MARKER-ASSISTED SELECTION

Current status and future perspectives in crops, livestock, forestry and fish
MARKER-ASSISTED SELECTION

Current status and future perspectives in crops, livestock, forestry and fish

Edited by

Elcio P. Guimarães
John Ruane
Beate D. Scherf
Andrea Sonnino
James D. Dargie
The designations employed and the presentation of material in this information
product do not imply the expression of any opinion whatsoever on the part
of the Food and Agriculture Organization of the United Nations concerning the
legal or development status of any country, territory, city or area or of its authorities,
or concerning the delimitation of its frontiers or boundaries. The mention of specific
companies or products of manufacturers, whether or not these have been patented, does
not imply that these have been endorsed or recommended by the Food and Agriculture
Organization of the United Nations in preference to others of a similar nature that are
not mentioned.

The views expressed in this publication are those of the authors and do not necessarily
reflect the views of the Food and Agriculture Organization of the United Nations.


All rights reserved. Reproduction and dissemination of material in this information
product for educational or other non-commercial purposes are authorized without
any prior written permission from the copyright holders provided the source is fully
acknowledged. Reproduction of material in this information product for resale or other
commercial purposes is prohibited without written permission of the copyright holders.
Applications for such permission should be addressed to:
Chief
Electronic Publishing Policy and Support Branch
Communication Division
FAO
Viale delle Terme di Caracalla, 00153 Rome, Italy
or by e-mail to:
copyright@fao.org

© FAO 2007
Contents

Acknowledgements vi
Foreword vii
Abbreviations and acronyms ix
Contributors xv

SECTION I – INTRODUCTION TO MARKER-ASSISTED SELECTION 1

Chapter 1
Marker-assisted selection as a tool for genetic improvement of crops, livestock, forestry and fish in developing countries: an overview of the issues 3
John Ruane and Andrea Sonnino

Chapter 2
An assessment of the use of molecular markers in developing countries 15
Andrea Sonnino, Marcelo J. Carena, Elcio P. Guimarães, Roswitha Baumung, Dafydd Pilling and Barbara Rischkowsky

SECTION II – MARKER-ASSISTED SELECTION IN CROPS – CASE STUDIES 27

Chapter 3
Molecular markers for use in plant molecular breeding and germplasm evaluation 29
Jeremy D. Edwards and Susan R. McCouch

Chapter 4
Marker-assisted selection in wheat: evolution, not revolution 51
Robert Koebner and Richard Summers

Chapter 5
Marker-assisted selection for improving quantitative traits of forage crops 59
Oene Dolstra, Christel Denneboom, Ab L.F. de Vos and E.N. van Loo

Chapter 6
Targeted introgression of cotton fibre quality quantitative trait loci using molecular markers 67
Jean-Marc Lacape, Trung-Bieu Nguyen, Bernad Hau and Marc Giband
Chapter 7
Marker-assisted selection in common beans and cassava 81
Mathew W. Blair, Martin A. Fregene, Steve E. Beebe and Hernán Ceballos

Chapter 8
Marker-assisted selection in maize: current status, potential, limitations and perspectives from the private and public sectors 117
Michel Ragot and Michael Lee

Chapter 9
Molecular marker-assisted selection for resistance to pathogens in tomato 151
Amalia Barone and Luigi Frusciante

SECTION III – MARKER-ASSISTED SELECTION IN LIVESTOCK – CASE STUDIES 165

Chapter 10
Strategies, limitations and opportunities for marker-assisted selection in livestock 167
Jack C.M. Dekkers and Julius H.J. van der Werf

Chapter 11
Marker-assisted selection in poultry 185
Dirk-Jan de Koning and Paul M. Hocking

Chapter 12
Marker-assisted selection in dairy cattle 199
Joel Ira Weller

Chapter 13
Marker-assisted selection in sheep and goats 229
Julius H.J. van der Werf

SECTION IV – MARKER-ASSISTED SELECTION IN FORESTRY – CASE STUDIES 249

Chapter 14
Marker-assisted selection in Eucalyptus 251
Dario Grattapaglia

Chapter 15
Marker-assisted selection in forestry species 283
Penny Butcher and Simon Southerton
SECTION V – MARKER-ASSISTED SELECTION IN FISH – CASE STUDIES 307

Chapter 16
Possibilities for marker-assisted selection in aquaculture breeding schemes 309
Anna K. Sonesson

Chapter 17
Marker-assisted selection in fish and shellfish breeding schemes 329
Victor Martinez

SECTION VI – SELECTED ISSUES RELEVANT TO APPLICATIONS OF MARKER-ASSISTED SELECTION IN DEVELOPING COUNTRIES 363

Chapter 18
Marker-assisted selection in crop and livestock improvement: how to strengthen national research capacity and international partnerships 365
Maurício Antônio Lopes

Chapter 19
Technical, economic and policy considerations on marker-assisted selection in crops: lessons from the experience at an international agricultural research centre 381
H. Manilal Williams, Michael Morris, Marilyn Warburton and David A. Hoisington

Chapter 20
Impacts of intellectual property rights on marker-assisted selection research and application for agriculture in developing countries 405
Victoria Henson-Apollonio

Chapter 21
Marker-assisted selection as a potential tool for genetic improvement in developing countries: debating the issues 427
Jonathan Robinson and John Ruane

Chapter 22
Marker-assisted selection: policy considerations and options for developing countries 441
James D. Dargie
Acknowledgements

Preparation of this major publication on marker-assisted selection would not have been possible without the committed support of a large number of individuals from the time that the idea of this publication was initially raised in early 2004. First and foremost among these are the members of the FAO Working Group on Biotechnology, in particular its former Chairperson, Jim Dargie, and his successor, Shivaji Pandey. Three of the members, Devin Bartley (FAO Aquaculture Management and Conservation Service), Pierre Sigaud (formerly of FAO Forest Resources Development Service) and Nuria Urquia (FAO Seed and Plant Genetic Resources Service), as well as Paul Boettcher (Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture) and Eric Hallerman (Virginia Polytechnic Institute and State University, United States of America) are particularly thanked for commenting on individual chapters in the book.

Thanks are also extended to Adrianna Gabrielli, Michela Paganini and Chrissi Redfern for their assistance with editorial and layout aspects of this publication.

Finally, special thanks go to all the authors for dedicating their time and efforts to the preparation of these chapters in an exemplary and timely fashion and for their meticulous attention to detail in the editing process.
Since almost the beginning of human civilization, exploiting variation in the characteristics of the plant and animal genetic resources that are used for producing food and other agricultural products through breeding has been at the heart of efforts to increase and diversify agricultural production and productivity, enhance food security and incomes, and adapt farming to changing environmental conditions and social needs. Initially, this was achieved simply by selecting and reproducing preferred individuals or spontaneous variants, and indeed this practice remains important today as the basis for producing new generations of cultivated landraces and indigenous breeds. However, the crops, trees, livestock and fish that are farmed today have arisen largely from the introduction of scientific breeding at the beginning of the twentieth century, with the inclusion of crosses into breeding schemes prior to artificial selection and application of Mendel’s laws of inheritance to improve both simple and quantitative traits providing the foundations for modern genetics.

Today, thanks to continuing investments made in research and technology development, the process of producing improved varieties, clones, breeds and strains of agriculturally important species has become progressively more accurate, reliable and efficient. Nevertheless, one of the continuing technical constraints to more effective breeding is that selecting material with one or a combination of the characteristics required by farmers, foresters, industry and consumers still relies mainly on physical and agronomic attributes (phenotype). Some of these characteristics are influenced by the environment and are therefore not necessarily a good guide to the actual heritable genetic composition (genotype) of the material in question. Others may not be visible or may only be detected in mature plants and animals. Others again may be difficult or very costly to screen, and many characters such as drought tolerance and milk composition are controlled by a large number of genes whose mode of action as well as their interaction with each other and with various environmental triggers is mainly unknown. Improving the identification, selection and monitoring of specific characters in plants and animals through breeding schemes is therefore a critical need to secure future improvements in genetic resources for food and agriculture.

Since the first description of DNA structure over 50 years ago, scientists have made tremendous strides in identifying genes and gene functions, making it increasingly possible to detect genetic differences (DNA polymorphisms) for traits among individual plants and animals in a much more direct way, thereby assisting in the selection of desired traits. The central technology involved is molecular marker-assisted selection (MAS), using sequences and/or banding patterns of DNA that have been shown through linkage mapping to be located in or near genes that affect the phenotype. These molecular markers can then be used to assist breeders track whether
the specific gene or chromosome segment(s) known to affect the phenotype of interest is present in the individuals or populations of interest.

Although the ultimate goal of identifying the location, function and most favourable alleles of each gene through genome sequence and post-genomics research, and then using markers to select for economically important genes in breeding programmes, is still decades away, in recent years the use of MAS in agriculture has moved progressively from theory to practical application. In the process, it has generated both high expectations for increasing genetic progress through breeding, and raised a number of unresolved challenges. These include: selection of the most appropriate methods and tools for MAS among the many now available for the task at hand, analysing and managing the data produced given the increasing trend towards high-throughput techniques and the constraints imposed by suboptimal levels of resources currently attached to breeding and science and technology including biotechnology, and dealing with intellectual property rights, especially in developing countries.

Since its foundation, FAO has recognized that the biological basis for sustainable agricultural production, fighting hunger and world food security lies in the genetic resources used for food and agriculture. It has also recognized the enormous contributions that have been made to the improvement of these resources through both traditional and more advanced breeding, as well as the ever-increasing role played by biotechnology in improving breeding processes and products. As a knowledge organization, one of FAO’s major roles is to provide its Members and their institutions with factual, comprehensive and current information relevant to sound stewardship of crops, livestock, forestry and fisheries, thereby ensuring its availability as a global public good. This book, by providing a comprehensive description and assessment of the use of MAS for increasing the rate of genetic gain in crops, livestock, forestry and farmed fish, including the related policy, organizational and resource considerations, continues the Organization’s tradition of dealing with issues of importance to agricultural and economic development in a multidisciplinary and cross-sectoral manner. As such it is hoped that the information and options presented and the suggestions made will provide valuable guidance to scientists and breeders in both the public and private sectors, as well as to government and institutional policy- and decision-makers.

