

METHODS IN MOLECULAR BIOLOGY

Series Editor

John M. Walker

School of Life and Medical Sciences

University of Hertfordshire

Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:

<http://www.springer.com/series/7651>

Evolutionary Genomics

Statistical and Computational Methods

Second Edition

Edited by

Maria Anisimova

*Institute of Applied Simulations, School of Life Sciences and Facility Management, Zurich University
of Applied Sciences (ZHAW), Wädenswil, Switzerland*

Swiss Institute of Bioinformatics, Lausanne, Switzerland

OPEN



Humana Press

Editor

Maria Anisimova
Institute of Applied Simulations
School of Life Sciences and Facility Management
Zurich University of Applied Sciences (ZHAW)
Wädenswil, Switzerland

Swiss Institute of Bioinformatics
Lausanne, Switzerland



ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-9073-3 ISBN 978-1-4939-9074-0 (eBook)
<https://doi.org/10.1007/978-1-4939-9074-0>

This book is an open access publication.

© The Editor(s) (if applicable) and The Author(s) 2012, 2019.

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Humana imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.

The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

This volume is a thoroughly revised second edition of *Evolutionary Genomics: Statistical and Computational Methods* published in 2012. Like the first edition, the new volume includes comprehensive reviews of the most recent and fundamental developments in bioinformatics methods for evolutionary genomics and related challenges associated with increasing data size, heterogeneity, and its inherent complexity.

Throughout the volume, prominent authors address the challenge of analyzing and understanding the dynamics of complex biological systems, and elaborate on some promising strategies that would bring us closer to the ultimate “holy grail” of biology—uncovering of the relationships between genotype and phenotype. Consequently, the presented collection of peer-reviewed articles also represents a synergy between theoretical and experimental scientists from a range of disciplines, working together towards a common goal. Once again, the revised volume reiterates the power of taking an evolutionary approach to study molecular data.

This book is intended for scientists looking for a compact overview of the cutting-edge statistical and computational methods in evolutionary genomics. The volume may serve as a comprehensive guide for both graduate and advanced undergraduate students planning to specialize in genomics and bioinformatics. Equally, the volume should be helpful for experienced researchers entering genomics from more fundamental disciplines, such as statistics, computer science, physics, and biology. In other words, the material presented here should suit both a novice in biology with strong statistics and computational skills and a molecular biologist with a good grasp of standard mathematical concepts. To cater to differences in reader backgrounds, *Part I* is composed of educational primers to help with fundamental concepts in genome biology (Chapter 1), probability and statistics (Chapter 2), and molecular evolution (Chapter 3). As these concepts reappear repeatedly throughout the book, the first three chapters will help the neophyte to stay “afloat”. The exercises and questions offered at the end of each chapter serve to deepen the understanding of the material.

Part II of this volume focuses on sequence homology and alignment—from aligning whole genomes (Chapter 4) to disentangling orthologs, paralogs, and transposable elements (Chapters 5 and 6). *Part III* includes chapters on phylogenetic methods to study genome evolution. Chapter 7 presents multispecies coalescent methods for reconciling phylogenetic discord between gene and species trees. However, a mathematically convenient “binary tree” model does not always live up to scrutiny as numerous evolutionary processes act in reticulate (network-like) fashion, complicating the statistical description of evolutionary models and increasing computational complexity, often to prohibitive levels. One simplification is to assume that some molecular sequence units (genes, gene segments) still evolve in a treelike manner. If so, Chapter 8 describes one practical approach to meaningfully summarize the binary tree distributions for a set of genomes as a “forest of trees”. Alternatively network-like phylogenetic relationships can be represented by graphs (Chapter 9). Dating methods for genome-scale data are discussed in Chapter 10, while Chapter 11 provides more examples of non-treelike processes in a comparative review of genome evolution in different breeding systems.

By disentangling different evolutionary forces acting on genomes, we hope to understand the origins of biological innovation, which is often thought to be coupled with natural selection. After all, how do we explain that, by the words of Darwin, “from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved”? This is the main topic of *Part IV* that discusses the methodology for evaluating selective pressures on genomic sequences (Chapters 12–14) and genomic evolution in light of protein domain architecture and transposable elements (Chapters 15 and 16). *Part V* of this book is dedicated to population genomics and other omics, with example applications to disease. Indeed, as evolution starts in populations, there is much interest in generating and studying population genome data for a wide range of species. Chapter 17 discusses models for genetic architectures of complex disease and genome-wide association studies for finding susceptibility variants. Chapter 18 reviews approaches to study ancestral population genomics. Chapters 19, 20 and 21 illustrate first principles of analyzing environmental sequences and applications to clinical trials and systems genetics. Finally, *Part VI* concludes the book by discussing current bottlenecks in handling and analyzing genomic data. Chapter 22 focuses on challenges and approaches for large and complex data representation and simultaneous querying of heterogeneous databases. Chapter 23 makes the case for using efficient high-performance computing strategies for computationally demanding phylogenetic analyses, in particular in the Bayesian framework. Solutions for scalable workflows and sharing programming resources are presented in Chapters 24 and 25.

