

# 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>



# **Plant Bioinformatics**

**Methods and Protocols**

**Second Edition**

Edited by

**David Edwards**

*University of Western Australia, School of Plant Biology, Perth, Australia*

 **Humana Press**

*Editor*

David Edwards  
University of Western Australia  
School of Plant Biology  
Perth, Australia

ISSN 1064-3745                      ISSN 1940-6029 (electronic)  
Methods in Molecular Biology  
ISBN 978-1-4939-3166-8              ISBN 978-1-4939-3167-5 (eBook)  
DOI 10.1007/978-1-4939-3167-5

Library of Congress Control Number: 2015951438

Springer New York Heidelberg Dordrecht London  
© Springer Science+Business Media New York 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

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.

Printed on acid-free paper

Humana Press is a brand of Springer  
Springer Science+Business Media LLC New York is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

---

## **Preface**

Bioinformatics is a rapidly growing field of research that is being driven by the requirement to manage and interrogate the vast quantities of data being generated by 'omics technologies. The term bioinformatics means different things to different people, and following the theme of this series, this volume focuses on applied bioinformatics with specific applications to crops and model plants.

This volume is aimed at plant biologists who have an interest in, or requirement for, accessing and manipulation of the huge amount of data being generated by high-throughput technologies. The volume would also be of interest to bioinformaticians and computer scientists who would benefit from an introduction to the different tools and systems available for plant research.

The scope of bioinformatics now extends from the genome to the phenome and is increasingly being applied outside of pure research and towards supporting the accelerated breeding of crop plants.

It is the integration of information relating to heritable agronomic traits, including important metabolic profiles, with the emerging genome and transcriptome data that will drive plant research, crop breeding, and bioinformatics developments in the future. One observation during the production of this volume is the requirement to manage the increasing volume and diversity of data from different plants and also the integration of multiple diverse forms of data. I expect this trend to continue as this field of research continues to develop.

Scientific research by its nature progresses and changes and this is especially true for bioinformatics. The rapid evolution of the tools and systems described in this volume may change during the lifetime of this edition and it is suggested that the reader consult the relevant web pages directly as they frequently host detailed list of updates and changes.

*Perth, WA, Australia*

*David Edwards*



---

# Contents

<i>Preface</i> . . . . .	<i>v</i>
<i>Contributors</i> . . . . .	<i>ix</i>
1 Using GenBank . . . . . <i>Eric W. Sayers and Ilene Karsch-Mizrachi</i>	1
2 UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View . . . . . <i>Emmanuel Boutet, Damien Lieberherr, Michael Tognolli, Michel Schneider, Parit Bansal, Alan J. Bridge, Sylvain Poux, Lydie Bougueleret, and Ioannis Xenarios</i>	23
3 KEGG Bioinformatics Resource for Plant Genomics and Metabolomics. . . . . <i>Minoru Kanehisa</i>	55
4 Plant Pathway Databases . . . . . <i>Pankaj Jaiswal and Björn Usadel</i>	71
5 The Plant Ontology: A Tool for Plant Genomics . . . . . <i>Laurel Cooper and Pankaj Jaiswal</i>	89
6 Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data . . . . . <i>Dan Bolser, Daniel M. Staines, Emily Pritchard, and Paul Kersey</i>	115
7 Gramene: A Resource for Comparative Analysis of Plants Genomes and Pathways. . . . . <i>Marcela Karey Tello-Ruiz, Joshua Stein, Sharon Wei, Ken Youens-Clark, Pankaj Jaiswal, and Doreen Ware</i>	141
8 PGSB/MIPS Plant Genome Information Resources and Concepts for the Analysis of Complex Grass Genomes . . . . . <i>Manuel Spannagl, Kai Bader, Matthias Pfeifer, Thomas Nussbaumer, and Klaus F.X. Mayer</i>	165
9 MaizeGDB: The Maize Genetics and Genomics Database . . . . . <i>Lisa Harper, Jack Gardiner, Carson Andorf, and Carolyn J. Lawrence</i>	187
10 WheatGenome.info: A Resource for Wheat Genomics Resource . . . . . <i>Kaitao Lai</i>	203
11 User Guidelines for the Brassica Database: BRAD . . . . . <i>Xiaobo Wang, Feng Cheng, and Xiaowu Wang</i>	215
12 TAG Sequence Identification of Genomic Regions Using TAGdb . . . . . <i>Pradeep Ruperao</i>	233
13 Short Read Alignment Using SOAP2 . . . . . <i>Bhavna Hurgobin</i>	241

14 Tablet: Visualizing Next-Generation Sequence Assemblies  
and Mappings . . . . . 253  
*Iain Milne, Micha Bayer, Gordon Stephen, Linda Cardle,  
and David Marshall*

15 Analysis of Genotyping-by-Sequencing (GBS) Data. . . . . 269  
*Sateesh Kagale, Chushin Koh, Wayne E. Clarke, Venkatesh Bollina,  
Isobel A.P. Parkin, and Andrew G. Sharpe*

16 Skim-Based Genotyping by Sequencing Using a Double  
Haploid Population to Call SNPs, Infer Gene Conversions,  
and Improve Genome Assemblies . . . . . 285  
*Philipp Emanuel Bayer*

17 Finding and Characterizing Repeats in Plant Genomes . . . . . 293  
*Jacques Nicolas, Pierre Peterlongo, and Sébastien Tempel*