Shivaji Pandey
Chairperson
FAO Working Group on Biotechnology
## Abbreviations and acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AATF</td>
<td>African Agricultural Technology Foundation</td>
</tr>
<tr>
<td>AB-QTL</td>
<td>Advanced backcross QTL</td>
</tr>
<tr>
<td>ACMV</td>
<td>African cassava mosaic virus</td>
</tr>
<tr>
<td>AFLP</td>
<td>Amplified fragment length polymorphism</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial insemination</td>
</tr>
<tr>
<td>AMBIONET</td>
<td>Asian Maize Biotechnology Network</td>
</tr>
<tr>
<td>AMMANET</td>
<td>African Molecular Marker Applications Network</td>
</tr>
<tr>
<td>AnGR</td>
<td>Animal genetic resources</td>
</tr>
<tr>
<td>ASARECA</td>
<td>Association for Strengthening Agricultural Research in Eastern and Central Africa</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial artificial chromosome</td>
</tr>
<tr>
<td>BCMNV</td>
<td>Bean common mosaic necrotic virus</td>
</tr>
<tr>
<td>BCMV</td>
<td>Bean common mosaic virus</td>
</tr>
<tr>
<td>BecA</td>
<td>Biosciences eastern and central Africa</td>
</tr>
<tr>
<td>BGYMV</td>
<td>Bean golden yellow mosaic virus</td>
</tr>
<tr>
<td>BIO-EARN</td>
<td>East African Regional Programme and Research Network for Biotechnology, Biosafety and Biotechnology Policy Development</td>
</tr>
<tr>
<td>BLUP</td>
<td>Best linear unbiased prediction</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>BSAPs</td>
<td>Biodiversity Strategies and Action Plans</td>
</tr>
<tr>
<td>Bt</td>
<td><em>Bacillus thuringensis</em></td>
</tr>
<tr>
<td>BTA</td>
<td><em>Bos taurus</em> chromosome</td>
</tr>
<tr>
<td>BYDV</td>
<td>Barley yellow dwarf virus</td>
</tr>
<tr>
<td>CAADP</td>
<td>Comprehensive Africa Agriculture Development Programme</td>
</tr>
<tr>
<td>CAPS</td>
<td>Cleaved amplified polymorphic sequences</td>
</tr>
<tr>
<td>CBB</td>
<td>Cassava bacterial blight</td>
</tr>
<tr>
<td>CBS</td>
<td>Cassava brown streak</td>
</tr>
<tr>
<td>CBSD</td>
<td>Cassava brown streak disease</td>
</tr>
<tr>
<td>CCN</td>
<td>Cereal cyst nematode</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CGIAR</td>
<td>Consultative Group on International Agricultural Research</td>
</tr>
<tr>
<td>CGM</td>
<td>Cassava green mite</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>CIAT</td>
<td>International Center for Tropical Agriculture (Centro Internacional de Agricultura Tropical)</td>
</tr>
<tr>
<td>CIMMYT</td>
<td>International Maize and Wheat Improvement Center (Centro Internacional de Mejoramiento de Maíz y Trigo)</td>
</tr>
<tr>
<td>CIP</td>
<td>International Potato Center Centro (Internacional de la Papa)</td>
</tr>
<tr>
<td>CIRAD</td>
<td>French Agricultural Research Centre for International Development (Centre de coopération internationale en recherche agronomique pour le développement)</td>
</tr>
<tr>
<td>cM</td>
<td>Centi-Morgan</td>
</tr>
<tr>
<td>CMD</td>
<td>Cassava mosaic disease</td>
</tr>
<tr>
<td>CMV</td>
<td>Cassava mosaic virus</td>
</tr>
<tr>
<td>CORPOICA</td>
<td>Colombian Agricultural Research Corporation (Corporación Colombiana de Investigación Agropecuaria)</td>
</tr>
<tr>
<td>CR</td>
<td>Country report</td>
</tr>
<tr>
<td>CT</td>
<td>Computer tomography</td>
</tr>
<tr>
<td>DArT</td>
<td>Diversity array technology</td>
</tr>
<tr>
<td>DFID</td>
<td>United Kingdom’s Department for International Development</td>
</tr>
<tr>
<td>DH</td>
<td>Double-haploid</td>
</tr>
<tr>
<td>DHPLC</td>
<td>Denaturing high pressure liquid chromatography</td>
</tr>
<tr>
<td>DMC</td>
<td>Dry matter content</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DYD</td>
<td>Daughter yield deviation</td>
</tr>
<tr>
<td>EACMV</td>
<td>East Africa cassava mosaic virus</td>
</tr>
<tr>
<td>EBV</td>
<td>Estimated breeding value</td>
</tr>
<tr>
<td>EC</td>
<td>European Commission</td>
</tr>
<tr>
<td>ECOSOC</td>
<td>Economic and Social Council of the United Nations</td>
</tr>
<tr>
<td>eQTL</td>
<td>Expressed gene QTL</td>
</tr>
<tr>
<td>EST</td>
<td>Expressed sequence tag</td>
</tr>
<tr>
<td>ESTP</td>
<td>Expressed sequence tagged polymorphism</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>EUCAGEN</td>
<td><em>Eucalyptus</em> Genome Network</td>
</tr>
<tr>
<td>F1</td>
<td>First filial generation</td>
</tr>
<tr>
<td>F2</td>
<td>Second filial generation</td>
</tr>
<tr>
<td>FAO</td>
<td>Food and Agriculture Organization of the United Nations</td>
</tr>
<tr>
<td>FAO-BioDeC</td>
<td>FAO Biotechnology in Developing Countries</td>
</tr>
<tr>
<td>FHB</td>
<td><em>Fusarium</em> head blight</td>
</tr>
</tbody>
</table>
FIVIMS  Food Insecurity and Vulnerability Information and Mapping Systems
FNP  Functional nucleotide polymorphism
FSC  Forest Stewardship Council
FSIL  Full-sib intercross line
GABI  Genome analysis of the plant biological system
GAS  Gene-assisted selection
GCA  General combining ability
GCP  Generation Challenge Programme
GDP  Gross domestic product
GE  Genetic engineering
GH  Growth hormone
GIS  Geographical information systems
GMOs  Genetically modified organisms
GRDC  Grains Research and Development Corporation
GRFA  Genetic resources for food and agriculture
GRM  Gametic relationship matrix
h²  Heritability
HIPC  Heavily indebted poor countries
HWE  Hardy-Weinberg equilibrium
IAC  InterAcademy Council
IAP  InterAcademy Panel
IARCs  International agricultural research centres
IBD  Identity by descent
ICAR  Indian Council for Agricultural Research
ICMV  Indian cassava mosaic virus
ICRISAT  International Crops Research Institute for the Semi-Arid Tropics
ICCSU  International Council for Science
IFPRI  International Food Policy Research Institute
IHN  Infectious haematopoietic necrosis
IITA  International Institute of Tropical Agriculture
ILRI  International Livestock Research Institute
INIFAP  National Institute for Forestry, Agriculture and Livestock Research (Instituto Nacional de Investigaciones Forestales y Agropecuarias)
IP  Intellectual property
IPGRI  International Plant Genetic Resource Institute
IPR  Intellectual property right
IRR  Internal rate of return
IRRI  International Rice Research Institute
ISAG  International Society for Animal Genetics
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISNAR</td>
<td>International Service for National Agricultural Research</td>
</tr>
<tr>
<td>ISSR</td>
<td>Inter-simple sequence repeats</td>
</tr>
<tr>
<td>ITPGRFA</td>
<td>International Treaty on Plant Genetic Resources for Food and Agriculture</td>
</tr>
<tr>
<td>JGI</td>
<td>Joint Genome Institute</td>
</tr>
<tr>
<td>KARI</td>
<td>Kenya Agricultural Research Institute</td>
</tr>
<tr>
<td>LD</td>
<td>Linkage disequilibrium</td>
</tr>
<tr>
<td>LDL</td>
<td>Linkage disequilibrium and linkage</td>
</tr>
<tr>
<td>LD-MAS</td>
<td>Linkage disequilibrium MAS</td>
</tr>
<tr>
<td>LE</td>
<td>Linkage equilibrium</td>
</tr>
<tr>
<td>LE-MAS</td>
<td>Linkage equilibrium MAS</td>
</tr>
<tr>
<td>LIMS</td>
<td>Laboratory information management system</td>
</tr>
<tr>
<td>LOD</td>
<td>Logarithm of the odds ratio</td>
</tr>
<tr>
<td>MABC</td>
<td>Marker-assisted back-crossing</td>
</tr>
<tr>
<td>MA-BLUP</td>
<td>Marker-assisted best linear unbiased prediction</td>
</tr>
<tr>
<td>MAI</td>
<td>Marker-assisted introgression</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix-assisted laser desorption/ionization-time of flight</td>
</tr>
<tr>
<td>MARS</td>
<td>Marker-assisted recurrent selection</td>
</tr>
<tr>
<td>MAS</td>
<td>Marker-assisted selection</td>
</tr>
<tr>
<td>MBL</td>
<td>Medical biotechnology laboratories</td>
</tr>
<tr>
<td>MC</td>
<td>Molecular characterization</td>
</tr>
<tr>
<td>MD</td>
<td>Marek’s disease</td>
</tr>
<tr>
<td>MDG</td>
<td>Millennium Development Goals</td>
</tr>
<tr>
<td>MFA</td>
<td>Microfibril angle</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>miRNA</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum likelihood</td>
</tr>
<tr>
<td>MoDAD</td>
<td>Measurement of domestic animal diversity</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MSV</td>
<td>Maize streak virus</td>
</tr>
<tr>
<td>MTA</td>
<td>Material Transfer Agreement</td>
</tr>
<tr>
<td>NARES</td>
<td>National agricultural research and extension systems</td>
</tr>
<tr>
<td>NARS</td>
<td>National agricultural research systems</td>
</tr>
<tr>
<td>NDA</td>
<td>Non-disclosure agreement</td>
</tr>
<tr>
<td>NEPAD</td>
<td>New Partnership for Africa’s Development</td>
</tr>
<tr>
<td>NGO</td>
<td>Non-governmental organization</td>
</tr>
<tr>
<td>NIRS</td>
<td>Near infrared reflectance spectroscopy</td>
</tr>
<tr>
<td>NPV</td>
<td>Net present value</td>
</tr>
<tr>
<td>NUE</td>
<td>Nitrogen use efficiency</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>OBM</td>
<td>Orange blossom midge</td>
</tr>
<tr>
<td>OECD</td>
<td>Organisation for Economic Co-operation and Development</td>
</tr>
<tr>
<td>OIE</td>
<td>World Organisation for Animal Health</td>
</tr>
<tr>
<td>OPV</td>
<td>Open-pollinated variety</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBRs</td>
<td>Plant breeders’ rights</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PGRFA</td>
<td>Plant genetic resources for food and agriculture</td>
</tr>
<tr>
<td>PIC</td>
<td>Polymorphic information content</td>
</tr>
<tr>
<td>PPB</td>
<td>Participatory plant breeding</td>
</tr>
<tr>
<td>PPD</td>
<td>Post-harvest physiological deterioration</td>
</tr>
<tr>
<td>PRSPs</td>
<td>Poverty reduction strategy papers</td>
</tr>
<tr>
<td>PT</td>
<td>Progeny test</td>
</tr>
<tr>
<td>PVP</td>
<td>Plant variety protection</td>
</tr>
<tr>
<td>QPM</td>
<td>Quality protein maize</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative trait loci (or locus)</td>
</tr>
<tr>
<td>QTL-NILs</td>
<td>Near isogenic lines for QTL</td>
</tr>
<tr>
<td>QTN</td>
<td>Quantitative trait nucleotide</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>Research and development</td>
</tr>
<tr>
<td>RAPD</td>
<td>Random amplified polymorphic DNA</td>
</tr>
<tr>
<td>rDNA</td>
<td>Ribosomal DNA</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction fragment length polymorphism</td>
</tr>
<tr>
<td>RGA</td>
<td>Resistance gene analogues</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RRA</td>
<td>Rapid rural appraisal</td>
</tr>
<tr>
<td>S&amp;T</td>
<td>Science and technology</td>
</tr>
<tr>
<td>SACMV</td>
<td>South African cassava mosaic virus</td>
</tr>
<tr>
<td>SAGE</td>
<td>Serial analysis of gene expression</td>
</tr>
<tr>
<td>SBMV</td>
<td>Soil-borne mosaic virus</td>
</tr>
<tr>
<td>SCA</td>
<td>Specific combining ability</td>
</tr>
<tr>
<td>SCAR</td>
<td>Sequence characterized amplified region</td>
</tr>
<tr>
<td>SCN</td>
<td>Soybean cyst nematode</td>
</tr>
<tr>
<td>SCS</td>
<td>Somatic cell score</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>siRNA</td>
<td>Short interfering RNA</td>
</tr>
<tr>
<td>SLS-MAS</td>
<td>Single large-scale MAS</td>
</tr>
<tr>
<td>SLU</td>
<td>Swedish University of Agricultural Sciences</td>
</tr>
<tr>
<td>SMA</td>
<td>Simple marker analysis</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td>SoW-AnGR</td>
<td>State of the World’s Animal Genetic Resources</td>
</tr>
<tr>
<td>SPS Agreement</td>
<td>WTO Agreement on the Application of Sanitary and Phytosanitary Measures</td>
</tr>
<tr>
<td>SSCP</td>
<td>Single strand conformation polymorphism</td>
</tr>
<tr>
<td>SSLP</td>
<td>Simple sequence length polymorphism</td>
</tr>
<tr>
<td>SSR</td>
<td>Simple sequence repeat (syn. microsatellite)</td>
</tr>
<tr>
<td>STB</td>
<td><em>Septoria tritici</em> blotch</td>
</tr>
<tr>
<td>STS</td>
<td>Sequence-tagged sites</td>
</tr>
<tr>
<td>SW</td>
<td>Seed weight</td>
</tr>
<tr>
<td>SWaps</td>
<td>Sector-wide approaches</td>
</tr>
<tr>
<td>TBT Agreement</td>
<td>WTO Agreement on Technical Barriers to Trade</td>
</tr>
<tr>
<td>TC</td>
<td>Tissue culture</td>
</tr>
<tr>
<td>TEs</td>
<td>Transposable elements</td>
</tr>
<tr>
<td>TMV</td>
<td>Tobacco mosaic virus</td>
</tr>
<tr>
<td>ToMV</td>
<td>Tomato mottle virus</td>
</tr>
<tr>
<td>TRIPS Agreement</td>
<td>WTO Agreement on Trade-Related Aspects of Intellectual Property Rights</td>
</tr>
<tr>
<td>TSWV</td>
<td>Tomato spotted wilt virus</td>
</tr>
<tr>
<td>TUA</td>
<td>Technology Use Agreement</td>
</tr>
<tr>
<td>TYLCV</td>
<td>Tomato yellow leaf curl virus</td>
</tr>
<tr>
<td>UN</td>
<td>United Nations</td>
</tr>
<tr>
<td>UPOV</td>
<td>International Union for the Protection of New Varieties of Plants</td>
</tr>
<tr>
<td>USAID</td>
<td>United States Agency for International Development</td>
</tr>
<tr>
<td>USDA</td>
<td>United States Department of Agriculture</td>
</tr>
<tr>
<td>WEC</td>
<td>Worm egg count</td>
</tr>
<tr>
<td>WFS</td>
<td>World Food Summit</td>
</tr>
<tr>
<td>WIPO</td>
<td>World Intellectual Property Organization</td>
</tr>
<tr>
<td>WRI</td>
<td>World Resources Institute</td>
</tr>
<tr>
<td>WSC</td>
<td>Wood specific consumption</td>
</tr>
<tr>
<td>WTO</td>
<td>World Trade Organization</td>
</tr>
<tr>
<td>YMV</td>
<td>Yellow mosaic virus</td>
</tr>
</tbody>
</table>
Contributors

Amalia Barone
Professor in Plant Genetics
University of Naples “Federico II”
Department of Soil, Plant, Environmental and Animal Production Sciences
Via Universita’ 100
80055 Portici, Italy
e-mail: ambarone@unina.it

Roswitha Baumung
Researcher, Division Livestock Sciences
University of Natural Resources and Applied Life Sciences (BOKU),
Gregor Mendelstr. 33
1180 Vienna, Austria
e-mail: roswitha.baumung@boku.ac.at

Steve E. Beebe
Bean Breeder
Bean Project Manager
Centro Internacional de Agricultura Tropical (CIAT)
Km 17, Recta Cali-Palmira
A.A. 6713, Cali, Colombia
e-mail: s.beebe@cgiar.org

Mathew W. Blair
Germplasm Specialist/Bean Breeder
Biotechnology Unit
Centro Internacional de Agricultura Tropical (CIAT)
Km 17, Recta Cali-Palmira
A.A. 6713, Cali, Colombia
e-mail: m.blair@cgiar.org

Penny Butcher
Conservation Research Scientist
Kings Park and Botanic Garden
Botanic Gardens and Parks Authority
Fraser Avenue, West Perth 6005
Western Australia, Australia
e-mail: pbutcher@bgpa.wa.gov.au

Marcelo J. Carena
Associate Professor/Director
Corn Breeding and Genetics
North Dakota State University
Department of Plant Sciences
Loftsgard Hall 166
Fargo, ND 58105-5051, USA
e-mail: marcelo.carena@ndsu.edu

Hernán Ceballos
Cassava Breeding
Cassava Breeding Project
Centro Internacional de Agricultura Tropical (CIAT)
Km 17, Recta Cali-Palmira
A.A. 6713, Cali, Colombia
and Universidad Nacional de Colombia
e-mail: h.ceballos@cgiar.org

James D. Dargie
Former Director, Joint FAO/IAEA
Division of Nuclear Techniques in Food and Agriculture
Brunnstubengasse 43
2102 Bisamberg, Austria
e-mail: j.dargie@aon.at
Jack C. M. Dekkers
Professor and Section Leader
Animal Breeding and Genetics
Department of Animal Science
239D Kildee Hall
Iowa State University
Ames, IA 50011-3150, USA
e-mail: jdekkers@iastate.edu

Dirk-Jan de Koning
Principal Investigator – Genetics and Genomics
The Roslin Institute
Roslin Biocentre
Roslin, Midlothian
EH25 9PS, UK
e-mail: DJ.deKoning@bbsrc.ac.uk

Christel Denneboom
Plant Research International
PO Box 16
6700 AA Wageningen
Netherlands
e-mail: christel.denneboom@wur.nl

Ab L.F. de Vos
Plant Research International
PO Box 16
6700 AA Wageningen
Netherlands
e-mail: ab.devos@wur.nl

Oene Dolstra
Senior Scientist
Plant Research International
PO Box 16
6700 AA Wageningen
Netherlands
e-mail: oene.dolstra@wur.nl

Jeremy D. Edwards
Research Associate
University of Arizona, Department of Plant Sciences
821B Marley Building, 1145 E 4 ST.
Tucson, AZ 85721, USA
e-mail: jdedw@cals.arizona.edu

Martin Fregene
Senior Scientist and Cassava Geneticist
Centro Internacional de Agricultura Tropical (CIAT)
Km 17, Recta Cali-Palmira
A.A. 6713, Cali
Colombia
e-mail: m.fregene@cgiar.org

Luigi Fruscianti
Professor in Plant Genetics
University of Naples “Federico II”
Department of Soil, Plant, Environmental and Animal Production Sciences
Via Universita’ 100
80055 Portici
Italy
e-mail: fruscian@unina.it

Marc Giband
Researcher – Cotton Molecular Genetics
CIRAD - UPR Systemes Cotonniers / EMBRAPA Algodão
Rua Osvaldo Cruz 1143
Centenário
58.107-720 Campina Grande, PB
Brazil
e-mail: marc.giband@cirad.fr

Dario Grattapaglia
Senior Scientist and Professor - Plant Genetics/Genomics
Brazilian Agricultural Research Corporation – Embrapa
Genetic Resources and Biotechnology Center
Av. W5 Norte - Final - Plano Piloto
70770-900 - Brasília, DF, Brazil
and
Genomic Science Program
Universidade Católica de Brasília
SGAN 916 Modulo B
70790-160 Brasília, DF, Brazil
e-mail: dario@cenargen.embrapa.br
Elcio Perpétuo Guimarães
Senior Officer - Cereals/Crops Breeding
Food and Agriculture Organization of the United Nations (FAO)
00153 Rome, Italy
e-mail: elcio.guimaraes@fao.org

Bernard Hau
Doctor in Plant breeding
CIRAD
UPR Systèmes Cotonniers
TA B-10/02
Avenue Agropolis
34980 Montpellier Cedex 5
France
e-mail: bernard.hau@cirad.fr

Victoria W.K. Henson-Apollonio
Senior Scientist, Project Manager
The CGIAR Central Advisory Service on Intellectual Property (CAS-IP)
Bioversity International
Via dei Tre Denari, 472/a
00057 Maccarese
Italy
e-mail: v.henson-apollonio@cgiar.org

Paul M. Hocking
Principal Investigator - Genetics and Genomics
The Roslin Institute
Roslin Biocentre, Roslin
Midlothian EH25 9PS, UK
e-mail: Paul.Hocking@bbsrc.ac.uk

David A. Hoisington
Global Theme Leader - Biotechnology
ICRISAT
Patancheru, 502 324 Andhra Pradesh, India
e-mail: d.hoisington@cgiar.org

Robert Koebner
Crop Breeding and Biotechnology Consultant
CropGen International
Mockbeggars, Townhouse Road
Old Costessey
Norwich NR8 5BX, UK
e-mail: mockbeggars@onetel.com

Jean-Marc Lacape
Cotton Molecular Breeding
CIRAD
UMR Développement et Amélioration des Plantes
TA A-96/03
Avenue Agropolis
34980 Montpellier Cedex 5, France
e-mail: marc.lacape@cirad.fr

Michael Lee
Professor of Plant Breeding and Genetics
Department of Agronomy
Iowa State University
Room 1553 Agronomy Hall
100 Osborn Drive
Ames, IA 50011-1010, USA
e-mail: mlee@iastate.edu

Maurício Antônio Lopes
Research Scientist - Plant Genetics
Brazilian Agricultural Research Corporation – Embrapa
Genetic Resources and Biotecnology Center
Av. W5 Norte - Final - Plano Piloto
70770-900 - Brasília, DF, Brazil
e-mail: mlopes@cenargen.embrapa.br

Victor Martinez
Assistant Professor
Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile
Avda. Santa Rosa 11735
La Pintana, Santiago, Chile
e-mail: vmartine@uchile.cl
Susan R. McCouch
Professor
Plant Breeding & Genetics
162 Emerson Hall
Cornell University
Ithaca, NY 14853-1901
USA
e-mail: srm4@cornell.edu

Michael Morris
Lead Agricultural Economist
The World Bank - MSN 6-602
1818 Street, NW.
Washington, DC 20433
USA
e-mail: mmorris3@worldbank.org

Trung-Bieu Nguyen
Cotton Breeding
CIRAD
UPR Systèmes Cotonniers
TA B-10/02
Avenue Agropolis
34980 Montpellier Cedex 5
France

Dafydd Pilling
Editor State of the World AnGR
Food and Agriculture Organization of the United Nations (FAO)
00153 Rome, Italy
e-mail: dafydd.pilling@fao.org

Michel Ragot
Head - Genetic Information Management
Syngenta Seeds
12 chemin de l’Hobit
31790 Saint-Sauveur
France
e-mail: michel.ragot@syngenta.com

Barbara Rischkowsky
At that time:
Coordinator State of the World AnGR
Food and Agriculture Organization of the United Nations (FAO)
00153 Rome, Italy
Currently:
Senior Livestock Scientist
International Center for Agricultural Research in the Dry Areas (ICARDA)
PO Box 5466
Aleppo, Syrian Arab Republic
e-mail: b.rischkowsky@cgiar.org

Jonathan Robinson
Consultant
Tick-aho, Joroisniemenkehätie
79600 Joroinen, Finland
e-mail: jrobinson@tiscali.it

John Ruane
Agricultural Officer (Biotechnology)
FAO Working Group on Biotechnology
Food and Agriculture Organization of the United Nations (FAO)
00153 Rome, Italy
e-mail: john.ruane@fao.org