On behalf of all authors, I hope that this book will become a source of inspiration and new ideas for our readers. Wishing you a pleasant reading!

*Wädenswil, Switzerland
Lausanne, Switzerland*

Maria Anisimova

Acknowledgements

This renewed edition of *Evolutionary Genomics: Statistical and Computational Methods* is a result of a dedicated effort by 94 co-authors of the book representing research institutions from nearly two dozen different countries. Special thanks go to almost 50 independent reviewers whose constructive and detailed comments have greatly contributed to improving the overall quality of the book chapters and the clarity of the presentation. As for the first edition of this book, the cover image was made by the author of Chapter 6 and a talented photography artist, Wojciech Makałowski, from the University of Münster, Germany.

By a mutual agreement between all authors of the book, all chapters are available *Open Access*. Swiss Institute of Bioinformatics (SIB) and Zurich University of Applied Sciences (ZHAW) have generously contributed to cover a part of the Open Access publication fees. Finally, I would like to thank my colleagues at the Institute of Applied Simulations and the School of Life Sciences and Facility Management of ZHAW (Zurich University of Applied Sciences) as well as my family for their support and encouragement.

Contents

<i>Preface</i>	<i>v</i>
<i>Acknowledgements</i>	<i>vii</i>
<i>Contributors</i>	<i>xiii</i>

PART I INTRODUCTION: BIOINFORMATICIAN’S PRIMERS

1 Introduction to Genome Biology and Diversity	3
<i>Noor Youssef, Aidan Budd, and Joseph P. Bielawski</i>	
2 Probability, Statistics, and Computational Science	33
<i>Niko Beerenwinkel and Juliane Siebourg</i>	
3 A Not-So-Long Introduction to Computational Molecular Evolution	71
<i>Stéphane Aris-Brosou and Nicolas Rodrigue</i>	

PART II GENOMIC ALIGNMENT AND HOMOLOGY INFERENCE

4 Whole-Genome Alignment	121
<i>Colin N. Dewey</i>	
5 Inferring Orthology and Paralogy	149
<i>Adrian M. Altenhoff, Natasha M. Glover, and Christophe Dessimoz</i>	
6 Transposable Elements: Classification, Identification, and Their Use As a Tool For Comparative Genomics	177
<i>Wojciech Makalowski, Valer Gotea, Amit Pande, and Izabela Makalowska</i>	

PART III PHYLOGENOMICS AND GENOME EVOLUTION

7 Modern Phylogenomics: Building Phylogenetic Trees Using the Multispecies Coalescent Model	211
<i>Liang Liu, Christian Anderson, Dennis Pearl, and Scott V. Edwards</i>	
8 Genome-Wide Comparative Analysis of Phylogenetic Trees: The Prokaryotic Forest of Life	241
<i>Pere Puigbò, Yuri I. Wolf, and Eugene V. Koonin</i>	
9 The Methodology Behind Network Thinking: Graphs to Analyze Microbial Complexity and Evolution	271
<i>Andrew K. Watson, Romain Lannes, Jananan S. Pathmanathan, Raphaël Méheust, Slim Karkar, Philippe Colson, Eduardo Corel, Philippe Lopez, and Eric Baptiste</i>	
10 Bayesian Molecular Clock Dating Using Genome-Scale Datasets	309
<i>Mariodos Reis and Ziheng Yang</i>	
11 Genome Evolution in Outcrossing vs. Selfing vs. Asexual Species	331
<i>Sylvain Glémin, Clémentine M. François, and Nicolas Galtier</i>	

PART IV NATURAL SELECTION AND INNOVATION IN GENOMIC SEQUENCES

12 Selection Acting on Genomes 373
Carolin Kosiol and Maria Anisimova

13 Looking for Darwin in Genomic Sequences: Validity and Success
 Depends on the Relationship Between Model and Data 399
Christopher T. Jones, Edward Susko, and Joseph P. Bielawski

14 Evolution of Viral Genomes: Interplay Between Selection,
 Recombination, and Other Forces 427
*Stephanie J. Spielman, Steven Weaver, Stephen D. Shank,
 Brittany Rife Magalis, Michael Li, and Sergei L. Kosakovsky Pond*

