Association Mapping in Plants

Association Mapping in Plants

Edited by

Nnadozie C. Oraguzie
The Horticulture and Food Research Institute of New Zealand Ltd (HortResearch)
Havelock North, New Zealand

Erik H.A. Rikkerink
The Horticulture and Food Research Institute of New Zealand Ltd (HortResearch)
Auckland, New Zealand

Susan E. Gardiner
The Horticulture and Food Research Institute of New Zealand Ltd (HortResearch)
Palmerston North, New Zealand

H. Nihal De Silva
The Horticulture and Food Research Institute of New Zealand Ltd (HortResearch)
Auckland, New Zealand

Springer

Dr Nnadozie C. Oraguzie
HortResearch
Hawkes Bay Research Centre
Private Bag 1401
Havelock North
New Zealand

Dr Erik H.A. Rikkerink
HortResearch
Mt. Albert Research Centre
Private Bag 92169
Auckland
New Zealand

Dr Susan E. Gardiner
HortResearch
Palmerston North Research Centre
Private Bag 11030
Palmerston North
New Zealand

Dr H. Nihal De Silva
HortResearch
Mt. Albert Research Centre
Private Bag 92169
Auckland
New Zealand

Library of Congress Control Number: 2006928327

ISBN-10: 0-387-35844-7 e-ISBN-10: 0-387-36011-5
ISBN-13: 978-0387-35844-4 e-ISBN-13: 978-0-387-36011-9

Printed on acid-free paper.

© 2007 Springer Science+Business Media, LLC.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

9 8 7 6 5 4 3 2 1

springer.com

Preface

The approach taken for locating the genes that underlie human diseases has shifted from pedigree-based linkage studies to population-based association studies. In both cases the proximity of a genetic marker to a susceptibility locus is inferred from statistical measures that reflect the number of recombination events between them: in a disease pedigree there are no more than a few hundred opportunities for recombination so that recombination rates less than about one percent cannot be estimated and genes can be located only coarsely on a genetic map with that approach. The linkage disequilibrium detected in an association study, however, reflects the actions of many thousands of recombination events since the initial disease mutation and the expectation is that susceptibility genes can then be mapped more accurately.

The editors of this volume have recognized the need for parallel activity in plant species. For the past 20 years, the genes that affect plant economic traits have usually been mapped with data collected from "pedigrees" of populations formed by crossing inbred lines. These Quantitative Trait Loci have been mapped on a coarse scale, and a QTL is likely to refer to several genes in a region. The move to population-based association studies was therefore as necessary in plants as it was in humans, and readers will find this book to be a useful review of the marker technology, statistical methodology, and progress to date. Although one of the authors fears that "plant genetics can be considered as less advanced than human genetics" the chapters suggest that if that is the case it will not be so for long.

The recent increased activity in association mapping in humans has rested on the development of efficient and affordable methods for discovering and employing Single Nucleotide Polymorphism markers. Plant geneticists cannot command the resources available to their human geneticist colleagues, but they can anticipate benefiting from the success of the International HapMap Project. The improvement in marker technology from such large projects will inevitably be imported to plant studies. The editors have provided helpful guides to the use of SNPs in association studies.

Along with the substantial increase in the volume of data when large numbers of individuals are typed at millions of SNPs there are substantial challenges in the statistical interpretation of the data. This book contains a valuable account of the issues of multiple testing and an accessible account of False Discovery Rates. The more basic concepts of linkage disequilibrium and case-control versus family-based association tests are also discussed. It is often the case that geneticists do not receive extensive statistical training and the coverage of the theory of estimation and testing is therefore welcome. Readers will notice a greater use of Bayesian methods than is usually found in statistical genetics books. Such methods are appearing more frequently in scientific papers.

I congratulate the editors and all the authors on this timely and comprehensive treatment of association mapping in plants. The importance of food and fiber for human welfare cannot be overstated, and progress in plant improvement will rest in no small part on the work described in these pages. On a personal level, I am delighted by the leadership shown by my fellow antipodeans.

B.S. Weir
Professor and Chair
Department of Biostatistics
University of Washington

