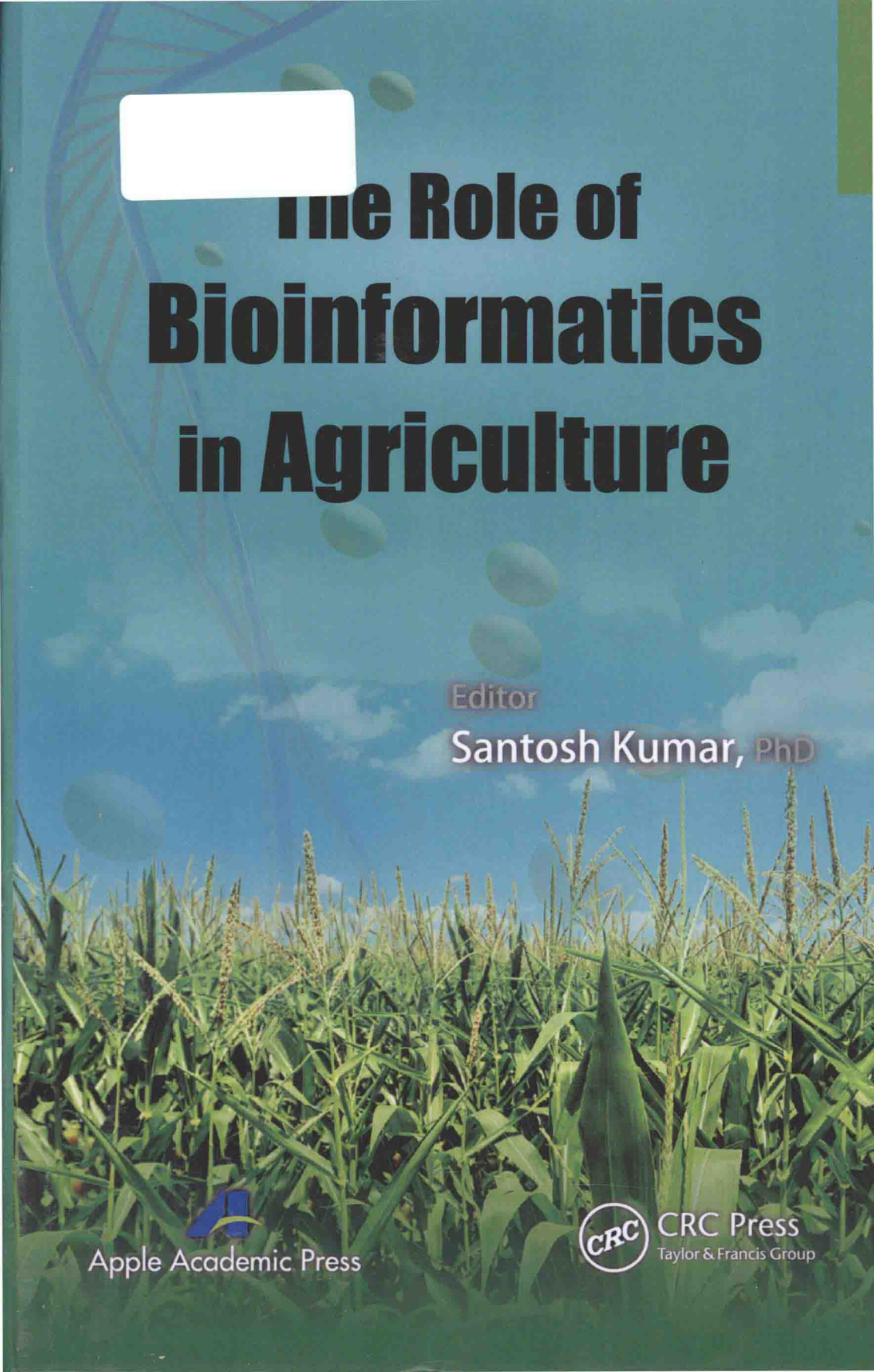
The Role of Bioinformatics in Agriculture

Editor
Santosh Kumar, PhD



Apple Academic Press

CRC Press
Taylor & Francis Group

The Role of Bioinformatics in Agriculture

Advances in information technology and next generation sequencing have propelled the use of bioinformatics in agriculture, especially in the area of crop improvement. An extremely large amount of genomics data is available from plants and animals due to tremendous improvements in the field. The challenge is now to make sense and use of this wealth of data.

