop was part of the four ALGO 2010 conference meetings, which, in addition to WABI, included ESA, ATMOS, and WAOA. WABI 2010 was hosted by the University of Liverpool Department of Computer Science, and sponsored by the European Association for Theoretical Computer Science (EATCS) and the International Society for Computational Biology (ISCB). See http://algo2010.csc.liv.ac.uk/wabi/ for more details.

The Workshop in Algorithms in Bioinformatics highlights research in algorithmic work for bioinformatics, computational biology and systems biology. The emphasis is mainly on discrete algorithms and machine-learning methods that address important problems in molecular biology, that are founded on sound models, that are computationally efficient, and that have been implemented and tested in simulations and on real datasets. The goal is to present recent research results, including significant work-in-progress, and to identify and explore directions of future research.

Original research papers (including significant work-in-progress) or state-of-the-art surveys were solicited for WABI 2010 in all aspects of algorithms in bioinformatics, computational biology and systems biology. In response to our call, we received 83 submissions for papers and 30 were accepted. In addition, WABI 2010 hosted distinguished lectures by Eran Halperin, of Tel Aviv University and ICSI, Berkeley, and, together with ESA, Paolo Ferragina of University of Pisa. We would like to sincerely thank the authors of all submitted papers and the conference participants. We also thank the Program Committee and their sub-referees for their hard work in reviewing and selecting papers for the workshop.

We would especially like to thank Bernard Moret and Tandy Warnow for all of their advice and support in carrying out the role of being Co-chairs.

Thanks once again to all who participated in making WABI’s 10th anniversary such a success. For us it has been an exciting and rewarding experience.

June 2010

Vincent Moulton
Mona Singh
# Organization

## Program Committee

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tatsuya Akutsu</td>
<td>Kyoto University, Japan</td>
</tr>
<tr>
<td>Bonnie Berger</td>
<td>MIT, USA</td>
</tr>
<tr>
<td>Tanya Berger-Wolf</td>
<td>University of Illinois, USA</td>
</tr>
<tr>
<td>Mathieu Blanchette</td>
<td>McGill University, Canada</td>
</tr>
<tr>
<td>Sebastian Böcker</td>
<td>University of Jena, Germany</td>
</tr>
<tr>
<td>Magnus Bordewich</td>
<td>University of Durham, UK</td>
</tr>
<tr>
<td>Mike Brudno</td>
<td>University of Toronto, Canada</td>
</tr>
<tr>
<td>Philipp Bucher</td>
<td>EPFL, Switzerland</td>
</tr>
<tr>
<td>Benny Chor</td>
<td>Tel Aviv University, Israel</td>
</tr>
<tr>
<td>Anne Condon</td>
<td>University of British Columbia, Canada</td>
</tr>
<tr>
<td>Lenore Cowen</td>
<td>Tufts University, USA</td>
</tr>
<tr>
<td>Keith Crandall</td>
<td>Brigham Young University, USA</td>
</tr>
<tr>
<td>Bhaskar Das Gupta</td>
<td>University of Illinois, USA</td>
</tr>
<tr>
<td>Nadia El-Mabrouk</td>
<td>University of Montreal, Canada</td>
</tr>
<tr>
<td>Liliana Florea</td>
<td>University of Maryland, USA</td>
</tr>
<tr>
<td>Olivier Gascuel</td>
<td>University of Montpellier, France</td>
</tr>
<tr>
<td>Barbara Holland</td>
<td>Massey University, New Zealand</td>
</tr>
<tr>
<td>Katharina Huber</td>
<td>University of East Anglia, UK</td>
</tr>
<tr>
<td>Daniel Huson</td>
<td>University of Tuebingen, Germany</td>
</tr>
<tr>
<td>Lydia Kavraki</td>
<td>Rice University, USA</td>
</tr>
<tr>
<td>Junhyong Kim</td>
<td>University Penn, USA</td>
</tr>
<tr>
<td>Carl Kingsford</td>
<td>University of Maryland, USA</td>
</tr>
<tr>
<td>Mehmet Koyuturk</td>
<td>Case Western, USA</td>
</tr>
<tr>
<td>Jens Lagergren</td>
<td>KTH, Sweden</td>
</tr>
<tr>
<td>Chris Langmead</td>
<td>CMU, USA</td>
</tr>
<tr>
<td>Ryan Lilien</td>
<td>University of Toronto, Canada</td>
</tr>
<tr>
<td>Ion Mandiou</td>
<td>University of Connecticut, USA</td>
</tr>
<tr>
<td>Joao Meidanis</td>
<td>Campinas University, Brazil</td>
</tr>
<tr>
<td>Satoru Miyano</td>
<td>Tokyo University, Japan</td>
</tr>
<tr>
<td>Bernard M.E. Moret</td>
<td>Swiss Federal Institute of Technology, Switzerland</td>
</tr>
<tr>
<td>Burkhard Morgenstern</td>
<td>University of Göttingen, Germany</td>
</tr>
<tr>
<td>Vincent Moulton</td>
<td>University of East Anglia, UK, Co-chair</td>
</tr>
<tr>
<td>Gene W. Myers</td>
<td>Janelia Farms, USA</td>
</tr>
<tr>
<td>Mihai Pop</td>
<td>University of Maryland, USA</td>
</tr>
<tr>
<td>Teresa Przytycka</td>
<td>NIH, USA</td>
</tr>
<tr>
<td>Cenk Sahinalp</td>
<td>Simon Fraser University, USA</td>
</tr>
<tr>
<td>David Sankoff</td>
<td>University of Montréal, Canada</td>
</tr>
</tbody>
</table>
Russell Schwartz  	CMU, USA
Joao Setubal  	Virginia Tech., USA
Mona Singh  	Princeton University, USA, Co-chair
Jens Stoye  	University of Bielefeld, Germany
Glenn Tesler  	UCSD, USA
Olga Vitek  
Purdue University, USA
Lusheng Wang  	City University Hong Kong, Hong Kong
Tandy Warnow  
University of Texas Austin, USA
Chris Workman  
Technical University of Denmark, Denmark
Louxin Zhang  
National University of Singapore, Singapore
## Table of Contents

