The Handbook of Plant Genome Mapping

Genetic and Physical Mapping

Edited by
Khalid Meksem and Günter Kahl

WILEY-VCH Verlag GmbH & Co. KGaA

The Handbook of Plant Genome Mapping

Edited by
Khalid Meksem and Günter Kahl

Further Titles of Interest

Günter Kahl

The Dictionary of Gene Technology

Genomics, Transcriptomics, Proteomics

2004
ISBN 3-527-30765-6

Christoph W. Sensen

Essentials of Genomics and Bioinformatics

2002
ISBN 3-527-30541-6

Michael R. Barnes and Ian C. Gray (Editors)

Bioinformatics for Geneticists

2003
ISBN 0470-84393-4

The Handbook of Plant Genome Mapping

Genetic and Physical Mapping

Edited by
Khalid Meksem and Günter Kahl

WILEY-VCH Verlag GmbH & Co. KGaA

Edited by

Prof. Dr. Khalid Meksem
Dept. Plant, Soil, Agriculture
Southern Illinois University
Carbondale, IL 629
USA

Prof. Dr. Günter Kahl
Plant Molecular Biology
Biocentre and GenXPro
Johann Wolfgang Goethe University
Marie-Curie-Strasse 9
60439 Frankfurt am Main
Germany

Cover Illustration:

The Title Page shows (from left to right) a linkage group (arrow-headed) from a genetic map, a blow-up of a specific region (in blue), the positions of molecular markers (RFLPs) along this region, (which is scaled in centiMorgans, cMs), and the two neighboring RFLP markers S10620 (red) and S14162 (green) at position 1.0 cM. These two RFLPs are radioactively labelled and hybridized to an array of BAC clones immobilized on a nylon membrane and (1) used to isolate genomic DNA containing sequences of interest (e. g. genes flanked by both markers) and (2) localized on chromosomal fibres by fibre fluorescent *in situ* hybridization (right). This technique is a variant of the conventional fluorescent *in situ* hybridization (FISH) method for the detection of target sequences, in which the two BAC clones are labelled with different fluorochromes, whose emission light is either in the red (S10620) or yellowish-green range, respectively (S14162). The chromosomal fibres are generated by molecular combing. For details, Chapter 6: "Physical mapping of Plant Chromosomes" by Barbara Hass-Jacobus and Scott A. Jackson

This book was carefully produced. Nevertheless, authors, editor and publisher do not warrant the information contained therein to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data:
A catalogue record for this book is available from the British Library.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

© 2005 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Printed in the Federal Republic of Germany.
Printed on acid-free paper.

Typesetting hagedorn kommunikation, Viernheim
Printing Strauss Offsetdruck GmbH, Mörlenbach
Bookbinding Litges & Dopf Buchbinderei GmbH, Heppenheim

ISBN-13: 978-3-527-31116-3
ISBN-10: 3-527-31116-5

We dedicate this book to the memory of late

Jozef (Jeff) Stefaan Schell,

who strongly inspired our way to think and to do research

Preface

At a time, when more than 150 bacterial and archaebacterial genomes, two plant genomes and a series of avertebrate and vertebrate genomes including the human genome have been deciphered base by base (some gaps notwithstanding), and more than 400 other genomes are in the mill, a Handbook of Plant Genome Mapping (Genetic and Physical Mapping) might seem a bit out of time. In fact, for many plants and animals genetic maps with vastly different densities are available and being improved continuously. Some of the scientific journals already begin to discourage authors to publish such increasingly dense maps and ask for more detailed informations such as genes isolated by map-based cloning.

In essence, genetic maps are by no means orphanized anymore. Also, if not yet available, genetic maps can be generated with speed and relative ease, provided a good selection of polymorphic parents, a wonderful and numerous segregating progeny, a highly resolving molecular marker system, powerful computer packages, a lot of people and enough money are at hand. So, in a not-too-far future, genetic maps will be commonplace given the relative ease of their generation.