Beate D. Scherf
Animal Production Officer
Animal Genetic Resources Group
Food and Agriculture Organization of the United Nations (FAO)
00153 Rome, Italy
e-mail: beate.scherf@fao.org

Anna K. Sonesson
Senior Scientist, Genetics and Breeding
AKVAFORSK (Institute of Aquaculture Research AS)
PO Box 5010
1432 Ås, Norway
e-mail: anna.sonesson@akvaforsk.no
Andrea Sonnino  
Senior Agricultural Research Officer  
Research and Technological Development Service  
Food and Agriculture Organization of the United Nations (FAO)  
00153 Rome, Italy  
e-mail: andrea.sonnino@fao.org

Simon Southerton  
Senior Research Scientist  
Ensis Genetics  
The joint forces of CSIRO and SCION  
PO Box E4008 Kingston ACT 2604, Australia  
e-mail: simon.southerton@ensisjv.com

Richard Summers  
Cereal Breeding Lead  
RAGT Seeds Ltd  
The Maris Centre  
45 Hauxton Road  
Trumpington  
Cambridge CB2 2LQ, UK  
e-mail: RSummers@ragt.fr

Julius H. J. van der Werf  
Professor in Animal Breeding and Genetics  
School of Rural Science and Agriculture  
University of New England  
Armidale, NSW 2351, Australia  
e-mail: julius.vanderwerf@une.edu.au

E.N. (Robert) van Loo  
Plant Research International  
PO Box 16  
6700 AA Wageningen, Netherlands  
e-mail: robert.vanloo@wur.nl

Marilyn Warburton  
Senior Scientist, Molecular Geneticist  
Genetic Resources Enhancement Unit  
International Maize and Wheat Improvement Center (CIMMYT)  
Apdo. Postal 6-641  
06600 Mexico, DF, Mexico  
e-mail: m.warburton@cgiar.org

Joel Ira Weller  
Research Scientist  
Department of Cattle and Genetics  
Institute of Animal Sciences  
Agricultural Research Organization  
The Volcani Center  
PO Box 6  
Bet Dagan 50250, Israel  
e-mail: weller@agri.huji.ac.il

H. Manilal William  
Senior Scientist, Global Wheat Program  
Genetic Resources Enhancement Unit  
International Maize and Wheat Improvement Center (CIMMYT)  
Apdo. Postal 6-641  
06600 México, DF, Mexico  
e-mail: m.william@cgiar.org
SECTION I

Introduction to marker-assisted selection
CHAPTER 1

Marker-assisted selection as a tool for genetic improvement of crops, livestock, forestry and fish in developing countries: an overview of the issues

John Ruane and Andrea Sonnino
SUMMARY
This chapter provides an overview of the techniques, current status and issues involved in using marker-assisted selection (MAS) for genetic improvement in developing countries. Molecular marker maps, the necessary framework for any MAS programme, have been constructed for the majority of agriculturally important species, although the density of the maps varies considerably among species. Despite the considerable resources that have been invested in this field and despite the enormous potential it still represents, with few exceptions, MAS has not yet delivered its expected benefits in commercial breeding programmes for crops, livestock, forest trees or farmed fish in the developed world. When evaluating the potential merits of applying MAS as a tool for genetic improvement in developing countries, some of the issues that should be considered are its economic costs and benefits, its potential benefits compared with conventional breeding or with application of other biotechnologies, and the potential impact of intellectual property rights (IPRs) on the development and application of MAS.
INTRODUCTION

The potential benefits of using markers linked to genes of interest in breeding programmes, thus moving from phenotype-based towards genotype-based selection, have been obvious for many decades. However, realization of this potential has been limited by the lack of markers. With the advent of DNA-based genetic markers in the late 1970s, the situation changed and researchers could, for the first time, begin to identify large numbers of markers dispersed throughout the genetic material of any species of interest and use the markers to detect associations with traits of interest, thus allowing MAS finally to become a reality. This led to a whole new field of academic research, including the milestone paper by Paterson et al. (1988). This showed that with the availability of large numbers of genetic markers for their species of interest (tomato), the effects and location of marker-linked genes having an impact on a number of quantitative traits (fruit traits in their case) could be estimated using an approach that could be applied to dissect the genetic make-up of any physiological, morphological and behavioural trait in plants and animals.

Most of the traits considered in animal and plant genetic improvement programmes are quantitative, i.e. they are controlled by many genes together with environmental factors, and the underlying genes have small effects on the phenotype observed. Milk yield and growth rate in animals or yield and seed size in plants are typical examples of quantitative traits. In classical genetic improvement programmes, selection is carried out based on observable phenotypes of the candidates for selection and/or their relatives but without knowing which genes are actually being selected. The development of molecular markers was therefore greeted with great enthusiasm as it was seen as a major breakthrough promising to overcome this key limitation. As Young (1999) wrote: “Before the advent of DNA marker technology, the idea of rapidly uncovering the loci controlling complex, multigenic traits seemed like a dream. Suddenly, it was difficult to open a plant genetics journal without finding dozens of papers seeking to pinpoint many, if not most, agriculturally relevant genes.” However, despite the considerable resources that have been invested in this field and despite the enormous potential it still represents, with few exceptions, MAS has not yet delivered its expected benefits in commercial breeding programmes for crops, livestock, forest trees or farmed fish in the developed world. In developing countries, where investments in molecular markers have been far smaller, delivery of benefits has lagged even further behind.

The focus of this chapter is on the use of molecular markers for genetic improvement of populations through MAS, including marker-assisted introgression. Its aim is to provide an easily understandable overview of the techniques, applications and issues involved in the use of DNA markers in MAS for genetic improvement of domestic plant and animal populations in developing countries. In the next section of the chapter, a brief description of the technical aspects of molecular markers and MAS is provided. The current status of the application of MAS in crops, forestry, livestock and fish is then summarized, while the final section

Note: This chapter is based on the Background Document to Conference 10 (on molecular marker-assisted selection as a potential tool for genetic improvement of crops, forest trees, livestock and fish in developing countries) of the FAO Biotechnology Forum, 17 November–14 December 2003 (available at www.fao.org/biotech/C10doc.htm).
highlights issues that might be important to applications of MAS in developing countries. Although molecular markers may be used for a wide range of different tasks, such as to quantify the genetic diversity and relationships within and between agricultural populations (e.g. livestock breeds), to investigate biological processes (such as mating systems, pollen movement or seed dispersal in plants) or to identify specific genotypes (e.g. cloned forest trees), these applications are not considered here.

**BACKGROUND TO MAS**

**Molecular markers**

All living organisms are made up of cells that are programmed by genetic material called DNA. This molecule is made up of a long chain of nitrogen-containing bases (there are four different bases – adenine [A], cytosine [C], guanine [G] and thymine [T]). Only a small fraction of the DNA sequence typically makes up genes, i.e. that code for proteins, while the remaining and major share of the DNA represents non-coding sequences, the role of which is not yet clearly understood. The genetic material is organized into sets of chromosomes (e.g. five pairs in *Arabidopsis thaliana*; 30 pairs in *Bos taurus* [cow]), and the entire set is called the genome. In a diploid individual (i.e. where chromosomes are organized in pairs), there are two alleles of every gene – one from each parent.

Molecular markers should not be considered as normal genes as they usually do not have any biological effect. Instead, they can be thought of as constant landmarks in the genome. They are identifiable DNA sequences, found at specific locations of the genome, and transmitted by the standard laws of inheritance from one generation to the next. They rely on a DNA assay, in contrast to morphological markers that are based on visible traits, and biochemical markers that are based on proteins produced by genes.

Different kinds of molecular markers exist, such as restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPDs) markers, amplified fragment length polymorphisms (AFLPs), microsatellites and single nucleotide polymorphisms (SNPs). They may differ in a variety of ways – such as their technical requirements (e.g. whether they can be automated or require use of radioactivity); the amount of time, money and labour needed; the number of genetic markers that can be detected throughout the genome; and the amount of genetic variation found at each marker in a given population. The information provided to the breeder by the markers varies depending on the type of marker system used. Each has its advantages and disadvantages and, in the future, other systems are likely to be developed. More details on the individual marker systems are provided in Chapter 3.

**From markers to MAS**

The molecular marker systems described above allow high-density DNA marker maps (i.e. with many markers of known location, interspersed at relatively short intervals throughout the genome) to be constructed for a range of economically important agricultural species, thus providing the framework needed for eventual applications of MAS.

Using the marker map, putative genes affecting traits of interest can then be detected by testing for statistical associations between marker variants and any trait of interest. These traits might be genetically simple – for example, many traits for disease resistance in plants are controlled by one or a few genes (Young, 1999). Alternatively,
they could be genetically complex quantitative traits, involving many genes (i.e. so-called quantitative trait loci [QTL]) and environmental effects. Most economically important agronomic traits tend to fall into this latter category. For example, using 280 molecular markers (comprising 134 RFLPs, 131 AFLPs and 15 microsatellites) and recording populations of rice lines for various plant water stress indicators, phenology, plant biomass, yield and yield components under irrigated and water stress conditions, Babu et al. (2003) detected a number of putative QTL for drought resistance traits.

Having identified markers physically located beside or even within genes of interest, in the next step it is now possible to carry out MAS, i.e. to select identifiable marker variants (alleles) in order to select for non-identifiable favourable variants of the genes of interest. For example, consider a hypothetical situation where a molecular marker M (with two alleles M1 and M2), identified using a DNA assay, is known to be located on a chromosome close to a gene of interest Q (with a variant Q1 that increases yield and a variant Q2 that decreases yield), that is, as yet, unknown. If a given individual in the population has the alleles M1 and Q1 on one chromosome and M2 and Q2 on the other chromosome, then any of its progeny receiving the M1 allele will have a high probability (how high depends on how close M and Q are to each other on the chromosome) of also carrying the favourable Q1 allele, and thus would be preferred for selection purposes. On the other hand, those that inherit the M2 allele will tend to have inherited the unfavourable Q2 allele, and so would not be preferred for selection. With conventional selection which relies on phenotypic values, it is not possible to use this kind of information.

The success of MAS is influenced by the relationship between the markers and the genes of interest. Dekkers (2004) distinguished three kinds of relationship:

- The molecular marker is located within the gene of interest (i.e. within the gene Q, using the example above). In this situation, one can refer to gene-assisted selection (GAS). This is the most favourable situation for MAS since, by following inheritance of the M alleles, inheritance of the Q alleles is followed directly. On the other hand, these kinds of markers are the most uncommon and are thus the most difficult to find.

- The marker is in linkage disequilibrium (LD) with Q throughout the whole population. LD is the tendency of certain combinations of alleles (e.g. M1 and Q1) to be inherited together. Population-wide LD can be found when markers and genes of interest are physically very close to each other and/or when lines or breeds have been crossed in recent generations. Selection using these markers can be called LD-MAS.

- The marker is not in linkage disequilibrium (i.e. it is in linkage equilibrium [LE]) with Q throughout the whole population. Selection using these markers can be called LE-MAS. This is the most difficult situation for applying MAS.

The universal nature of DNA, molecular markers and genes means that MAS can, in theory, be applied to any agriculturally important species. Indeed, active research programmes have been devoted to building molecular marker maps and detecting QTLs for potential use in MAS programmes in a whole range of crop, livestock, forest tree and fish species. In addition, MAS can be applied to support existing conventional breeding programmes. These programmes use strategies such as: recurrent selection (i.e.
using within-breed or within-line selection, important in livestock); development of crossbreds or hybrids (by crossing several improved lines or breeds) and introgression (where a target gene is introduced from, for example, a low-productive line or breed (donor) into a productive line (recipient) that lacks the target gene (a strategy especially important in plants). See Dekkers and Hospital (2002) for more details. MAS can be incorporated into any one of these strategies (e.g. for marker-assisted introgression by using markers to accelerate introduction of the target gene). Alternatively, novel breeding strategies can be developed to harness the new possibilities that MAS raises.

CURRENT STATUS OF APPLICATIONS OF MAS IN AGRICULTURE

Below is a brief summary of the current status regarding application of MAS in the different agricultural sectors. For more details, a number of case studies for crops are presented in Section II of the book and for livestock, forestry and fish in Sections III, IV and V, respectively.

Crops

The promise of MAS has possibly been greeted with the most enthusiasm and expectation in this particular agricultural sector, stimulating tremendous investments in the development of molecular marker maps and research to detect associations between phenotypes and markers. Molecular marker maps have been constructed for a wide range of crop species. Information on major plant projects (such as the sequencing of the entire rice genome) can be found at www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html.

In a recent review, however, Dekkers and Hospital (2002) noted that “as theoretical and experimental results of QTL detection have accumulated, the initial enthusiasm for the potential genetic gains allowed by molecular genetics has been tempered by evidence for limits to the precision of the estimates of QTL effects”, and that “overall, there are still few reports of successful MAS experiments or applications.” They reported that marker-assisted introgression of known genes was widely used in plants, particularly by private breeding companies, whereas marker-assisted introgression of unknown genes had often proved to be less useful in practice than expected. As Young (1999) wrote: “even though marker-assisted selection now plays a prominent role in the field of plant breeding, examples of successful, practical outcomes are rare. It is clear that DNA markers hold great promise, but realizing that promise remains elusive.”

There is also considerable divergence with respect to the applications of MAS among different crop species. For example, Koebner (2003) highlighted the relatively fast uptake of MAS in maize compared with wheat and barley, arguing that this largely reflected the breeding structure. Thus, whereas maize breeding is dominated in industrialized countries by a small number of large private companies that produce F_1 hybrids, a system allowing protection from farm-saved seed and competitor use, breeding for the other major cereal species is primarily by public sector organizations and most varieties are inbred pure breeding lines, a system allowing less protection over the released varieties. Progress in arable crops is nevertheless quite advanced compared with horticultural crop species such as apples and pears, where development of molecular marker maps has been slow and only few QTL have been detected (Tartarini, 2003), even if MAS can potentially be very
useful for genetic improvement of such long-cycle plants.

**Livestock**

Again, much effort has been put into the development of molecular marker maps in this sector. The first reported map in livestock was for chicken in 1992, which was quickly followed by the publication of maps for cattle, pigs and sheep. Since then, the search for useful markers has continued and further species have been targeted, including goat, horse, rabbit and turkey (see www.thearkdb.org/ for the current status regarding some major livestock species). Microsatellite markers have been of major importance.

Dekkers (2004) recently reviewed commercial applications of MAS in livestock and noted that several gene or marker tests are available on a commercial basis in different species and for different traits, and that the majority of uses involve GAS, where an important gene (e.g. responsible for a congenital defect) has been identified or, to a lesser degree, LD-MAS. He pointed out that documentation is poor since, although several genetic tests are available, the extent to which they are used in commercial applications is unclear, as is the manner in which they are used and whether their use leads to greater responses to selection. He concluded that “opportunities for the application of MAS exist, in particular for GAS and LD-MAS and, to a lesser degree, for LE-MAS because of greater implementation requirements. Regardless of the strategy, successful application of MAS requires a comprehensive integrated approach with continued emphasis on phenotypic recording programmes to enable QTL detection, estimation and confirmation of effects, and use of estimates in selection. Although initial expectations for the use of MAS were high, the current attitude is one of cautious optimism.”

**Forestry**

As for crops, extensive efforts have been devoted to construction of molecular marker maps for the major commercial genera, such as eucalypts, pines and acacia. RFLPs, RAPDs, microsatellites and AFLPs have been extensively used. The Web site http://dendrome.ucdavis.edu/index.php provides updated information on the status regarding molecular marker maps in forestry.

Molecular maps have been used to locate markers associated with variation in forestry traits of commercial interest, such as growth, frost tolerance, wood properties, vegetative propagation, leaf oil composition and disease resistance. Since MAS allows early selection before traits of interest (e.g. wood quality) are expressed, a major incentive for using molecular techniques in tree breeding is to improve the rate of genetic gain by reducing the long generation interval. However, Butcher (2003) noted that “MAS has yet to be incorporated in operational breeding programmes for plantation species” and she referred to the high costs of genotyping, the large family sizes required to detect QTL and the lack of knowledge of QTL interactions with genetic background, tree age and environment as explanatory factors.

In a recent review of biotechnology in forestry, Yanchuk (2002) also highlighted the potential advantage of early selection using MAS, but again pointed out that MAS is not yet being applied routinely in tree breeding programmes, largely “because of economic constraints (i.e. the additional genetic gains are generally not large enough to offset the costs of applying the technology). Thus it is likely that MAS will only be applied for a handful of species and situations, e.g. a few
of the major commercially used pine and Eucalyptus species. Molecular markers are therefore primarily an information tool and are used to locate DNA/genes that can be of interest for genetic transformation, or information on population structure, mating systems and pedigree confirmation."

**Fish**

Molecular marker maps have been constructed for a number of aquaculture species, e.g. tilapia, catfish, giant tiger prawn, kuruma prawn, Japanese flounder and Atlantic salmon, although their density is generally low. Density is high for the rainbow trout, where the map published in 2003 has over 1 300 markers spread throughout the genome – the vast majority are AFLPs but it also includes over 200 microsatellite markers (Nichols et al., 2003). Some QTLs of interest have been detected (e.g. for cold and salinity tolerance in tilapia and for specific diseases in rainbow trout and salmon). In a recent review of MAS in fish breeding schemes, Sonesson (2003) suggested that MAS would be especially valuable for traits that are impossible to record on the candidates for selection such as disease resistance, fillet quality, feed efficiency and sexual maturation, and concluded that MAS is not used in fish breeding schemes today and that the lack of dense molecular maps is the limiting factor.

**Conclusions**

Molecular marker maps, the necessary framework for any MAS programme, have been constructed for the majority of agriculturally important species but the density of the maps varies considerably among species. Currently, MAS does not play a major role in genetic improvement programmes in any of the agricultural sectors. Enthusiasm and optimism remain concerning the potential contributions that MAS offers for genetic improvement. However, this seems to be tempered by the realization that it may be more difficult and therefore take longer than originally thought before genetic improvement of quantitative traits using MAS is realized. The conclusions from the review by Dekkers and Hospital (2002) are a good reflection of this: “Further advances in molecular technology and genome programmes will soon create a wealth of information that can be exploited for the genetic improvement of plants and animals. High-throughput genotyping, for example, will allow direct selection on marker information based on population-wide LD. Methods to effectively analyse and use this information in selection are still to be developed. The eventual application of these technologies in practical breeding programmes will be on the basis of economic grounds, which, along with cost-effective technology, will require further evidence of predictable and sustainable genetic advances using MAS. Until complex traits can be fully dissected, the application of MAS will be limited to genes of moderate-to-large effect and to applications that do not endanger the response to conventional selection. Until then, observable phenotype will remain an important component of genetic improvement programmes, because it takes account of the collective effect of all genes.”

**Some Factors Relevant to Applying MAS in Developing Countries**

In the debate on the role or value of MAS as a potential tool for genetic improvement in developing countries, some of the potential factors that should be considered are described briefly below, as they may influence applications of the technology.
Economic factors
As with any new technology promising increased benefits, the costs of application must also be considered. According to Dekkers and Hospital (2002), “economics is the key determinant for the application of molecular genetics in genetic improvement programmes. The use of markers in selection incurs the costs that are inherent to molecular techniques. Apart from the cost of QTL detection, which can be substantial, costs for MAS include the costs of DNA collection, genotyping and analysis.” For example, Koebner (2003) suggested that the current costs of MAS would need to fall considerably before it would be used widely in wheat and barley breeding. In practice, therefore, although MAS may lead to increased genetic responses, decision-makers need to consider whether it may be cost-effective or whether the money and resources spent on developing and applying MAS might instead be more efficiently used on improving existing conventional breeding programmes or adopting other new technologies.

Little consideration has been given to this issue. Some results have, however, been published recently from studies at the International Maize and Wheat Improvement Center (CIMMYT) in Mexico on the relative cost-effectiveness of conventional selection and MAS for different maize breeding applications. One application considered by Morris et al. (2003) was the transfer of an elite allele at a single dominant gene from a donor line to a recipient line. Here, conventional breeding is less expensive but MAS is quicker. For situations like this, where the choice between conventional breeding and MAS involves a trade-off between time and money, they suggested that the cost-effectiveness of using MAS depends on four parameters: the relative cost of phenotypic versus marker screening; the time saved by MAS; the size and temporal distribution of benefits associated with accelerated release of improved germplasm and, finally, the availability to the breeding programme of operating capital. They conclude that “all four of these parameters can vary significantly between breeding projects, suggesting that detailed economic analysis may be needed to predict in advance which selection technology will be optimal for a given breeding project.”

In the applications considered by CIMMYT, the costs of developing molecular markers associated with the trait of interest were not considered, as it was assumed that they were already available. There is a distinction between development costs (e.g. identifying molecular markers on the genome, detecting associations between markers and the traits of interest) and running costs (typing individuals for the appropriate markers in the selection programme) of MAS. Development costs can be considerable, so developing countries need to consider whether to develop their own technology or, alternatively, to import the technology developed elsewhere, if available.

Another aspect to be considered is how to evaluate the economic benefits of MAS. For a publicly-funded breeding programme, it should include economic benefits to farmers from genetic improvement of their plants or animals. For private companies on the other hand, the impacts of using MAS on their market share, and not on rates of genetic improvement, would be of greatest interest.

The economics of MAS are considered in more detail later, in particular in Chapter 19.

MAS versus conventional methods
Although conventional breeding programmes that rely on phenotypic records
have their limitations, they have shown over time that they can be highly successful. Application of MAS will not occur in a vacuum and the potential benefits (genetic, economic, etc.) of using MAS need to be compared with those achieved or expected from any existing conventional breeding programmes.

In the different agricultural sectors, this question has received much attention from researchers. There seems to be general consensus that the relative success of MAS compared with conventional breeding may depend on the kind of trait (or traits) to be genetically improved. If the trait is difficult to record or is not routinely recorded in conventional programmes, MAS will offer more advantages than if it is routinely recorded. Similarly, if the trait is sex-limited or can only be measured late in life then MAS is favoured, as marker information can be used in both sexes and at any age.

In considering the merits of MAS versus conventional breeding, it is also important to keep in mind that the existence of a strong breeding programme is a prerequisite for the application of advanced molecular technologies such as MAS. In situations where the infrastructure and capacity are insufficient to support a successful conventional breeding programme, MAS will not provide a shortcut to genetic improvement.

**MAS versus other biotechnologies for genetic improvement**

The relative costs and benefits of applying MAS should be compared not only with conventional breeding but also with the use of other new technologies that can potentially improve agricultural populations genetically. These include tissue culture in crops and forest trees, reproductive technologies (e.g. embryo transfer or cloning) in livestock and triploidization or sex-reversal in farmed fish. They also include genetic modification, a technology that can be applied to all sectors. Compared with genetic modification, regulation of MAS, be it at the level of research and development, field testing, commercial release or import/export of developed products, is more relaxed; in addition, public acceptance of the technology is not an issue.

**Intellectual property rights issues**

As discussed in Conference 6 of the FAO Biotechnology Forum (FAO, 2001), the issue of intellectual property rights (IPRs) is playing an ever greater role in food and agriculture in developing countries. Participants in that conference, inter alia, suggested that this issue was having a generally negative influence on the quality of agricultural research carried out and on the nature of research collaborations between the public and private sector and between developing and developed countries.

It is therefore obvious that IPRs may also have an impact on the development and application of MAS in developing countries. For example, the AFLP molecular marker mapping technique is patented. Molecular markers can be patented, although this can often be overcome by using other markers near the gene of interest. Individual genes can also be patented. With IPRs, however, there is nevertheless public disclosure of the invention or information. Non-disclosure of information, where patents are not sought but the information on markers or detected QTL is nevertheless kept secret, can also have negative impacts, by denying developing countries access to potentially useful information.

More details on IPRs and MAS can be found in Chapter 20.
REFERENCES


CHAPTER 2

An assessment of the use of molecular markers in developing countries

Andrea Sonnino, Marcelo J. Carena, Elcio P. Guimarães, Roswitha Baumung, Dafydd Pilling and Barbara Rischkowsky
SUMMARY

Four different sources of information were analysed to assess the current uses of molecular markers in crops, forest trees and livestock in developing countries: the FAO Biotechnology in Developing Countries (FAO-BioDeC) database of biotechnology in developing countries; country reports evaluating the current status of applied plant breeding and related biotechnologies; country reports on animal genetic resources management for preparing the First Report on the State of the World’s Animal Genetic Resources (SoW-AnGR); and the results of a questionnaire survey on animal genetic diversity studies. Even if still largely incomplete, the current data show that molecular markers are widely used for plant breeding in the developing world and most probably their use will increase in the future. In the animal sector the use of molecular markers seems less developed and limited or absent in most developing countries. Major differences exist among and within regions regarding the application of molecular marker techniques in plant and animal breeding and genetics. These can be explained by the relatively high investments in infrastructure and human resources necessary to undertake research in these fields. The spectrum of application of molecular markers in crop plants is quite wide, covering many plants relevant to the enhancement of food security, but other important plant species are still neglected. The practical results of marker-assisted selection (MAS) in the field are disappointingly modest, possibly due to: low levels of investment; limited coordination between biotechnologists and practical breeders; instable, non-focused or ill-addressed research projects; and the lack of linkages between research and farmers. Partnerships between developed and developing countries may be a means of better realizing the potential of molecular marker techniques for improving both animal and crop production.
INTRODUCTION

Assessments relating to the use of molecular markers in crop plants are based on two sources of information: (i) FAO-BioDeC, a searchable database of biotechnology products and techniques in use and in the pipeline in developing and transition countries (available at www.fao.org/biotech/inventory_admin/dep/default.asp); and (ii) FAO country reports produced by national agricultural research systems (NARS) as part of a survey of country information and trends in resources allocated for applied plant breeding and related biotechnology, with the aim of raising awareness, evaluating opportunities for investment and designing national, regional and/or global strategies to strengthen the capacity of national plant breeding programmes (Guimarães, Kueneman and Carena, 2006).

As the FAO-BioDeC database contains little information on the use of molecular markers in relation to animals, it is even more difficult to give a comprehensive overview of the situation with respect to livestock in developing countries than it is for crops. However, information on the use of molecular markers was drawn from the country reports on animal genetic resources (AnGR) management submitted to FAO as part of the preparation of the First Report on the State of the World’s Animal Genetic Resources (SoW-AnGR) and from a questionnaire survey on genetic diversity studies. The country reports covered a wide variety of aspects of AnGR management and contain only quite general information about the role of molecular techniques. The questionnaire survey looked specifically at the use of molecular markers in livestock genetic diversity studies and was directed to researchers involved in such studies. As such, it gives an indication of where genetic diversity studies are being undertaken and which markers are primarily used, but it does not provide a complete picture.

While this book focuses on the use of markers to assist in genetic selection (MAS), it is often difficult to obtain specific information on the extent to which markers are used for this purpose in developing countries. For this reason, some of the data presented in this chapter cover the overall use of molecular markers in developing countries and do not allow discrimination between molecular markers used for selection from uses for other purposes, such as the descriptive studies of genetic diversity within populations or genetic distance between populations. Other data presented here describe the use of molecular markers for measuring genetic diversity only. In this case, the information can be considered as an indicator of the human capacity and infrastructure available for use of markers in MAS. For these reasons, and due to the incomplete nature of some of the information available, this overview should be considered preliminary, but still meaningful.

FAO-BioDeC

At the time of writing (September 2006), FAO-BioDeC includes 2,336 entries related to crops and 829 entries related to forest trees. The database currently covers 74 developing countries, including countries with economies in transition.

No quantitative information is available concerning the human capacity or funding involved in any research initiative. Activities carried out in developed countries or at international research centres, such as those that are part of the Consultative Group on International Agricultural Research (CGIAR), are not considered.

To compile the data in FAO-BioDeC, several sources of information were
consulted (for a complete description see FAO, 2005). In particular, information on plant biotechnology products and techniques was gathered from a survey undertaken in Latin America by the International Service for National Agricultural Research (ISNAR) and from country biotechnology status assessment reports prepared for FAO in South and Southeast Asia, Africa and transition countries in Eastern Europe. Other information was obtained from country reports and published literature.

The initial biotechnology application data obtained was classified on a country/regional/continental basis, by species, trait analysed or technique used, and by whether the application was in the research or field testing phases or was already commercially released.

FAO-BioDeC currently contains 677 entries related to the use of molecular marker techniques, 489 of which are associated with crop plants and 188 with forest trees. Table 1 suggests that early generation DNA-based molecular markers such as randomly amplified polymorphic DNAs (RAPDs) are more widely used than the more recently developed markers, e.g. amplified fragment length polymorphisms (AFLPs), while isozymes are still largely used in the forestry sector.

Only in five cases have the research initiatives reported reached the final stage of development, giving rise to commercialized products (Table 2). These are one variety of an unspecified ornamental plant released in Brazil; one variety of rice commercialized in Indonesia; one strain of *Rhizobium etli*, the soil bacterium inducing the formation of nitrogen-fixing nodules on the roots of a common bean obtained in Mexico; one rice variety containing pyramided genes for bacterial leaf blight resistance obtained in the Netherlands Antilles; and one variety of an unspecified forest tree in Burundi. In 115 cases (107 in the crop sector and eight

---

**Table 1**

<table>
<thead>
<tr>
<th>Markers</th>
<th>Crop</th>
<th>Forestry</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFLP</td>
<td>61</td>
<td>9</td>
<td>70</td>
</tr>
<tr>
<td>RAPD</td>
<td>158</td>
<td>15</td>
<td>173</td>
</tr>
<tr>
<td>SSRs/Microsatellites</td>
<td>68</td>
<td>19</td>
<td>87</td>
</tr>
<tr>
<td>AFLP</td>
<td>65</td>
<td>3</td>
<td>68</td>
</tr>
<tr>
<td>Isozymes</td>
<td>2</td>
<td>50</td>
<td>52</td>
</tr>
<tr>
<td>Chloroplast DNA markers</td>
<td>0</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>rDNA (ribosomal DNA sequences)</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Other or not specified</td>
<td>135</td>
<td>77</td>
<td>212</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>489</td>
<td>188</td>
<td>677</td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>Phase</th>
<th>Crop</th>
<th>Forestry</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental phase</td>
<td>344</td>
<td>179</td>
<td>523</td>
</tr>
<tr>
<td>Field tests</td>
<td>107</td>
<td>8</td>
<td>115</td>
</tr>
<tr>
<td>Commercial phase</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Unspecified</td>
<td>34</td>
<td>0</td>
<td>34</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>489</td>
<td>188</td>
<td>677</td>
</tr>
</tbody>
</table>
for forest trees), the research initiatives have reached the field test stage, while in 523 cases (344 of which are related to the crop sector), they are at earlier stages.

The use of molecular markers is widespread in Latin America and the Caribbean with molecular research being reported from ten countries. Special emphasis is on the crop sector and includes applications on Andean local roots and tubers, sugar cane, rice, cocoa, banana, bean and maize (Table 3). In the Asia and Pacific region, research activities with molecular markers focus on forest trees, sugar cane, rice, jute, banana, coconut and wheat. The FAO-BioDeC database shows that, while research involving molecular markers in Africa is under way in only a few countries including Ethiopia, Nigeria, South Africa and Zimbabwe, the crops under study range from traditional commodities to tropical fruits. Molecular research in the Near East and North Africa is reported for only six countries and focuses on date palm, durum and bread wheat, rice, barley and olive trees. In transition countries of Eastern Europe, molecular markers target several crop plants including wheat, maize, pulses, vegetables and tobacco across seven countries.

Table 4 shows that most attention focuses on cereals, especially durum and bread wheat, barley, maize and rice. Other important cereal or pseudo-cereal species such as sorghum, amaranthus and buckwheat receive less attention and no research initiatives are reported for teff or millets. Among the pulses, molecular research projects are reported for beans (18), chickpea (5), cowpea (9) and soybean (7) and little or no attention is dedicated to lentil, pigeon pea, faba bean and other locally important leguminous plants such as bambara groundnut. Among root and tuber crops, potato, sweet potato and cassava attract the most research effort involving molecular markers, but some research is also undertaken on Andean roots and tubers. Few or no records are available for root and tuber species important for food security in many developing countries such as yam, taro (or dasheen), cocoyam and other aroids. Research on fruit trees involving molecular markers includes tropical fruit trees such as banana, cocoa, coconut and papaya, as well as plants more typical of temperate climates such as strawberry and apple, while less
research was reported for citrus, mango, pineapple and many other fruit trees largely cultivated in developing countries. Several research initiatives are applying molecular markers to industrial crop species, such as sugar cane, cotton, rubber, jute, coffee, flax and oil palm.

**FAO PLANT BREEDING AND RELATED BIOTECHNOLOGY CAPACITY ASSESSMENT**

In 2002, a draft questionnaire was designed to gather country information on resource allocation trends in plant breeding and biotechnology related activities. Later in the same year, a group of experts including representatives from CGIAR centres, the public and private sectors and non-governmental organizations (NGOs), met at FAO headquarters to discuss the nature of the information to be collected and the procedure for its collection. This resulted in a questionnaire being developed and sent to all public and private applied plant breeding programmes as well as to biotechnology laboratories in developing countries and countries in transition. Among other issues, the survey gathered information on the number of full-time equivalent plant breeders and biotechnologists available during each five-year period beginning from 1985. The questionnaire also requested information concerning trends of resources allocated to biotechnology as well as to germplasm improvement (pre-breeding), line development and line evaluation.

One of the objectives of the survey was to assess the gap between biotechnology tools and their successful deployment in applied breeding programmes (Guimarães, Kueneman and Carena, 2006). The survey therefore also concentrated on priorities for breeding, potential international support to strengthen national breeding programmes, the number of varieties released and the factors that are most likely to limit the success of applied plant breeding programmes, including the current status of biotechnology. The work of gathering the information and preparing a technical report on the current status of national plant breeding and related biotechnology was assigned to a well-known and respected national plant breeding scientist. This has been the key to identifying gaps in order to develop strategies for strengthening efforts directed at the sustainable use of plant genetic resources for food and agriculture (PGRFA) in national programmes.

For the purposes of this chapter, biotechnology data were gathered from 25 countries to complement the preliminary assessments based on FAO-BioDeC on the use of molecular markers in developing countries (Table 5). The data gathered indicate that tissue culture is the most common biotechnology technique as it was used in 88 percent of all cases, followed by MAS (44 percent), the double-haploid technique (32 percent), interspecific crosses (28 percent), molecular characterization (24 percent) and genetic engineering (12 percent).

Applications of molecular markers include a number of categories within biotechnology such as MAS, molecular characterization, facilitating genetic engineering and tracking desirable chromosome segments when making wide crosses (e.g. interspecific crosses). The results in Table 5 suggest that molecular markers might be an integral part of developing country agricultural efforts. MAS seems to be the second most utilized biotechnology tool applied after tissue culture, implying that emphasis should be given to the development of molecular markers to make selection more efficient. However, rapid and efficient
advancement of plant breeding efforts might not be achieved through MAS because of the complexity encountered in multitrait and multistage selection for economically important traits. Consequently, today in the developed world, molecular markers do not have a prominent role in breeding programmes (Hallauer, 1999).

USE OF MOLECULAR TECHNIQUES IN AnGr MANAGEMENT
FAO invited 188 countries to participate in the preparation of the First Report on the SoW-AnGR. One hundred and sixty-nine country reports (CR) on AnGR were submitted (available at www.fao.org/dad-is/).

The countries were offered guidelines for the preparation of the country reports, one section of which was to be devoted to reviewing the state of national capacities and assessing future capacity building requirements (FAO, 2001). Countries were assigned to seven regions on the basis of the regional classification established by FAO for the purpose of preparing the SoW-AnGR. This analysis considered 148 country reports available by July 2005, of which 42 were from Africa, 25 from Asia, 39 from Europe and the Caucasus, 22 from Latin America and the Caribbean, 7 from the Near and Middle East, 2 from North America and 11 from the Southwest Pacific (Pilling et al., 2007).

TABLE 5
Biotechnology applications in plant genetic resources for food and agriculture in use in 25 developing countries

<table>
<thead>
<tr>
<th>Country</th>
<th>TC</th>
<th>MAS</th>
<th>IC</th>
<th>DH</th>
<th>MC</th>
<th>GE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N1</td>
<td>N</td>
</tr>
<tr>
<td>Angola</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Armenia</td>
<td>X</td>
<td>N</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>Cameroon</td>
<td>X</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>X</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>X</td>
<td>X</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Georgia</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>Ghana</td>
<td>X</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Mali</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Kenya</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N</td>
<td>X</td>
</tr>
<tr>
<td>Malawi</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Moldova</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Mozambique</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>X</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>Niger</td>
<td>X</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Nigeria</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>Senegal</td>
<td>X</td>
<td>N</td>
<td>X</td>
<td>X</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Sudan</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Tunisia</td>
<td>X</td>
<td>X</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Uzbekistan</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>X</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>Zambia</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>X</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

1 One or more institutions in the country are using the tool. However, this does not measure its impact.
2 Not in use.

TC = tissue culture; MAS = marker-assisted selection; IC = interspecific crosses; DH = double-haploid technology; MC = molecular characterization; GE = genetic engineering
Not surprisingly, the information provided by the country reports indicates that there is a large gap between developed and developing countries in terms of capacity to utilize molecular markers for the study and management of AnGR (Table 6). Compared with other developing regions, a higher percentage of countries from Asia and Latin America and the Caribbean reported their use. In Africa, the Southwest Pacific (excluding Australia), the Near and Middle East, and Eastern Europe and the Caucasus, very few countries report the use of these technologies, the prominent exception in the last case being Ukraine which has carried out molecular characterization and genetic distance studies on a number of livestock species (CR Ukraine, 2004).

In Africa, only four countries describe the existence of characterization or genetic distance studies based on the use of molecular markers and in all cases the studies relate to local breeds. One country report indicates that local breeds of goat, pig and chicken are the subject of molecular characterization carried out abroad. In no case is the use of MAS reported from this region.

Excluding Japan, seven Asian countries (out of 15 providing information on whether or not the technologies are used) report molecular marker studies, of which five specify genetic distance studies and one mentions research into MAS (CR Malaysia, 2003). In Latin America and the Caribbean, 11 countries out of the 15 that provided information indicate some use of molecular markers. Among nine countries providing information on the species involved in molecular characterization studies, seven mentioned cattle while smaller numbers mention sheep, pigs, chickens, horses, goats, buffaloes, llamas, alpacas, vicuñas or guanacos. Several countries indicate the inclusion of locally adapted breeds in such studies, but there was little indication that molecular markers have been incorporated within breeding programmes. However, the report from Colombia (2003) noted the potential significance of MAS programmes for utilizing the genes of the Blanco Orejinegro cattle breed, which is reported to show resistance to brucellosis and which has been the subject of molecular characterization.

### Table 6

<table>
<thead>
<tr>
<th>Region</th>
<th>Number providing information</th>
<th>Reporting use of molecular markers %</th>
<th>Number with information on species</th>
<th>Reporting use of molecular markers in cattle %</th>
<th>In other species %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>29</td>
<td>83</td>
<td>18</td>
<td>89</td>
<td>100</td>
</tr>
<tr>
<td>Africa</td>
<td>29</td>
<td>14</td>
<td>3</td>
<td>100</td>
<td>33</td>
</tr>
<tr>
<td>Asia</td>
<td>16</td>
<td>50</td>
<td>7</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td>Latin America and the Caribbean</td>
<td>15</td>
<td>73</td>
<td>9</td>
<td>78</td>
<td>89</td>
</tr>
<tr>
<td>Southwest Pacific</td>
<td>9</td>
<td>11</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>North America</td>
<td>2</td>
<td>100</td>
<td>1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Near and Middle East</td>
<td>5</td>
<td>40</td>
<td>2</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

A range of species are the subject of molecular characterization, the most common being cattle, chickens, sheep, goats and pigs; however, some studies involving buffaloes, ducks, horses, camels or deer are also reported. Systematic studies of Asian breeds are being conducted by the Society for Research on Native Livestock in Japan, including analysis of genetic relationships based on mitochondrial DNA polymorphisms and other DNA markers (CR Japan, 2003).
Apart from Australia, no countries in the Southwest Pacific region report the use of molecular markers.

In the Near and Middle East one report (CR Jordan, 2003) refers to molecular characterization and genetic distance studies in indigenous goats, while another (CR Egypt, 2003) notes that molecular studies of buffalo, sheep and goats had recently been initiated with the aid of regional and international organizations.

SURVEY ON THE USE OF MOLECULAR MARKERS IN GENETIC DISTANCE STUDIES IN LIVESTOCK

More specific and detailed information on the use of molecular markers in AnGR research was obtained from a questionnaire study launched in 2003. One hundred and thirty-two questionnaires were sent out via e-mail to research teams that had been involved in genetic distance studies during the past ten years. The researchers were identified through a literature search and enquiry via several Internet discussion groups. The points covered in the survey were: number of breeds and sample sizes; number and type of markers used; additional breed information such as phenotypic traits or geographic spread; and the mathematical and statistical methods chosen for measuring genetic distance. The study also aimed to verify the degree of familiarity and acceptance of measurement of domestic animal diversity (MoDAD) recommendations, which had been proposed as standards for genetic diversity studies by the International Society for Animal Genetics (ISAG) and FAO about ten years earlier (FAO, 1998a; b). Compliance with the recommendations was seen as important as it would enable the compilation of results from different genetic distance studies.

Information on 87 genetic distance studies was obtained from 57 researchers. The studies covered breeds from 13 mammalian and avian species and investigated samples from 93 countries; the largest number of countries was in Europe, followed by those in Asia and the Pacific (Table 7). Most of the studies focused on ruminants. The size of the projects varied between one and 120 breeds originating from up to 33 countries. However, a large number of national projects focused on breeds within a specific country or region. There were also a few large international projects involving cattle and goats (Table 8). A smaller number of pig and chicken projects were implemented. No feedback was received regarding breeds of llamas, ducks, turkeys or geese.

With regard to compliance with the recommendations of the FAO/ISAG advisory group, 95 percent of all projects aimed to fulfil the minimum requirement of sampling 25 animals per breed. Although microsatellite markers were used in 90 percent of the studies, in only 23 percent were all markers taken from the recommended marker list. In about 57 percent of studies some recommended microsatellites were used. The degree of acceptance of the recommendations was highest in pigs and lowest in chickens. More detailed information on the results is given by Baumung, Simianer and Hoffmann (2004) and FAO (2004).

<table>
<thead>
<tr>
<th>FAO region</th>
<th>Number of countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>13</td>
</tr>
<tr>
<td>Asia and the Pacific</td>
<td>19</td>
</tr>
<tr>
<td>Europe</td>
<td>37</td>
</tr>
<tr>
<td>Latin America and the Caribbean</td>
<td>10</td>
</tr>
<tr>
<td>Near East</td>
<td>9</td>
</tr>
<tr>
<td>North America</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>93</strong></td>
</tr>
</tbody>
</table>

Table 7  
Number of countries where samples were collected for AnGR genetic distance studies

TABLE 7  
Number of countries where samples were collected for AnGR genetic distance studies

In about 57 percent of studies some recommended microsatellites were used. The degree of acceptance of the recommendations was highest in pigs and lowest in chickens. More detailed information on the results is given by Baumung, Simianer and Hoffmann (2004) and FAO (2004).
Conclusions

Even if still largely incomplete, the current data allow some general conclusions to be drawn regarding the use of molecular markers in agricultural research and development in developing countries.

Molecular markers are widely utilized in the plant production sector of the developing world even if the present uptake of molecular marker technologies does not reflect their actual potential. It might therefore be speculated that a significant increase in their utilization might be expected in the near future. However, it is recommended that each technique is carefully assessed for its actual potential for improving the efficiency of plant breeding and germplasm characterization. Until this is demonstrated, the use of molecular markers would be a costly investment with limited returns. Publishing all marker research that has not been successful is also strongly encouraged in order to avoid potential failures and/or importing inappropriate technologies from developed countries.

Major differences exist between regions (and within regions) regarding the application of molecular marker techniques in plant breeding and genetics. While some countries have developed quite extensive research programmes, vast geographical areas, particularly in Africa, remain excluded from these technological advancements or can count only on minimal activities. This can be explained by the relatively high investments in infrastructure and human resources necessary to undertake research in this field. High costs can also be indicated as a cause of the low technological level of genetic marker research in many countries, which focus on isozymes or on restriction fragment length polymorphisms (RFLPs) and have not yet adopted the more advanced polymerase chain reaction (PCR)-based markers. However, the life span of PCR-based markers is very short and it might be better to wait until improved markers such as single nucleotide polymorphisms (SNPs) become available. The spectrum of application of molecular markers to crop plants in developing countries is quite wide and covers many plant species that are relevant for the enhancement of food security or for the improvement of farmers’ incomes.

### Table 8

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of projects</th>
<th>Number of countries</th>
<th>FAO region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffalo</td>
<td>3</td>
<td>9</td>
<td>Africa, Asia and the Pacific, Europe, Latin America and the Caribbean</td>
</tr>
<tr>
<td>Cattle</td>
<td>24</td>
<td>40</td>
<td>Africa, Asia and the Pacific, Europe, Latin America and the Caribbean</td>
</tr>
<tr>
<td>Goat</td>
<td>11</td>
<td>28</td>
<td>Africa, Asia and the Pacific, Europe, Latin America and the Caribbean</td>
</tr>
<tr>
<td>Sheep</td>
<td>19</td>
<td>56</td>
<td>Africa, Asia and the Pacific, Europe, Latin America and the Caribbean</td>
</tr>
<tr>
<td>Pig</td>
<td>6</td>
<td>19</td>
<td>Africa, Asia and the Pacific, Europe</td>
</tr>
<tr>
<td>Ass</td>
<td>1</td>
<td>1</td>
<td>Europe</td>
</tr>
<tr>
<td>Horse</td>
<td>5</td>
<td>25</td>
<td>Africa, Asia and the Pacific, Europe, Latin America and the Caribbean, North America</td>
</tr>
<tr>
<td>Bactrian camel</td>
<td>1</td>
<td>2</td>
<td>Asia and the Pacific</td>
</tr>
<tr>
<td>Dromedary</td>
<td>2</td>
<td>7</td>
<td>Africa, Near East</td>
</tr>
<tr>
<td>Alpaca</td>
<td>3</td>
<td>2</td>
<td>Near East, Latin America and the Caribbean</td>
</tr>
<tr>
<td>Rabbit</td>
<td>1</td>
<td>19</td>
<td>Africa, Asia and the Pacific, Europe</td>
</tr>
<tr>
<td>Chicken</td>
<td>8</td>
<td>34</td>
<td>Africa, Asia and the Pacific, Europe, Latin America and the Caribbean, Near East</td>
</tr>
<tr>
<td>Yak</td>
<td>2</td>
<td>8</td>
<td>Asia and the Pacific, Europe, Near East</td>
</tr>
</tbody>
</table>
in tropical areas. However, other important plant species are still neglected by the ongoing research initiatives.

According to the data reported in FAO-BioDeC, only five products obtained through the use of molecular markers have been commercially released to date in developing countries. Even if more commercial products have been released but are missing from the database, such as those reported by Toenniessen, O’Toole and DeVries (2003) or others obtained by the international agricultural research centres or the private sector, the totality of practical results obtained from using molecular markers is disappointingly modest compared with the declared potential of the approach. The reasons for the poor results to date are multiple and include: the low level of investments in both biotechnology research and applied plant breeding; the limited coordination between biotechnology laboratories and plant breeding programmes; managerial and political frailties leading to unstable, unfocused or ill-addressed research projects; legal, infrastructural or technical weaknesses of the seed production and commercialization systems; and the lack of linkages between research and practical application of research products by farmers.

Applied plant breeding should continue to be the foundation for the application of molecular markers. Focusing useful molecular techniques on the right traits will build a strong linkage between genomics and plant breeding in order to produce new and better cultivars. Therefore, more than ever, there is the need for better communication and cooperation among scientists in plant breeding and biotechnology. Public plant breeding and biotechnology programmes in developing countries are being seriously eroded through lack of funding.

This loss of public support affects breeding continuity and objectivity and, equally importantly, the training of future plant breeders and biotechnologists and the utilization and improvement of plant genetic resources currently available. The fact that poor farmers rely on public and private breeding institutions for solving long-term challenges should influence policy-makers to reverse the trend of reduced funding. Cooperation between industry and public institutions is a promising approach to follow. Ensuring strong applied breeding programmes incorporating the application of molecular markers will be essential in ensuring the sustainable use and enhancement of plant genetic resources.

AnGR management shows a similar pattern to the use of MAS in plant breeding management in terms of the differences that exist among regions in the use of molecular marker techniques. Within several regions there are also differences between more and less developed countries. The reasons are similar to those mentioned above, namely a lack of financial, human and technical resources. In particular, human capacities in animal genetics and breeding are much smaller than those existing in the crop sector. Consequently, the use of molecular techniques to evaluate genetic resources, to plan conservation efforts, or to facilitate the achievement of desired breeding objectives is limited or absent in most developing countries.

Nevertheless, country reports expressed a strong desire to develop greater capacity to carry out molecular studies of national AnGR, and the responses to the FAO questionnaire also indicated a high level of interest in doing so. For the near future, microsatellite loci will remain the most useful type of genetic marker for genetic distance studies and for genetic improvement
programmes but SNPs were singled out as promising markers for the future. With partnerships between developed and developing countries within or across regions, genetic diversity studies may be a means of realizing the potential of molecular marker techniques to improve decision-making on breed development and the prioritization of breeds for conservation programmes.

The successful application of MAS in animal breeding necessitates a high level of expenditure in terms of establishment and maintenance costs and requires skilled human resources, equipment, laboratories and supportive infrastructure. As such, the cost-effectiveness of these strategies has to be carefully evaluated before promoting them in resource-poor environments.

REFERENCES


FAO. Country reports on the state of animal genetic resources (available at www.fao.org/dad-is/).


SECTION II

Marker-assisted selection in crops – case studies
Molecular markers for use in plant molecular breeding and germplasm evaluation

Jeremy D. Edwards and Susan R. McCouch
SUMMARY
A number of molecular marker technologies exist, each with different advantages and disadvantages. When available, genome sequence allows for the development of greater numbers and higher quality molecular markers. When genome sequence is limited in the organism of interest, related species may serve as sources of molecular markers. Some molecular marker technologies combine the discovery and assay of DNA sequence variations, and therefore can be used in species without the need for prior sequence information and up-front investment in marker development. As a prerequisite for marker-assisted selection (MAS), there must be a known association between genetic markers and genes affecting the phenotype to be modified. Comparative databases can facilitate the transfer of knowledge of genetic marker-phenotype association across species so that discoveries in one species may be applied to many others. Further genomics research and reductions in the costs associated with molecular markers will continue to provide new opportunities to employ MAS.
INTRODUCTION
Molecular markers are valuable tools for the classification of germplasm and in MAS. The purpose of this chapter is to provide guidance in selecting appropriate molecular marker systems based on the availability of technological resources in various species and to provide some examples of MAS applications. One of the many benefits of the increasing amount of DNA sequence information in many organisms is the expanding opportunity for the development of new molecular markers. As the full genome sequence will not be available for most species of interest in the near future, it is important to find strategies for developing and using molecular markers when sequence resources are limited. This chapter describes several technologies that exist for developing molecular markers without DNA sequence information. It also draws on some examples from rice (Oryza sativa L.) to illustrate how molecular marker development was influenced by the addition of each layer of sequence information, culminating in the present status of rice as the first crop with nearly complete genome sequence information.

MOLECULAR MARKER TECHNOLOGIES
Restriction fragment length polymorphisms
Restriction fragment length polymorphisms (RFLPs) were the first DNA-based molecular markers. An application of Southern analysis (Southern, 1975), RFLPs exploit the ability of single stranded DNA to bind (hybridize) to DNA with a complementary sequence. RFLP markers detect variation in DNA sequences at the same loci in different individuals or accessions. Technically, RFLP technology involves the hybridization of cloned DNA to restriction fragments of differing molecular weights from restriction enzyme-digested genomic DNA. The digested DNA fragments are size-separated on agarose gels by electrophoresis and transferred as denatured (single stranded) arrays of fragments to filters through capillary action. The filters are then incubated with specific labelled probes (genes or anonymous fragments of single stranded DNA), washed and exposed to x-ray film. To identify polymorphisms between individuals or accessions, the genomic DNA extracted from each individual is digested with a series of restriction enzymes to find enzymes that produce fragments (bands) that differ in molecular weight between accessions and can be distinguished by hybridization with a given probe. To ensure that probes hybridize to single fragments on a gel, the DNA used as a probe should be from a single or low copy (non-repetitive) region of the genome. Probes may represent genes (i.e. derived from complementary DNA [cDNA]) or they may represent anonymous sequences derived from genomic DNA. Genomic probes are generated by shearing or digesting DNA and cloning the fragments into a plasmid vector that allows for amplification of the cloned fragment in a suitable host. To increase the frequency of low copy clones in a genomic library, the DNA may be digested with a methylation-sensitive enzyme, such as PstI. The repetitive regions of a genome are typically heavily methylated and thus produce fragments >25 kb when digested with a methylation-sensitive enzyme. As a result, these fragments do not clone efficiently into plasmid vectors and consequently are effectively filtered out of the analysis. Thus, use of methylation-sensitive enzymes increases the representation of unmethylated and typically low copy gene sequences in RFLP analysis. Sharing of anonymous,
unsequenced RFLP markers among researchers requires an infrastructure for the maintenance and distribution of cloned probes for use by multiple researchers. However, if end-sequence or full-clone sequence information is available, the probes can be amplified readily from genomic DNA via the polymerase chain reaction (PCR), and the cumbersome aspects of clone maintenance and distribution are avoided. The polymorphisms detected by RFLPs may result from single base changes causing a loss of restriction sites or a gain of new restriction sites, or from insertions and deletions (indels) between restriction sites (McCouch et al., 1988; Edwards, Lee and McCouch, 2004).

PCR-based markers
Many advances in molecular marker technology have come through applications of the PCR method (Mullis et al., 1986). In PCR, a thermo-stable DNA polymerase enzyme makes copies of a target sequence beginning from two small pieces of synthetically produced DNA (primers) that are complementary to sequences bracketing the target. Through iterations of the process with heating to separate the double stranded DNA molecules and cooling to allow the primers to re-anneal, the target sequence is exponentially amplified. Polymerase chain reaction-based markers require much less DNA per assay than RFLPs and are more compatible with automated high-throughput genotyping (i.e. the ability to process large numbers of samples quickly and efficiently).

Randomly amplified polymorphic DNA markers
Randomly amplified polymorphic DNA markers (RAPDs) use PCR to amplify stretches of DNA between single primers of arbitrary sequence (Williams et al., 1990; Welsh and McClelland, 1990). Amplification occurs only where sequences complementary to the primers are in close enough proximity for successful PCR. The typical oligonucleotide used for RAPDs is ten bases long and will amplify many loci simultaneously, allowing multiple markers to be assayed in a single PCR reaction and a single lane on an agarose gel. As the primers are arbitrary, RAPD technology can be applied directly to any species with no prior sequence knowledge. This technology is particularly useful when there is a need to assay loci across the entire genome. The polymorphisms are detected only as the presence or absence of a band of a particular molecular weight, and it is not possible to differentiate between homozygous and heterozygous markers. RAPDs are notoriously unreliable because, aside from sequence differences, the amplification or failure of amplification of any band may be sensitive to any number of factors, including DNA template quality, PCR conditions, reagents and equipment.

Amplified fragment length polymorphisms
Amplified fragment length polymorphisms (AFLPs) are molecular markers derived from the selective amplification of restriction fragments (Vos et al., 1995). Genomic DNA is digested with a pair of restriction enzymes and oligonucleotide adaptors are ligated to the ends of each restriction fragment. The fragments are amplified using primers that anneal to the adaptor sequence and extend into the restriction fragment. Only a portion of restriction fragments will be within the range of sizes than can be amplified by PCR and visualized on polyacrylamide gels (between 50 and 350 bp). For large genomes, additional selective bases
can be added to the primers to reduce the number of co-amplified bands. AFLPs have many of the advantages of RAPDs, but have much better reproducibility. AFLP technology requires greater technical skill than RAPDs and, because AFLPs run on polyacrylamide gels instead of agarose, they also require a larger investment in equipment than RAPDs. Using manual gels, AFLP bands are detectable using silver stain, or by labelling of the primers with a radioactive isotope. Alternatively, for higher throughput, AFLPs can be detected with an automated DNA sequencer by using fluorescently labelled primers.

Diversity array technology (DArT) is a modification of the AFLP procedure using a microarray platform (Jaccoud et al., 2001) that greatly increases throughput. In DArT, DNA fragments from one sample are arrayed and used to detect polymorphisms for the fragments in other samples by differential hybridization (Wenzl et al., 2004).

DEVELOPING MOLECULAR MARKERS WITH DNA SEQUENCE INFORMATION

When the DNA sequence is available, it is possible to design primers to amplify across a specific locus. However not all loci will be polymorphic. Targeting highly variable sequence features increases the likelihood of detecting polymorphism. These highly variable features include tandem repeats such as microsatellites, and dispersed complex repeats such as transposable elements.

Microsatellites

Simple sequence length polymorphisms (SSLPs), also known as simple sequence repeats (SSRs), or microsatellites, consist of tandemly repeated di-, tri- or tetra-nucleotide motifs and are a common feature of most eukaryotic genomes. The number of repeats is highly variable because slipped strand mis-pairing causes frequent gain or loss of repeat units. With their high level of allelic diversity, microsatellites are valuable as molecular markers, particularly for studies of closely related individuals.

PCR-based markers are designed to amplify fragments that contain a microsatellite using primers complementary to unique sequences surrounding the repeat motif (Weber and May, 1989). Differences in the number of tandem repeats are readily assayed by measuring the molecular weight of the resulting PCR fragments. As the differences may be as small as two base pairs, the fragments are separated by electrophoresis on polyacrylamide gels or using capillary DNA sequencers that provide sufficient resolution.

Without prior sequence knowledge, microsatellites can be discovered by screening libraries of clones. Clones containing the repeat motif must be sequenced to find unique sites for primer design flanking the repeats. Microsatellite marker development from pre-existing sequence is far more direct. Good reviews of microsatellite marker development include those of McCouch et al. (1997) and Zane, Bargelloni and Atarnello (2002). Microsatellites discovered in non-coding sequence often have a higher rate of polymorphism than microsatellites discovered in genes. However, in some species such as spruce (Picea spp.) with highly repetitive genomes, SSR markers developed from gene sequences have fewer instances of null alleles, i.e. failure of PCR amplification (Rungus et al., 2004).

Microsatellite markers have several advantages. They are co-dominant; the heterozygous state can be discerned from the homozygous state. The markers are easily automated using florescent primers on an automated sequencer and it is possible
to multiplex (combine) several markers with non-overlapping size ranges on a single electrophoresis run. The results are highly reproducible, and the markers are easily shared among researchers simply by distributing primer sequences. Although SSRs are abundant in most eukaryotic genomes, their genomic distribution may vary. Uneven distributions of microsatellites limit their usefulness in some species.

Inter-SSRs (ISSRs) are another type of molecular marker that makes use of microsatellite sequences. ISSRs use PCR primers anchored in the termini of the repeats extending into the flanking sequence by several nucleotides (Zietkiewicz, Rafalski and Labuda, 1994). PCR products are produced for each pair of microsatellites that are in sufficient proximity for PCR to occur, or may be generated by anchoring one primer in the SSR motif and using a second “universal” primer corresponding to a sequence that has been ligated onto the ends of restriction fragments (as in the AFLP technique described above, where genomic DNA is first digested with a restriction enzyme and oligonucleotide adaptors are ligated to the ends of each restriction fragment, except that one primer resides in an SSR motif that is bracketed by the restriction sites) (Gupta et al., 1994; Goodwin, Aitken and Smith, 1997). Markers at multiple loci are assayed as the presence or absence of bands of particular sizes. ISSRs can be visualized on agarose gels, on silver stained polyacrylamide gels or fluorescently labelled for detection with an automated DNA sequencer.

Transposable element-based markers
Transposable elements (TEs) are another rapidly changing feature of the genome that can be exploited as a source of variability for molecular markers. Discovery of TE sequences is a prerequisite for their use as markers. While TEs may be discovered as mutations in alleles of genes conferring mutant phenotypes, they have also been discovered directly in genomic sequence (reviewed by Feschotte, Jiang and Wessler, 2002). Transposon display is a modified AFLP procedure that differs only in that one of the two primers is designed within the consensus sequence of a TE family so that amplification depends on the presence of a TE insertion within a restriction fragment (Casa et al., 2000). Using this approach, the presence or absence of a TE can be assayed simultaneously at many loci throughout the genome. To assay for a TE insertion at a specific locus, single copy “anchor markers” can be designed with primers located in unique sequences flanking the region of interest. A size polymorphism indicates the presence or absence of the TE in that particular location. Anchor markers are advantageous because they are co-dominant, can be run on a simple agarose gel system and are biologically informative in that they provide evidence of both complete, or incomplete, insertion or excision events. This methodology can also be applied to any known indel feature regardless of whether or not it is derived from a TE.

Single nucleotide polymorphisms
Single nucleotide polymorphisms (SNPs) are an abundant source of sequence variants that can be targeted for molecular marker development. Of all the molecular marker technologies available today, SNPs provide the greatest marker density. SNPs are often the only option for finding markers very near or within a gene of interest, and can even be used to detect a known functional nucleotide polymorphism (FNP). Discovery of SNPs
requires obtaining an initial DNA sequence in a reference individual followed by some form of re-sequencing in other varieties to find variable base pairs. In addition to direct sequencing, SNPs can be discovered through ecotilling with the CEL I enzyme (Comai et al., 2004) or by denaturing high pressure liquid chromatography (DHPLC) to measure small conformational differences when PCR amplified sequences are hybridized to a reference sequence (Kwok, 2001). In addition to SNP discovery, both DHPLC and ecotilling are viable technologies for SNP detection. There is a myriad of other SNP assay technologies in development and to date no single method stands out as superior to the others. Table 1 lists some examples of SNP allele discrimination methods and detection systems that can be combined in various ways (see reviews by Kwok, 2001 and Gut, 2001). The benefits of SNP assays include increased speed of genotyping, lower cost and the parallel assay of multiple SNP.

**TABLE 1**

<table>
<thead>
<tr>
<th>SNP technologies</th>
<th>Detection methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele discrimination</td>
<td>Gel separation</td>
</tr>
<tr>
<td>Hybridization</td>
<td>Arrays</td>
</tr>
<tr>
<td>Primer extension</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>Ligation</td>
<td>Plate readers</td>
</tr>
</tbody>
</table>

**Special Considerations for Diversity Studies and Germplasm Evaluation**

The interpretation of molecular marker data for germplasm classification and diversity can be confounded by uncertainty about the underlying sources of the polymorphisms and by homoplasy (false homology). For RFLPs in rice, indels can account for as much or more of the polymorphism as changes in the restriction sites themselves (Edwards, Lee and McCouch, 2004). AFLPs and RAPDs can also be sensitive to both indels and base changes. The ratio of indels to base changes is important for diversity studies because, when molecular markers are used to estimate nucleotide divergence, the divergence will be overestimated if indel-derived polymorphisms are common (Upholt, 1977; Nei and Miller, 1990; Innan et al., 1999). The greatest certainty of the underlying polymorphism comes from SNP technologies that directly assay for single base changes.

For SSR markers among closely related individuals, most polymorphism should be caused by expansion or contraction of the number of repeat units. However, as genetic distance between the varieties increases, there is an increasing chance that indel events will cause additional size

_Single feature polymorphisms and microarray-based genotyping_

Indel polymorphisms, also known as single feature polymorphisms (SFPs), are particularly amenable to microarray-based genotyping. These assays are done by labeling genomic DNA (target) and hybridizing to arrayed oligonucleotide probes that are complementary to indel loci. Each SFP is scored by the presence or absence of a hybridization signal with its corresponding oligonucleotide probe on the array. Both spotted oligonucleotides (Barrett et al., 2004) and Affymetrix-type arrays (Borevitz et al., 2003) have been used in these assays. The SFPs can be discovered through sequence alignments or by hybridization of genomic DNA with whole genome microarrays. The advantage of microarray platforms for genotyping is that they are highly parallel, and they are well suited for applications such as quantitative trait loci (QTL) analysis, where whole genome coverage with many markers is desirable.
polymorphism (Chen, Cho and McCouch, 2002). Thus, the use of stepwise SSR mutation models would be inappropriate for highly diverged populations. Homoplasy is also a problem in SSR markers because the hyper-variability leads to some shared allele sizes through parallelism, convergence and reversion (Doyle et al., 1998). Homoplasy from reversions can affect transposon-based markers or any markers with polymorphisms potentially derived from Class II DNA transposable elements. This class of TE has a cut and paste mechanism of transposition, so a TE may insert onto a locus and later excise.

In RAPDs, ISSRs and AFLPs, homoplasy can occur when two or more loci produce PCR fragments of similar molecular weight. Although it is desirable to have high numbers of bands to maximize the amount of information per lane, this must be balanced against the increasing risk of homoplasy as more loci are represented.

**SPECIAL CONSIDERATIONS FOR MARKER-ASSISTED SELECTION**

Quality markers for use in MAS should be reliable and easily shared among researchers. Co-dominant markers are preferred to avoid the need for progeny testing. Sometimes less desirable markers for MAS such as RAPDs, ISSRs and AFLPs are useful for finding markers linked to the desired allele. Once such a marker is found, it is possible to extract and sequence the corresponding band. This sequence can be used to develop co-dominant markers such as cleaved amplified polymorphic sequences (CAPS) (Konieczny and Ausubel, 1993) or to sequence characterized polymorphic regions (SCARs) (Paran and Michelmore, 1993). SCAR and CAPS markers are co-dominant and simplify the screening of large numbers of individuals.

When a genetic map exists, markers can be positioned on the map and other linked markers can be substituted. The additional markers are useful for higher resolution mapping to find markers more closely linked to the desired allele or ultimately for positional cloning of the underlying gene.

**Reproducibility of molecular marker data**

For orphan species, clearly there is a huge value to the anonymous primer approaches (AFLP, DArTs, ISSRs and RAPDs) that do not require sequence information or much up-front investment. However, the data can be difficult to score, and reproducibility requires a lot of technical skill. Technologies that depend on the presence or absence of PCR amplified bands are susceptible to changes in PCR conditions and the quality of sample DNA, and the data from separate experiments may differ. Further, in any method that depends on accurate measurement of molecular weight differences between bands (e.g. SSRs), the exact molecular weights assigned to each allele may be different in each analysis because of differences in labelling of PCR products, rounding of allele molecular weight estimates and binning of alleles. Without controls for each allele encountered, it is difficult or impossible to merge separate sets of data. Despite discrepancies in the exact data derived from molecular markers, the results and conclusions should be consistent within independent experiments. For reliability in making inferences across independent data-sets, SNP markers are preferred. SNP data-sets can be easily integrated based on sequence, and SNPs have properties (such as a low mutation rate) that are particularly valuable for evolutionary inference (Nielsen, 2000).
Choosing a Molecular Marker Technology

Clearly there is no single best choice of molecular marker for all situations. Factors influencing the decision may include the objectives of the study, availability of organism specific sequences, equipment and technical resources, and biological features of the species. Several important advantages/disadvantages for each type of molecular marker discussed are summarized in Table 2 (see review by Powell et al., 1996).

If available, microsatellite or SNP markers are often the best choice. The rate of adoption of SSR markers can be facilitated, and the costs reduced, by preparing “kits” of selected SSR markers for certain species to provide a reliable set of markers with good amplification, reasonable polymorphism and good genome coverage. This was done in the early days of the rice SSR effort and SSR kits were distributed at very low cost through Research Genetics (called Rice-Pairs; McCouch et al., 1997). Similarly, for SNPs, there is a need to develop useful sets of markers that are widely available and can be mass-produced (at reduced cost) for distribution to the international community. SNP kits would also have a clear benefit for databasing and analysing datasets obtained from multiple laboratories. In addition to kits of markers, there is a need to distribute sets of “control genotypes” as samples, particularly to address the problem surrounding the difficulties in integrating SSR datasets. When SNPs or SSRs are not available, it is sometimes possible to transfer molecular markers from closely related species (Gupta et al., 2003; La Rota et al., 2005; Zhang et al., 2005). When financial resources are restricted, RAPDs, AFLPs and ISSRs can provide large numbers of markers with a limited investment. AFLPs, SSRs and ISSRs can provide high throughput using an automated sequencer, while RAPDs and ISSRs can be run on agarose gels with minimal investment in equipment. The effectiveness of each method may vary by species and by application. Therefore, it is reasonable to try to use more than one method, particularly at the early stages of research.