This book acquaints readers with state-of-the-art sequencing technologies, recent developments in computing algorithms, and certain biological perspectives that influence development of bioinformatics tools by giving specific examples from model plant species.

The book covers a wide range of topics in this field, including

- functioning, accuracy, and cost of sequencing from second generation sequencers
- development of several protocols for the application of next generation sequencing technologies
- advances in wheat, maize, and rice genomics
- a concise list of plant genome databases and other resources for individual crop improvement
- computational tools that are required to understand the large amount of complex small RNA data generated through high throughput sequencing
- the heritability of epigenetic changes and the potential to drive natural variation
- more

About The Editor

Dr. Santosh Kumar obtained his BSc agriculture degree in plant breeding from Punjab Agricultural University, Ludhiana, India. He obtained his MSc working on wheat phytosiderophores from the Indian Agricultural Research Institute, New Delhi, India, and PhD working on control of flowering and germination in barley from the University of Manitoba, Canada. At present, Dr. Kumar is stationed at the Cereal Research Centre, Manitoba, Canada, working for the University of Manitoba. Dr. Kumar has studied various disciplines in plant sciences and received many awards throughout his academic career. His current interests include crop genomics and bioinformatics. Dr. Kumar is member of various plant science societies in Canada and the US. Dr. Kumar has publications that include research articles, a review paper, and a book chapter in areas of plant physiology, genetics, and bioinformatics.

ISBN: 978-1-771880-03-9
90000
9 781771 880039

Apple Academic Press
www.appleacademicpress.com

Kumar

The Role of Bioinformatics in Agriculture

THE ROLE OF BIOINFORMATICS IN AGRICULTURE

Edited by

Santosh Kumar, PhD

Apple Academic Press

TORONTO NEW JERSEY

Apple Academic Press Inc.
3333 Mistwell Crescent
Oakville, ON L6L 0A2
Canada

Apple Academic Press Inc.
9 Spinnaker Way
Waretown, NJ 08758
USA

Printed in the United States of America on acid-free paper

International Standard Book Number-13: 978-1-77188-003-9 (Hardcover)

Library of Congress Control Number: 2013958424

Library and Archives Canada Cataloguing in Publication

The role of bioinformatics in agriculture/edited by Santosh Kumar, PhD.

Includes bibliographical references and index.
ISBN 978-1-77188-003-9 (bound)
1. Agriculture--Data processing. 2. Bioinformatics.
3. Agricultural informatics. I. Kumar, Santosh, 1974-, editor of compilation

S494.5.D3R64 2013 630.285 C2013-907167-9

Apple Academic Press also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic format. For information about Apple Academic Press products, visit our website at **www.appleacademicpress.com** and the CRC Press website at **www.crcpress.com**

THE ROLE OF BIOINFORMATICS IN AGRICULTURE

ABOUT THE EDITOR

Santosh Kumar, PhD
Research Associate in Plant Science, University of Manitoba, Canada

Dr. Santosh Kumar obtained his BSc Agriculture degree in plant breeding from Punjab Agricultural University, Ludhiana, India. He obtained his MSc working on wheat phytosiderophores from the Indian Agricultural Research Institute, New Delhi, India, and PhD working on control of flowering and germination in barley from University of Manitoba, Canada. At present, Dr. Kumar is stationed at the Cereal Research Centre, Manitoba, Canada, working for the University of Manitoba. Dr. Kumar has studied various disciplines in plant sciences and received many awards throughout his academic career. His current interests include crop genomics and bioinformatics. Dr. Kumar is member of various plant science societies in Canada and the US. Dr. Kumar has publications that include research articles, a review paper, and a book chapter in areas of plant physiology, genetics, and bioinformatics.