The situation is quite different for physical maps of genomes. First of all, a physical map traditionally depends on the availability of a genetic map. Despite other approaches, the most practical method to establish a physical map still requires a large-insert clone bank on one hand, and a preferably highly dense genetic map on the other. And each and every marker, or marker bundle, that allows easy identification of the underlying DNA clones of whatever make-up (BAC clones seem to have won the race by now, and YAC clones lost because of chimerism and redundancies) is welcome. A highly resolving physical map, however, still requires a lot more input in labour, time, knowledge and funds than a genetic map. It is for this reason, that physical maps are available only for relatively few higher organisms, though common for prokaryota, whose chromosomes are sequenced and directly aligned into an ultimate physical map. Such complete genomic sequences (i. e. complete physical maps) are still an exception for eukaryotes. And, with one single exception, techniques used for the assembly of whole genome sequences still makes use of genetic and physical maps. The exception to this rule is the HAPPY mapping procedure, an ingenious tool with which physical maps can directly and happily be generated. This exception aside, the treadmill for many postdocs is and will be for the foreseeable future, the establishment of ge-

netic maps of the target organism as a prerequisite, including the production of BAC libraries and the physical alignment of the thousands of clones into contigs, at least for a region of interest. And genetically mapped markers will still serve to guide the way.

In appreciation of these facts, we set out to invite internationally renowned and highly competent plant researchers with an undisputed scientific reputation to portray their contributions to the genetic and physical mapping of plant genomes. The present "Handbook of Plant Genome Mapping: Genetic Mapping, Physical Mapping" is the most complete, up-to-date and competently written and compiled treatise of this complex topic. All the authors have striven to report the latest achievements and developments in their fields and did not spare any pains to introduce their areas of research, to detail methodological aspects and to present the state-of-the-art and future perspectives as well. This Handbook reflects the quality of worldwide research on plant genetic and physical mapping.

The editors most cordially appreciate the various contributions to make this book a standard for plant genome mapping for the foreseeable future.

November 2004

Khalid Meksem
Carbondale (IL, USA)

Günter Kahl
Frankfurt am Main (Germany)

Contents

List of Contributors

Bartoš, Jan
Laboratory of Molecular Cytogenetics and Cytometry
Institute of Experimental Botany
Sokolovska 6
77200 Olomouc
Czech Republic

Doležel, Jaroslav
Laboratory of Molecular Cytogenetics and Cytometry
Institute of Experimental Botany
Sokolovska 6
77200 Olomouc
Czech Republic

Gresshoff, Peter M.
ARC Centre of Excellence for Integrative Legume Research and School of Life Sciences
The University of Queensland,
St. Lucia
Brisbane Qld 4072
Australia

Hass-Jacobus, Barbara
Department of Agronomy
Agricultural Genomics
Lilly Hall of Life Sciences
915 W. State Street
Purdue University
West Lafayette, IN 47907-2054
USA

Ishihara, Hirofumi
Department of Plant, Soil and General Agriculture
Agriculture Bldg. Room 176
Southern Illinois University at Carbondale
Carbondale, IL 62901-4415
USA

Jackson, Scott A.
Department of Agronomy
Agricultural Genomics
Lilly Hall of Life Sciences
915 W. State Street
Purdue University
West Lafayette, IN 47907-2054
USA

Jesse, Tacco
Keygene N. V., Agro Business Park 90
P. O. Box 216
6700 AE Wageningen
The Netherlands

Kahl, Günter
Plant Molecular Biology, Biocenter and GenXPro
University of Frankfurt am Main
Marie-Curie-Strasse 9
60439 Frankfurt am Main
Germany

Kubaláková, Marie
Laboratory of Molecular Cytogenetics and Cytometry
Institute of Experimental Botany
Sokolovska 6
77200 Olomouc
Czech Republic

Lee, Mi-Kyung
Laboratory for Plant Genomics & GENEfinder Genomic Resources
Department of Soil & Crop Sciences and Institute for Plant Genomics & Biotechnology
Texas A&M University
College Station, TX 77843-2123
USA

Macas, Jiří
Laboratory of Molecular Cytogenetics
Institute of Plant Molecular Biology
Branisovska 31
37005 České Budějovice
Czech Republic

Mast, Andrea
Third Wave Technologies
502 South Rosa Road
Madison, WI 53719-1256
USA

Meksem, Khalid
Department of Plant, Soil and General Agriculture
Agriculture Bldg. Room 176
Southern Illinois University at Carbondale
Carbondale, IL 62901-4415
USA

