

B.D. Singh · A.K. Singh

Marker-Assisted Plant Breeding: Principles and Practices

 Springer

Marker-Assisted Plant Breeding: Principles and Practices

B.D. Singh • A.K. Singh

Marker-Assisted Plant Breeding: Principles and Practices

 Springer

B.D. Singh
School of Biotechnology
Banaras Hindu University
Varanasi, UP, India

A.K. Singh
Division of Genetics
Indian Agricultural Research Institute
New Delhi, Delhi, India

ISBN 978-81-322-2315-3 ISBN 978-81-322-2316-0 (eBook)
DOI 10.1007/978-81-322-2316-0

Library of Congress Control Number: 2015943502

Springer New Delhi Heidelberg New York Dordrecht London

© Author(s) 2015

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

Springer (India) Pvt. Ltd is part of Springer Science+Business Media (www.springer.com)

*To
Prof. M.S. Swaminathan,
who reshaped agricultural research in India
and inspired a whole generation of plant scientists*

Foreword

Plant breeding is the discipline that fashioned our crop plants out of the wild weedy species and continues its endeavor to modify their genotypes to enhance their performance and usefulness to the changing human needs and climate conditions. In the past, the new genotypes developed by plant breeders have been considerably successful in keeping pace with the growing global food needs and consumer preferences. For example, the evolution of hybrid varieties and semi-dwarf cereal genotypes has contributed to quantum jumps in crop productivity, and the latter was responsible for the ‘green revolution’ that made countries like India virtually self-sufficient in their food grain requirements within a short span of a few years.

The world population is increasing at a rapid rate and is expected to go past the nine billion mark by the year 2042. In addition, the nature and the relevance of both abiotic and biotic stresses are undergoing unrelenting changes in the wake of the environmental alterations engendered by climate change and global warming. In view of these, it is necessary not only to continue to evolve crop genotypes with higher yield potential and tolerance to the various prevailing stresses but also to develop them at a much faster pace. The plant breeders thus face unprecedented challenges of harnessing the reservoirs of genetic variability present in the unadapted germplasm with the minimum investment of time and in a highly precise and predictable manner.

Traditional plant breeding methods rely on phenotype-based selection, but phenotypic evaluation of many traits is problematic, unreliable or expensive. Also, the usefulness of trait phenotypes of individuals/lines in predicting the performance of their progeny is questionable. In addition, the conventional breeding methods do not allow the use of desirable genes from related species in an efficient manner, and there is always the risk of linkage drag. Plant breeders have always been trying to develop breeding strategies that would make their selections more effective and reliable and that would facilitate the utilization of unadapted germplasm with the minimum risk of linkage drag. One of the options that was pursued with some success was the use of simply inherited traits for an indirect selection for complex traits. This effort led to the discovery of protein-based markers and, eventually, the DNA-based markers.

Since the deployment of RFLPs in biological studies, several user-friendly DNA markers like SSRs and SNPs have been developed. The current

emphasis is on technologies that permit low-cost, high-throughput genotyping using molecular markers. Markers are being increasingly used for marker-assisted selection to facilitate gene introgression and for accelerated recurrent selection with the use of off-season nurseries and greenhouse facilities. In addition, markers have found applications in many other plant breeding activities like diversity analysis, germplasm characterization, hybrid seed lot genetic purity determination, elucidation of heterosis loci, etc. In view of the increasing integration of markers in plant breeding programs, many universities have introduced courses on marker-based plant breeding. There is, therefore, an urgent need for a book covering the various aspects relevant to the use of markers in plant breeding.

The book 'Marker-Assisted Plant Breeding' is designed to provide up-to-date information on molecular markers and their applications. The authors have attempted to provide sufficient basic information in an easily understandable narrative so that even the beginners have little difficulty in following the subject. This book will also be useful to teachers, breeders and research workers since it makes available at one place the current information on the various aspects of the subject. The development of different molecular markers and their various applications are described in a simple language, and in a clear and easily comprehensible manner. In the first chapter, the field of marker-assisted plant breeding is introduced and placed in the proper perspective in relation to plant breeding. The next three chapters describe the various molecular marker systems, while mapping populations and mapping procedures, including high-throughput genotyping and association mapping, are discussed in the subsequent five chapters. Four chapters are devoted to various applications of markers, while the last two chapters provide information about relevant bioinformatics tools and phenomics.

The authors deserve compliments for conceiving this book and for developing this concept into a useful and informative book. I am confident that the students, teachers and the professional plant breeders will find this book to be of considerable usefulness as it provides a wealth of information at one place. The book assumes contemporary relevance and importance, since varieties breed with the help of marker-assisted selection are eligible for certification under organic farming.

M S Swaminathan
Research Foundation
Third Cross Street
Taramani Institutional Area
Chennai 600 113, India

Prof. M.S. Swaminathan

Preface

Improved genotypes developed by plant breeding remain pivotal to global food security. In the wake of ever-increasing human population, declining agricultural resources and the stresses generated by climate change, plant breeding is expected to make larger contributions in increasingly shorter time frames. Therefore, plant breeding methods and schemes would have to be made more efficient and capable of accelerated variety development, say, by making efficient use of off-season nurseries, greenhouse facilities and innovative breeding methods. One of the chief limitations of plant breeding is the low effectiveness of phenotypic selection for many traits, particularly the quantitative traits. Further, selection for many other traits is tedious, problematic, time consuming and/or poorly reliable due to threshold requirements, difficulties in assay procedures and phenotype measurement, etc. Breeders have long been searching for tools that would permit effective indirect selection for such traits. Oligogenic phenotypic traits were the first to be used for this purpose, followed by protein-based/isozyme markers. However, the chief limitation of the above marker systems was the limited availability of good informative markers closely linked to the traits of interest.