15 Evolution of Protein Domain Architectures 469
Sofia K. Forslund, Mateusz Kaduk, and Erik L. L. Sonnhammer

16 New Insights on the Evolution of Genome Content: Population
 Dynamics of Transposable Elements in Flies and Humans 505
Lain Guio and Josefa González

PART V POPULATION GENOMICS AND OMICS IN LIGHT OF DISEASE
 AND EVOLUTION

17 Association Mapping and Disease: Evolutionary Perspectives 533
*Søren Besenbacher, Thomas Mailund, Bjarni J. Vilhjálmsson,
 and Mikkel H. Schierup*

18 Ancestral Population Genomics 555
Julien Y. Dutheil and Asger Hobolth

19 Introduction to the Analysis of Environmental Sequences:
 Metagenomics with MEGAN 591
Caner Başçı, Sina Beier, Anna Górski, and Daniel H. Huson

20 Multiple Data Analyses and Statistical Approaches
 for Analyzing Data from Metagenomic Studies and Clinical Trials 605
Suparna Mitra

21 Systems Genetics for Evolutionary Studies 635
*Pjotr Prins, Geert Smant, Danny Arends, Megan K. Mulligan,
 Rob W. Williams, and Ritsert C. Jansen*

PART VI HANDLING GENOMIC DATA: RESOURCES AND COMPUTATION

22 Semantic Integration and Enrichment of Heterogeneous
 Biological Databases 655
*Ana Claudia Sima, Kurt Stockinger, Tarcisio Mendes de Farias,
 and Manuel Gil*

23 High-Performance Computing in Bayesian Phylogenetics
 and Phylodynamics Using BEAGLE 691
*Guy Baele, Daniel L. Ayres, Andrew Rambaut,
 Marc A. Suchard, and Philippe Lemey*

24 Scalable Workflows and Reproducible Data Analysis for Genomics 723
*Francesco Strozzi, Roel Janssen, Ricardo Wurmus, Michael R. Crusoe,
George Githinji, Paolo Di Tommaso, Dominique Belbachemi,
Steffen Möller, Geert Smant, Joepde Ligt, and Pjotr Prins*

25 Sharing Programming Resources Between Bio* Projects 747
*Raoul J. P. Bonnal, Andrew Yates, Naohisa Goto, Laurent Gautier,
Scooter Willis, Christopher Fields, Toshiaki Katayama, and Pjotr Prins*

Index 767

Contributors

- ADRIAN M. ALTENHOFF • *Computer Science Department, ETH Zurich, Zurich, Switzerland; Swiss Institute of Bioinformatics, Lausanne, Switzerland*
- CHRISTIAN ANDERSON • *Advantage Testing of Boston, Newton Centre, MA, USA*
- MARIA ANISIMOVA • *Institute of Applied Simulation, School of Life Sciences and Facility Management, Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland; Swiss Institute of Bioinformatics, Lausanne, Switzerland*
- DANNY ARENDS • *Animal Breeding Biology and Molecular Genetics, Albrecht Daniel Thaer-Institute for Agricultural and Horticultural Sciences, Humboldt University zu Berlin, Berlin, Germany*
- STÉPHANE ARIS-BROSOU • *Department of Biology, University of Ottawa, Ottawa, ON, Canada; Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada*
- DANIEL L. AYRES • *Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD, USA*
- GUY BAELE • *Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium*
- CANER BAĞCI • *Algorithms in Bioinformatics, Faculty of Computer Science, University of Tübingen, Tübingen, Germany*
- ERIC BAPTESTE • *Sorbonne Universités, Institut de Biologie Paris-Seine, UPMC Université Paris 6, Paris, France*
- NIKO BEERENWINKEL • *Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland*
- SINA BEIER • *Algorithms in Bioinformatics, Faculty of Computer Science, University of Tübingen, Tübingen, Germany*
- DOMINIQUE BELHACHEMI • *Life Technologies, Waltham, MA, USA*
- SØREN BESENBACHER • *Department of Clinical Medicine (MOMA), Aarhus University, Aarhus, Denmark*
- JOSEPH P. BIELAWSKI • *Department of Biology, Dalhousie University, Halifax, NS, Canada; Department of Mathematics & Statistics, Dalhousie University, Halifax, NS, Canada*
- RAOUL J. P. BONNAL • *Istituto Nazionale Genetica Molecolare INGM Romeo ed Enrica Invernizzi, Milan, Italy*
- AIDAN BUDD • *Structural and Computational Biology (SCB) Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany*
- PHILIPPE COLSON • *Fondation Institut Hospitalo-Universitaire Méditerranée Infection, Pôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie-Hygiène-Virologie, Centre Hospitalo-Universitaire Tione, Assistance Publique-Hôpitaux de Marseille, Marseille, France; Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE) UM63, CNRS 7278, IRD 198, INSERM U1095, Aix-Marseille University, Marseille, France*
- EDUARDO COREL • *Sorbonne Universités, Institut de Biologie Paris-Seine, UPMC Université Paris 6, Paris, France*
- MICHAEL R. CRUSOE • *Common Workflow Language Project, Vilnius, Lithuania*