### Biomolecular Structure: RNA, Protein and Molecular Comparison

- **A Worst-Case and Practical Speedup for the RNA Co-folding Problem Using the Four-Russians Idea** .................................................. 1  
  *Yelena Frid and Dan Gusfield*

- **Sparse Estimation for Structural Variability** ............................ 13  
  *Raghavendra Hosur, Rohit Singh, and Bonnie Berger*

- **Data Structures for Accelerating Tanimoto Queries on Real Valued Vectors** ............................................................ 28  
  *Thomas G. Kristensen and Christian N.S. Pedersen*

- **Sparsification of RNA Structure Prediction Including Pseudoknots** ........ 40  
  *Mathias Möhl, Raheleh Salari, Sebastian Will, Rolf Backofen, and S. Cenk Sahinalp*

- **Prediction of RNA Secondary Structure Including Kissing Hairpin Motifs** .............................................................. 52  
  *Corinna Theis, Stefan Janssen, and Robert Giegerich*

- **Reducing the Worst Case Running Times of a Family of RNA and CFG Problems, Using Valiant’s Approach** .................. 65  
  *Shay Zakov, Dekel Tsur, and Michal Ziv-Ukelson*

### Comparative Genomics

- **Reconstruction of Ancestral Genome Subject to Whole Genome Duplication, Speciation, Rearrangement and Loss** ............ 78  
  *Denis Bertrand, Yves Gagnon, Mathieu Blanchette, and Nadia El-Mabrouk*

- **Genomic Distance with DCJ and Indels** ............................... 90  
  *Marília D.V. Braga, Eyla Willing, and Jens Stoye*

- **Listing All Sorting Reversals in Quadratic Time** .................... 102  
  *Krister M. Swenson, Ghada Badr, and David Sankoff*

### Haplotype and Genotype Analysis

- **Discovering Kinship through Small Subsets** .......................... 111  
  *Daniel G. Brown and Tanya Berger-Wolf*
Table of Contents

Fixed-Parameter Algorithm for Haplotype Inferences on General Pedigrees with Small Number of Sites ........................................... 124  
Duong D. Doan and Patricia A. Evans

Haplotypes versus Genotypes on Pedigrees .................................. 136  
Bonnie Kirkpatrick

Haplotype Inference on Pedigrees with Recombinations and Mutations ................................................................. 148  
Yuri Pirola, Paola Bonizzoni, and Tao Jiang

High-throughput Data Analysis: Next Generation Sequencing and Flow Cytometry

Identifying Rare Cell Populations in Comparative Flow Cytometry ...... 162  
Ariful Azad, Johannes Langguth, Youhan Fang, Alan Qi, and Alex Pothen

Fast Mapping and Precise Alignment of AB SOLiD Color Reads to Reference DNA .................................................. 176  
Miklós Csúrós, Szilveszter Juhas, and Attila Bérces

Design of an Efficient Out-of-Core Read Alignment Algorithm ........... 189  
Arun S. Konagurthu, Lloyd Allison, Thomas Conway, Bryan Beresford-Smith, and Justin Zobel

Estimation of Alternative Splicing Isoform Frequencies from RNA-Seq Data .......................................................... 202  
Marius Nicolae, Serghei Mangul, Ion Mândoiu, and Alex Zelikovsky

Networks

Improved Orientations of Physical Networks .................................. 215  
Iftah Gamzu, Danny Segev, and Roded Sharan

Enumerating Chemical Organisations in Consistent Metabolic Networks: Complexity and Algorithms ........................................ 226  
Paulo Vieira Milreu, Vicente Acuña, Etienne Birmelé, Pierluigi Crescenzi, Alberto Marchetti-Spaccamela, Marie-France Sagot, Leen Stougie, and Vincent Lacroix

Efficient Subgraph Frequency Estimation with G-Tries .................... 238  
Pedro Ribeiro and Fernando Silva
# Phylogenetics

Accuracy Guarantees for Phylogeny Reconstruction Algorithms Based on Balanced Minimum Evolution .............................................. 250  
*Magnus Bordewich and Radu Mihaescu*

The Complexity of Inferring a Minimally Resolved Phylogenetic Supertree .................................................................................. 262  
*Jesper Jansson, Richard S. Lemence, and Andrzej Lingas*

Reducing Multi-state to Binary Perfect Phylogeny with Applications to Missing, Removable, Inserted, and Deleted Data ................. 274  
*Kristian Stevens and Dan Gusfield*

An Experimental Study of Quartets MaxCut and Other Supertree Methods .................................................................................... 288  
*M. Shel Swenson, Rahul Suri, C. Randal Linder, and Tandy Warnow*

An Efficient Method for DNA-Based Species Assignment via Gene Tree and Species Tree Reconciliation ........................................ 300  
*Louxin Zhang and Yun Cui*

## Sequences, Strings and Motifs

Effective Algorithms for Fusion Gene Detection ................................. 312  
*Dan He and Eleazar Eskin*

Swiftly Computing Center Strings ..................................................... 325  
*Franziska Hufsky, Léon Kuchenbecker, Katharina Jahn, Jens Stoye, and Sebastian Böcker*

Speeding Up Exact Motif Discovery by Bounding the Expected Clump Size .................................................................................. 337  
*Tobias Marschall and Sven Rahmann*

Pair HMM Based Gap Statistics for Re-evaluation of Indels in Alignments with Affine Gap Penalties .............................................. 350  
*Alexander Schönhuth, Raheleh Salari, and S. Cenk Sahinalp*

Quantifying the Strength of Natural Selection of a Motif Sequence ...... 362  
*Chen-Hsiang Yeang*

## Author Index

.......................................................... 375