In 1980, Botstein and coworkers proposed the use of restriction fragment length polymorphism (RFLP) for linkage mapping in humans. RFLP soon emerged as the first DNA-based molecular marker system, and it was used for the preparation of marker linkage maps and for the mapping of several traits of interest in many crops. The greater abundance and other desirable features of RFLPs as compared to phenotypic and protein markers, prompted the development of other relatively more convenient DNA marker systems like random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), etc. Single nucleotide polymorphism (SNP) has emerged as the most abundant molecular marker that is amenable to high-throughput genotyping. Each of these marker systems offers some advantages and suffers from certain limitations.

Molecular markers provide a tool for identifying genomic regions involved in the control of traits of interest. They also facilitate selection for the target genomic regions on the basis of marker genotype rather than the phenotype of the concerned trait. The reliability of such indirect selection depends mainly on the strength of linkage between the marker and the genomic region of interest. Therefore, markers located within the genes of interest, particularly those associated with allelic differences with respect to

trait phenotype (the functional nucleotide polymorphism, FNP, being the ultimate), would be the most informative and useful. However, the practical usefulness of MAS will be primarily determined by the relative cost of marker development, identification of trait-linked markers and marker genotyping in the breeding populations as compared to the direct trait phenotype-based selection.

The first step in the use of markers for MAS is the identification of markers tightly linked to the traits of interest. Ordinarily, a suitable mapping population needs to be constructed to identify the linked markers by linkage mapping. Several different types of mapping populations, ranging from simple F_2 through recombinant inbred lines to multi-parent advanced generation intercross (MAGIC) and interconnected populations can be used for linkage analyses. Alternatively, a collection of germplasm lines/individuals from natural populations can be used for linkage disequilibrium-based association mapping. In addition, the rich genomic resources that are now becoming available for most crops of interest can be analysed for marker identification.

Molecular mapping of oligogenes is relatively simple, while that of quantitative trait loci (QTLs) poses many problems, and the results from mapping studies are affected by a variety of factors, including the genetic model and the statistical algorithm used for QTL analysis. Generally, different QTLs governing the same trait are identified from different studies and consensus QTLs need to be identified by QTL meta-analysis. In addition, for a reliable detection of association/linkage between markers and traits, the trait phenotypes have to be measured precisely and reliably; a discipline called phenomics devoted to large-scale precision phenotyping is currently the area of intensive research activity.

Molecular markers tightly linked to the desired traits can be used for MAS to select for the genes governing the concerned traits, recover the recurrent parent genotype in backcross programs as well as to eliminate linkage drag, wherever required. Innovative breeding schemes are being designed to facilitate an efficient utilization of resources and to maximize gains from the marker technology. For example, marker-assisted recurrent selection (MARS) is being used for the improvement of quantitative traits, including yield, and up to three generations can be raised in a single season using off-season nurseries or a phytotron. The comprehensive scheme of genomic selection has been proposed for the selection of all genomic regions influencing the traits of interest, whether or not they show significant association with the trait phenotype. An ambitious breeding scheme, breeding by design, has been proposed to accumulate all the positive alleles for all the relevant traits into a single genotype that may be expected to have an outstanding performance. Similarly, a reverse breeding scheme for isolation of complementing inbred pairs from any heterotic hybrid combination has been patented.

Molecular markers have found a variety of other applications, including genetic diversity analysis, phylogenetic studies and construction of heterotic pools. Markers enable unambiguous identification of lines/varieties and facilitate seed certification and PBR (plant breeder rights) implementation. Tightly linked markers provide the basis for fine mapping and positional cloning of genes, which enables generation of information on gene function and regulation, as well as production of transgenic lines expressing the traits of interest.

A successful integration of molecular marker technology in plant breeding would require a low-cost, user-friendly marker systems amenable to high-throughput marker genotyping. Considerable effort is currently focused on the development of low-cost marker identification and genotyping platforms, including genotyping strategies that reduce the volume of genotyping work and/or combine marker discovery with marker genotyping without greatly sacrificing the amount of information obtained. The exciting developments in the above areas are generating new information and concepts/ideas with concomitant creation of specialized terms and phrases that together constitute the discipline of 'marker-assisted plant breeding'.

The chief constraints that limit the integration of molecular markers as a common tool in plant breeding are relatively higher cost of marker genotyping, and the fact that marker technology may appear unfamiliar to those trained in conventional plant breeding. There is continuous generation of new information, concepts/ideas and, inevitably, terminology related to molecular markers and their applications for achieving plant breeding objectives. Further, the marker technology has triggered innovations in breeding strategies and methods and has necessitated the creation of statistical and bioinformatics tools for data processing to facilitate their use for timely decision making. Plant breeding students need to be exposed to the various concepts, procedures and techniques relevant to the field in order to be able to appreciate the opportunities and the limitations of various options offered by the marker technology. It is encouraging that most educational institutions are introducing courses devoted either fully or partly to molecular markers.