Acknowledgements

The motivation to write this book came from numerous correspondences and finally a meeting with Keri Witman, former senior editor in plant sciences at Springer Science+Business Media, New York, USA (who continually stressed the need for a book in this field of research), in Tulln Austria in 2004 at an EUCARPIA Quantitative genetics conference. The recommendation and suggestions by anonymous referees consulted by Springer to review the original proposal kept up the initial enthusiasm. This book would not have been possible without HortResearch's support. In particular, we would like to thank the following people; Drs Vincent Bus (Science Leader, Pipfruit and Summerfruit breeding), Andrew Granger (Future Fruit Group Leader) and Bruce Campbell (General manager science operations) for their encouragement and moral support, Stuart Ritchie for assistance with contract negotiation, Sharlene Cookson for assistance with book cover design, Dr David Chagné for his immense contribution right from the proposal stage of the book till it went into press, and the SPU team especially, Dr Anne Gunson and Christine Lamont, for editorial and technical assistance. We are grateful to the following for permission to reproduce copyright material: Trustees of the Royal Botanic Gardens, Kew; Elsevier (Figure 1 in Trends in Genetics 11: 83-90(2002) and Tables 1 & 2 in Trends in Genetics 20(2): 105(2004)); Swedish National Biobanking program, Wallenberg Consortium North (http://www.meb.ki.se/genestat.htm/-pairwise D` for 45 SNPs within a linked region); The University of Chicago press (Table 3, American Journal of Human Genetics 60(3): 676-690, 1997); Annual Reviews (Table 1 in Annual Review of Plant Biology 54, 2003); Nature Publishing group (Table 3 in Nature Genetics 28:286-289); and MacMillan publishing Inc. (Tables 2, 3, 4, 5 & 6 in Genetics 170:859-873, 2005). Finally we thank the following colleagues who reviewed the manuscripts and made suggestions that significantly improved the quality of the chapters: Drs David B. Neale (Professor of Forest Genetics, Dept of Plant Sciences, UCLA), Mark E. Sorrells (Professor, Dept of Plant Breeding and Genetics, Cornell University, Ithaca, NY), Trudy F.C. Mackay (WNR Distinguished Professor of Genetics, North Carolina State

University), Christophe Plomion (Molecular Geneticist, INRA, Cedex, France), David Pot (Molecular Geneticist, CIRAD, Montpellier, Cedex, France), Pauline Garnier-Géré (Forest Genetics Group, INRA, Cedex, France), Fred van Eeuwijk (Associate Professor in Statistics, Laboratory of Plant Breeding, Wageningen University and Research center, Wageningen, The Netherlands), Rasmus Neilsen (Ole RØmer Fellow/Professor in Statistical Genomics, University of Copenhagen, Denmark), Diane Mather (Professorial Research Fellow, School of Agriculture and Wine Science, University of Adelaide/Program Leader-New Molecular technologies, Molecular plant breeding cooperative Research Center, Plant Genomics Center, Glen Osmond SA, Australia), Martin Lascoux (Professor-Program in Evolutionary Functional Genomics, Evolutionary Biology Center, Uppsala University, Sweden), Hans van Buijtenen (Professor emeritus-Genetics, Dept of Forest Science, Texas A & M University), and Slade Lee (Associate Professor/Deputy Director, Center for Conservation Genetics/Sub-Program Leader (Phenomics), Grain Foods CRC Limited, Southern Cross University, Lismore NSW, Australia.

Contents

Contributors ..... **xi**

Introduction ..... **xiii**

1. An Overview of Association Mapping **1**
Nnadozie C. Oraguzie and Phillip L. Wilcox

2. Linkage Disequilibrium **11**
Nnadozie C. Oraguzie, Phillip L. Wilcox, Erik H.A. Rikkerink, and H. Nihal de Silva

3. What Are SNPs? **41**
David Edwards, John W. Forster, David Chagné, and Jacqueline Batley

4. Single Nucleotide Polymorphism Discovery **53**
David Edwards, John W. Forster, Noel O.I. Cogan, Jacqueline Batley, and David Chagné

5. Single Nucleotide Polymorphism Genotyping in Plants **77**
David Chagné, Jacqueline Batley, David Edwards, and John W. Forster

6. SNP Applications in Plants **95**
Jacqueline Batley and David Edwards

7. Linkage Disequilibrium Mapping Concepts **103**
H. Nihal de Silva and Roderick D. Ball

8. Statistical Analysis and Experimental Design 133
Roderick D. Ball

9. Linkage Disequilibrium-Based Association Mapping in Forage Species 197
Mark P. Dobrowolski and John W. Forster

10. Gene-Assisted Selection
Applications of Association Genetics for Forest Tree Breeding 211
Phillip L. Wilcox, Craig E. Echt, and Rowland D. Burdon

11. Prospects of Association Mapping in Perennial Horticultural Crops 249
Erik H.A. Rikkerink, Nnadozie C. Oraguzie, and Susan E. Gardiner

Index 271

Contributors

Roderick D. Ball
Ensis Wood Quality, Scion (New Zealand Forest Research Institute Limited), 49 Sala Street, Private Bag 3020, Rotorua, New Zealand

Jacqueline Batley
Primary Industries Research Victoria, Victorian AgriBiosciences Centre, La Trobe R&D Park, Bundoora, Victoria 3083, Australia

Rowland D. Burdon
Ensis Genetics, Scion (New Zealand Forest Research Institute Limited), 49 Sala Street, Private Bag 3020, Rotorua, New Zealand

David Chagné
HortResearch, Plant Gene Mapping group, Private Bag 11030, Palmerston North, New Zealand

Noel O.I. Cogan
Primary Industries Research Victoria, Victorian AgriBiosciences Centre, La Trobe R&D Park, Bundoora, Victoria 3083, Australia

Craig E. Echt
USDA Forest Service, Southern Institute of Forest Genetics 23332 MS Highway 67, Saucier, MS 39574, USA