Impact of the Rice Genome Sequence: A Case History

DNA sequence information greatly accelerates the development of molecular markers. This is evident in the history of rice microsatellite marker proliferation coinciding with the release of data from rice genome sequencing projects. Figure 1

TABLE 2

Key features of common molecular marker technologies

<table>
<thead>
<tr>
<th>Marker type</th>
<th>PCR-based</th>
<th>Uses restriction enzymes</th>
<th>Polymorphism</th>
<th>Abundance</th>
<th>Co-dominant</th>
<th>Automation</th>
<th>Loci per assay</th>
<th>Specialized equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFLP</td>
<td>no</td>
<td>yes</td>
<td>moderate</td>
<td>moderate</td>
<td>yes</td>
<td>no</td>
<td>1 to few</td>
<td>Radioactive isotope</td>
</tr>
<tr>
<td>RAPD</td>
<td>yes</td>
<td>no</td>
<td>moderate</td>
<td>moderate</td>
<td>no</td>
<td>yes</td>
<td>many</td>
<td>Agarose gels</td>
</tr>
<tr>
<td>AFLP</td>
<td>yes</td>
<td>no</td>
<td>moderate</td>
<td>moderate</td>
<td>no</td>
<td>yes</td>
<td>many</td>
<td>Polycrylamide gels/capillary</td>
</tr>
<tr>
<td>ISSR</td>
<td>yes</td>
<td>no</td>
<td>moderate</td>
<td>moderate</td>
<td>no</td>
<td>yes</td>
<td>many</td>
<td>Agarose/polycrylamide gels</td>
</tr>
<tr>
<td>DaRT</td>
<td>yes</td>
<td>yes</td>
<td>moderate</td>
<td>moderate</td>
<td>yes</td>
<td>yes</td>
<td>single</td>
<td>Agarose gels</td>
</tr>
<tr>
<td>CAPS</td>
<td>yes</td>
<td>yes</td>
<td>variable</td>
<td>moderate</td>
<td>yes</td>
<td>yes</td>
<td>single</td>
<td>Agarose gels</td>
</tr>
<tr>
<td>SCAR</td>
<td>yes</td>
<td>no</td>
<td>low</td>
<td>moderate</td>
<td>yes</td>
<td>yes</td>
<td>single</td>
<td>Agarose gels</td>
</tr>
<tr>
<td>SSR</td>
<td>yes</td>
<td>no</td>
<td>low</td>
<td>moderate</td>
<td>yes</td>
<td>yes</td>
<td>1 to about 20</td>
<td>Polycrylamide gels/capillary</td>
</tr>
<tr>
<td>TE-Anchor</td>
<td>yes</td>
<td>no</td>
<td>variable</td>
<td>variable</td>
<td>yes</td>
<td>yes</td>
<td>single</td>
<td>Agarose gels</td>
</tr>
<tr>
<td>SNP</td>
<td>yes</td>
<td>no</td>
<td>highest</td>
<td>highest</td>
<td>yes</td>
<td>yes</td>
<td>1 to thousands</td>
<td>Variable</td>
</tr>
</tbody>
</table>
tracks the publication of rice microsatellite markers derived from screening libraries of clones and from the various categories of sequences deposited in public databases. The earliest method of developing microsatellite markers in rice was by using microsatellite sequences as probes to isolate clones from genomic libraries (Zhao and Kochert, 1993; Wu and Tanksley, 1993; Panaud, Chen and McCouch, 1996; Akagi et al., 1996; Chen et al., 1997; Temnykh et al., 2000). In 1996, Akagi et al. used microsatellite repeats found in rice sequences from database searches to develop 35 new markers and in 2000, Temnykh et al. published 91 new microsatellite markers developed from expressed sequence tag (EST) sequences. Temnykh et al. (2001) developed 200 new markers, mostly from end sequences of rice bacterial artificial chromosomes (BACs). However, the most dramatic increase in microsatellite markers (2,240 new markers in 2002 and 25,000 in 2004) was made possible primarily through the use of whole genome shotgun sequences (McCouch et al., 2002; G. Wilson, personal communication).

Complete genome sequence provides an additional advantage in electronically determining the position of new markers on genetic and physical maps. However, full genomic sequence is not a requirement for microsatellite marker development, and there are a number of microsatellite markers that have been developed for a wide array of crop species (Table 3) without the benefit of full genomic sequence.

**MARKER-ASSISTED SELECTION STRATEGIES AND EXAMPLES**

MAS in a breeding context involves scoring indirectly for the presence or absence of a desired phenotype or phenotypic component based on the sequences or banding patterns of molecular markers located in or near the genes controlling the phenotype. The sequence polymorphism or banding pattern of the molecular marker is indicative of the presence or absence of a specific gene or chromosomal segment that is known to carry a desired allele.

DNA markers can increase screening efficiency in breeding programmes in a
number of ways. For example, they provide:

- the ability to screen in the juvenile stage for traits that are expressed late in the life of the organism (i.e. grain or fruit quality, male sterility, photoperiod sensitivity);
- the ability to screen for traits that are extremely difficult, expensive or time consuming to score phenotypically (i.e. quantitatively inherited or environmentally sensitive traits such as root morphology, resistance to quarantined pests or to specific races or biotypes of diseases or insects, tolerance to certain abiotic stresses such as drought, salt and mineral deficiencies or toxicities);
- the ability to distinguish the homozygous from the heterozygous condition of many loci in a single generation without the need for progeny testing (as molecular markers are co-dominant);
- the ability to perform simultaneous MAS for several characters at one time (or to combine MAS with phenotypic or biochemical evaluation).

This section provides examples of how molecular markers are being used in breeding and germplasm evaluation. While these examples are drawn mostly from rice, they illustrate applications of MAS techniques that are used in other species.

Before molecular markers can be used for selection purposes, their association with genes or traits of interest must be firmly established. While the number of economically important genetic loci that have been cloned or tagged via linkage to molecular markers is still limited in most species, work towards this end is accelerating rapidly. This is particularly true in rice, due to the availability of complete genome sequence information.

Nonetheless, a great deal of time and effort is required to identify the genetic loci and specific allelic variants that are responsible for the tremendous array of

### Table 3

Examples of SSR markers available across different plant species

<table>
<thead>
<tr>
<th>Common name</th>
<th>Species</th>
<th>Number of SSRs</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rice</td>
<td>Oryza sativa</td>
<td>2240</td>
<td>McCouch et al., 2002</td>
</tr>
<tr>
<td>Maize</td>
<td>Zea mays</td>
<td>1669</td>
<td>MapPairs (mp.invitrogen.com)</td>
</tr>
<tr>
<td>Soybean</td>
<td>Glycine max</td>
<td>597</td>
<td>MapPairs (mp.invitrogen.com)</td>
</tr>
<tr>
<td>Cassava</td>
<td>Manihot esculenta</td>
<td>318</td>
<td>MapPairs (mp.invitrogen.com)</td>
</tr>
<tr>
<td>Arabidopsis</td>
<td>Arabidopsis thaliana</td>
<td>290</td>
<td>MapPairs (mp.invitrogen.com)</td>
</tr>
<tr>
<td>Cotton</td>
<td>Gossypium spp.</td>
<td>217</td>
<td>MapPairs (mp.invitrogen.com)</td>
</tr>
<tr>
<td>Sugar cane</td>
<td>Saccharum spp.</td>
<td>200</td>
<td><a href="http://www.intl-pag.org/pag/9/abstracts/W30_04.html">www.intl-pag.org/pag/9/abstracts/W30_04.html</a></td>
</tr>
<tr>
<td>Wheat</td>
<td>Triticum aestivum</td>
<td>193</td>
<td>MapPairs (mp.invitrogen.com)</td>
</tr>
<tr>
<td>Grape</td>
<td>Vitis vinifera</td>
<td>152</td>
<td>no</td>
</tr>
<tr>
<td>Groundnut</td>
<td>Arachis hypogaea</td>
<td>110</td>
<td>Ferguson et al., 2004</td>
</tr>
<tr>
<td>Cucumber</td>
<td>Cucumis sativus</td>
<td>110</td>
<td>Fazio, Staub and Chung, 2002</td>
</tr>
<tr>
<td>Peach</td>
<td>Prunus persica</td>
<td>109</td>
<td>Aranzana et al., 2004</td>
</tr>
<tr>
<td>Kiwifruit</td>
<td>Actinidia spp.</td>
<td>105</td>
<td>Testolin et al., 2001</td>
</tr>
<tr>
<td>Barley</td>
<td>Hordeum vulgare</td>
<td>44</td>
<td>MapPairs (mp.invitrogen.com)</td>
</tr>
<tr>
<td>Potato</td>
<td>Solanum tuberosum</td>
<td>31</td>
<td>Ghislain et al., 2004</td>
</tr>
<tr>
<td>Pine trees</td>
<td>Pinus spp.</td>
<td>28</td>
<td>MapPairs (mp.invitrogen.com)</td>
</tr>
<tr>
<td>Banana</td>
<td>Musa spp.</td>
<td>28</td>
<td>MapPairs (mp.invitrogen.com)</td>
</tr>
<tr>
<td>Sweet potato</td>
<td>Ipomoea batatas</td>
<td>26</td>
<td>MapPairs (mp.invitrogen.com)</td>
</tr>
<tr>
<td>Sugar beet</td>
<td>Beta vulgaris</td>
<td>25</td>
<td><a href="http://www.intl-pag.org/pag/10/abstracts/PAGX_W306.html">www.intl-pag.org/pag/10/abstracts/PAGX_W306.html</a></td>
</tr>
<tr>
<td>Eggplant</td>
<td>Solanum melongena</td>
<td>23</td>
<td><a href="http://www.intl-pag.org/pag/11/abstracts/P3b_P181_XI.html">www.intl-pag.org/pag/11/abstracts/P3b_P181_XI.html</a></td>
</tr>
</tbody>
</table>

From: Thomson, Septiningsih and Sutrisno, 2003 (reprinted with permission of author)
Marker-assisted selection – Current status and future perspectives in crops, livestock, forestry and fish

characters that breeders are concerned about in population or variety improvement programmes. Given the complexity of quantitative traits, many different lines or crosses must be carefully analysed over different years and environments to unravel important components of gene interaction. In a breeding context, understanding the genetic basis of genotype by genotype interaction (G x G) and genotype by environment interaction (G x E) is critical as the basis for predicting how QTL are likely to behave. Information from a large number of studies addressing each of these points must then be assembled into a database that offers easy access to users and allows many different kinds of data to be integrated with a simple query.

The Gramene database represents a beginning in the quest to serve this user community. Gramene is a comparative genome database for grasses and currently offers a complete inventory of all published QTL that have been identified in rice (www.gramene.org/qtl/index.html), allowing users to find information about where along the chromosome a QTL is located, what phenotype is associated with the QTL, how it was measured, what germplasm was used, what molecular markers reside nearby, what the corresponding position is on a comparative map of another grass species and with what statistical significance the QTL was detected. The database also provides a link to the published article so that users can readily find more information on the subject. Similar inventories and databases are being assembled for other families of plants and are critical to the implementation of effective molecular breeding strategies.

Comparative genome methods take advantage of the fact that some species have more developed genetic systems than others. Examples of well-studied “model” organisms with available genomic sequence include species such as Arabidopsis and rice for plants, Populus (Taylor, 2002) and Eucalyptus (Poke et al., 2005) specifically for forestry, and Fugu (Aparicio et al., 2002) and zebrafish (Guryev et al., 2006) for fisheries. Relying heavily on the use of comparative maps and comparative sequence analysis, genome databases allow researchers to make predictions about the location and phenotypic consequences of homologous genes in related species. Thus, understanding how a gene or QTL behaves in one species can potentially shortcut the process of identifying a related gene or QTL in the genetic system of another species. This approach underscores the search for QTL associated with abiotic stress tolerance in cereals. A global effort to identify loci associated with drought tolerance has recently been initiated under the umbrella of the Generation Challenge Programme (www.generationcp.org).

Markers associated with tolerance for a variety of environmental stresses rank as important targets for molecular MAS in cereal breeding because these complex traits are often prohibitively difficult to screen using classical selection techniques. Efforts to identify QTL associated with tolerance to drought, salt and mineral deficiencies or toxicities (Champoux et al., 1995; Flowers et al., 2000; Nguyen et al., 2002; Kamoshita et al., 2002; Price et al., 2002; Gregorio, 2002) in a number of genetic backgrounds represent an important first step towards achieving this goal. Additional studies have specifically addressed the problems associated with G x G and G x E (Zheng et al., 2000; Li et al., 2003; Hittalmani et al., 2003).

In the area of biotic stress, several genes have been cloned and characterized for resistance to major diseases such as bacterial blight and blast (Song et al., 1995;
Yoshimura *et al.*, 1998; Wang *et al.*, 1999; Bryan *et al.*, 2000; Sun *et al.*, 2004) and many other genes for disease resistance have been tagged with linked markers. This opens the door for targeted approaches to MAS (Valent *et al.*, 2001). While the disease resistance literature is too vast to summarize here, it is important to note that advances in this area are having an impact on varietal improvement programmes (www.syix.com/rrb/98rpt/MarkerAssist.htm). Pyramiding of resistance genes into a single variety and the construction of multiline varieties, each with one or more R genes (resistance genes) that can be used in various combinations, are under way to develop more durable forms of disease and insect resistance (Yoshimura *et al.*, 1992; Yoshimura *et al.*, 1995; Hittalmani *et al.*, 1995; Blair and McCouch, 1997; Ndjiondjop *et al.*, 1999; Davierwala *et al.*, 2001; Su *et al.*, 2002; Conaway-Bormans *et al.*, 2003; Lorieux *et al.*, 2003; Hayashi *et al.*, 2004).

Marker-based selection is also helpful in attempts to transfer genes from exotic germplasm into cultivated lines. In rice, several workers have used RFLP and SSR markers to monitor introgression of brown planthopper resistance from *O. officinalis* (Kochert, Jena and Zhao, 1990), bacterial blight resistance from *O. longistaminata* (Ronald *et al.*, 1992), aluminum tolerance or yield and quality-related traits from *O. rufipogon* (Nguyen *et al.*, 2002; Thomson *et al.*, 2003; Septiningsih *et al.*, 2003a, b) or from other wild species such as *O. glumaepatula* (Brondani *et al.*, 2002) or *O. glaberrima* (Jones *et al.*, 1997; Lorieux *et al.*, 2003) into cultivated *O. sativa* backgrounds. Marker-assisted introgression strategies have also been used in a number of livestock breeding programmes but, because of longer generation intervals and lower reproductive rates, this is generally feasible for genes of large effect (Dekkers, 2004; Chapter 10). Identifying the recombinants with the least amount of donor DNA flanking the genes of interest is enhanced by the use of molecular markers (Monna *et al.*, 2002; Takeuchi *et al.*, 2003; Blair, Panaud and McCouch, 2003). In these examples, MAS offers a powerful strategy for making efficient use of the wealth of useful genetic variation that exists in the early landraces and wild species of cultivated food crops (Tanksley and McCouch, 1997).

As this kind of information accumulates, MAS permits rapid identification of individuals that may contain only one genetic component of a complex trait. Once identified, such an individual can be crossed with another individual in a breeding programme so that multiple, complementary genes are combined to optimize a quantitatively inherited trait. Individuals containing only one gene of interest often defy accurate phenotypic identification where polygenic traits are concerned because various types of epistasis, or gene interaction, may be required to generate the phenotype of interest (Yamamoto *et al.*, 2000; Zheng *et al.*, 2000).

Linkage disequilibrium (LD) mapping is another marker-assisted approach that provides important information that is immediately relevant to breeding programmes (Remington, Ungerer and Purugganan, 2001; Flint-Garcia, Thornsberry and Buckler, 2003). Using collections of distantly related germplasm accessions rather than populations derived from bi-parental crosses allows researchers to explore the relationship between phenotype and genotype in materials that have been amply tested over years and environments, often as part of an applied breeding programme. This provides
critical information about how specific combinations of genes and alleles interact in relevant varietal backgrounds and allows breeders to compare the phenotypic effect of genes or chromosomal segments that have been inherited from a common ancestor and selected in multiple-cross combinations.

In addition to the use of MAS in traditional crossing and selection programmes, breeders also have opportunities to adjust particular traits or phenotypes via the introduction of genes using a transgenic approach (Ye et al., 2000; James, 2003; Nuffield Council on Bioethics, 2004). Once introduced into the gene pool, a transgene can be tracked with the aid of molecular markers (designed to tag the transgene sequence itself) through subsequent crosses, just as would be done for any other gene of interest in a breeding programme.