The book 'Marker-Assisted Plant Breeding, Principles to Practice' is designed for such students who have had little or no exposure to molecular markers, but have a basic knowledge of genetics and plant breeding, and some exposure to molecular biology. This book will also be useful for teachers, research workers and practicing plant breeders. We have attempted to explain the basic principles, procedures and techniques of marker technology and provide, in brief, the up-to-date information on various aspects in a clear and easily comprehensible manner. Figures and line drawings are provided to highlight the chief features of important procedures/schemes/concepts with a view to facilitate their understanding by the students. In the first chapter, the field of marker-assisted plant breeding is introduced and placed in the proper perspective in relation to plant breeding. The next three chapters describe the various molecular marker systems, while mapping populations and procedures, including high-throughput genotyping and phenotyping, are discussed in the following five chapters. Four chapters are devoted to various applications of molecular markers, including MAS, diversity analysis, positional cloning, etc. The last two chapters provide information about relevant bioinformatics tools and phenomics.

Varanasi, UP, India
New Delhi, Delhi, India
November 25, 2014

Brahma Deo Singh
Ashok Kumar Singh

Acknowledgements

We wish to acknowledge the valuable help received from several of our colleagues and many of our research scholars. Prof. Umesh Singh and Prof. A.M. Kayastha, Varanasi, Dr. Anjana Jajoo, Indore, Dr. Sanjeev Kumar and Dr. Kusum Yadav Lucknow, Dr. K.K. Vinod, Aduthurai, Dr. Vinay, Kumar, Hyderabad, Dr. J.K. Joy, Mohali, Dr. Gopala Krishnan, Dr. Shailesh Tripathi, Dr. Neelu Jain, Dr. M. Vignesh, Dr. Shailendra K. Jha, Dr. Ramya Kurian, Dr. Renu Singh and Ms. Prachi Yadav, New Delhi, and Dr R.K. Sharma, Palampur read parts of the manuscript and/or proof, and suggested very useful additions and alterations. Our research scholars Mallikarjuna, Haritha, Ranjith, Sateesh, Fiyaz, Niranjana, and Naresh, New Delhi, and Vinay K. Singh and Reena Deshmukh, Varanasi, assisted us in a variety of ways, including finding of a suitable flavour for the text, checking for format, style spelling, grammatical errors, etc.

Dr. Kusum Yadav, Lucknow, provided PC scatter plots and the gel image, Dr. Balram Marathi provided the LOD score curves for SIM and CIM, and Reena Deshmukh provided images of multiple sequence alignment and rooted and unrooted trees; these contributions have been suitably acknowledged at the appropriate places. The efforts of Ranjith deserve a special mention as he has meticulously generated and later corrected the line drawings for the entire book.

Several of our colleagues and numerous students suggested that we should develop a book on marker-assisted plant breeding; this was one of the reasons for our decision to undertake this effort. We appreciate that Prof. Kole encouraged us to contact Springer for the publication of this book. We would also like to record our happiness for Springer's decision to publish this book and are thankful to all those involved in this process.

In the end, we are highly appreciative of the affection, support and encouragement we have always received from our family members, including our wives, sons and daughters.

Contents

Part I General

1	Introduction to Marker-Assisted Crop Improvement	3
1.1	Introduction	3
1.2	Domestication: The Evolution of Crop Plants	3
1.3	Plant Breeding	4
1.3.1	Major Developments in Plant Breeding	4
1.3.2	The Genotype and Phenotype	5
1.3.3	Genetic Variation: Qualitative and Quantitative Inheritance	5
1.3.4	Contributions: Pure Line Varieties	7
1.3.5	Contributions: Hybrid Varieties	7
1.3.6	Contributions: Clones	8
1.3.7	Limitations of Phenotype-Based Plant Breeding	8
1.4	The Growing Food Needs	9
1.5	The Transgenic Technology: Lukewarm Social Response	10
1.6	Molecular Markers: Selection Made Easy and More Reliable	12
1.7	Designer Crops	13
1.8	Some Notable Achievements of Marker-Assisted Plant Breeding	14
1.9	Future Prospects of Marker-Assisted Plant Breeding	14

Part II Genetic Markers

2	Hybridization-Based Markers	19
2.1	Introduction	19
2.2	Genetic Markers	19
2.2.1	Visible/Morphological Markers	20
2.2.2	Protein-Based Markers	20
2.2.3	DNA Markers	22
2.2.4	Concluding Remarks on Genetic Markers	24
2.3	Random, Gene-Based, and Functional Markers	24
2.4	Isolation and Purification of DNA from Plants	26

2.5	Restriction Fragment Length Polymorphism	27
2.5.1	Restriction Enzymes	28
2.5.2	Southern Hybridization	29
2.5.3	Probes	31
2.5.4	Polymorphisms Detected by RFLP Markers	33
2.5.5	Genetic Aspects of RFLPs	34
2.5.6	Advantages of RFLPs	35
2.5.7	Limitations of RFLPs	35
2.5.8	Conversion of RFLP Markers into PCR-Based Markers	35
2.6	Diversity Array Technology	36
2.7	Variable Number of Tandem Repeats	39
2.8	Single Feature Polymorphisms	39
2.9	Restriction-Site-Associated DNA Markers	41
	Appendices	42
	Appendix 2.1: Isolation and Purification of DNA from Plants	42
	Appendix 2.2: Genomic and cDNA Libraries	44
	Appendix 2.3: Microarrays	45
3	Polymerase Chain Reaction-Based Markers	47
3.1	Introduction	47
3.2	Oligonucleotides	47
3.3	Polymerase Chain Reaction	48
3.3.1	Generalized Procedure for PCR	48
3.3.2	Separation of PCR Amplification Products	50
3.3.3	Multiplex PCR	51
3.3.4	Applications of PCR	51
3.3.5	Advantages and Limitations of PCR	52
3.4	PCR-Based Markers	52
3.5	Randomly Amplified Polymorphic DNAs	52
3.6	DNA Amplification Fingerprinting	54
3.7	Arbitrary-Primed PCR	55
3.8	Sequence-Characterized Amplified Regions	55
3.9	Amplified Fragment Length Polymorphisms	55
3.9.1	The Procedure of AFLP	57
3.9.2	Features of AFLP	57
3.9.3	Modifications of the AFLP Technique	58
3.9.4	Conversion of AFLP Markers	59
3.10	Sequence-Tagged Sites	59
3.11	Microsatellites or Simple Sequence Repeats	59
3.12	Simple Sequence Repeat Markers	60
3.12.1	Discovery of SSR Markers	60
3.12.2	Increasing the Throughput of SSR Markers	60
3.12.3	Merits of SSR Markers	62
3.12.4	Limitations of SSR Marker System	62
3.13	Inter-Simple Sequence Repeats	63
3.13.1	Modifications of ISSR	63
3.13.2	Merits and Limitations of ISSR Markers	64

3.14	Cleaved Amplified Polymorphic Sequences	64
3.15	Single-Strand Conformation Profile/Polymorphism	65
3.16	Denaturing/Temperature Gradient Gel Electrophoresis . .	66
3.17	Sequence-Related Amplification Polymorphism	68
3.18	Target Region Amplification Polymorphism	69
3.19	Transposable Element-Based Markers	69
3.20	Conserved Orthologous Set of Markers	70
3.21	Start Codon-Targeted Polymorphism	71
3.22	CAAT Box-Derived Polymorphism	72
3.23	Conserved DNA-Derived Polymorphism	72
3.24	Conserved Region Amplification Polymorphism	73
3.25	Intron-Targeting Polymorphism	73
3.26	RNA-Based Molecular Markers	74
	Appendices	74
	Appendix 3.1: The Number of RAPD Bands	
	Theoretically Expected from a DNA Sample	74
	Appendix 3.2: Polymerase Chain Reaction	
	and Randomly Amplified Polymorphic DNAs	75
4	Sequence-Based Markers	77
4.1	Introduction	77
4.2	DNA Sequencing	77
4.2.1	First-Generation DNA Sequencing Methods . .	78
4.2.2	Next-Generation DNA Sequencing Methods . .	79
4.2.3	The Third-Generation DNA Sequencing	
	Methods	87
4.2.4	Comparison Between NGS and TGS	
	Sequencers	92
4.3	RNA Sequencing	92
4.3.1	RNA-Seq	92
4.3.2	Single-Molecule Direct RNA Sequencing	94
4.4	Single-Nucleotide Polymorphisms	94
4.4.1	Types of SNPs	95
4.5	Methods for Discovery of SNPs	96
4.5.1	Amplicon Sequencing	96
4.5.2	SNP Mining	97
4.5.3	Transcriptome Sequencing	97
4.5.4	Whole-Genome Sequencing	98
4.5.5	Reduced Representation Approaches	99
4.5.6	Sequence Capture	100
4.5.7	Validation of Discovered SNPs	101
4.6	Methods for SNP Genotyping	101
4.6.1	Allele-Specific PCR	101
4.6.2	5'-Nuclease Assay (TaqMan® Assay)	103
4.6.3	Molecular Beacons	103
4.6.4	Microarray-Based SNP Genotyping	105
4.6.5	Bead-Based Techniques	107
4.6.6	Primer Extension	108

4.6.7	Pyrosequencing	110
4.6.8	Oligonucleotide Ligation Assay	110
4.6.9	Dynamic Allele-Specific Hybridization	111
4.6.10	Denaturing High-Performance Liquid Chromatography	111
4.6.11	InDels as Molecular Markers	114
4.7	Epigenetic Markers	114
4.8	Use of Genomics, Transcriptomics, Proteomics, and Metabolomics in Marker Development	114
4.9	Polymorphic Information Content of Marker Loci	116
4.10	Marker System Selection	118

Part III Linkage Maps

5	Mapping Populations	125
5.1	Introduction	125
5.2	Mapping Populations	125
5.3	Selection of Parents for Developing a Mapping Population	126
5.4	F_2 Population	127
5.5	F_2 -Derived F_3 Population	129
5.6	Backcross Population	130
5.7	Doubled Haploids	130
5.8	Recombinant Inbred Lines	131
5.9	Immortalized F_2 Population	135
5.10	Near-Isogenic Lines	136
5.11	Chromosomal Segment Substitution Lines	139
5.12	Backcross Inbred Lines	141
5.13	Advanced Intercross Lines	141
5.14	Recurrent Selection Backcross Population	141
5.15	Interconnected Mapping Populations	142
5.16	Multiparent Advanced Generation Intercross Populations	143
5.17	Nested Association Mapping Population	145
5.18	Mapping Populations for Cross-Pollinated Species	145
5.19	Linkage Mapping in Polyploid Species	145
5.20	Chromosome-Specific Genetic Stocks	147
5.21	Natural Populations and Germplasm/Breeding Lines	147
5.22	Segregation Ratios in Mapping Populations	147
5.23	Characterization of Mapping Populations	148
5.24	Problems in Mapping Studies	148
5.25	Size of Mapping Population	149
5.26	Choice of Mapping Population	149
6	Linkage Mapping of Molecular Markers and Oligogenes	151
6.1	Introduction	151
6.2	Genetic Maps	151
6.2.1	Linkage Maps	151
6.2.2	Cytogenetic Maps	152
6.2.3	Physical Maps	152

6.3	Estimation of Recombination Rates	153
6.4	Genetic Distance	153
6.4.1	The Haldane Distance	155
6.4.2	The Kosambi Distance	156
6.4.3	Variation in Genetic Distance	156
6.4.4	Relationship Between Genetic and Physical Distances	157
6.5	General Procedure for Linkage Mapping of Molecular Markers and Oligogenes	158
6.6	Mapping of the Loci Present in a Chromosome	158
6.7	Strategies for Mapping of Oligogenes	159
6.7.1	Use of Near-Isogenic Lines	159
6.7.2	Bulked Segregant Analysis	160
6.7.3	Mapping of Recessive Morphological Mutants by a Two-Step Procedure	162
6.7.4	Bulked Segregant RNA-Seq	163
6.7.5	The MutMap Technique	164
6.8	LOD Score and LOD Score Threshold	165
6.9	A Complete Linkage Map	167
6.10	Integration or Merger of Linkage Maps	168
6.11	Confirmation and Validation	169
6.12	Comparative Mapping	169
6.13	Fine Mapping (High-Resolution Mapping)	171
6.14	Software for Mapping of Oligogenes/Molecular Markers	173
6.14.1	MapMaker/Exp	173
6.14.2	RI Plant Manager	174
6.14.3	G-MENDEL	174
6.14.4	MultiMap	174
6.14.5	AntMap	174
6.14.6	JoinMap	175
6.14.7	MergeMap	175
6.14.8	ActionMap	175
6.14.9	TetraploidMap for Windows	176
6.14.10	MultiPool	176
6.14.11	Mutation Mapping Analysis Pipeline for Pooled RNA-Seq	176
6.14.12	MapPop	177
6.14.13	Next-Generation Mapping	177
6.15	Selective Mapping and Selective Genotyping	177
6.16	Pooled DNA Analysis	179
6.17	Physical Mapping of Molecular Markers	180
6.18	Sources of Errors in Linkage Mapping	181
6.19	The Significance of Genetic Maps	182
7	Mapping of Quantitative Trait Loci	185
7.1	Introduction	185
7.2	Quantitative Trait Loci	185
7.3	The General Procedure for QTL Mapping	186

7.4	Marker and Quantitative Trait Data Structure	187
7.5	Methods for QTL Detection and Mapping	187
	7.5.1 Single QTL Mapping	188
	7.5.2 Multiple QTL Mapping	193
	7.5.3 Some Remarks on QTL Mapping	197
7.6	Bulked Segregant Analysis for QTL Mapping	197
7.7	Multiple Trait QTL Mapping	199
7.8	LOD Score and LOD Score Threshold	200
7.9	QTL Confidence/Support Interval	201
7.10	Confirmation and Validation of QTL Mapping Results	202
7.11	QTL Fine Mapping	203
	7.11.1 Homozygous Lines Derived from Near-Isogenic Lines	203
	7.11.2 Intercross Recombinant Inbred Lines	203
	7.11.3 Recurrent Selection Backcross QTL Mapping	204
	7.11.4 Genetically Heterogeneous Stocks	204
	7.11.5 Multiparent Advanced Generation Intercross Population	204
	7.11.6 Reverse QTL Mapping	204
	7.11.7 Combination of QTL Mapping and Transcriptome Profiling	205
7.12	QTL Meta-Analysis	205
7.13	Inconsistent Estimates of QTL Effects	207
	7.13.1 Segregation of Different QTLs in Different Populations	207
	7.13.2 QTL \times Genetic Background Interaction	207
	7.13.3 QTL \times Environment Interaction	208
	7.13.4 The Beavis Effect	208
7.14	QTL Detection Power and Precision of QTL Mapping	208
7.15	Factors Affecting Results from QTL Mapping	209
	7.15.1 Genetic Properties of QTLs	209
	7.15.2 Genetic Background	209
	7.15.3 Type and Size of Mapping Population	210
	7.15.4 Environmental Effects on QTL Expression	211
	7.15.5 Experimental Error	211
7.16	Advantages of QTL Linkage Mapping	211
7.17	Limitations of QTL Mapping	211
7.18	Nature and Function of Polygenes	212
7.19	Software for QTL Mapping	213
	7.19.1 MapMaker/QTL	213
	7.19.2 PLABQTL	213
	7.19.3 QTL Cartographer	213
	7.19.4 MapManager QT/QTX	214
	7.19.5 R/QTL	214
	7.19.6 R/QTLBIM	214
	7.19.7 QTL Express	214

	7.19.8	FlexQTL	215
	7.19.9	INTERQTL	215
	7.19.10	MCQTL	215
	7.19.11	QGene	216
	7.19.12	Some Other Software Programs	216
8		Association Mapping	217
	8.1	Introduction	217
	8.2	The General Procedure for Association Mapping	217
	8.3	Phenotyping	220
	8.4	Genome-Wide and Candidate Gene Approaches for Association Mapping	220
	8.5	Populations Used for Association Mapping in Plants . . .	222
	8.5.1	Population-Based Association Panels	222
	8.5.2	Family-Based Association Panels: NAM Population	224
	8.5.3	Family-Based Association Panels: MAGIC Population	224
	8.6	Linkage Disequilibrium for Biallelic Loci	226
	8.7	Measures of Linkage Disequilibrium	228
	8.7.1	Two Biallelic Loci	230
	8.7.2	Two Loci with Multiple Alleles	232
	8.7.3	Multiple Locus Methods	232
	8.8	Graphic Representation of LD	233
	8.9	Useful LD	234
	8.10	The Extent of LD in Plant Species	234
	8.11	Uses of LD in Plant Molecular Biology	235
	8.12	Experimental Designs and Models for Association Mapping	236
	8.12.1	Case and Control Approach	236
	8.12.2	Family-Based Designs	237
	8.12.3	Structured Association Model	237
	8.12.4	Mixed Linear Models	238
	8.12.5	Joint Linkage-Association Mapping	241
	8.12.6	Multilocus Mixed Model	241
	8.12.7	Multitrait Mixed Model	242
	8.13	Significance Tests for Marker-Trait Associations	242
	8.14	Controlling “False Discovery” Rate	243
	8.15	Relevance of Marker Systems in LD Estimation	244
	8.16	Factors Affecting LD and Association Mapping	245
	8.16.1	Mating Pattern in the Population	245
	8.16.2	Selection	246
	8.16.3	Population Structure	247
	8.16.4	Admixture	247
	8.16.5	Genomic Region	248
	8.16.6	Kinship	248
	8.16.7	Genetic Drift and Bottleneck	248

8.16.8	Gene Conversion	248
8.16.9	Ascertainment Bias	249
8.16.10	Marker Mutation Rate	249
8.16.11	Errors in Genotyping	249
8.17	Conclusions About LD Patterns in Plant Species	249
8.18	LD Maps	250
8.19	Mapping of Expression Quantitative Trait Loci	250
8.20	Power of Association Mapping	250
8.21	Confirmation of Marker-Trait Associations Through Replication Studies	251
8.22	The tagSNP Strategy of SNP Genotyping	251
8.23	Software for LD Studies	252
8.24	Conclusions from Association Mapping Studies	252
8.25	Current Issues in Association Mapping	253
8.26	Future Perspectives	254
8.27	Merits of Association Mapping	254
8.28	Limitations of Association Mapping	255

Part IV Applications

9	Marker-Assisted Selection	259
9.1	Introduction	259
9.2	Marker-Assisted Characterization of Germplasm and Genetic Purity	260
9.3	Marker-Assisted Backcrossing	260
9.3.1	Foreground Selection	261
9.3.2	Background Selection	261
9.3.3	Recombinant Selection	264
9.3.4	A Four-Step Comprehensive Selection Strategy	266
9.4	A Theory for Background Selection During MABC	266
9.5	MABC for Transfer of Oligogenic Traits	267
9.6	MABC for Transfer of Quantitative Trait Loci	269
9.7	MABC for Gene Pyramiding	271
9.7.1	Strategy for Gene Pyramiding	271
9.7.2	Pyramiding of Oligogenes	273
9.7.3	Pyramiding of QTLs with Oligogenes Governing the Same Trait	274
9.7.4	Transgene Pyramiding	275
9.8	Multitrait Introgression	275
9.9	Combined Marker-Assisted Selection	275
9.10	Marker-Assisted Recurrent Selection	277
9.10.1	MARS in Cross-Pollinated Crops	278
9.10.2	F_2 Enrichment and MARS in Self-Pollinated Crops	279

9.11	Innovative Breeding Schemes for Effective Use of MAS	280
9.11.1	Inbred Enhancement and QTL Mapping	280
9.11.2	Advanced Backcross QTL Analysis	283
9.11.3	Single Large-Scale MAS	284
9.11.4	Pedigree MAS	285
9.11.5	Single Backcross-Doubled Haploid Scheme	285
9.11.6	Breeding by Design	286
9.11.7	Mapping as You Go	287
9.11.8	Marker-Evaluated Selection for Adaptation and Agronomic Performance	287
9.12	Integration of MAS in Breeding Programs	287
9.13	Advantages of MAS	288
9.14	Limitations of MAS	289
9.15	Present Constraints and Future Directions	289
9.16	Achievements	292
10	Genomic Selection	295
10.1	Introduction	295
10.2	Genome-Wide Selection	295
10.3	A Generalized Procedure for Genomic Selection	296
10.4	Training Population	297
10.4.1	Genetic Composition	297
10.4.2	Population Size	298
10.4.3	Marker Density	299
10.5	Computation of Genomic Estimated Breeding Values	299
10.5.1	Stepwise Regression	299
10.5.2	Ridge Regression	300
10.5.3	Bayesian Approach	300
10.5.4	Semi-parametric Regression Methods	301
10.5.5	Machine Learning Methods	301
10.6	Factors Affecting the Accuracy of GEBV Estimates	302
10.6.1	The Method of Estimation of Marker Effects	302
10.6.2	The Polygenic Effect Term Based on Kinship	302
10.6.3	The Method of Phenotypic Evaluation of Training Population	303
10.6.4	The Marker Type and Density	303
10.6.5	Trait Heritability and the Number of QTLs Affecting the Trait	303
10.6.6	The Breeding Population	304
10.7	Effects of Genomic Selection on Genetic Diversity	304
10.8	Integration of Genomic Selection in Breeding Programs	305

10.9	Effectiveness of Genomic Selection	308
10.10	Advantages of Genomic Selection	308
10.11	Limitations of Genomic Selection	310
10.12	Future Directions	311
11	Phylogenetic Relationships and Genetic Diversity	313
11.1	Introduction	313
11.2	Estimation of Genetic Distance/Similarity	313
11.2.1	Estimation of Genetic Distance from Morphological Trait Data	313
11.2.2	Estimation of Genetic Distance from Molecular Marker Data	314
11.2.3	Estimation of Genetic Distance from Populations	315
11.2.4	Choice of the Genetic Distance Measure	315
11.3	Genetic Diversity Analysis: Phylogenetic Relationships	316
11.3.1	Cluster Analysis	317
11.3.2	Principal Component Analysis	318
11.3.3	Principal Coordinate Analysis	319
11.3.4	Multidimensional Scaling	320
11.3.5	Determination of the Optimal Number of Clusters	321
11.3.6	Choice of Clustering Method	321
11.3.7	Use of Diverse Datasets	322
11.3.8	Resampling Techniques	322
11.4	Genetic Diversity Analysis: Conservation of Genetic Resources	323
11.4.1	Germplasm Conservation	323
11.4.2	Applications of Molecular Markers in Germplasm Conservation	323
11.4.3	Conservation of Wild Species	327
11.5	Genetic Diversity Analysis: Prediction of Heterotic Pools and Heterotic Combinations	327
11.5.1	Genetic Basis of Heterosis	329
11.5.2	Molecular Basis of Heterosis	330
11.5.3	Identification/Prediction of Heterotic Pools and Heterotic Cross Combinations	333
11.5.4	Molecular Markers in Resolution of the Genetic Basis of Heterosis	333
11.5.5	Molecular Markers for Identification/Prediction of Heterotic Pools and Heterotic Cross Combinations	334
12	Fingerprinting and Gene Cloning	341
12.1	Introduction	341
12.2	DNA Fingerprinting	341

12.3	Characterization of Lines and Hybrids for Intellectual Property Rights Protection	342
12.3.1	Plant Breeder's Rights	342
12.3.2	Description of Plant Varieties	344
12.3.3	Limitations of Molecular Markers	345
12.4	Assessment of Genetic Purity of Lines and Hybrids	346
12.5	In Silico Gene Prediction	348
12.6	Chromosome Walking	351
12.7	Chromosome Jumping	353
12.8	Positional Gene Cloning	355
12.8.1	The Three Steps of Positional Cloning	355
12.8.2	Positional Cloning of Some Plant Genes	358
12.8.3	Some Useful Tips for Positional Gene Cloning	359
12.8.4	Problems in Positional Cloning	360
12.9	Chromosome Landing	360
12.10	Positional Cloning of Quantitative Trait Loci	361
12.11	cDNA Sequencing in Positional Cloning	362
12.12	Achievements	363
13	High-Throughput SNP Genotyping	367
13.1	Introduction	367
13.2	High-Throughput Genotyping of Known SNP Loci	367
13.2.1	The Invader Technology	368
13.2.2	Pyrosequencing	370
13.2.3	KASP™ Genotyping Assay	371
13.2.4	TaqMan OpenArray Genotyping System	373
13.2.5	SNP Analysis by MALDI-TOF MS (The Homogeneous MassEXTEND Assay)	373
13.2.6	Nanofluidic Dynamic Array-Based Assays	376
13.2.7	The Array Tape Technology	376
13.2.8	The Illumina GoldenGate SNP Genotyping Platform	376
13.2.9	Molecular Inversion Probe Technology	381
13.2.10	Whole-Genome-Based Microarray Platforms	382
13.3	High-Throughput SNP Discovery and Genotyping	386
13.4	Reduced Representation Sequencing	386
13.4.1	Reduced Representation Libraries	386
13.4.2	Complexity Reduction of Polymorphic Sequences	389
13.5	Restriction Site-Associated DNA Sequencing	390
13.6	Low-Coverage Genotyping	393
13.6.1	Genotyping by Sequencing	394
13.6.2	Multiplexed Shotgun Genotyping	395
13.7	Applications of NGS-Based Marker Discovery and Genotyping Methods	396

13.8	A Comparison of NGS and Other SNP Genotyping Approaches	397
13.9	Reduced Representation Versus Whole-Genome Sequencing	397
13.10	SNP Discovery in Polyploids	398
13.11	Bioinformatics Tools for Marker Discovery from NGS Sequence Data	398
13.11.1	PoPoolation	398
13.11.2	RADtools 1.2.4	398
13.11.3	Stacks	398
13.11.4	TASSEL	399
13.11.5	SAMtools/BCFtools	399
13.12	Future Directions	399
14	Bioinformatics Tools and Databases for Genomics Research	401
14.1	Introduction	401
14.2	Representation of Nucleotide and Amino Acid Sequences	401
14.3	Bioinformatics Tools	402
14.3.1	AutoSNP	403
14.3.2	SNP2CAPS	403
14.3.3	TASSEL	404
14.3.4	STRUCTURE	405
14.3.5	Microarray Software	405
14.3.6	A C. Elegans Database (AceDB)	406
14.3.7	MAPMAN	406
14.3.8	GenScan	407
14.3.9	ClustalW	407
14.4	Bioinformatics Databases	409
14.4.1	GenBank	410
14.4.2	Phytozome	411
14.4.3	European Molecular Biology Laboratory Nucleotide Sequence Database	411
14.4.4	Swiss-Prot	412
14.4.5	UniProt Knowledgebase (UniProtKB)	412
14.4.6	Gramene	413
14.4.7	GrainGenes	413
14.4.8	MaizeGDB	414
14.4.9	RiceGeneThresher	414
14.4.10	Microarray Databases (ArrayExpress and Gene Expression Omnibus)	414
14.4.11	HarVEST	415
14.5	Sources of Multiple Databases and Tools	415
14.5.1	National Center for Biotechnology Information	415
14.5.2	Kyoto Encyclopedia of Genes and Genomes	416

14.5.3	Molecular Biology Database Collection	419
14.5.4	Architecture for Metabolomics (ArMet)	420
14.5.5	Database Search and Analysis Tools	421
14.5.6	Genamics SoftwareSeek	427
14.5.7	Sequence Manipulation Suite	427
14.5.8	PHYLIP	428
15	Phenomics	431
15.1	Introduction	431
15.2	Phenomics	431
15.3	The Imaging Technology	433
15.4	Advantages of Image-Based Phenotyping	434
15.5	Reflectance Imaging	435
15.5.1	Visual Imaging	435
15.5.2	Near Infrared Imaging	436
15.6	Infrared Imaging	436
15.7	Fluorescence Imaging	437
15.7.1	Chlorophyll Fluorescence	438
15.7.2	Green Fluorescence Protein	439
15.8	Magnetic Resonance Imaging	440
15.9	Multi-sensor Monitoring Approaches	440
15.10	Field-Based Phenomics	440
15.11	Morphological and Growth Analyses	443
15.11.1	Dynamic Measurement of Leaf Area	443
15.11.2	Plant Biomass Estimation	444
15.11.3	Basic Plant Growth Analysis	445
15.11.4	Assessment of Structure/Development	446
15.11.5	Measurement of Senescence/Necrosis	446
15.11.6	Analysis of Root Systems	446
15.11.7	Seed and Fruit Phenotyping	447
15.11.8	Laser Scanning: 3-D Plant Morphology	447
15.12	Analyses of Chemical and Physiological Parameters	448
15.12.1	Estimation of Relative Chlorophyll Content	448
15.12.2	Monitoring Photosynthesis	449
15.12.3	Assessment of Water Use	449
15.12.4	Estimation of Soil Water Content	450
15.12.5	Analysis of Chemical Composition	451
15.13	Biotic Stress Detection	451
15.14	Monitoring Drought Stress	452
15.14.1	Stomatal Conductance	452
15.14.2	Leaf/Canopy Temperature	453
15.14.3	Visible Imaging	453
15.14.4	IR Thermography	453
15.14.5	Chlorophyll Fluorescence	454
15.14.6	Estimation of Tissue Water Content	454

15.15	Molecular Biomarkers	455
15.16	Image Analysis	455
15.17	Image Analysis Software	456
15.17.1	ImageJ	456
15.17.2	HTPheno	456
15.17.3	Rosette Tracker	457
15.17.4	Martrack Leaf	457
15.17.5	HPGA (High-throughput Plant Growth Analysis)	457
15.17.6	Root System Analyzer	458
15.17.7	SmartRoot	458
15.17.8	RootReader2D	458
15.17.9	RootReader3D	459
15.18	Applications of Phenomics	459
15.19	Achievements	459
15.20	Future Directions	460
Glossary		463
References		485
Author Index		501
Subject Index		507

About the Authors

Brahma Deo Singh Brahma Deo Singh is currently Emeritus Professor at School of Biotechnology, Banaras Hindu University, Varanasi, India. He obtained his bachelor's degree in agriculture from Allahabad Agricultural Institute, Allahabad, India, and master's degree in Agricultural Botany from Government Agricultural College, Kanpur, India, with first position in the university and was awarded the University Gold Medal. He earned his Ph.D. degree from University of Saskatchewan, Saskatoon, Canada. Prof. Singh has 40 years of teaching and research experience in the areas of genetics and breeding of pulse crops, plant tissue culture, biological nitrogen fixation and molecular markers. He has published over 150 research papers in reputed journals and authored several books in genetics, plant breeding and biotechnology. He was awarded the First Prize of the Dr. Rajendra Prasad Puraskar in 1987 and 1990 by the Indian Council of Agricultural Research, New Delhi, for the books *PadapPrajanan* and *Anuvanshiki*, respectively.

Ashok Kumar Singh Ashok Kumar Singh, Fellow of National Academy of Agricultural Sciences, India, is the present Head, Division of Genetics at the prestigious Indian Agricultural Research Institute, New Delhi. He completed his bachelor's and master's degrees from Banaras Hindu University, Varanasi, India, and earned his Ph.D. degree from the institute where he is currently working as a dedicated teacher and rice breeder. He has been associated with the development of 11 Basmati rice varieties, including the first superfine grain aromatic rice hybrid Pusa RH 10, which combine earliness with higher yield and higher per day productivity with excellent grain and cooking quality. He has successfully integrated marker-assisted selection for incorporating resistances to bacterial blight, blast, brown plant hopper drought, salinity and submergence in rice varieties. His current research interests include TILLING, bio-prospecting for genes and novel alleles and marker-assisted breeding in rice. He is well recognized for his contributions to Basmati rice breeding and marker-assisted breeding. He has