Nelson, James C.
Department of Plant Pathology
Kansas State University
Manhattan, KS 66506-5502
USA

Nelson, William
Arizona Genome Computational Laboratory
Institute for Biomedical Science and Biotechnology
303 Forbes Building,
University of Arizona
Tucson, AZ 85721
USA

Nguyen, Henry T.
Department of Agronomy,
Plant Sciences Unit
1-87 Agriculture Building
University of Missouri-Columbia
Columbia, MO 65211
USA

Peleman, Johan D.
KEYGENE GENETICS
P. O. Box 216
6700 AE Wageningen
The Netherlands

Peterson, Daniel
Mississippi Genome Exploration Laboratory (MGEL)
Dept. of Plant and Soil Sciences
117 Dorman Hall, Box 9555
Mississippi State University
Mississippi State, MS 39762
USA

Ren, Chengwei
Laboratory for Plant Genomics & GENEfinder Genomic Resources
Department of Soil & Crop Sciences and Institute for Plant Genomics & Biotechnology
Texas A&M University
College Station, TX 77843-2123
USA

Rouppe van der Voort, Jeroen N. A.M.
KEYGENE GENETICS
P. O. Box 216
6700 AE Wageningen
The Netherlands

Santos, Teofila S.
Laboratory for Plant Genomics &
GENEfinder Genomic Resources
Department of Soil & Crop Sciences
and Institute for Plant Genomics &
Biotechnology
Texas A&M University
College Station, TX 77843-2123
USA

Schneider, Katharina
Max-Planck-Institut für Züchtungs-
forschung,
Carl-von-Linne-Weg 10
50829 Cologne
Germany

Shen, Richard
Illumina, Inc.
9885 Towne Centre Drive
San Diego, CA 92121-1975
USA

Scheuring, Chantel
Department of Soil and Crop Sciences
and
Institute for Plant Genomics and Bio-
technology
2123 TAMU
Texas A&M University
College Station, TX 77843-2123
USA

Soderlund, Carol
Arizona Genome Computational
Laboratory
Institute for Biomedical Science and
Biotechnology
303 Forbes Building
University of Arizona
Tucson, AZ 85721
USA

Sørensen, Anker P.
KEYGENE GENETICS
P. O. Box 216
6700 AE Wageningen
The Netherlands

Sun, Shuku
Laboratory for Plant Genomics &
GENEfinder Genomic Resources
Department of Soil & Crop Sciences
and Institute for Plant Genomics &
Biotechnology
Texas A&M University
College Station, TX 77843-2123
USA

Tooke, Nigel
Biotage AB
Kungsgatan 76
753 18 Uppsala
Sweden

Town, Christopher D.
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850
USA

van den Boom, Dirk
SEQUENOM, Inc.
3595 John Hopkins Court
San Diego, CA 92121
USA

Wu, Chengcang
Laboratory for Plant Genomics &
GENEfinder Genomic Resources
Department of Soil & Crop Sciences
Crop Biotechnology Center
Texas A&M University
College Station, TX 77843-2123
USA

Wu, Xiaolei
Department of Agronomy,
Plant Sciences Unit
1-87 Agriculture Building
University of Missouri-Columbia
Columbia, MO 65211
USA

Xu, Zhanyou
Laboratory for Plant Genomics &
GENEfinder Genomic Resources
Department of Soil & Crop Sciences and
Institute for Plant Genomics &
Biotechnology
Texas A&M University
College Station, TX 77843-2123
USA

Zhang, Hong-Bin
Laboratory for Plant Genomics &
GENEfinder Genomic Resources
Department of Soil & Crop Sciences and
Institute for Plant Genomics & Biotech-
nology
Texas A&M University
College Station, TX 77843-2123
USA

Part I
Genetic Mapping

1
Mapping Populations and Principles of Genetic Mapping

Katharina Schneider

Overview

Mapping populations consist of individuals of one species, or in some cases they derive from crosses among related species where the parents differ in the traits to be studied. These genetic tools are used to identify genetic factors or loci that influence phenotypic traits and to determine the recombination distance between loci. In different organisms of the same species, the genes, represented by alternate allelic forms, are arranged in a fixed linear order on the chromosomes. Linkage values among genetic factors are estimated based on recombination events between alleles of different loci, and linkage relationships along all chromosomes provide a genetic map of the organism. The type of mapping population to be used depends on the reproductive mode of the plant to be analyzed. In this respect, the plants fall into the main classes of self-fertilizers and self-incompatibles. This chapter illustrates the molecular basis of recombination, summarizes the different types of mapping populations, and discusses their advantages and disadvantages for different applications.