- TARCISIO MENDES DE FARIAS • *University of Lausanne, Lausanne, Switzerland; SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland*
- JOEP DE LIGT • *Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands*
- CHRISTOPHE DESSIMOZ • *Swiss Institute of Bioinformatics, Lausanne, Switzerland; Department of Computational Biology, University of Lausanne, Lausanne, Switzerland; Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; Department of Genetics, Evolution and Environment, University College London, London, UK; Department of Computer Science, University College London, London, UK*
- COLIN N. DEWEY • *Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA*
- PAOLO DI TOMMASO • *Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, Spain*
- MARIO DOS REIS • *School of Biological and Chemical Sciences, Queen Mary University of London, London, UK*
- JULIEN Y. DUTHEIL • *Department of Evolutionary Genetics, Max Planck Institute of Evolutionary Biology, Plön, Germany*
- PETER EBERT • *Max Planck Institute for Informatics, Saarbrücken, Saarland, Germany*
- SCOTT V. EDWARDS • *Department of Organismic and Evolutionary Biology & Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA*
- CHRISTOPHER FIELDS • *Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA*
- SOFIA K. FORSLUND • *EMBL Heidelberg, Heidelberg, Germany; Max Delbrück Centre for Molecular Medicine, Berlin, Germany*
- CLÉMENTINE M. FRANÇOIS • *Institut des Sciences de l'Evolution, UMR5554, Université Montpellier II, Montpellier, France*
- NICOLAS GALTIER • *Institut des Sciences de l'Evolution, UMR5554, Université Montpellier II, Montpellier, France*
- LAURENT GAUTIER • *DMAC, Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kongens Lyngby, Denmark*
- MANUEL GIL • *ZHAW Zurich University of Applied Sciences, Winterthur, Switzerland; SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland*
- GEORGE GITHINJI • *KEMRI Wellcome Trust Research Programme, Kilifi, Kenya*
- SYLVAIN GLÉMIN • *Institut des Sciences de l'Evolution, UMR5554, Université Montpellier II, Montpellier, France*
- NATASHA M. GLOVER • *Swiss Institute of Bioinformatics, Lausanne, Switzerland; Department of Computational Biology, University of Lausanne, Lausanne, Switzerland; Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland*
- JOSEFA GONZÁLEZ • *Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain*
- ANNA GÓRSKA • *Algorithms in Bioinformatics, Faculty of Computer Science, University of Tübingen, Tübingen, Germany*
- VALER GOTEA • *National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA*
- NAOHISA GOTO • *Department of Genome Informatics, Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan*
- LAIN GUIO • *Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain*

- ASGER HOBOLTH • *Bioinformatics Research Center (BiRC), Aarhus University, Aarhus, Denmark*
- DANIEL H. HUSON • *Algorithms in Bioinformatics, Faculty of Computer Science, University of Tübingen, Tübingen, Germany*
- RITSERT C. JANSSEN • *Groningen Bioinformatics Centre, GBB, University of Groningen, Groningen, Netherlands*
- ROEL JANSSEN • *Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands*
- CHRISTOPHER T. JONES • *Department of Mathematics and Statistics, Dalhousie University, Halifax, NS, Canada*
- MATEUSZ KADUK • *Department of Biochemistry and Biophysics, Stockholm Bioinformatics Centre, Science for Life Laboratory, Stockholm University, Solna, Sweden*
- SLIM KARKAR • *Sorbonne Universités, Institut de Biologie Paris-Seine, UPMC Université Paris 6, Paris, France; Department of Ecology, Evolution, and Natural Resources, School of Environmental and Biological Sciences, Rutgers, The State University of NJ, New Brunswick, NJ, USA*
- TOSHIAKI KATAYAMA • *Database Center for Life Science, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Chiba, Japan*
- EUGENE V. KOONIN • *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*
- SERGEI L. KOSAKOVSKY POND • *Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, USA*
- CAROLIN KOSIOL • *Centre of Biological Diversity, School of Biology, University of St Andrews, Fife, UK; Institut für Populationsgenetik, Vetmeduni Vienna, Wien, Austria*
- ROMAIN LANNES • *Sorbonne Universités, Institut de Biologie Paris-Seine, UPMC Université Paris 6, Paris, France*
- PHILIPPE LEMEY • *Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium*
- MICHAEL LI • *Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, USA*
- LIANG LIU • *Department of Statistics, University of Georgia, Athens, GA, USA*
- PHILIPPE LOPEZ • *Sorbonne Universités, Institut de Biologie Paris-Seine, UPMC Université Paris 6, Paris, France*
- BRITTANY RIFE MAGALIS • *Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, USA*
- THOMAS MAILUND • *Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark*
- IZABELA MAKALOWSKA • *Institute of Anthropology, Adam Mickiewicz University, Poznań, Poland*
- WOJCIECH MAKALOWSKI • *Institute of Bioinformatics, University of Muenster, Muenster, Germany*
- RAPHAËL MÉHEUST • *Sorbonne Universités, Institut de Biologie Paris-Seine, UPMC Université Paris 6, Paris, France*
- SUPARNA MITRA • *Leeds Institute of Medical Research, University of Leeds, Microbiology, Old Medical School, Leeds General Infirmary, Leeds LS1 3EX, West Yorkshire, UK*
- STEFFEN MÖLLER • *Institute for Biostatistics and Informatics in Medicine and Ageing Research (IBIMA), Rostock University Medical Center, Rostock, Germany*
- MEGAN K. MULLIGAN • *Department of Genetics, Genomics and Informatics, The University of Tennessee Health Science Center, Memphis, TN, USA*