ACKNOWLEDGMENT AND HOW TO CITE

The chapters in this book were previously published in various places and in various formats. By bringing them together here in one place, we offer the reader a comprehensive perspective on recent investigations into the role of bioinformatics in agriculture. Each chapter is added to and enriched by being placed within the context of the larger investigative landscape. Specifically:

- Chapter 1 provides a description of hands-on experience with functioning, accuracy, and cost of sequencing from second generation sequencers. The authors also discuss the applications of data generated through high throughput second generation sequencers.
- Chapter 2 focuses on the typical challenges of plant genomes that explains why plant genomics is less developed than animal genomics. It provides a discussion on factors hindering accurate plant genome assembly.
- Chapter 3 sheds light on the development of several protocols for the application of next generation sequencing technologies to genome walking.
- Chapter 4 focuses on technologies and strategies that may allow an in-depth analysis of polyploid genomes. Origin and genetics of polyploids as well as the main tools available for genome and gene expression analysis are discussed. The implications of next generation sequencing in study of polyploids is also discussed.
- Chapter 5 discusses the advances in wheat genomics and also describes the available resources which can be used for future genomics research; this review is important because wheat has a large genome and is a relatively difficult for genomics research.
- Chapter 6 focuses on producing accurate genome sequences of rice and studying its functional and evolutionary implications for comparative genomics
- Chapter 7 is a concise list of plant genome databases, resources for individual crop improvement and potential course of action for future improvement of genome databases.
- Chapter 8 surveys the RNA-Seq methods and the bioinformatics challenges that follow in analyzing the sequence data.
- Chapter 9 provides insights into power and accuracy of differential gene estimation based on different experimental designs.

- Chapter 10 sheds light on the computational tools that are required to understand the large amount of complex small RNA data generated through high throughput sequencing.
- Chapter 11 focuses on evaluation of genome diversity in maize by analyzing expression profile of genes. Maize is one of the most important cereals used as human and animal feed throughout the world. For this reason, maize has evolved to be genetically and phenotypically diverse. Not only genes but gene expression products also contribute to genetic and phenotypic plasticity.
- Chapter 12 discusses the heritability of epigenetic changes and its potential to drive natural variation. The epigenetic modifications alter gene expression without any change in the nucleotide sequence but by changing structural conformation of chromatin. They are key to adaptations occurring in short time spans.
- Chapter 13 summarizes the current protocols in metagenomics, experimental designs, guidance on sample processing, sequencing technology and sequence assembly, annotation, statistical analysis, data management and dissemination.
- Chapter 14 discusses the historical perspective of SNP development and the current experimental designs for SNP discovery and validation using specific examples. Molecular markers are regarded as foundation of genetic analysis and are indispensable for efficient breeding. The single nucleotide polymorphic sites are the ideal markers as they can be identified on the genome wide scale by a single experiment. The next generation sequencing has simplified SNP identification and validation.
- Chapter 15 summarizes an approach where sequencing data is used simultaneously to discover and validate SNPs thereby bypassing the entire marker development stages. With more and more genomes being sequenced through next generation sequencing, this is a timely review that summarizes the current state of genotyping-by-sequencing on genome wide scale.

We wish to thank the authors who made their research available for this book, whether by granting permission individually or by releasing their research as open source articles. When citing information contained within this book, please do the authors the courtesy of attributing them by name, referring back to their original articles, using the credits provided at the beginning of each chapter.

LIST OF CONTRIBUTORS

Rajat Aggarwal
Department of Trait Genetics and Technologies, Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268, USA

Claudia Angelini
Istituto per le Applicazioni del Calcolo "Mauro Picone", IAC-CNR, 80131 Naples, Italy

Baltazar A. Antonio
Division of Genome and Biodiversity Research, National Institute of Agrobiological Sciences, 2-1-2, Kannondai, Tsukuba, Ibaraki 3058602, Japan

Riccardo Aversano
Department of Soil, Plant, Environmental and Animal Production Sciences, University of Naples Federico II, Via Università 100, Portici 80055, Italy

Rocío Bautista
Bioinformatics Andalusian Platform, Bio-innovation Building, University of Malaga, 29590 Málaga, Spain

Hicham Benzerki
Department of Molecular Biology and Biochemistry, Faculty of Sciences, University of Malaga, 29071 Málaga, Spain and Bioinformatics Andalusian Platform, Bio-innovation Building, University of Malaga, 29590 Málaga, Spain

C. Robin Buell
Department of Plant Biology, Michigan State University, East Lansing, Michigan, United States of America and Department of Energy Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, Michigan, United States of America

Conrad J. Burden
Mathematical Sciences Institute, Australian National University, Canberra, Australia

Ramesh Buyyarapu
Department of Trait Genetics and Technologies, Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268, USA

Domenico Carputo
Department of Soil, Plant, Environmental and Animal Production Sciences, University of Naples Federico II, Via Università 100, Portici 80055, Italy

Immacolata Caruso
Department of Soil, Plant, Environmental and Animal Production Sciences, University of Naples Federico II, Via Università 100, Portici 80055, Italy

Alfredo Ciccodicola
Institute of Genetics and Biophysics "A. Buzzati-Traverso", IGB-CNR, 80131 Naples, Italy

Luigi R. Ceci
Institute of Biomembranes and Bioenergetics, Italian National Research Council (CNR), Via Amendola 165/A, 70126 Bari, Italy

Valerio Costa
Institute of Genetics and Biophysics "A. Buzzati-Traverso", IGB-CNR, 80131 Naples, Italy

Alessandra Costanza
Department of Biosciences, Biotechnology and Pharmacological Sciences, University of Bari, Via Amendola 165/A, 70126 Bari, Italy

Elizabeth S. Dennis
Commonwealth Scientific and Industrial Research Organisation (CSIRO) Plant Industry, Canberra ACT 2601, Australia

Stéphane Deschamps
DuPont Agricultural Biotechnology, Experimental Station, PO Box 80353, 200 Powder Mill Road, Wilmington, DE 19880-0353, USA

Italia De Feis
Istituto per le Applicazioni del Calcolo "Mauro Picone", IAC-CNR, 80131 Naples, Italy

Natalia de Leon
Department of Agronomy, University of Wisconsin-Madison, Madison, Wisconsin, United States of America and Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin, United States of America

David Edwards
School of Agriculture and Food Sciences, University of Queensland, Brisbane, QLD 4072, Australia and Australian Centre for Plant Functional Genomics, University of Queensland, Brisbane, QLD 4072, Australia

Maria Raffaella Ercolano
Department of Soil, Plant, Environmental and Animal Production Sciences, University of Naples Federico II, Via Università 100, Portici 80055, Italy

Immacolata Fanizza
Department of Biosciences, Biotechnology and Pharmacological Sciences, University of Bari, Via Amendola 165/A, 70126 Bari, Italy

Carlo Fasano
Department of Soil, Plant, Environmental and Animal Production Sciences, University of Naples Federico II, Via Università 100, Portici 80055, Italy

Noé Fernández-Pozo
Department of Molecular Biology and Biochemistry, Faculty of Sciences, University of Malaga, 29071 Málaga, Spain

Ryo Fujimoto
Graduate School of Science and Technology, Niigata University, Nishi-ku, Niigata 950-2181, Japan

Jack Gilbert
Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA and Department of Ecology and Evolution, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

Manuel Gonzalo Claros
Department of Molecular Biology and Biochemistry, Faculty of Sciences, University of Malaga, 29071 Málaga, Spain and Bioinformatics Andalusian Platform, Bio-innovation Building, University of Malaga, 29590 Málaga, Spain

Darío Guerrero-Fernández
Bioinformatics Andalusian Platform, Bio-innovation Building, University of Malaga, 29590 Málaga, Spain

P. K. Gupta
Molecular Biology Laboratory, Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut 250 004, India

Candice N. Hansey
Department of Plant Biology, Michigan State University, East Lansing, Michigan, United States of America and Department of Energy Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, Michigan, United States of America

Yimin He
NGS Sequencing Department, Beijing Genomics Institute (BGI), 4th Floor, Building 11, Beishan Industrial Zone, Yantian District, Guangdong, Shenzhen 518083, China

Ni Hu
NGS Sequencing Department, Beijing Genomics Institute (BGI), 4th Floor, Building 11, Beishan Industrial Zone, Yantian District, Guangdong, Shenzhen 518083, China

Ryo Ishikawa
Laboratory of Plant Breeding, Graduate School of Agricultural Science, Kobe University, Nada, Kobe 657-8510, Japan and Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK

Shawn M. Kaeppler
Department of Agronomy, University of Wisconsin-Madison, Madison, Wisconsin, United States of America and Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin, United States of America

Takahiro Kawanabe
Watanabe Seed Co., Ltd, Machiyashiki, Misato-cho, Miyagi 987-8607, Japan

J. Kumar
Molecular Biology Laboratory, Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut 250 004, India

Siva Kumpatla
Department of Biotechnology Regulatory Sciences, Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268, USA

Kaitao Lai
School of Agriculture and Food Sciences, University of Queensland, Brisbane, QLD 4072, Australia and Australian Centre for Plant Functional Genomics, University of Queensland, Brisbane, QLD 4072, Australia

Maggie Law
NGS Sequencing Department, Beijing Genomics Institute (BGI), 4th Floor, Building 11, Beishan Industrial Zone, Yantian District, Guangdong, Shenzhen 518083, China

Claudia Leoni
Department of Biosciences, Biotechnology and Pharmacological Sciences, University of Bari, Via Amendola 165/A, 70126 Bari, Italy

Lei Li
Department of Biology, University of Virginia, Charlottesville VA 22904, USA

Siliang Li
NGS Sequencing Department, Beijing Genomics Institute (BGI), 4th Floor, Building 11, Beishan Industrial Zone, Yantian District, Guangdong, Shenzhen 518083, China

Yinhu Li
NGS Sequencing Department, Beijing Genomics Institute (BGI), 4th Floor, Building 11, Beishan Industrial Zone, Yantian District, Guangdong, Shenzhen 518083, China

Danni Lin
NGS Sequencing Department, Beijing Genomics Institute (BGI), 4th Floor, Building 11, Beishan Industrial Zone, Yantian District, Guangdong, Shenzhen 518083, China

Lin Liu
NGS Sequencing Department, Beijing Genomics Institute (BGI), 4th Floor, Building 11, Beishan Industrial Zone, Yantian District, Guangdong, Shenzhen 518083, China

Victor Llaca
DuPont Agricultural Biotechnology, Experimental Station, PO Box 80353, 200 Powder Mill Road, Wilmington, DE 19880-0353, USA

Michał T. Lorenc
School of Agriculture and Food Sciences, University of Queensland, Brisbane, QLD 4072, Australia

Lihua Lu
NGS Sequencing Department, Beijing Genomics Institute (BGI), 4th Floor, Building 11, Beishan Industrial Zone, Yantian District, Guangdong, Shenzhen 518083, China

Jafar Mammadov
Department of Trait Genetics and Technologies, Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268, USA

Takashi Matsumoto
Division of Genome and Biodiversity Research, National Institute of Agrobiological Sciences, 2-1-2, Kannondai, Tsukuba, Ibaraki 3058602, Japan

Gregory D. May
DuPont Pioneer, 7300 NW 62nd Ave., P.O. Box 1004, Johnston, IA 50131-1004, USA

Folker Meyer
Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA and Computation Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

R. R. Mir
Molecular Biology Laboratory, Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut 250 004, India

A. Mohan
Molecular Biology Laboratory, Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut 250 004, India

Kenji Osabe
Commonwealth Scientific and Industrial Research Organisation (CSIRO) Plant Industry, Canberra ACT 2601, Australia

Antonio Placido
Department of Biosciences, Biotechnology and Pharmacological Sciences, University of Bari, Via Amendola 165/A, 70126 Bari, Italy

Ray Pong
NGS Sequencing Department, Beijing Genomics Institute (BGI), 4th Floor, Building 11, Beishan Industrial Zone, Yantian District, Guangdong, Shenzhen 518083, China

Sumaira E. Qureshi
Mathematical Sciences Institute, Australian National University, Canberra, Australia

José A. Robles
CSIRO Plant Industry, Black Mountain Laboratories, Canberra, Australia

Daniele Rosellini
Department of Applied Biology, University of Perugia, Borgo XX Giugno 74, Perugia 06121, Italy

Taku Sasaki
Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Dr. Bohrgasse 3, Vienna 1030, Austria

Takuji Sasaki
Division of Genome and Biodiversity Research, National Institute of Agrobiological Sciences, 2-1-2, Kannondai, Tsukuba, Ibaraki 3058602, Japan

Rajandeep S. Sekhon
Department of Agronomy, University of Wisconsin-Madison, Madison, Wisconsin, United States of America and Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin, United States of America

Pedro Seoane
Department of Molecular Biology and Biochemistry, Faculty of Sciences, University of Malaga, 29071 Málaga, Spain

Stuart J. Stephen
CSIRO Plant Industry, Black Mountain Laboratories, Canberra, Australia

Jennifer M. Taylor
CSIRO Plant Industry, Black Mountain Laboratories, Canberra, Australia