

# DIMACS

Series in Discrete Mathematics  
and Theoretical Computer Science

---

Volume 61

## Bioconsensus

DIMACS Working Group Meetings on  
Bioconsensus.

October 25–26, 2000 and

October 2–5, 2001

DIMACS Center

M. F. Janowitz

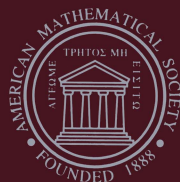
F.-J. Lapointe

F. R. McMorris

B. Mirkin

F. S. Roberts

Editors



---

American Mathematical Society

# Bioconsensus

*This page intentionally left blank*

# DIMACS

Series in Discrete Mathematics  
and Theoretical Computer Science

---

Volume 61

## Bioconsensus

DIMACS Working Group Meetings on  
Bioconsensus  
October 25–26, 2000 and  
October 2–5, 2001  
DIMACS Center

M. F. Janowitz  
F.-J. Lapointe  
F. R. McMorris  
B. Mirkin  
F. S. Roberts  
Editors

Center for Discrete Mathematics  
and Theoretical Computer Science  
A consortium of Rutgers University, Princeton University,  
AT&T Labs–Research, Bell Labs (Lucent Technologies),  
NEC Laboratories America, and Telcordia Technologies  
(with partners at Avaya Labs, IBM Research, and Microsoft Research)



---

American Mathematical Society

This DIMACS volume contains papers from working group meetings on Bioconsensus, held October 25–26, 2000 and October 2–5, 2001.

2000 *Mathematics Subject Classification*. Primary 05C05, 05C65, 06A99, 62H30, 91B10, 91F99, 92B05, 92B10, 92C40, 92D15.

---

#### Library of Congress Cataloging-in-Publication Data

Bioconsensus / M. F. Janowitz... [et al.], editors.

p. cm. — (DIMACS series in discrete mathematics and theoretical computer science, ISSN 1052-1798 ; v. 61)

Papers from two workshops held Oct. 25–26, 2000 and Oct. 2–5, 2001. With additional papers. Includes bibliographical references.

ISBN 0-8218-3197-6 (alk. paper)

1. Biomathematics. 2. Trees (Graph theory) I. Janowitz, M. F. (Melvin F.), 1929– II. DIMACS (Group) III. Series.

QH323.5.B454 2003  
570'.15'1—dc21

2003040387

---

**Copying and reprinting.** Material in this book may be reproduced by any means for educational and scientific purposes without fee or permission with the exception of reproduction by services that collect fees for delivery of documents and provided that the customary acknowledgment of the source is given. This consent does not extend to other kinds of copying for general distribution, for advertising or promotional purposes, or for resale. Requests for permission for commercial use of material should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294, USA. Requests can also be made by e-mail to [reprint-permission@ams.org](mailto:reprint-permission@ams.org).

Excluded from these provisions is material in articles for which the author holds copyright. In such cases, requests for permission to use or reprint should be addressed directly to the author(s). (Copyright ownership is indicated in the notice in the lower right-hand corner of the first page of each article.)

© 2003 by the American Mathematical Society. All rights reserved.  
The American Mathematical Society retains all rights  
except those granted to the United States Government.  
Printed in the United States of America.

∞ The paper used in this book is acid-free and falls within the guidelines  
established to ensure permanence and durability.  
Visit the AMS home page at <http://www.ams.org/>

10 9 8 7 6 5 4 3 2 1 08 07 06 05 04 03

## Contents

Foreword	vii
Preface	ix

### Part I. Axiomatic Considerations

Axiomatics in group choice and bioconsensus WILLIAM H. E. DAY AND F. R. MCMORRIS	3
The Arrovian program from weak orders to hierarchical and tree-like relations F. R. MCMORRIS AND R. C. POWERS	37
Consensus $n$ -trees, weak independence, and veto power R. C. POWERS	47
The size of a maximum agreement subtree for random binary trees DAVID BRYANT, ANDY MCKENZIE, AND MIKE STEEL	55
An injective set representation of closed systems of sets G. D. CROWN AND M. F. JANOWITZ	67

### Part II. Data Analysis Considerations

Consensus list colorings of graphs and physical mapping of DNA N. V. R. MAHADEV AND FRED S. ROBERTS	83
A top-down method for building genome classification trees with linear binary hierarchies BORIS MIRKIN AND EUGENE KOONIN	97
An application of seriation to agent development consensus: A genetic algorithm approach MICHAEL L. GARGANO, WILLIAM EDELSON, AND JOSEPH DECICCO	113
Achieving consensus of long genomic sequences with the $W$ -curve DOUGLAS J. CORK	123
Flipping: A supertree construction method D. CHEN, L. DIAO, O. EULENSTEIN, D. FERNÁNDEZ-BACA, AND M. SANDERSON	135

**Part III. Practical Considerations**

A classification of consensus methods for phylogenetics DAVID BRYANT	163
A view of supertree methods JOSEPH L. THORLEY AND MARK WILKINSON	185
Reduced consensus MARK WILKINSON AND JOSEPH L. THORLEY	195
How good can a consensus get? Assessing the reliability of consensus trees in phylogenetic studies FRANÇOIS-JOSEPH LAPOINTE AND GUY CUCUMEL	205
Increasing phylogenetic accuracy with global congruence CLAUDINE LEVASSEUR AND FRANÇOIS-JOSEPH LAPOINTE	221
MRP supertree construction in the consensus setting OLAF R. P. BININDA-EMONDS	231

## Foreword

This volume is based on two DIMACS working group meetings on “Bioconsensus” held on October 25–26, 2000 and October 2–5, 2001. We would like to express our appreciation to Mel Janowitz, François-Joseph Lapointe, Fred R. McMorris, Boris Mirkin and Fred Roberts for their efforts to organize and plan these two successful meetings.

The meetings were part of the Special Focus on Computational Molecular Biology. We extend our thanks to Martin Farach-Colton, Craig Nevill-Manning, Wilma Olson and Fred Roberts for their work as special focus organizers.

A “working group” is a new concept at DIMACS, and the Bioconsensus meetings constituted our very first effort at organizing such an activity. A working group is typically interdisciplinary, as was this one. Its meetings stress informal presentations (with a few more formal ones mixed in, possibly an associated “public workshop”) and a lot of time for interaction. The group or subgroups come back to DIMACS for further meetings, discussions, and collaborations. This particular working group emphasized the uses in the biological sciences of methods of consensus and social choice originally developed mostly by mathematical social scientists. Because of the biological content of the meeting and the level of activity in the field, emphasis was placed on phylogenetic trees. Attention was also given to the role that ideas from bioconsensus might apply back to problems in the social sciences.

The group brought together mathematical scientists (mathematicians, computer scientists, statisticians) and biological scientists for a pair of lively meetings that have come a long way toward establishing the new field of “bioconsensus.” This volume, however, is more than just a report of two meetings. In particular, it includes some historical background, as well as a substantial introduction to the axiomatic foundations of the field of bioconsensus, some practical applications of consensus methods to real data, and it includes papers by some individuals who were not present at the group meetings.

DIMACS gratefully acknowledges the generous support that makes these programs possible. The special focus on Computational Molecular Biology has been primarily supported by the Alfred P. Sloan Foundation and the National Science Foundation, with supplementary support from the New Jersey Commission on Science and Technology, Celera Genomics, SmithKline Beecham, and IBM, and was also supported by DIMACS’s partners at Rutgers, Princeton, AT&T Labs-Research, Bell Labs, NEC Laboratories America, and Telcordia Technologies.

Fred S. Roberts  
Director  
Robert Tarjan  
Co-Director for Princeton



*This page intentionally left blank*

## Preface

Consensus methods developed in the context of voting, decision making, and other areas of the social and behavioral sciences have begun to have a variety of applications in the biological sciences, originally in taxonomy and evolutionary biology, and more recently in molecular biology. Typically, several alternatives (such as alternative phylogenetic trees, molecular sequences, or alignments) are produced using different methods or under different models and then one needs to find a consensus solution. There are, already, several hundred papers in this developing field of “Bioconsensus.” We hope that this volume will provide a valuable introduction and reference to various aspects of the subject.

This volume has its genesis in two working group meetings on Bioconsensus that were held at DIMACS on October 25–26, 2000 and October 2–5, 2001. The goals of the meetings were to gather together a diverse group of researchers from Mathematics, Statistics, Biology, and Computer Science with the idea of exploring which biological concepts could benefit from concepts and methods of consensus theory, as well as the extent to which consensus methods from social choice theory might apply to problems in evolutionary biology and other areas of the biological sciences. The formal sessions were kept to a minimum so as to allow ample time for discussions and actual research. We should mention, however, that this volume is more than just a Proceedings of two workshops. It includes invited papers that were not part of the working group meetings, as well as expansions of talks that were presented at the two meetings. It is anticipated that the two survey papers will provide a valuable guide to the already vast literature in this area.

Issues such as the following were typical of those that were discussed: computational complexity of various consensus algorithms, the role of supertrees in evolutionary biology, the extent to which techniques from social choice theory can be modified to be useful in molecular biology, study of axiomatic properties of consensus methods, the importance of choosing consensus methods based on biological rather than mathematical properties, the extent to which consensus methods even make sense in connection with the reconstruction of evolutionary history.

There are a total of 16 papers in the volume. They are difficult to classify by area of author expertise because many of them are joint papers written by authors from different fields. We chose instead to divide them into three general areas. The section on axiomatic considerations is more theoretical than the rest of the volume, and contains five papers. The papers leaning toward computer science approaches can be found in the next section on data analysis, and consists of another five papers. The final part deals more with phylogenetic techniques and has in it the remaining six papers.

The working group meetings that formed the basis for this volume were part of the DIMACS 2000–2003 Special Focus on Computational Molecular Biology. Funding was provided by the Alfred P. Sloan Foundation, Celera Genomics, the National Science Foundation, the New Jersey Commission on Science and Technology, IBM, and SmithKline Beecham. The editors wish to express their thanks for this support.

Melvin F. Janowitz  
François-Joseph Lapointe  
Fred R. McMorris  
Boris Mirkin  
Fred S. Roberts

## Titles in This Series

- 61 **M. F. Janowitz, F.-J. Lapointe, F. R. McMorris, B. Mirkin, and F. S. Roberts, Editors**, Bioconsensus
- 60 **Saugata Basu and Laureano Gonzalez-Vega, Editors**, Algorithmic and Quantitative Real Algebraic Geometry
- 59 **Michael H. Goldwasser, David S. Johnson, and Catherine C. McGeoch, Editors**, Data Structures, Near Neighbor Searches, and Methodology: Fifth and Sixth DIMACS Implementation Challenges
- 58 **Simon Thomas, Editor**, Set Theory: The Hajnal Conference
- 57 **Eugene C. Freuder and Richard J. Wallace, Editors**, Constraint Programming and Large Scale Discrete Optimization
- 56 **Alexander Barg and Simon Litsyn, Editors**, Codes and Association Schemes
- 55 **Ding-Zhu Du, Panos M. Pardalos, and Jie Wang, Editors**, Discrete Mathematical Problems with Medical Applications
- 54 **Erik Winfree and David K. Gifford, Editors**, DNA Based Computers V
- 53 **Nathaniel Dean, D. Frank Hsu, and R. Ravi, Editors**, Robust Communication Networks: Interconnection and Survivability
- 52 **Sanguthevar Rajasekaran, Panos Pardalos, and D. Frank Hsu, Editors**, Mobile Networks and Computing
- 51 **Pierre Hansen, Patrick Fowler, and Maolin Zheng, Editors**, Discrete Mathematical Chemistry
- 50 **James M. Abello and Jeffrey Scott Vitter, Editors**, External Memory Algorithms
- 49 **Ronald L. Graham, Jan Kratochvíl, Jaroslav Nešetřil, and Fred S. Roberts, Editors**, Contemporary Trends in Discrete Mathematics
- 48 **Harvey Rubin and David Harlan Wood, Editors**, DNA Based Computers III
- 47 **Martin Farach-Colton, Fred S. Roberts, Martin Vingron, and Michael Waterman, Editors**, Mathematical Support for Molecular Biology
- 46 **Peng-Jun Wan, Ding-Zhu Du, and Panos M. Pardalos, Editors**, Multichannel Optical Networks: Theory and Practice
- 45 **Marios Mavronicolas, Michael Merritt, and Nir Shavit, Editors**, Networks in Distributed Computing
- 44 **Laura F. Landweber and Eric B. Baum, Editors**, DNA Based Computers II
- 43 **Panos Pardalos, Sanguthevar Rajasekaran, and José Rolim, Editors**, Randomization Methods in Algorithm Design
- 42 **Ding-Zhu Du and Frank K. Hwang, Editors**, Advances in Switching Networks
- 41 **David Aldous and James Propp, Editors**, Microsurveys in Discrete Probability
- 40 **Panos M. Pardalos and Dingzhu Du, Editors**, Network Design: Connectivity and Facilities Location
- 39 **Paul W. Beame and Samuel R. Buss, Editors**, Proof Complexity and Feasible Arithmetics
- 38 **Rebecca N. Wright and Peter G. Neumann, Editors**, Network Threats
- 37 **Boris Mirkin, F. R. McMorris, Fred S. Roberts, and Andrey Rzhetsky, Editors**, Mathematical Hierarchies and Biology
- 36 **Joseph G. Rosenstein, Deborah S. Franzblau, and Fred S. Roberts, Editors**, Discrete Mathematics in the Schools
- 35 **Dingzhu Du, Jun Gu, and Panos M. Pardalos, Editors**, Satisfiability Problem: Theory and Applications
- 34 **Nathaniel Dean, Editor**, African Americans in Mathematics
- 33 **Ravi B. Boppana and James F. Lynch, Editors**, Logic and Random Structures

TITLES IN THIS SERIES

- 32 **Jean-Charles Grégoire, Gerard J. Holzmann, and Doron A. Peled, Editors,** The SPIN Verification System
- 31 **Neil Immerman and Phokion G. Kolaitis, Editors,** Descriptive Complexity and Finite Models
- 30 **Sandeep N. Bhatt, Editor,** Parallel Algorithms: Third DIMACS Implementation Challenge
- 29 **Doron A. Peled, Vaughan R. Pratt, and Gerard J. Holzmann, Editors,** Partial Order Methods in Verification
- 28 **Larry Finkelstein and William M. Kantor, Editors,** Groups and Computation II
- 27 **Richard J. Lipton and Eric B. Baum, Editors,** DNA Based Computers
- 26 **David S. Johnson and Michael A. Trick, Editors,** Cliques, Coloring, and Satisfiability: Second DIMACS Implementation Challenge
- 25 **Gilbert Baumslag, David Epstein, Robert Gilman, Hamish Short, and Charles Sims, Editors,** Geometric and Computational Perspectives on Infinite Groups
- 24 **Louis J. Billera, Curtis Greene, Rodica Simion, and Richard P. Stanley, Editors,** Formal Power Series and Algebraic Combinatorics/Séries Formelles et Combinatoire Algébrique, 1994
- 23 **Panos M. Pardalos, David I. Shalloway, and Guoliang Xue, Editors,** Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding
- 22 **Panos M. Pardalos, Mauricio G. C. Resende, and K. G. Ramakrishnan, Editors,** Parallel Processing of Discrete Optimization Problems
- 21 **D. Frank Hsu, Arnold L. Rosenberg, and Dominique Sotteau, Editors,** Interconnection Networks and Mapping and Scheduling Parallel Computations
- 20 **William Cook, László Lovász, and Paul Seymour, Editors,** Combinatorial Optimization
- 19 **Ingemar J. Cox, Pierre Hansen, and Bela Julesz, Editors,** Partitioning Data Sets
- 18 **Guy E. Blelloch, K. Mani Chandy, and Suresh Jagannathan, Editors,** Specification of Parallel Algorithms
- 17 **Eric Sven Ristad, Editor,** Language Computations
- 16 **Panos M. Pardalos and Henry Wolkowicz, Editors,** Quadratic Assignment and Related Problems
- 15 **Nathaniel Dean and Gregory E. Shannon, Editors,** Computational Support for Discrete Mathematics
- 14 **Robert Calderbank, G. David Forney, Jr., and Nader Moayeri, Editors,** Coding and Quantization: DIMACS/IEEE Workshop
- 13 **Jin-Yi Cai, Editor,** Advances in Computational Complexity Theory
- 12 **David S. Johnson and Catherine C. McGeoch, Editors,** Network Flows and Matching: First DIMACS Implementation Challenge
- 11 **Larry Finkelstein and William M. Kantor, Editors,** Groups and Computation
- 10 **Joel Friedman, Editor,** Expanding Graphs
- 9 **William T. Trotter, Editor,** Planar Graphs
- 8 **Simon Gindikin, Editor,** Mathematical Methods of Analysis of Biopolymer Sequences
- 7 **Lyle A. McGeoch and Daniel D. Sleator, Editors,** On-Line Algorithms
- 6 **Jacob E. Goodman, Richard Pollack, and William Steiger, Editors,** Discrete and Computational Geometry: Papers from the DIMACS Special Year
- 5 **Frank Hwang, Fred Roberts, and Clyde Monma, Editors,** Reliability of Computer and Communication Networks

Consensus methods developed in the context of voting, decision making, and other areas of the social and behavioral sciences have a variety of applications in the biological sciences, originally in taxonomy and evolutionary biology, and more recently in molecular biology. Typically, several alternatives (such as alternative phylogenetic trees, molecular sequences, or alignments) are produced using different methods or under different models, and then one needs to find a consensus solution.

This volume is based on two DIMACS meetings on "Bioconsensus". It provides a valuable introduction and reference to the various aspects of this rapidly developing field.

The book includes some historical background, as well as a substantial introduction to the axiomatic foundations of the field of bioconsensus and some practical applications of consensus methods to real data. Also included are contributed papers from experts who were not at the meetings. The book is intended for mathematical biologists, evolutionary biologists, and computer scientists.

ISBN 0-8218-3197-6



9 780821 831977

DIMACS/61

AMS *on the Web*  
[www.ams.org](http://www.ams.org)