Abstract

In genetics and breeding, mapping populations are the tools used to identify the genetic loci controlling measurable phenotypic traits. For self-pollinating species, F_2 populations and recombinant inbred lines (RILs) are used; for self-incompatible, highly heterozygous species, F_1 populations are mostly the tools of choice. Backcross populations and doubled haploid lines are a possibility for both types of plants. The inheritance of specific regions of DNA is followed by molecular markers that detect DNA sequence polymorphisms. Recombination frequencies between traits and markers reveal their genetic distance, and trait-linked markers can be anchored, when necessary, to a more complete genetic map of the species. For map-based cloning of a gene, populations of a large size provide the resolution required.

Due to intensive breeding and pedigree selection, genetic variability within the gene pools of relevant crops is at risk. Interspecific crosses help to increase the size of the gene pool, and the contribution of wild species to this germ plasm in the form of introgression lines is of high value, particularly with respect to traits like disease resistance. The concept of exotic libraries with near-isogenic lines, each harboring a DNA fragment from a wild species, implements a systematic scan of the gene pool of a wild species.

To describe the complexity of genome organization, genetic maps are not sufficient because they are based on recombination, which is largely different along all genomes. However, genetic maps, together with cytogenetic data, are the basis for the construction of physical maps. An integrated map then provides a detailed view on genome structure and enforces positional cloning of genes and ultimately the sequencing of complete genomes.

1.1 Introduction

Since Mendel formulated his laws of inheritance in 1865, it is a core component of biology to relate genetic factors to functions visible as phenotypes. At Mendel's time, genetic analysis was restricted to visual inspection of the plants. Pea (*Pisum sativum* [Fabaceae]) was already a model plant at the time, and Mendel studied visible traits such as seed and pod color, surface structure of seeds and pods (smooth versus wrinkled), and plant height. These traits are, in fact, the first genetic markers used in biology. In 1912 Vilmorin and Bateson described the first work on linkages in *Pisum*. However, the concept of linkage groups representing chromosomes was not clear in *Pisum* until 1948, when Lamprecht described the first genetic map with 37 markers distributed on 7 linkage groups (summarized in Swiecicki et al. 2000). Large collections of visible markers are today available for several crop species and for *Arabidopsis thaliana* (Koornneef et al. 1987; Neuffer et al. 1997).

In the process of finding more and more genetic markers, the first class of characters scored at the molecular level was isoenzymes. These are isoforms of proteins that vary in amino acid composition and charge and that can be distinguished by electrophoresis. The technique is applied to the characterization of plant populations and breeding lines and in plant systematics, but it is also used for genetic mapping of variants, as shown particularly in maize (Frei et al. 1986, Stuber et al. 1972). However, due to the small number of proteins for which isoforms exist and that can be separated by electrophoresis, the number of isoenzyme markers is limited.

The advance of molecular biology provided a broad spectrum of technologies to assess the genetic situation at the DNA level. The first DNA polymorphisms described were restriction fragment length polymorphism (RFLP) markers (Botstein et al. 1980). This technique requires the hybridization of a specific probe to restricted genomic DNA of different genotypes. The whole genome can be covered

by RFLP and, depending on the probe, coding or non-coding sequences can be analyzed. The next generation of markers was based on PCR: rapid amplified polymorphic DNA (RAPD) (Williams et al. 1990; Welsh and McClelland 1990) and amplified fragment length polymorphism (AFLP) (Vos et al. 1995). Recently, methods have been developed to detect single nucleotide polymorphisms (summarized in Rafalski 2002). Because these methods have the potential for automatization and multiplexing, they allow the establishment of high-density genetic maps.

Whereas RAPD and AFLP analyses are based on anonymous fragments, RFLP and SNP analyses allow the choice of expressed genes as markers. Genes of a known sequence and that putatively influence the trait of interest can be selected and mapped. In this way function maps can be constructed (Chen et al. 2001; Schneider et al. 2002). Phenotypic data of the segregating population, correlated to marker data, prove or disprove potential candidate genes supporting mono- and polygenic traits.

The basis for genetic mapping is recombination among polymorphic loci, which involves the reaction between homologous DNA sequences in the meiotic prophase. Currently, the double-strand-break repair model (Szostak et al. 1983) is acknowledged to best explain meiotic reciprocal recombination (Figure 1.1). In this model, two sister chromatids break at the same point and their ends are resected at the 5′ ends. In the next step the single strands invade the intact homologue and pair with their complements. The single-strand gaps are filled in using the intact strand as template. The resulting molecule forms two Holliday junctions. Upon resolution of the junction, 50 % of gametes with recombinant lateral markers and 50 % non-recombinants are produced. In the non-recombinants, genetic markers located within the region of strand exchange may undergo gene conversion, which can result in nonreciprocal recombination, a problem interfering in genetic mapping. In plants, gene conversion events were identified by Büschges et al. (1997) when cloning the *Mlo* resistance gene from barley.

The likelihood that recombination events occur between two points of a chromosome depends in general on their physical distance: the nearer they are located to each other, the more they will tend to stay together after meiosis. With the increase of the distance between them, the probability for recombination increases and genetic linkage tends to disappear. This is why genetic linkage can be interpreted as a measure of physical distance. However, taking the genome as a whole, the fre-

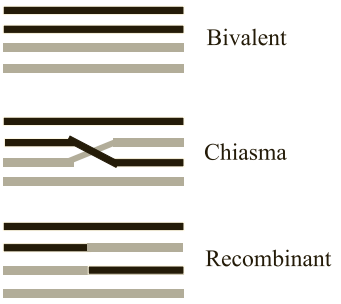

Figure 1.1. Generation of recombinants by chiasma formation. In the meiotic prophase, two sister chromatids of each parent (labeled in red and green, respectively) align to form a bivalent. A chiasma is formed by a physical strand exchange between two non-sister chromatids. Breakage and reunion of reciprocal strands leads to the generation of recombinants.

quency of recombination is not constant because it is influenced by chromosome structure. An example is the observation that recombination is suppressed in the vicinity of heterochromatin: here, the recombination events along the same chromatid appear to be reduced, an observation called positive interference. It reduces the number of double recombinants when, for example, three linked loci are considered.

Linkage analysis based on recombination frequency and the order of linked loci is evaluated statistically using maximum likelihood equations (Fisher 1921; Haldane and Smith 1947; Morton 1955). Large amounts of segregation data are routinely processed by computer programs to calculate a genetic map; among the most popular are JoinMap (Stam 1993) and MAPMAKER (Lander et al. 1987).

1.2 Mapping Populations

The trait to be studied in a mapping population needs to be polymorphic between the parental lines. Additionally, a significant trait heritability is essential. It is always advisable to screen a panel of genotypes for their phenotype and to identify the extremes of the phenotypic distribution before choosing the parents of a mapping population. It is expected that the more the parental lines differ, the more genetic factors will be described for the trait in the segregating population and the easier their identification will be. This applies to monogenic as well as to polygenic traits.

A second important feature to be considered when constructing a mapping population is the reproductive mode of the plant. There are two basic types. On the one hand are plants that self naturally, such as *Arabidopsis thaliana*, tomato, and soybean, or that can be manually selfed, such as sugar beet and maize; on the other hand are the self-incompatible, inbreeding-sensitive plants such as potato. Self-incompatible plants show high genetic heterozygosity, and for these species it is frequently not possible to produce pure lines due to inbreeding depression. Usually only self-compatible plants allow the generation of lines displaying a maximum degree of homozygosity. In conclusion, the available plant material determines the choice of a mapping population. Other factors are the time available for the construction of the population and the mapping resolution required. Based on these concepts, this section will be divided into seven parts:

1. mapping populations suitable for self-fertilizing plants,
2. mapping populations for cross-pollinating species,
3. two-step strategies for mapping mutants and DNA fragments,
4. chromosome-specific tools for mapping,
5. mapping in natural populations/breeding pools,
6. mapping genes and mutants to physically aligned DNA, and
7. specific mapping problems.