- AMIT PANDE • *Institute of Bioinformatics, University of Muenster, Muenster, Germany*
- JANANAN S. PATHMANATHAN • *Sorbonne Universités, Institut de Biologie Paris-Seine, UPMC Université Paris 6, Paris, France*
- DENNIS PEARL • *Department of Statistics, Pennsylvania State University, University Park, PA, USA*
- PJOTR PRINS • *Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; Department of Genetics, Genomics and Informatics, The University of Tennessee Health Science Center, Memphis, TN, USA; Laboratory of Nematology, Department of Plant Science, Wageningen University, Wageningen, The Netherlands*
- PERE PUIGBÓ • *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA; Division of Genetics and Physiology, Department of Biology, University of Turku, Turku, Finland*
- ANDREW RAMBAUT • *Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK*
- NICOLAS RODRIGUE • *Department of Biology, Carleton University, Ottawa, ON, Canada; Institute of Biochemistry, Carleton University, Ottawa, ON, Canada; School of Mathematics and Statistics, Carleton University, Ottawa, ON, Canada*
- MIKKEL H. SCHIERUP • *Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark*
- STEPHEN D. SHANK • *Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, USA*
- JULIANE SIEBOURG • *Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland*
- ANA CLAUDIA SIMA • *ZHAW Zurich University of Applied Sciences, Winterthur, Switzerland; University of Lausanne, Lausanne, Switzerland*
- GEERT SMANT • *Laboratory of Nematology, Department of Plant Science, Wageningen University, Wageningen, the Netherlands*
- ERIK L. L. SONNHAMMER • *Department of Biochemistry and Biophysics, Stockholm Bioinformatics Centre, Science for Life Laboratory, Stockholm University, Solna, Sweden*
- STEPHANIE J. SPIELMAN • *Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, USA*
- KURT STOCKINGER • *ZHAW Zurich University of Applied Sciences, Winterthur, Switzerland*
- FRANCESCO STROZZI • *Enterome Bioscience, Paris, France*
- MARC A. SUCHARD • *Department of Human Genetics and Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA*
- EDWARD SUSKO • *Department of Mathematics and Statistics, Dalhousie University, Halifax, NS, Canada*
- BJARNI J. VILHJÁLMSSON • *Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark*
- ANDREW K. WATSON • *Sorbonne Universités, Institut de Biologie Paris-Seine, UPMC Université Paris 6, Paris, France*
- STEVEN WEAVER • *Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, USA*
- ROB W. WILLIAMS • *Department of Genetics, Genomics and Informatics, The University of Tennessee Health Science Center, Memphis, TN, USA*
- SCOOTER WILLIS • *Department of Computer & Information Science & Engineering, University of Florida, Gainesville, FL, USA*

YURI I. WOLF • *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*

RICARDO WURMUS • *BIMSB Scientific Bioinformatics Platform, Max Delbrück Center for Molecular Medicine, Berlin, Germany*

ZIHENG YANG • *Department of Genetics, Evolution and Environment, University College London, London, UK*

ANDREW YATES • *European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK*

NOOR YOUSSEF • *Department of Biology, Dalhousie University, Halifax, NS, Canada*