David Edwards
Primary Industries Research Victoria, Victorian AgriBiosciences Centre,
La Trobe R&D Park, Bundoora, Victoria 3083, Australia

John W. Forster
Primary Industries Research Victoria, Victorian AgriBiosciences Centre,
La Trobe R&D Park, Bundoora, Victoria 3083, Australia

Mark P. Dobrowolski
Primary Industries Research Victoria, Plant Genetics and Genomics Platform, Hamilton Centre, Mt Napier Road Hamilton, Victoria 3300, Australia

Susan E. Gardiner
HortResearch, Palmerston North Research Centre, Private Bag 11030, Palmerston North, New Zealand

Nnadozie C. Oraguzie
HortResearch, Hawkes Bay Research Centre, Cnr. Crosses and St. George's Roads, Private Bag 1401, Havelock North, New Zealand

Erik H.A. Rikkerink
HortResearch, Mt Albert Research Centre, 120 Mt Albert Road, Private Bag 92169, Auckland, New Zealand

H. Nihal De Silva
HortResearch, Mt Albert Research Centre, 120 Mt Albert Road, Private Bag 92169, Auckland, New Zealand

Phillip L. Wilcox
Cell wall Biotechnology Centre, Scion (New Zealand Forest Research Institute Limited), 49 Sala Street, Private Bag 3020, Rotorua, New Zealand

Introduction

Most traits we deal with on a daily basis have complex inheritance patterns that complicate the ability of existing mapping technologies to detect the underlying genetic factors. In the last decade or so, we have seen the successful use of conventional map-based strategies in identification and cloning of quantitative trait loci (QTLs) in model plant species including tomato and Arabidopsis. However, efficient gene discovery with this method will probably continue to be largely limited to those loci that have large effects on quantitative trait variation. Techniques are also needed to more rapidly identify genes that play a modest role in regulating quantitative trait variation. Association mapping via linkage disequilibrium or LD (non-random association of alleles at different loci) offers promise in this area. The traditional approach of linkage/QTL mapping reliant on developing large mapping populations continues to suffer from lack of mapping resolution inherent in samples with limited meiotic cross-over events. These problems are exacerbated in tree crops, where very large populations are impractical from a plant management point of view. In association mapping, there may not be any need to make crosses initially to generate segregating populations. The natural variation that exists in the available germplasm can be utilized for mapping straightaway.

Association genetics via LD mapping is an emerging field of genetic mapping that has the potential for resolution to the level of individual genes (alleles) underlying quantitative traits. LD mapping is a technology that can take full advantage of the phenomenal leaps and bounds in technology development in the area of molecular biology and marry it with our increasing understanding of the molecular basis of inheritance and molecular tools recently developed in terms of molecular markers and genetic maps in a way that could have a significant practical impact on breeding. The convergence of improved statistical methods, availability of growing plant genomics databases and improvements in the affordability and potential scale of sequencing and

genotyping, suggests that this technology will probably be more widely adopted for mapping and gene discovery in plants in the near future.

This book provides a basic understanding of association mapping and an awareness of population genomics tools available to facilitate mapping and identification of the underlying causes of quantitative trait variation, as well as an analysis of the prospects of applying this technology to plants. In the book, we discuss how technological advances have recently brought association mapping into the realm of possibility for plants, particularly, the second and third tier crops (which include a number of long-lived tree species), which normally lag some way behind the first tier of crops (including annuals - cereals mainly) in technology development. For convenience, the book can be divided into 4 sections; there are two chapters in Section 1 which introduce association mapping and present the basic principles of association genetics in relation to the concept of linkage disequilibrium (LD); Section 2 comprises four chapters which deal with technology development in relation to SNP discovery through to SNP applications; Section 3 consists of 2 chapters providing a detailed discussion on statistical methodology and experimental design issues necessary for the successful application of association genetics; and lastly Section 4 contains 3 chapters which deal with specific issues and applications of association genetics using the crop groupings of forage, forestry and horticultural species as case examples. Application of LD mapping in model organisms including humans, Drosophila, Arabidopsis and maize is discussed in chapter 2 (in section 1).

The book brings together all the information on association genetics and linkage disequilibrium published in different journals in one volume and will be of interest to advanced breeding/genetics students, researchers, professional plant breeders and university lecturers. Breeders will find it particularly useful as a guide for making decisions on breeding strategies that will facilitate identification of ‘superior’ parents for development of new improved varieties. Difficult statistical concepts and tools are presented with detailed illustrations in a way readers can comprehend hence; the book could also serve as a teaching aid for postgraduate students. A very comprehensive comparison of statistical approaches/methodologies and guidelines on optimal study design, as well as the comparison of the relative benefits of association mapping and conventional QTL mapping will be particularly useful to geneticists wanting to set up studies on gene/genome mapping.

Association mapping now stands at the cross-roads of application to a large number of species and situations. Over the next decade it will become more apparent just how much influence this technology will have on increasing our fundamental understanding of the genetic basis of variation in plants and the practical outcomes of plant breeding in the future.