18 Analysis of RNA-Seq Data Using TopHat and Cufflinks . . . . . 339  
*Sreya Ghosh and Chon-Kit Kenneth Chan*

*Index*. . . . . 363



---

## Contributors

- CARSON ANDORF • *Maize Genetics and Genomics Database, USDA-ARS, Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA, USA*
- KAI BADER • *Plant Genome and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany*
- PARIT BANSAL • *Swiss Institute of Bioinformatics, Centre Medical Universitaire, Geneva, Switzerland*
- MICHA BAYER • *Information & Computational Sciences, The James Hutton Institute, Invergowrie, Dundee, Scotland, UK*
- PHILIPP EMANUEL BAYER • *School of Plant Biology, University of Western Australia, Perth, WA, Australia*
- VENKATESH BOLLINA • *Agriculture and Agri-Food Canada, Saskatoon, SK, Canada*
- DAN BOLSER • *European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, UK*
- LYDIE BOUGUELERET • *Swiss Institute of Bioinformatics, Centre Medical Universitaire, Geneva, Switzerland*
- EMMANUEL BOUTET • *Swiss Institute of Bioinformatics, Centre Medical Universitaire, Geneva, Switzerland*
- ALAN J. BRIDGE • *Swiss Institute of Bioinformatics, Centre Medical Universitaire, Geneva, Switzerland*
- LINDA CARDLE • *Information & Computational Sciences, The James Hutton Institute, Invergowrie, Dundee, Scotland, UK*
- CHON-KIT KENNETH CHAN • *School of Plant Biology, University of Western Australia, Crawley, Perth, WA, Australia*
- FENG CHENG • *Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China*
- WAYNE E. CLARKE • *Agriculture and Agri-Food Canada, Saskatoon, SK, Canada*
- LAUREL COOPER • *Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA*
- JACK GARDINER • *Maize Genetics and Genomics Database, USDA-ARS, Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA, USA*
- SREYA GHOSH • *Department of Biotechnology, National Institute of Technology, Durgapur, West Bengal, India*
- LISA HARPER • *Maize Genetics and Genomics Database, USDA-ARS, Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA, USA*
- BHAVNA HURGOBIN • *University of Queensland, St. Lucia, QLD, Australia; School of Plant Biology, University of Western Australia, Perth, WA, Australia*
- PANKAJ JAISWAL • *Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA*
- SATEESH KAGALE • *National Research Council Canada, Saskatoon, SK, Canada*
- MINORU KANEHISA • *Institute for Chemical Research, Kyoto University, Kyoto, Japan*

- ILENE KARSCH-MIZRACHI • *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*
- PAUL KERSEY • *European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, UK*
- CHUSHIN KOH • *National Research Council Canada, Saskatoon, SK, Canada*
- KAITAO LAI • *School of Agriculture and Food Science, University of Queensland, St. Lucia, QLD, Australia*
- CAROLYN J. LAWRENCE • *Department of Genetics, Development and Cell Biology, Roy J Carver Co-Laboratory, Iowa State University, Ames, IA, USA*
- DAMIEN LIEBERHERR • *Swiss Institute of Bioinformatics, Centre Medical Universitaire, Geneva, Switzerland*
- DAVID MARSHALL • *Information & Computational Sciences, The James Hutton Institute, Invergowrie, Dundee, Scotland, UK*
- KLAUS F.X. MAYER • *Plant Genome and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany; School of Life Sciences Weihenstephan, Technical University Munich, Neuherberg, Germany*
- IAIN MILNE • *Information & Computational Sciences, The James Hutton Institute, Invergowrie, Dundee, Scotland, UK*
- MARCELA KAREY TELLO-RUIZ • *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA*
- JACQUES NICOLAS • *Dyliss Team, Irisa/Inria Centre de Rennes Bretagne Atlantique, Rennes Cedex, France*
- THOMAS NUSSBAUMER • *Plant Genome and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany*
- ISOBEL A.P. PARKIN • *Agriculture and Agri-Food Canada, Saskatoon, SK, Canada*
- PIERRE PETERLONGO • *Irisa/Inria Centre de Rennes Bretagne Atlantique, Rennes Cedex, France*
- MATTHIAS PFEIFER • *Plant Genome and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany*
- SYLVAIN POUX • *Swiss Institute of Bioinformatics, Centre Medical Universitaire, Geneva, Switzerland*
- EMILY PRITCHARD • *European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, UK*
- PRADEEP RUPERAO • *School of Agriculture and Food Sciences, University of Queensland, St. Lucia, QLD, Australia; School of Plant Biology, University of Western Australia, Perth, WA, Australia*
- ERIC W. SAYERS • *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*
- MICHEL SCHNEIDER • *Swiss Institute of Bioinformatics, Centre Medical Universitaire, Geneva, Switzerland*
- ANDREW G. SHARPE • *National Research Council Canada, Saskatoon, SK, Canada*
- MANUEL SPANNAGL • *Plant Genome and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany*
- DANIEL M. STAINES • *European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, UK*
- JOSHUA STEIN • *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA*
- GORDON STEPHEN • *Information & Computational Sciences, The James Hutton Institute, Invergowrie, Dundee, Scotland, UK*

- SÉBASTIEN TEMPEL • *LCB, CNRS UMR 7283, Marseille Cedex, France*
- MICHAEL TOGNOLLI • *Swiss Institute of Bioinformatics, Centre Medical Universitaire, Geneva, Switzerland*
- BJÖRN USADEL • *IBMG-Institute for Biology I, RWTH Aachen University, Aachen, Germany; Forschungszentrum Jülich IBG-2 Plant Sciences, Jülich, Germany*
- XIAOBO WANG • *Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China*
- XIAOWU WANG • *Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China*
- DOREEN WARE • *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA; USDA-ARS NAA Plant, Soil & Nutrition Laboratory Research Unit, Cornell University, Ithaca, NY, USA*
- SHARON WEI • *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA*
- IOANNIS XENARIOS • *Swiss Institute of Bioinformatics, Centre Medical Universitaire, Geneva, Switzerland; University of Lausanne, Lausanne, Switzerland*
- KEN YOUENS-CLARK • *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA*