

# Molecular Plant Breeding

Yunbi Xu



# Molecular Plant Breeding

---

**Yunbi Xu**

*International Maize and Wheat Improvement Center (CIMMYT)  
Apdo Postal 6-641  
06600 Mexico, DF  
Mexico*



**CABI is a trading name of CAB International**

CABI Head Office  
Nosworthy Way  
Wallingford  
Oxfordshire OX10 8DE  
UK

CABI North American Office  
875 Massachusetts Avenue  
7th Floor  
Cambridge, MA 02139  
USA

Tel: +44 (0)1491 832111  
Fax: +44 (0)1491 833508  
E-mail: [cabi@cabi.org](mailto:cabi@cabi.org)  
Web site: [www.cabi.org](http://www.cabi.org)

Tel: +1 617 395 4056  
Fax: +1 617 354 6875  
E-mail: [cabi-nao@cabi.org](mailto:cabi-nao@cabi.org)

©CAB International 2010. All rights reserved. No part of this publication may be reproduced in any form or by any means, electronically, mechanically, by photocopying, recording or otherwise, without the prior permission of the copyright owners.

A catalogue record for this book is available from the British Library, London, UK.

**Library of Congress Cataloging-in-Publication Data**

Xu, Yunbi.

Molecular plant breeding / Yunbi Xu.

p. cm.

ISBN 978-1-84593-392-0 (alk. paper)

1. Crop improvement. 2. Plant breeding. 3. Crops--Molecular genetics. 4. Crops--Genetics. I. Title.

SB106.147X8 2010

631.5'233--dc22

20009033246

ISBN: 978 1 84593 392 0

Typeset by SPi, Pondicherry, India.

Printed and bound in the UK by MPG Books Group.

## **In Memoriam**

Norman Ernest Borlaug  
(25 March 1914–12 September 2009)

Norman Borlaug was one of the greatest men of our times – a steadfast champion and spokesman against hunger and poverty. He dedicated his 95 richly lived years to filling the bellies of others, and is credited by the United Nations' World Food Program with saving more lives than any other man in history.

An American plant pathologist who spent most of his years in Mexico, it was Dr Borlaug's high-yielding dwarf wheat varieties that prevented wide-spread famine in South Asia, specifically India and Pakistan, and also in Turkey. Known as the 'Green Revolution', this feat earned him the Nobel Peace Prize in 1970. He was instrumental in establishing the International Maize and Wheat Improvement Center, known by its Spanish acronym CIMMYT, and later the Consultative Group of International Agricultural Research (CGIAR), a network of 15 agricultural research centres.

Dr Borlaug spent time as a microbiologist with DuPont before moving to Mexico in 1944 as a geneticist and plant pathologist to develop stem rust resistant wheat cultivars. In 1966 he became the director of CIMMYT's Wheat Program, seconded from the Rockefeller Foundation. His full-time employment with the Center ended in 1979, although he remained a part-time consultant until his death. In 1984 he began a new career as a university professor and went on to establish the World Food Prize, which honours the achievements of individuals who have advanced human development by improving the quality, quantity or availability of food in the world. In 1986, he joined forces with former US President Jimmy Carter and the Nippon Foundation of Japan, under the chairmanship of Ryoichi Sasakawa, to establish Sasakawa Africa Association (SAA) to address Africa's food problems. Since then, more than 1 million small-scale African farmers in 15 countries have been trained by SAA in improved farming techniques.

Dr Borlaug influenced the thinking of thousands of agricultural scientists. He was a path-breaking wheat breeder and, equally important, his stature enabled him to influence politicians and leaders around the world. His legacy and his work ethic – to get things done and not mind getting your hands dirty – influenced us all and remain CIMMYT guiding principles.

We will honor Dr Borlaug's memory by carrying forward his mission and spirit of innovation: applying agricultural science to help smallholder farmers produce more and better-quality food using fewer resources. At stake is no less than the future of humanity, for, as Borlaug said: 'The destiny of world civilization depends upon providing a decent standard of living for all'. His presence will never really leave CIMMYT; it is embedded in our soul.

**Thomas A. Lumpkin**  
Director General, CIMMYT

**Marianne Bänziger**  
Deputy Director General for Research and Partnerships, CIMMYT

**Hans-Joachim Braun**  
Director for Global Wheat Program, CIMMYT

# CONTENTS

## **In memoriam- Dr. Norman Ernest Borlaug**

by Drs. Thomas A. Lumpkin, Marianne Banziger, and Hans-Joachim Braun

## **Preface**

## **Foreword by Dr. Norman Borlaug**

## **Foreword by Dr. Ronald L. Phillips**

## **1 Introduction 1**

- 1.1 Domestication of Crop Plants 1
- 1.2 Early Efforts at Plant Breeding 3
- 1.3 Major Developments in the History of Plant Breeding 4
  - 1.3.1 Breeding and hybridization
  - 1.3.2 Mendelian genetics
  - 1.3.3 Selection
  - 1.3.4 Breeding types and polyploidy
  - 1.3.5 Genetic diversity and germplasm conservation
  - 1.3.6 Quantitative genetics and genotype-by-environment interaction
  - 1.3.7 Heterosis and hybrid breeding
  - 1.3.8 Refinement of populations
  - 1.3.9 Cell totipotency, tissue culture and somaclonal variation
  - 1.3.10 Genetic engineering and gene transfer
  - 1.3.11 DNA markers and genomics
- 1.4 Genetic Variation 9
  - 1.4.1 Crossover, genetic drift and gene flow
  - 1.4.2 Mutation
- 1.5 Quantitative Traits: Variance, Heritability and Selection Index 10
  - 1.5.1 Qualitative and quantitative traits
  - 1.5.2 The concept of allelic and genotypic frequencies
  - 1.5.3 Hardy-Weinberg equilibrium (HWE)
  - 1.5.4 Population means and variances
  - 1.5.5 Heritability
  - 1.5.6 Response to selection
  - 1.5.7 Selection index and selection for multiple traits
  - 1.5.8 Combining ability
  - 1.5.9 Recurrent selection
- 1.6 The Green Revolution and the Challenges Ahead 16
- 1.7 Objectives of Plant Breeding 17
- 1.8 Molecular Breeding 18

## **2 Molecular Breeding Tools: Markers and Maps 21**

- 2.1 Genetic Markers 21
  - 2.1.1 Classical markers
    - Morphological markers
    - Cytological markers
    - Protein markers

- 2.1.2 DNA markers
  - RFLP
  - RAPD
  - AFLP
  - SSR
  - SNP
  - Diversity array technology
  - Genic and functional markers
- 2.2 Molecular Maps 43
  - 2.2.1 Chromosome theory and linkage
  - 2.2.2 Genetic linkage mapping
    - Developing mapping populations
    - Interference and mapping function
    - Segregation and linkage tests
    - Maximum likelihood estimation (MLE) of recombinant frequency
    - Likelihood ratio and linkage test
    - Multi-point analysis and ordering a set of markers
    - Linkage mapping in the presence of genotyping errors
    - Molecular maps in plants
  - 2.2.3 Integration of genetic maps
    - Integration of conventional and molecular maps
    - Integration of multiple molecular maps
    - Integration of genetic and physical maps

### **3 Molecular Breeding Tools: Omics and Arrays 59**

- 3.1 Molecular Techniques in Omics 59
  - 3.1.1 2-Dimensional gel electrophoresis
  - 3.1.2 Mass spectrometry
  - 3.1.3 Yeast two-hybrid system
  - 3.1.4 Serial analysis of gene expression
  - 3.1.5 Quantitative real-time PCR
  - 3.1.6 Subtraction suppressive hybridization
  - 3.1.7 *In situ* hybridization
- 3.2 Structural Genomics 68
  - 3.2.1 Genome organization
    - Major differences among various genomes
    - Sequence complexity
  - 3.2.2 Physical mapping
    - DNA libraries
    - Construction of large insert genomic libraries
    - Physical mapping
  - 3.2.3 Genome sequencing
    - Technical developments in DNA sequencing
    - Sequencing strategies
    - Genome filtering strategies
    - Plant genome sequences
  - 3.2.4 cDNA sequencing
    - Why cDNA sequencing
    - cDNA libraries
    - cDNA sequencing
- 3.3 Functional Genomics 81
  - 3.3.1 Transcriptomics

- 3.3.2 Proteomics
  - Protein extraction
  - Protein identification and quantification
  - Protein profiling
  - Protein-protein interactions
  - Post-translation modifications
- 3.3.3 Metabolomics
- 3.4 Phenomics 91
  - 3.4.1 Importance of phenotypes in genomics
  - 3.4.2 Phenomics in plants
- 3.5 Comparative Genomics 93
  - 3.5.1 Comparative maps
    - Importance of comparative maps
    - An example: *Arabidopsis*-tomato comparative map
  - 3.5.2 Collinearity
    - Orthology and paralogy
    - Macrocollinearity
    - Microcollinearity
    - Importance of genome collinearity
- 3.6 Array Technologies in Omics 100
  - 3.6.1 Production of arrays
    - Types of arrays
    - Source of arrays
    - Array content
    - Slide substrates
    - Arrays and spotting pins
  - 3.6.2 Experimental design
  - 3.6.3 Sample preparation
  - 3.6.4 Labelling
  - 3.6.5 Hybridization and post-hybridization washes
  - 3.6.6 Data acquisition and quantification
  - 3.6.7 Statistical analysis and data mining
  - 3.6.8 Protein microarrays and others
  - 3.6.9 Universal chip or microarray
  - 3.6.10 Whole-genome analysis using tiling microarrays
  - 3.6.11 Array-based genotyping

#### **4 Populations in Genetics and Breeding 113**

- 4.1 Properties and Classification of Populations 113
  - 4.1.1 Genetic constitution-based classification
  - 4.1.2 Genetic maintenance-based classification
  - 4.1.3 Genetic background-based classification
  - 4.1.4 Origin-based classification
    - Populations of natural cultivars
    - Populations formed by planned matings
    - Inbreeding populations
- 4.2 Doubled Haploids (DHs) 116
  - 4.2.1 Haploid production
    - Chromosome or genome elimination
    - Ovary culture or gynogenesis
    - Anther culture or androgenesis
    - Semigamy

- Inducer-based approach
  - 4.2.2 Diploidization of haploid plants
  - 4.2.3 Evaluation of DH lines
    - Randomness
    - Stability
  - 4.2.4 Quantitative genetics of DHs
    - Expected gain from selection
    - Implications of epistatic effects
  - 4.2.5 Applications of DH populations in genomics
  - 4.2.6 Applications of DH populations in plant breeding
  - 4.2.7 Limitations and future prospects
- 4.3 Recombinant Inbred Lines (RILs) 131
  - 4.3.1 Inbreeding and its genetic effects
  - 4.3.2 Development of RILs
    - Single seed descent (SSD) method
    - Advantages and disadvantages of SSD procedures
  - 4.3.3 Map distance and recombinant fraction in RIL populations
  - 4.3.4 Construction of genetic maps using RILs
  - 4.3.5 Interrelated RILs and nested RIL populations
    - Interrelated RILs
    - Multi-way or nested RIL populations
- 4.4 Near-isogenic Lines (NILs) 138
  - 4.4.1 Backcrossing and its genetic effects
  - 4.4.2 Other methods for production of NILs
  - 4.4.3 Introgression line libraries
    - Rice ILs
    - The *Lycopersicon pennelli* ILs
  - 4.4.4 Gene tagging strategy using NILs
  - 4.4.5 Theoretical considerations in genetic mapping using NILs
    - Backcross introgression without selection for the RP phenotype
    - Backcross introgression with selection for the RP phenotype
  - 4.4.6 Application of NILs in gene tagging
- 4.5 Cross-population Comparison: Recombination Frequency and Selection 145
  - 4.5.1 Recombination frequencies across populations
  - 4.5.2 Unintentional selection during the process of population development
    - Selection pressure associated with DH development
    - Selection pressure involved in RIL development
    - Genetics of selection associated segregation distortion
    - Implications for genetics and plant breeding

## **5 Plant Genetic Resources: Management, Evaluation and Enhancement 151**

- 5.1 Genetic Erosion and Potential Vulnerability 152
  - 5.1.1 Genetic erosion
  - 5.1.2 Genetic vulnerability
- 5.2 The Concept of Germplasm 155
  - 5.2.1 A generalized concept of germplasm
  - 5.2.2 Classical germplasm
  - 5.2.3 Artificial or synthetic germplasm
  - 5.2.4 *In situ* and *ex situ* conservation
- 5.3 Collection/Acquisition 161
  - 5.3.1 Several issues on germplasm collections
  - 5.3.2 Core collections



- 5.4 Maintenance, Rejuvenation and Multiplication 166
  - 5.4.1 *In vitro* storage techniques
  - 5.4.2 Cryopreservation
  - 5.4.3 Synthetic seeds and storage of DNA
  - 5.4.4 Rejuvenation and multiplication
- 5.5 Evaluation 171
  - 5.5.1 Marker-assisted germplasm evaluation
  - 5.5.2 *In vitro* evaluation
  - 5.5.3 Genetic diversity
    - Factors impacting genetic diversity
    - Measurement of diversity
    - Germplasm classification
    - Phylogenetics
  - 5.5.4 Collection redundancies and gaps
  - 5.5.5 Genetic drafts/shifts and gene flow
  - 5.5.6 Unique germplasm
  - 5.5.7 Allele mining
- 5.6 Germplasm Enhancement 186
  - 5.6.1 Purification of germplasm collections
  - 5.6.2 Tissue culture and transformation in germplasm enhancement
  - 5.6.3 Gene introgression in germplasm enhancement
- 5.7 Information Management 188
  - 5.7.1 Information system
  - 5.7.2 Standardization of data collection
  - 5.7.3 Information integration and utilization
- 5.8 Future Prospects 192

## **6 Molecular Dissection of Complex Traits: Theory 195**

- 6.1 Single Marker-based Approaches 197
  - 6.1.1 Assumptions
  - 6.1.2 Comparison of marker means
  - 6.1.3 ANOVA
  - 6.1.4 Regression approach
  - 6.1.5 Likelihood approach
- 6.2 Interval Mapping 202
  - 6.2.1 Assumptions
  - 6.2.2 Likelihood approach
- 6.3 Composite Interval Mapping 205
  - 6.3.1 Basis
  - 6.3.2 Model
  - 6.3.3 Likelihood analysis
  - 6.3.4 Hypothesis test
  - 6.3.5 Selection of markers as cofactors
  - 6.3.6 Inclusive composite interval mapping
- 6.4 Multiple Interval Mapping 209
  - 6.4.1 Multiple interval mapping model and likelihood analysis
  - 6.4.2 Model selection
    - Pre-model selection
    - Model selection using multiple interval mapping
    - Stop rules
  - 6.4.3 Estimating genotypic values and variance components of QTL effects
- 6.5 Multiple Populations/Crosses 214

- 6.5.1 Experimental designs
- 6.5.2 QTL for multiple crosses
- 6.5.3 Pooled analysis
- 6.6 Multiple QTL 217
  - 6.6.1 Reality of multiple QTL
  - 6.6.2 Selecting a class of QTL models
  - 6.6.3 Multiple QTL with epistasis
- 6.7 Bayesian Mapping 219
  - 6.7.1 Advantages of Bayesian mapping
  - 6.7.2 Bayesian mapping statistics: a brief overview
  - 6.7.3 Bayesian mapping methods
    - Bayesian shrinkage estimation (BSE) method
    - Model selection
    - Penalized maximum likelihood (PML) method
- 6.8 Linkage Disequilibrium Mapping 223
  - 6.8.1 Why linkage disequilibrium mapping?
  - 6.8.2 Measurement of linkage disequilibrium
  - 6.8.3 Factors affecting linkage disequilibrium
  - 6.8.4 Methods for linkage disequilibrium mapping
    - The transmission disequilibrium test and derivatives
    - Structured association
    - Principal component analysis
    - Mixed models
    - Quantitative inbred pedigree disequilibrium
    - Bayesian methods
  - 6.8.5 Applications of linkage disequilibrium mapping
- 6.9 Meta-analysis 233
  - 6.9.1 Meta-analysis of QTL locations
  - 6.9.2 Meta-analysis of QTL maps
  - 6.9.3 Meta-analysis of QTL effects
  - 6.9.4 Examples of meta-analysis
- 6.10 *In Silico* Mapping 237
  - 6.10.1 Pros and cons
  - 6.10.2 Mixed-model approach
  - 6.10.3 Statistical power
- 6.11 Sample Size, Power and Thresholds 239
  - 6.11.1 Power and sample sizes
    - Factors determining the required sample size
    - Effects of dominance
    - Effect of linkage: multiple linked QTL
  - 6.11.2 Cross validation and sample size
  - 6.11.3 Confidence interval of QTL location
  - 6.11.4 QTL thresholds
    - QTL thresholds for interval mapping
    - Genome-wide thresholds
    - Permutation and thresholds
  - 6.11.5 False discovery rate
- 6.12 Summary and Prospects 247

## **7 Molecular Dissection of Complex Traits: Practice 249**

- 7.1 QTL Separating 249
  - 7.1.1 Mapping approaches

- Fine mapping
  - Minor QTL mapping
  - Regional mapping
  - 7.1.2 Screening for allele dispersion
- 7.2 QTL for Complicated Traits 258
  - 7.2.1 Trait components
  - 7.2.2 Correlated traits
  - 7.2.3 Qualitative-quantitative traits
  - 7.2.4 Seed traits
- 7.3 QTL Mapping across Species 262
- 7.4 QTL across Genetic Backgrounds 264
  - 7.4.1 Homogeneous genetic backgrounds
  - 7.4.2 Heterogeneous genetic backgrounds
  - 7.4.3 Epistasis
    - Importance of epistasis
    - Statistical methods for epistatic QTL
    - Population strategies for epistatic QTL analysis
  - 7.4.4 Multiple alleles at a locus
- 7.5 QTL across Growth and Developmental Stages 270
  - 7.5.1 Dynamic traits
  - 7.5.2 Dynamic mapping
  - 7.5.3 Statistical methods for dynamic mapping
- 7.6 Multiple Traits and Gene Expression 274
  - 7.6.1 Features of gene expression
  - 7.6.1 Examples of eQTL in plants
- 7.7 Selective Genotyping and Pooled DNA Analysis 277
  - 7.7.1 Major-gene controlled traits
  - 7.7.2 Quantitative traits
  - 7.7.3 The power of selective genotyping and pooled DNA analysis
  - 7.7.4 Use of selective genotyping and pooled DNA analysis
    - 'All-in-one plate'- genetic mapping of all target traits in one step
    - Genome-wide association mapping
    - Integration with selective genotyping

## **8 Marker-assisted Selection: Theory 286**

- 8.1 Components of Marker-assisted Selection 288
  - 8.1.1 Genetic markers and maps
  - 8.1.2 Marker characterization
    - Allele number
    - Polymorphism Information Content (PIC) value
    - Informative markers
  - 8.1.3 Validation of marker-trait associations
  - 8.1.4 Genotyping and high-throughput genotyping systems
  - 8.1.5 Data management and delivery
- 8.2 Marker-assisted Gene Introgression 293
  - 8.2.1 Marker-assisted foreground selection
    - Selection using single markers
    - Selection using bracket markers
    - Selection using multiple markers for multiple traits
  - 8.2.2 Marker-assisted background selection
    - Concept of graphical genotypes
    - Requirements for deducing graphical genotypes

- Assumptions employed in developing graphical genotypes
- Application of graphical genotypes
- 8.2.3 Donor genome content in BC generations
- 8.2.4 Linkage drag in gene introgression
- 8.2.5 Effect of genome size on gene introgression
- 8.2.6 Background selection at carrier chromosome
- 8.2.7 Whole genome selection for genetic background
- 8.2.8 Multiple gene introgression by repeated backcrossing
- 8.3 Marker-assisted Gene Pyramiding 308
  - 8.3.1 Gene-pyramiding schemes
    - Definition
    - Pedigree height
    - Number of pedigrees
    - Gene transmission probability through a pedigree
    - Minimum population sizes necessary to obtain ideotype
    - A case study
  - 8.3.2 Crossing and selection strategies
    - Comparing biparental, back- and top-crosses
    - Minimizing the total number of marker assays within sequential culling
    - Enrichment of favourable alleles at early generations
  - 8.3.3 Gene-pyramiding for different traits
  - 8.3.4 Marker-assisted recurrent selection versus genome-wide selection
- 8.4 Selection for Quantitative Traits 318
  - 8.4.1 Selection based on phenotypic values
  - 8.4.2 Selection based on marker scores
  - 8.4.3 Index selection
  - 8.4.4 Genotypic selection
  - 8.4.5 Integrated marker-assisted selection
  - 8.4.6 Response to marker-assisted selection
- 8.5 Long-term Selection 327
  - 8.5.1 Long-term selection on maize
    - Procedure
    - Limits to selection
    - Explanation of progress
    - Marker-assisted evaluation
    - Application to plant breeding
  - 8.5.2 Divergent selection in rice

## 9 Marker-assisted Selection: Practice 336

- 9.1 Selection Schemes for Marker-assisted Selection 337
  - 9.1.1 Selection without testcrossing or progeny test
  - 9.1.2 Selection independent of environments
  - 9.1.3 Selection without laborious field or intensive laboratory work
  - 9.1.4 Selection as an early breeding stage
  - 9.1.5 Selection for multiple genes and multiple traits
  - 9.1.6 Whole genome selection
- 9.2 Bottlenecks in Application of Marker-assisted Selection 339
  - 9.2.1 Effective marker-trait association
  - 9.2.2 Cost-effective and high-throughput genotyping systems
  - 9.2.3 Phenotyping and sample tracking
  - 9.2.4 Epistasis and genotype-by-environment interaction
- 9.3 Reducing Costs and Increasing Scale and Efficiency 344

- 9.3.1 Cost-benefit analysis
- 9.3.2 Seed DNA-based genotyping and MAS system
- 9.3.3 Integrated diversity analysis, genetic mapping and MAS
- 9.3.4 Developing breeding strategies for simultaneous improvement of multiple traits
- 9.4 Traits Most Suitable for MAS 350
  - 9.4.1 Traits requiring testcrossing or progeny testing
    - Cytoplasmic male sterility and fertility restoration
    - Outcrossing
    - Wide compatibility
    - Heterosis
  - 9.4.2 Environment-dependent traits
    - Photoperiod/temperature sensitivity
    - Environment-induced genic male sterility
    - Biotic and abiotic stresses
  - 9.4.3 Seed and quality traits
    - Seed traits
    - Hybrid seed traits
    - Quality traits
- 9.5 Marker-assisted Gene Introgression 356
  - 9.5.1 Marker-assisted gene introgression from wild relatives
  - 9.5.2 Marker-assisted gene introgression from elite germplasm
  - 9.5.3 Marker-assisted gene introgression for drought tolerance
  - 9.5.4 Marker-assisted gene introgression for quality traits
    - Rice
    - Wheat
    - Maize
    - Barley
- 9.6 Marker-assisted Gene Pyramiding 363
  - 9.6.1 Gene pyramiding for major genes
  - 9.6.2 Gene pyramiding through marker-assisted recurrent selection
- 9.7 Marker-assisted Hybrid Prediction 367
  - 9.7.1 Genetic basis of heterosis
    - QTL for heterosis
    - Gene expression analysis of heterosis
    - Prospects on genetic basis of heterosis
  - 9.7.2 Heterotic groups
    - Construction of heterotic groups based on hybrid performance
    - Construction of heterotic groups using molecular marker information
    - Future direction
  - 9.7.3 Marker-assisted hybrid prediction
    - Genome-wide heterozygosity and hybrid prediction
    - Hybrids are more predictable within than between heterotic groups
    - Heterosis-associated markers and hybrid prediction
    - Favourable alleles combination and hybrid prediction
    - Conclusions and prospects
- 9.8 Opportunities and Challenges 378
  - 9.8.1 Molecular tools and breeding systems
  - 9.8.2 Crop-specific issues
  - 9.8.3 Quantitative traits
  - 9.8.4 Genetic networks
  - 9.8.5 Marker-assisted selection in developing countries

## **10 Genotype-by-environment Interaction 381**

- 10.1 Multi-environment Trials 383
  - 10.1.1 Experimental design
  - 10.1.2 Basic data analysis and interpretation
    - Genotype  $\times$  location interaction
    - Genotype  $\times$  year interaction
    - Genotype  $\times$  year  $\times$  location interaction
- 10.2 Environmental Characterization 386
  - 10.2.1 Classification of environments
  - 10.2.2 GIS and environment characterization
  - 10.2.3 Selection of locations for testing
- 10.3 Stability of Genotype Performance 394
  - 10.3.1 Linear-bilinear models for studying GEI
  - 10.3.2 GGE bi-plot analysis
  - 10.3.3 Mixed model
- 10.4 Molecular Dissection of GEI 402
  - 10.4.1 Partition of environmental factors
  - 10.4.2 QTL mapping across environments
  - 10.4.3 QTL mapping with incorporated GEI
    - Mixed models
    - Factorial regression model
    - Structural equation model
    - QEI mapping examples
  - 10.4.4 Utilization of MET and genotypic data
- 10.5 Breeding for GEI 410
  - 10.5.1 Breeding for resource-limited environments
  - 10.5.2 Breeding for adaptation and stability
  - 10.5.3 Assessment of GEI in breeding programmes
    - Measure interaction at intermediate growth stages
    - Multi-environmental testing at early stages of breeding
    - Unbalanced data
  - 10.5.4 MAS for GEI
- 10.6 Future Perspectives 414

## **11 Isolation and Functional Analysis of Genes 417**

- 11.1 *In Silico* Prediction 419
  - 11.1.1 Evidence-based gene prediction
  - 11.1.2 Homology-based gene prediction
    - Comparison with EST/cDNA databases
    - Comparison with protein sequence databases
    - Comparison of a translated genomic sequence with translated nucleotide databases
    - Comparison of homologous genomic sequences
  - 11.1.3 *Ab initio* gene prediction
    - Splice site prediction
    - Exon prediction
    - Gene modeling
  - 11.1.4 Gene prediction by integrated methods
  - 11.1.5 Detecting protein function from genomic sequences
    - Intrinsic methods
    - Homology search methods
    - Gene context methods

- 11.2 Comparative Approaches for Gene Isolation 426
  - 11.2.1 Genomic bases of comparative approaches
  - 11.2.2 Experimental procedures involved in comparative analysis
    - Basic procedures
    - Examples
  - 11.2.3 Cloning QTL facilitated by related major genes
- 11.3 Cloning Based on cDNA Sequencing 431
  - 11.3.1 Generation of ESTs
  - 11.3.2 Generation of full-length cDNAs
  - 11.3.3 Full-length cDNA sequencing
  - 11.3.4 Directed EST screens to identify specific genes
  - 11.3.5 Full-length cDNAs for the discovery and annotation of genes
- 11.4 Positional Cloning 435
  - 11.4.1 Theoretical consideration of positional cloning
  - 11.4.2 Examples of positional cloning
    - Preliminary genetic mapping
    - Fine mapping and candidate gene identification
    - Complementation tests
    - Exploration of ORFX identify
    - Characterization of ORFX
- 11.5 Identification of Genes by Mutagenesis 441
  - 11.5.1 Generation of mutant populations
  - 11.5.2 Insertional mutagenesis
    - T-DNA tagging
    - Transposon tagging
    - Retrotransposon tagging
    - Activation tagging
    - Entrapment/enhancer/promoter tagging
  - 11.5.3 Non-tagging mutagenesis
    - Point mutations
    - Deletion mutagenesis
  - 11.5.4 RNA interference
  - 11.5.5 Gene isolation via mutagenesis
- 11.6 Other Approaches for Gene Isolation 454
  - 11.6.1 Gene expression analysis
  - 11.6.2 Using homologous probes

## **12 Gene Transfer and Genetically Modified Plants 458**

- 12.1 Plant Tissue Culture and Genetic Transformation 458
  - 12.1.1 Plant tissue culture
  - 12.1.2 Genetic transformation
  - 12.1.3 Development of core plant transformation facilities
- 12.2 Transformation Approaches 461
  - 12.2.1 *Agrobacterium*-mediated transformation
    - Agrobacterium* strains
    - Application in cereals
    - Host plant genes involved in *Agrobacterium*-mediated transformation
  - 12.2.2 Particle bombardment
    - Particle bombardment facilitates a wide range of transformation strategies
    - Particle bombardment has no biological constrains or host limitations
    - Diverse cell types can be targeted efficiently for foreign DNA delivery
    - Vectors are not required for particle bombardment

- High molecular weight DNA delivery into plant cells
- Particle bombardment is the most convenient way to achieve organelle transformation
- Comparison with other methods
- 12.2.3 Electroporation and other direct gene transfer approaches
- 12.3 Expression Vectors 468
  - 12.3.1 Binary vectors
  - 12.3.2 Gateway-based binary vectors
  - 12.3.3 Choice of transformation vectors
- 12.4 Selectable Marker Genes 473
  - 12.4.1 Functions of selection marker genes
  - 12.4.2 Selectable marker genes for plants
    - Antibiotic resistance genes
    - Herbicide tolerance genes
    - Engineering detoxification of herbicides that inhibits glutamine synthase
    - Engineering tolerance to and detoxification of herbicides that inhibits 5-enol-pyruvylshikimate -3-phosphate synthase
  - 12.4.3 Positive selection
  - 12.4.4 Elimination of selectable marker genes from transgenic plants
    - Use of co-transformation
    - Removal of marker genes and other unnecessary segments by recombination
    - Use of transposons
    - Use of homologous recombination
    - Use of positive markers
- 12.5 Transgene Integration, Expression and Localization 480
  - 12.5.1 Transgene integration
  - 12.5.2 Transgene expression
  - 12.5.3 Confirmation of transgene and analysis of gene expression in transgenic plants
  - 12.5.4 Reporter genes
    - $\beta$ -Glucuronidase
    - Luciferase
    - Anthocyanin biosynthetic pathway genes
    - Green fluorescent protein
  - 12.5.5 Promoters
    - Promoters for constitutive transgene expression
    - Promoters for non-constitutive transgene expression
  - 12.5.6 Transgene inactivation
- 12.6 Transgene Stacking 487
  - 12.6.1 Sexual crosses
  - 12.6.2 Co-transformation via plasmids
  - 12.6.3 Co—transformation via particle bombardment
- 12.7 Transgenic Crop Commercialization 492
  - 12.7.1 Commercial targets
  - 12.7.2 Current status of transgenic crop commercialization
  - 12.7.3 Regulating transgenic crops
    - Risk assessment
    - Regulatory systems
  - 12.7.4 Product release and marketing strategies
  - 12.7.5 Monitoring transgenes
- 12.8 Perspectives 499

## 13 Intellectual Property Rights and Plant Variety Protection 501



- 13.1 Intellectual Property and Plant Breeders' Rights 502
  - 13.1.1 Basic aspects of intellectual property
  - 13.1.2 Intellectual property rights in plant breeding
- 13.2 Plant Variety Protection: Needs and Impacts 504
  - 13.2.1 Needs for protection of crop cultivars
    - The exercise of the breeders' rights
    - The derived benefits
  - 13.2.2 Impacts of plant variety protection
    - Plant breeding
    - Agricultural production and trade
    - Transfer of technology and know-how
    - Breeding strategies
    - NARI organizations
    - International agricultural research
- 13.3 International Agreements Affecting Plant Breeding 509
  - 13.3.1 The UPOV Convention and UPOV
    - Distinctness, uniformity and stability (DUS)
    - Essentially derived varieties/cultivars
    - Farmer's privilege
    - Breeders' rights
    - Breeders' exemption
  - 13.3.2 The 1983 International Undertaking on Plant Genetic Resources
  - 13.3.3 The 1992 Convention on Biological Diversity
  - 13.3.4 The 1994 TRIPS Agreement
  - 13.3.5 The 2001 International Treaty on Plant Genetic Resources for Food and Agriculture
- 13.4 Plant Variety Protection Strategies 518
  - 13.4.1 Plant variety protection for plant breeders' rights
  - 13.4.2 Patents
  - 13.4.3 Biological protection
  - 13.4.4 Seed laws
  - 13.4.5 Contract law
  - 13.4.6 Brands and trademarks
  - 13.4.7 Trade secrets
- 13.5 Intellectual Property Rights Affecting Molecular Breeding 524
  - 13.5.1 Genetic transformation technologies
    - Transformation methods
    - Genes and DNA sequences
    - Selection and identification of transformants
    - Promoters and other regulatory elements
    - 'Golden Rice' as an example for freedom-to-operate
  - 13.5.2 Marker-assisted plant breeding
    - Selection of microsatellite primers and the PCR
    - Analysis of PCR products
    - Marker-assisted breeding methods
  - 13.5.3 Product development and commercialization
- 13.6 Use of Molecular Techniques in Plant Variety Protection 535
  - 13.6.1 DUS testing
  - 13.6.2 Essentially derived varieties
  - 13.6.3 Cultivar identification
  - 13.6.4 Seed certification
  - 13.6.5 Seed purification
- 13.7 Plant Variety Protection Practice 541

- 13.7.1 Plant variety protection in the EU
- 13.7.2 Plant variety protection in the USA
- 13.7.3 Plant variety protection in Canada
- 13.7.4 Plant variety protection in developing countries
- 13.7.5 Participatory plant breeding and plant variety protection
- 13.8 Future Perspectives 543
  - 13.8.1 Extension and enforcement
  - 13.8.2 Administrative challenges for implementing PVP
  - 13.8.3 The need to update UPOV
  - 13.8.4 Collaboration in use of genetic resources
  - 13.8.5 Technology and intellectual property interaction
  - 13.8.6 Seed saving and plant variety protection
  - 13.8.6 Other plant products

#### **14 Breeding Informatics 550**

- 14.1 Information-driven Plant Breeding 550
  - 14.1.1 Basics of informatics
  - 14.1.2 Gaps between bioinformatics and plant breeding
  - 14.1.3 A universal system for information management and data analysis
  - 14.1.4 Transforming information to new cultivars
- 14.2 Information Collection 554
  - 14.2.1 Data collection procedure
  - 14.2.2 Germplasm information
    - Passport data
    - Pedigree and genealogy
    - Genetic stocks
  - 14.2.3 Genotypic information
    - Molecular markers
    - Sequences
    - Expression information
  - 14.2.4 Phenotypic information
  - 14.2.5 Environmental information
- 14.3 Information Integration 562
  - 14.3.1 Data standardization
  - 14.3.2 Development of generic databases
  - 14.3.3 Use of controlled vocabularies and ontologies
  - 14.3.4 Interoperable query system
  - 14.3.5 Redundant data condensing
  - 14.3.6 Database integration
  - 14.3.7 Tool-based information integration
- 14.4 Information Retrieval and Mining 568
  - 14.4.1 Information retrieval
    - Bibliographic databases
    - Abstracts
    - Full text of research articles
    - Books and text-rich web sites
  - 14.4.2 Information mining
    - Data mining
    - Comparative informatics
    - Sequence similarity analysis
- 14.5 Information Management Systems 572

- 14.5.1 Laboratory information management systems
- 14.5.2 Breeding information management systems
- 14.5.3 International Crop Information System
- 14.5.4 Other information tools
- 14.5.5 Other needs for informatics tools
- 14.6 Plant Databases 579
  - 14.6.1 Sequence databases
    - Nucleotide sequence databases
    - Protein sequence databases
  - 14.6.2 General genomics and proteomics databases
  - 14.6.3 General plant databases
  - 14.6.4 Individual plant databases
- 14.7 Future Prospects for Breeding Informatics 595

## **15 Decision Support Tools 599**

- 15.1 Germplasm and Breeding Population Management and Evaluation 600
  - 15.1.1 Germplasm management and evaluation
  - 15.1.2 Breeding population management
    - Establishing heterotic patterns
    - Predicting hybrid performance
- 15.2 Genetic Mapping and Marker–Trait Association Analysis 605
  - 15.2.1 Genetic map construction
  - 15.2.2 Linkage-based QTL mapping
  - 15.2.3 eQTL mapping
  - 15.2.4 Linkage-disequilibrium based QTL mapping
    - LD-based QTL mapping
    - Genome-wide association mapping
    - Integrated haplotype and LD analysis
  - 15.2.5 Genotype-by-environment interaction analysis
  - 15.2.6 Comparative mapping and consensus maps
- 15.3 Marker-assisted Selection 613
  - 15.3.1 MAS methodologies and implementation
  - 15.3.2 Marker-assisted inbred and synthetic creation
- 15.4 Simulation and Modelling 615
  - 15.4.1 Importance of simulation and modeling
  - 15.4.2 Genetic models used in simulation
  - 15.4.3 A simulation module for genetics and breeding: QuLine
  - 15.4.5 The future of simulation and modelling
- 15.5 Breeding by Design 621
  - 15.5.1 Parental selection
  - 15.5.2 Breeding product prediction
  - 15.5.3 Selection method evaluation
- 15.6 Future Perspectives 623

**References 627-716**

**Index 717-734**

**The colour plate section can be found following p. 270**

# Preface

---

The genomics revolution of the past decade has greatly enhanced our understanding of the genetic composition of living organisms including many plant species of economic importance. Complete genomic sequences of *Arabidopsis* and several major crops, together with high-throughput technologies for analyses of transcripts, proteins and mutants, provide the basis for understanding the relationship between genes, proteins and phenotypes. Sequences and genes have been used to develop functional and biallelic markers, such as single nucleotide polymorphism (SNP), that are powerful tools for genetic mapping, germplasm evaluation and marker-assisted selection.

The road from basic genomics research to impacts on routine breeding programmes has been long, windy and bumpy, not to mention scattered with wrong turns and unexpected blockades. As a result, genomics can be applied to plant breeding only when an integrated package becomes available that combines multiple components such as high-throughput techniques, cost-effective protocols, global integration of genetic and environmental factors and precise knowledge of quantitative trait inheritance. More recently, the end of the tunnel has come in sight, and the multinational corporations have ramped up their investments in and expectations from these technologies. The challenge now is to translate and integrate the new knowledge from genomics and molecular biology into appropriate tools and methodologies for public-sector plant breeding programmes, particularly those in low-income countries. It is expected that harnessing the outputs of genomics research will be an important component in successfully addressing the challenge of doubling world food production by 2050.

## **What does *Molecular Plant Breeding* include?**

The term ‘molecular plant breeding’ has been much used and abused in the literature, and thus loved or maligned in equal measure by the readership. In the context of this book, the term is used to provide a simple umbrella for the multidisciplinary field of modern plant breeding that combines molecular tools and methodologies with conventional approaches for improvement of crop plants. This book is intended to provide comprehensive coverage of the components that should be integrated within plant breeding programmes to develop crop products in a more efficient and targeted way.

The first chapter introduces some basic concepts that are required for understanding fundamentally important issues described in subsequent chapters. The concepts include crop domestication, critical events in the history of plant breeding, basics of quantitative genetics (variance, heritability and selection index), plant breeding objectives and molecular breeding goals. Chapters 2 and 3 introduce the key genomics tools that are used in molecular breeding programmes, including molecular markers, maps, 'omics' technologies and arrays. Different types of molecular markers are compared and construction of molecular maps is discussed. Chapter 4 describes common types of populations that have been used in genetics and plant breeding, with a focus on recombinant inbred lines, doubled haploids and near-isogenic lines. Chapter 5 provides an overview of marker-assisted germplasm evaluation, management and enhancement. Chapters 6 and 7 discuss the theory and practice, respectively, of using molecular markers to dissect complex traits and locate quantitative trait loci (QTL). Chapters 8 and 9 cover the theory and practice, respectively, of marker-assisted selection. Genotype-by-environment interaction (GEI) is discussed in Chapter 10, including multi-environment trials, stability of genotype performance, molecular dissection of GEI and breeding for optimum GEI. Chapter 11 provides a summary of gene isolation and functional analysis approaches, including *in silico* prediction of genes, comparative approaches for gene isolation, gene cloning based on cDNA sequencing, positional cloning and identification of genes by mutagenesis. Chapter 12 describes the use of isolated and characterized genes for gene transfer and the generation of genetically modified plants, focusing on the vital elements of expression vectors, selectable marker genes, transgene integration, expression and localization, transgene stacking and transgenic crop commercialization. Chapter 13 is devoted to intellectual property rights and plant variety protection, including plant breeders' rights, international agreements affecting plant breeding, plant variety protection strategies, intellectual property rights affecting molecular breeding and the use of molecular techniques in plant variety protection. The last two chapters (14 and 15) discuss supporting tools that are required in molecular breeding for information management and decision making, including data collection, integration, retrieval and mining and information management systems. Decision support tools are described for germplasm and breeding population management and evaluation, genetic mapping and marker-trait association analysis, marker-assisted selection, simulation and modelling, and breeding by design.

### **Intended audience and guidance for reading and using this book**

This book is intended to provide a handbook for biologists, geneticists and breeders, as well as a textbook for final year undergraduates and graduate students specializing in agronomy, genetics, genomics and plant breeding. Although the book has attempted to cover all relevant areas of molecular breeding in plants, many examples have been drawn from the genomics research and molecular breeding of major cereal crops. It is hoped that the book can also serve as a resource for training courses as described below. As each chapter covers a complete story on a special topic, readers can choose to read chapters in any order.

Advanced Course on Quantitative Genetics: Chapters 1, 2, 4, 6, 7, 10 and 14, which cover all molecular marker-based QTL mapping, including markers, maps, populations, statistics and genotype-by-environment interaction.

Comprehensive Course on Marker-assisted Plant Breeding: Chapters 1, 2, 3, 4, 5, 8, 9, 10, 13, 14 and 15, which cover basic theories, tools, methodologies about markers, maps, omics, arrays, informatics and support tools for marker-assisted selection.

Short Course on Genetic Transformation: Chapters 1, 11, 12 and 13, which provide a brief introduction to gene isolation, transformation techniques, genetic-transformation-related intellectual property and genetically modified organism (GMO) issues.

---

Introductory Course on Breeding Informatics: Chapters 1, 2, 3, 4, 5, 10, 14 and 15, which cover bioinformatics, focusing on plant breeding-related applications, including basic concepts in plant breeding, markers, maps, omics, arrays, population and germplasm management, environment and geographic information system (GIS) information, data collection, integration and mining, and bioinformatics tools required to support molecular breeding. Additional introductory information can be found in other chapters.

### History of writing this book

This book has been almost a decade in preparation. In fact, the initial idea for the book was stimulated by the impact from my previous book *Molecular Quantitative Genetics* published by China Agriculture Press (Xu and Zhu, 1994), which was well received by colleagues and students in China and used as a textbook in many universities. Preliminary ideas related to the book were developed in a review article on QTL separation, pyramiding and cloning in *Plant Breeding Reviews* (Xu, 1997). Much of the hopeful thinking described in this paper has fortunately come true during the following 10 years, and the manipulation of QTL has been revolutionized and become mainstream. As complete sequences for several plant genomes have become available and with more anticipated, as shown by numerous genes and QTL that have been separated and cloned individually, some of them have been pyramided for plant breeding through genetic transformation or marker-assisted selection.

I started making tangible progress on this book while working as a molecular breeder for hybrid rice at RiceTec, Inc., Texas (1998–2003). This experience shaped my thinking about how an applied breeding programme could be integrated with molecular approaches. With numerous QTL accumulating for a model crop, taking all the QTL into consideration becomes necessary. Initial thoughts on this were described in ‘Global view of QTL...’, published in the proceedings on quantitative genetics and plant breeding, which considered various genetic background effects and genotype-by-environment interaction (Xu, Y., 2002). Hybrid rice breeding, which involves a three-line system, requires a large number of test-crosses in order to identify traits that perform well in seed and grain production. My experience in development of marker-assisted selection strategies for breeding hybrid rice was then summarized in a review article in *Plant Breeding Reviews* (Xu, Y., 2003), which also covered general strategies for other crops using hybrids.

Moving on to research at Cornell University with Dr Susan McCouch helped me to better understand how molecular techniques could facilitate breeding of complex traits such as water-use efficiency, which is a difficult trait to measure and requires strong collaboration among researchers across many disciplines. In addition, this experience with rice as a model crop raised the issue of how we can use rice as a reference genome for improvement of other crops, which was discussed in an article published in a special rice issue of *Plant Molecular Biology* (Xu *et al.*, 2005).

With over 20 years’ experience in rice, I decided to shift to another major crop by working for the International Maize and Wheat Improvement Center (CIMMYT) as the principle maize molecular breeder. CIMMYT has given me exposure to an interface connecting basic research with applied breeding for developing countries and the resource-poor. Comparing public- and private-sector breeding programmes has given me an intense understanding of the importance of making the type of breeding systems that have been working well for the private sector a practical reality for the public sector, particularly in developing countries. This has been addressed in a recent review paper published in *Crop Science* (Xu and Crouch, 2008), which discussed the critical issues for achieving this translation. My most recent research has focused on the development of various molecular breeding

platforms that can be used to facilitate breeding procedures through seed DNA-based genotyping, selective and pooled DNA analysis, and chip-based large-scale germplasm evaluation, marker–trait association and marker-assisted selection (see Xu *et al.*, 2009b for further details). Thus, my career has evolved alongside the transition from molecular biology research to routine molecular plant breeding applications and I strongly believe that now is the right time for a mainstream publication providing comprehensive coverage of all fields relevant for a new generation of molecular breeders.

## Acknowledgements

### Assistance and professional support

The dream of writing this book could not have become reality without the wonderful support of Dr Susan McCouch at Cornell University and Dr Jinhua Xiao, now at Monsanto, who have both fully supported my proposal since 2002. Their support and consistent encouragement has greatly motivated me throughout the process. While working with Susan, she allowed me so much flexibility in my research projects and working hours so that I could continue to make progress on the writing of this book. At the same time the Cornell libraries were an indispensable source of the major references cited throughout the book. Susan's encouragement provided the impetus to keep working on the book through a very difficult time in my life. I also extend my appreciation to Dr Jonathan Crouch, the Director of the Germplasm Resources Program at CIMMYT, where I received his full understanding and support so that I could complete the second half of the book. Jonathan's guidance and contribution to my research projects and publications while at CIMMYT has significantly impacted the preparation of the book.

I would also like to thank the chief editors of the three journals for which I have served on the editorial boards during the preparation of this book: Dr Paul Christou for *Molecular Breeding*, Dr Albrecht Melchinger for *Theoretical and Applied Genetics*, and Dr Hongbin Zhang for *International Journal of Plant Genomics*. I thank them for their patience, support and flexibility with my editorial responsibilities during the preparation of the book. In addition, Drs Christou and Melchinger also reviewed several chapters in their respective fields.

My appreciation also goes to Yanli Lu (a graduate student from Sichuan Agricultural University of China) and Dr Zhuanfang Hao (a visiting scientist from the Chinese Academy of Agricultural Sciences) who helped prepare some figures and tables during their work in my lab at CIMMYT, Mexico. I would like to give special thanks to Dr Rodomiro Ortiz at CIMMYT for his consistent information sharing and stimulating discussions during our years together at CIMMYT. Finally, I would like to thank my colleagues at CIMMYT, particularly Drs Kevin Pixley, Manilal William, Jose Crossa and Guy Davenport, who provided useful discussions on various molecular breeding-related issues.

## Forewords

I am greatly indebted to Dr Norman E. Borlaug, visioned plant breeder and Nobel laureate for his role in the Green Revolution, and Dr Ronald L. Phillips, Regents Professor and McKnight Presidential Chair in Genomics, University of Minnesota, who each contributed a foreword for the book. Their contributions emphasized the importance of molecular breeding in crop improvement and the role that this book will play in molecular breeding education and practice.

---

## Reviewers

Each chapter of the book has undergone comprehensive peer review and revision before finalization. The constructive comments and critical advice of these reviewers have greatly improved this book. The reviewers were selected for their active expertise in the field of the respective chapter. Reviewers come from almost all continents and work in various fields including plant breeding, quantitative genetics, genetic transformation, intellectual property protection, bioinformatics and molecular biology, many of whom are CIMMYT scientists and managers. Considering that each chapter is relatively large in content, reviewers had to contribute a lot of time and effort to complete their reviews. Although these inputs were indispensable, any remaining errors remain my sole responsibility. The names and affiliations of the reviewers (alphabetically) are:

Raman Babu (Chapters 7 and 9), CIMMYT, Mexico  
Paul Christou (Chapter 12), Lleida, Spain  
Jose Crossa (Chapter 10), CIMMYT, Mexico  
Jonathan H. Crouch (Chapters 13 and 15), CIMMYT, Mexico  
Jedidah Danson (Chapters 7 and 9), African Center for Crop Improvement, South Africa  
Guy Davenport (Chapter 14), CIMMYT, Mexico  
Yuqing He (Chapter 8), Huazhong Agricultural University, Wuhan, China  
Gurdev S. Khush (Chapter 1), IRRI, Philippines  
Alan F. Krivanek (Chapter 4), Monsanto, Illinois, USA  
Huihui Li (Chapter 6), Chinese Academy of Agricultural Sciences, China  
George H. Liang (Chapter 12), San Diego, California, USA  
Christopher Graham McLaren (Chapter 14), GCP/CIMMYT, Mexico  
Kenneth L. McNally (Chapter 5), IRRI, Philippines  
Albrecht E. Melchinger (Chapter 8), University of Hohenheim, Germany  
Rodomiro Ortiz (Chapters 12, 13 and 15), CIMMYT, Mexico  
Edie Paul (Chapter 14), GeneFlow, Inc., Virginia, USA  
Kevin V. Pixley (Chapters 1, 4 and 5), CIMMYT, Mexico  
Trushar Shah (Chapter 14), CIMMYT, Mexico  
Daniel Z. Skinner (Chapter 12), Washington State University, USA  
Debra Skinner (Chapter 11), University of Illinois, USA  
Michael J. Thomson (Chapters 2 and 3), IRRI, Philippines  
Bruce Walsh (Chapters 1, 6 and 8), University of Arizona, USA  
Marilyn L. Warburton (Chapter 5), USDA/Mississippi State University, USA  
Huixia Wu (Chapter 12), CIMMYT, Mexico  
Rongling Wu (Chapter 1), University of Florida, Gainesville, USA  
Weikai Yan (Chapter 10), Agriculture and Agri-Food Canada, Ottawa, Canada  
Qifa Zhang (Chapters 8 and 12), Huazhong Agricultural University, Wuhan, China  
Wanggen Zhang (Chapter 12), Syngenta, Beijing, China  
Yuhua Zhang (Chapter 12), Rothamsted Research, UK

## Publishers and development editors

Several editors at CABI have been working with me over the years: Tim Hardwick (2002–2006), Sarah Hulbert (2006–2007), Stefanie Gehrig (2007–2008), Claire Parfitt (2008–2009), Meredith Carroll (2009) and Tracy Head (2009). These editors and their associates have done a superb job of converting a series of manuscripts into a useable and coherent book. I thank them for their effort, consideration and cooperation.



### **Research grants**

During the preparation of the book, my research on genomic analysis of plant water-use efficiency at Cornell University was supported by the National Science Foundation (Plant Genome Research Project Grant DBI-0110069). My molecular breeding research at CIMMYT has been supported by the Rockefeller Foundation, the Generation Challenge Programme (GCP), Bill and Melinda Gates Foundation and the European Community, and through other attributed or unrestricted funds provided by the members of the Consultative Group on International Agricultural Research (CGIAR) and national governments of the USA, Japan and the UK.

### **Family**

It is difficult to imagine writing a book without the full support and understanding of one's family. My greatest thanks go to my wife, Yu Wang, who has given me her wholehearted and unwavering support, and to my sons, Sheng, Benjamin and Lawrence, who have retained great patience during this long adventure. And finally to my parents, for their love, encouragement and vision that unveiled in me from my earliest years the desire to thrive on the challenge of always striving to reach the highest mountain in everything I do.

# Foreword

DR NORMAN E. BORLAUG

---

The past 50 years have been the most productive period in world agricultural history. Innovations in agricultural science and technology enabled the 'Green Revolution', which is reputed to have spared one billion people the pains of hunger and even starvation. Although we have seen the greatest reductions in hunger in history, it has not been enough. There are still one billion people who suffer chronic hunger, with more than half being small-scale farmers who cultivate environmentally sensitive marginal lands in developing countries.

Within the next 50 years, the world population is likely to increase by 60–80%, requiring global food production to nearly double. We will have to achieve this feat on a shrinking agricultural land base, and most of the increased production must occur in those countries that will consume it. Unless global grain supplies are expanded at an accelerated rate, food prices will remain high, or be driven up even further.

Spectacular economic growth in many newly industrializing developing countries, especially in Asia, has spurred rapid growth in global cereal demand, as more people eat better, especially through more protein-heavy diets. More recently, the subsidized conversion of grains into biofuels in the USA and Europe has accelerated demand even faster. On the supply side, a slowing in research investment in the developing world and more frequent climatic shocks (droughts, floods) have led to greater volatility in production.

Higher food prices affect everyone, but especially the poor, who spend most of their disposable income on food. Increasing supply, primarily through the generation and diffusion of productivity-enhancing new technologies, is the best way to bring food prices down and secure minimum nutritional standards for the poor.

Today's agricultural development challenges are centred on marginal lands and in regions that have been bypassed during the Green Revolution, such as Africa and resource-poor parts of Asia, and are experiencing the ripple effects of food insecurity through hunger, malnutrition and poverty.

Despite these serious and daunting challenges, there is cause for hope. New science and technology – including biotechnology – have the potential to help the world's poor and food insecure. Biotechnologies have developed invaluable new scientific methodologies and products for more productive agriculture and added-value food. This journey deeper into the genome to the molecular level is the consequence of our progressive understanding of the workings of nature. Genomics-based methods have enabled breeders greater precision in selecting and transferring genes, which has not only reduced the time needed to

eliminate undesirable genes, but has also allowed breeders to access useful genes from distant species.

Bringing the power of science and technology to bear on the challenges of these riskier environments is one of the great challenges of the 21st century. With the new tools of biotechnology, we are poised for another explosion in agricultural innovation. New science has the power to increase yields, address agroclimatic extremes and mitigate a range of environmental and biological challenges.

*Molecular Plant Breeding*, authored by my CIMMYT colleague Yunbi Xu, is an outstanding review and synthesis of the theory and practice of genetics and genomics that can drive progress in modern plant breeding. Dr Xu has done a masterful job in integrating information about traditional and molecular plant breeding approaches. This encyclopedic handbook is poised to become a standard reference for experienced breeders and students alike. I commend him for this prodigious new contribution to the body of scientific literature.

# Foreword

DR RONALD L. PHILLIPS

---

## The New Plant Breeding Roadmap

The road is long from basic research findings to final destinations reflecting important applications – but it is a road that can ultimately save time and money. There may be obstacles along the way that delay building that road but they are generally overcome by careful thought and timely considerations. A new road may involve the former road but with some widening and the filling in of certain potholes. We seldom look back and think that the improvements were not useful.

The road to improved varieties by traditional plant breeding has and continues to serve society well. That approach has been based on careful observation, evaluation of multiple genotypes (parents and progenies), selection at various generational levels, extensive testing and the sophisticated utilization of statistical analyses and quantitative genetics. About 50% of the increased productivity of new varieties is generally attributed to genetic improvements, with the remaining 50% due to many other factors such as time of planting, irrigation, fertilizer, pesticide applications and planting densities.

The statistical genetics associated with traditional plant breeding can now be supplemented by extensive genomic information, gene sequences, regulatory factors and linked genetic markers. We can now draw on a broader genetic base, the identification of major loci controlling various traits and expression analyses across the entire genome under various biotic and abiotic conditions. One can anticipate a future when the networking of genes, genotype-by-environment ( $G \times E$ ) interactions, and even hybrid vigour will be better understood and lead to new breeding approaches. The importance of *de novo* variation may modify much of our current interpretation of breeding behaviour; *de novo* variation such as mutation, intragenic recombination, methylation, transposable elements, unequal crossing over, generation of genomic changes due to recombination among dispersed repeated elements, gene amplification and other mechanisms will need to be incorporated into plant breeding theory.

This book calls for an integration of approaches – traditional and molecular – and represents a theoretical/practical handbook reflecting modern plant breeding at its finest. I believe the reader will be surprised to find that that this single-authored book is so full of information that is useful in plant genetics and plant breeding. Students as well as established researchers wanting to learn more about molecular plant breeding will be

well-served by reading this book. The information is up-to-date with many current references. Even many of the tables are packed with information and references. A good representation of international and domestic breeding is reflected through many examples. The importance of  $G \times E$  interactions is clearly demonstrated. Various statistical models are provided as appropriate. The importance of defining mega-environments for varietal development is made clear. The role of core germplasm collections, appropriate population sizes, major databases and data management issues are all integrated with various plant breeding approaches. Marker-assisted selection receives considerable attention, including its requirements and advantages, along with the multitude of quantitative trait locus (QTL) analysis methods. Transformation technologies leading to the extensive use of transgenic crops are reviewed along with the increased use of trait stacking. The procurement of intellectual property that, in part, is driving the application of molecular genetics in plant breeding provides the reader with an understanding of why private industry is now more involved and why some common crops represent new business opportunities.

*Molecular Plant Breeding* is not like other plant breeding books. The interconnecting road that it depicts is one where you can look at the beautiful new scenery and appreciate the current view, yet see the horizon down the road.

# Molecular Plant Breeding

Yunbi Xu

Recent advances in plant genomics and molecular biology have revolutionized our understanding of plant genetics, providing new opportunities for more efficient and controllable plant breeding. Modern plant breeding involves development and utilization of various genetic populations, molecular tools, statistical methods, breeding informatics and decision support tools. Successful techniques therefore require a solid understanding of the underlying molecular biology as well as experience in applied plant breeding.

Bridging the gap between developments in biotechnology and its applications in plant improvement, *Molecular Plant Breeding* provides an integrative overview of issues from basic theories to their applications to crop improvement. Chapters include discussions of breeding methodology, quantitative genetics, genomics and bioinformatics and present statistical issues related to gene mapping, marker-assisted selection and genotype-by-environment interactions in clear and concise language. Providing an integrated profile of molecular breeding in plants, this book will be an essential resource for researchers and students involved in plant biology and breeding, genetics and applied genomics.

## Related titles

### **Molecular Plant–Microbe Interactions**

Edited by K. Bouarab, N. Brisson and F. Daayf  
2009 368 pages 978 1 84593 574 0

### **Environmental Impact of Genetically Modified Crops**

Edited by N. Ferry and A. Gatehouse  
2009 432 pages 978 1 84593 409 5

### **Insect-resistant Maize**

J. Bürgi  
2009 304 pages 978 1 84593 569 6

### **RNA Interference: Methods for Plants and Animals**

Edited by T. Doran and C. Helliwell  
2008 272 pages 978 1 84593 410 1

For further information on these titles and other publications, see our website at [www.cabi.org](http://www.cabi.org)

### **CABI Head Office**

Nosworthy Way, Wallingford, Oxfordshire, OX10 8DE, UK

### **CABI North American Office**

875 Massachusetts Avenue, 7th Floor, Cambridge, MA 02139, USA

ISBN 978-1-84593-392-0



9 781845 933920