Another use of molecular markers in variety improvement involves marker-assisted germplasm evaluation (Xu, Ishii and McCouch, 2003). Population structure analysis offers insight about how diversity is partitioned within a species and can help define clusters, or subpopulations, of germplasm that are likely to contain high frequencies of particular alleles (Garris, McCouch and Kresovich, 2003). This type of analysis can also guide allele mining efforts aimed at identifying valuable accessions in a germplasm collection for use as parents in a breeding programme. Such approaches have the potential to make parental selection more efficient, to expand the gene pool of modern cultivars and ultimately to speed up the development of productive new varieties. As information is generated about which genes and alleles are associated with phenotypic characters of agronomic importance, and as the complex interactions among genes are enumerated in the context of specific gene pools and the environments to which they are adapted, breeders are increasingly empowered to make predictions about how to combine diverse alleles productively.

To exploit molecular breeding strategies fully, information resources must be developed so that the overwhelming amount of information about genes, alleles and natural genetic variation can be funneled into a useful tool for breeding applications. This will involve a very different approach to information resources than currently employed by the large genome databases, which are oriented towards genomics researchers and molecular biologists rather than the breeding community. Nonetheless, a few examples offer beacons of inspiration in this area, including the emerging International Rice Information System (IRIS) database (Bruskiewich et al., 2003; www.icis.cgiar.org/), the GeneFlow database (www.geneflow.com), the marker-assisted selection wheat (MASwheat) database (http://maswheat.ucdavis.edu/) and software such as Real Time QTL (http://zamir.sgn.cornell.edu/Qtl/Html/home.htm).

In conclusion, genomics research is generating information about the location and phenotypic consequences of specific genes and alleles in a wide range of species. This information can be translated into tools for breeders. Molecular marker technology can benefit breeding objectives by increasing the efficiency and reliability of selection and by providing essential insights into how genes behave in different environments and in different genetic backgrounds. Once genes and QTL are identified, markers allow interesting alleles to be traced through the pedigrees of breeding programmes or mined out of germplasm collections to serve as the basis for future varietal improvement. Using markers in combination with both QTL and association approaches, the
effect of specific alleles on a phenotype can be monitored with relative precision. As all this information is assembled and organized in databases that provide easy querying capabilities for plant breeders, breeders will take advantage of the power that comes from the application of genome based strategies for plant improvement.

REFERENCES


avrPto induces disease resistance in tomato cultivars containing the Pto resistance gene. *J. Bacteriol.* 
174: 1604–1611.

Rungus, D., Berube, Y., Zhang, J., Ralph, S., Ritland, C.E., Ellis, B.E., Douglas, C., Bohlmann, J. 
& Ritland, K. 2004. Robust simple sequence repeat markers for spruce (Picea spp.) from expressed 

McCouch, S.R. 2003a. Identification of quantitative trait loci for yield and yield components in 
an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative 

quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza 


Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel 


Sun, X., Cao, Y., Yang, Z., Xu, C., Li, X., Wang, S. & Zhang, Q. 2004. Xa26, a gene conferring 


Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, 
length variation, transposon associations, and genetic marker potential. *Genome Res.* 11: 
1441–1452.

(*Actinidia* spp.) linkage map based on microsatellites and integrated with AFLP markers. *Theor. 

Thomson M., Septiningsih E. and Sutrisno. 2003. Workshop on IPR in relation to Plant Breeding, 
Cipanas, Indonesia, 10–18 November.

C.P. & McCouch, S.R. 2003. Mapping quantitative trait loci for yield, yield components and


Yoshimura, S., Yamanouchi, U., Katayose, Y., Toki, S., Wang, Z.X., Kono, I., Kurata, N., Yano, M.,
Iwata, N. & Sasaki, T. 1998. Expression of Xa1, a bacterial blight-resistance in rice, is induced by bac-

Ecol. 11: 1–16.

Zhao, X. & Kochert, G. 1993. Phylogenetic distribution and genetic mapping of a (GGC)n


Quantitative trait loci for root-penetration ability and root thickness in rice: comparison of genetic

CHAPTER 4

Marker-assisted selection in wheat: evolution, not revolution

Robert Koebner and Richard Summers
SUMMARY
This chapter reviews the uptake of marker-assisted selection (MAS) in wheat in a European context. Although less intense than the scale of its application in maize, reflecting the fact that maize varieties are predominantly F₁ hybrids, the use of MAS in wheat has grown over the last few years. This growth has been encouraged by an increase in the number of amenable target traits, but more significantly by a combination of technological improvements, particularly in the areas of DNA acquisition, laboratory management systems and integration into the breeding cycle, which together have served to reduce the per unit cost of each data point. Microsatellites (simple sequence repeats [SSRs]) are, and will likely remain for some time, the marker of choice because of their flexibility and the knowledge base associated with them. Some current examples are provided of the use of MAS in a major United Kingdom commercial breeding programme.
INTRODUCTION
Wheat is a very important world staple crop. The 2005 United States Department of Agriculture (USDA) estimates for the global production of wheat (both bread and durum) and maize are, respectively, 627 million tonnes and 708 million tonnes. In Europe, bread wheat is without doubt the most important broad-acre crop, with a production in the extended European Union of 25 states of 115 million tonnes (maize 48 million tonnes). The largest production and highest productivity of bread wheat are achieved in northwest Europe. Historically, wheat has been bred largely by government-sponsored national and regional programmes, but the introduction of plant variety rights into Europe in the 1960s encouraged participation by the private sector. Currently, wheat breeding in northwest Europe is almost exclusively carried out by private companies, with some research underpinning by the public sector. Breeders continue to be successful in the production of high-yielding, disease-resistant, high-quality varieties and, in the United Kingdom at least, genetic advances for yield have been running at between 0.5 to 1 percent per annum for many years.

Wheat is a naturally inbreeding species, and although a level of heterosis can be demonstrated, difficulties in enforcing cross-pollination in a reliable and cost-effective way have hindered the development of any significant contribution of F₁ hybrids to the variety pool. Most varietal development programmes are therefore based on versions of the long-established pedigree breeding system, where large F₂ populations are generated and conventional phenotypic selection is carried out in early generations for highly heritable, qualitative traits (such as disease resistance) and in later ones for quantitative traits (primarily yield and quality). Thus, most varieties are bred and grown as inbred, pure breeding lines. As a result, the unit value of seed and economic margins for breeders are low. By contrast, maize is a naturally out-crossing species that shows highly significant levels of heterosis. This has resulted in the majority of maize breeding being geared to the production of F₁ hybrids. In industrialized countries, maize hybrid breeding has for some time been dominated by a small number of large private sector companies that are able to sustain profitability through their control over the genotype of their varieties. No revenue is lost as a result of the use of farm-saved seed, and the inbred components of a successful hybrid are not available to competitors to use as parental material for their own varietal improvement programmes. This has far-reaching implications on the feasibility of MAS in maize, and largely explains the lead that maize enjoys over wheat in the deployment of MAS technology.

The continuing development of molecular marker technology over the last decade has been a happy by-product of “big biology” genomics research. As recently as 1996, the definition of 5 000 SSR loci in the human genome merited a major publication in Nature (Dib et al., 1996), but the number of known human single nucleotide polymorphisms (SNPs) now runs into millions. Thus, although marker availability, potentially at least, is no longer limiting in crops, and the clear potential benefits of marker deployment to plant breeding are undisputed, only relatively recently has it begun to make more than a marginal impact on breeding methodology. Even in maize, where the level of DNA marker polymorphism is high, large-scale deployment of MAS did not gather any significant momentum until more than 15 years after the publication of the first restriction frag-
ment length polymorphism (RFLP)-based maize genetic map. In the less genetically variable cereals, prominently wheat, the level of polymorphism is not now in practice likely to represent the major constraint to MAS uptake, although in the past it was argued that this was the case. What has changed in recent times is that current marker technology, and systems of DNA acquisition, laboratory management and integration into the breeding cycle, have all developed to the extent where the benefits of MAS can be increasingly realized in actual practice. As many of these improvements are incremental rather than sudden, we argue that the trends in MAS application in wheat are characteristically evolutionary rather than revolutionary.

**TARGET TRAITS FOR MAS IN NORTHWEST EUROPEAN WINTER WHEAT BREEDING**

The use of MAS to date has a history of about 20 years, and until recently involved the exploitation of just two non-DNA-based assays. The first, which has been retained with only slight modifications since its inception, exploits a correlation between bread-making quality and allelic status at the Glu-1 (endosperm storage protein subunit) loci. It uses electrophoretic profiles obtained by the straightforward, robust and cheap procedure sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) from crude seed protein extracts, which have been shown to be partially predictive of end-use quality. The second is predictive for the presence of the gene Pch1, which confers a high level of resistance to eyespot, a stem base disease that is difficult to screen using conventional pathology methods. Both these targets have in the meanwhile become assayable by polymerase chain reaction (PCR)-based assays, although SDS-PAGE remains in routine use thanks to its flexibility and cost effectiveness. In recent years, the number of loci for which DNA-based assays have been generated has increased dramatically, the majority using PCR as a technology platform. Over 50 of these are described (specifically in a United States of America context) at http://maswheat.ucdavis.edu/, which reports the output of an ongoing United States Department of Agriculture (USDA)-funded programme. The focus is heavily on disease and pest resistance, reflecting the generally simple inheritance of genes conferring these traits.

Some of the above traits are of sufficient relevance to the United Kingdom context that identical or equivalent assays have been incorporated in a number of breeding programmes, where they are used as guides to parental selection and/or in early generation selection. Prominent among these are markers for the genes Rht-1 (responsible for the “Green Revolution” semi-dwarfism), Pinb (grain texture), Pch1, Lr37/Yr17 (a gene complex conferring resistance to two of the most important leaf fungal pathogens) and the wheat/rye translocation 1B/1R (which is associated with high levels of yield). Emerging MAS targets are necessarily programme-dependent, but the broad focus is on quantitative trait locus (QTL) targets that could have a major impact on breeding efficiency. In the United Kingdom, as elsewhere worldwide, current focus is on resistance to the diseases Fusarium head blight (FHB), Septoria tritici blotch (STB) and barley yellow dwarf virus (BYDV), and on durable resistance to yellow rust. Other current targets, more specific to the United Kingdom and northwest European context, but in routine use, are resistance to the insect pest orange blossom midge (OBM), and soil-borne mosaic virus (SBMV).
**FHB**

The importance of FHB is less in its effect on yield reduction, but rather on the potentially damaging reduction in grain quality associated with infected grain, which can be heavily contaminated by the fungal trichotheclin toxins. An important source of FHB resistance originates from the Chinese variety Sumai 3, and a major component of this resistance (up to 50 percent) has been associated with a single QTL (Waldron et al., 1999; Anderson et al., 2001; Buerstmayr et al., 2002). While this QTL is largely effective in preventing the spread of the pathogen following infection, a further QTL that gives a significant degree of protection against initial infection has been mapped to a different chromosome (Buerstmayr et al., 2003). Selection for FHB resistance by conventional means is complicated both by the quantitative nature of the Sumai 3 resistance and by difficulties in ensuring even and reliable artificial infections in breeding nurseries. However, SSR-based MAS protocols have been developed for both QTL (see http://maswheat.ucdavis.edu/ and Buerstmayr et al., 2003), and the urgency of breeding for resistance has ensured that increasing use is being made of such assays. Both these QTL in concert do not explain all the genetic resistance of Sumai 3 to FHB, but the remainder appears to be determined by QTL of minor effects and/or pleiotropic effects associated with an ear morphology, which is inconsistent with a northwest European winter wheat ideotype.

**STB**

STB of wheat is caused by the fungus *Mycosphaerella graminicola* (syn. *Septoria tritici*), and in recent years has become the major leaf disease of wheat in many regions of the world. In past years, good levels of control were achieved by the application of strobilurin fungicides, but their heavy use has led to the emergence of pathogen strains that cannot be so easily controlled by chemical means. A number both of major genes giving near-complete resistance to specific races of the pathogen and of quantitative race non-specific resistances with polygenic inheritance have been defined, and one of the former, *Stb6*, which maps close to the SSR locus *Xgwm369* on chromosome 3A (Chartrain, Brading and Brown, 2004), is common in many gene pools. This ensures that the gene has been retained in elite materials, and its known map position has made it relatively straightforward to use a marker assay to track its presence in breeding populations.

**BYDV**

Significant grain yield losses are attributable to natural infections of BYDV, and no major source of resistance has been identified to date in wheat. Control is achieved in the absence of genetic resistance by insecticidal spray, which is associated with both an economic and an environmental cost. However, a potent resistance is present in the related species *Thinopyrum intermedium*. It is possible to generate sexual hybrids between wheat and this grass, but the F1 plants are self-sterile and either have to be rescued by chromosome doubling or back-crossed to wheat. By this route, a distal segment of the grass chromosome that carries the BYDV resistance gene **Bdv2** has been introduced into wheat. As this introgression comprises a significant length of non-wheat chromosome, it has been relatively straightforward to generate markers suitable for MAS use (Ayala et al., 2001a; Zhang et al., 2004). A MAS approach for screening is attractive because artificial inoculation involves the propagation of virus-bearing aphids, while natural infections...
are unreliable. Interestingly, unlike the experience with many alien introgression segments, no obvious negative effects of its presence on agronomic performance have yet been detected either in International Maize and Wheat Improvement Center (CIMMYT) trials (Ayala et al., 2001b) or at RAGT Seeds (Cambridge, UK).

**Durable resistance to yellow rust**

Yellow rust is historically the most damaging of the leaf fungal pathogens in temperate Europe. Control has been achieved in the past largely by a combination of fungicide application and of combinations of major seedling resistance genes, of which a significant number have been described in the literature. However, like most race-specific resistances, most of these major genes have lost their effectiveness, and this has led to a renewed effort in the definition of partial or adult plant resistances to this disease. The French variety Cappelle-Desprez dominated the wheat crop across France and the United Kingdom during the 1960s and 1970s, and maintained its level of adult resistance to yellow rust over the whole of this period. A major part of the genetic basis for this durable resistance was located to a translocated wheat chromosome (Law and Worland, 1997), and this has been confirmed by a rigorous QTL analysis (Mallard et al., 2005), which has provided a number of informative SSR markers for this effect. Other independent sources of adult resistance have been identified in French and Eastern European germplasm at RAGT Seeds, and the major QTL responsible have been defined and marked.

**OBM**

OBM larvae feed on developing grain and heavy infestations result in a significant reduction in grain quality and some loss in yield. As for many sporadic pests, phenotypic screening is unreliable and an indirect means of selection would be valuable. The gene Sm1 confers resistance to OBM (*Sitodiplosis mosellana*) by the expression of an antibiotic that kills or slows the development of larvae. Thomas et al. (2005) defined the map position of Sm1 and proposed a close linkage with an SSR locus Xbarc35. This linkage remains to be validated in United Kingdom breeding populations, as it remains unclear whether the antibiotic effect shown by a few United Kingdom wheat varieties is conferred by Sm1.

**SBMV**

SBMV is one of two known viral pathogens transmitted by the soil fungus *Polymyxa graminis* (another one being yellow mosaic virus [YMV]), and can be an important agent of yield loss in some areas. Chemical control is not feasible, and once soil is infected by the virus-bearing host, the only solutions possible are to abandon wheat culture or to use resistant varieties. Phenotyping is particularly difficult as plant infection is environmentally sensitive, and the detection of infection is laborious and prone to error. A proprietary assay for resistance to SBMV originating from European germplasm has been in routine use at RAGT Seeds since 2000 with a very high level of marker/phenotype association. More recently, a bulk segregant analysis along with a QTL approach has allowed the definition of a resistance locus to YMV from Chinese germplasm, and a number of linked SSR markers have been identified (Liu et al., 2005).

**HIGH-THROUGHPUT INFRASTRUCTURES**

Technical considerations of DNA acquisition, laboratory information management
system (LIMS), laboratory automation and data capture and analysis are generic for any MAS set-up, and these are well covered elsewhere in this volume. The limitations affecting MAS deployment in wheat flow from the restricted revenue generated by breeding a self-pollinated, homozygous, non-hybrid product. As a result, the volume of capital investment affordable in maize is not available to a wheat MAS programme. Financial constraints also affect the development of marker platforms. It is well known that the predictive ability of a linked marker will be disrupted by recombination, and therefore that “perfect” markers are more desirable than linked ones. However, the development of genome-wide gene-based markers, pre-eminently SNPs, which are particularly suited to high-throughput genotyping on automated platforms, is still some way off. At present, an insufficient number of such assays has been established (grain hardness, semi-dwarfness and grain texture) to consider adjusting the present major genotyping methodology, which is founded on SSRs. Doubts have been raised that SNP frequency in exon sequence will be high enough to generate informative assays for many critical genes, but early experience suggests that sequence polymorphism is more than adequate in introns and other untranslated regions of wheat genes. At present, the consensus is that there is plenty of mileage left in SSR technology, and wheat maps continue to be refined by the addition of new SSR loci.

CONCLUSION

In 1999, Young set out his “cautiously optimistic vision” for MAS. Seven years on, the situation continues to crystallize. The technology itself is no longer limiting. With respect to marker availability, SSRs remain useful and SSR-based genetic maps are becoming increasingly densely populated, while SNPs may eventually represent a source of plentiful perfect markers for genes of defined function. The “big biology” spawned by the genomics revolution has brought miniaturization and automation to biological assays so that levels of throughput relevant to the wheat breeding process are becoming attainable. The issue that remains unresolved is the affordability of large-scale MAS. As wheat is a broad-acre commodity product, its value is low, and this impedes the ability of the industry to invest in MAS infrastructure to the extent that is possible for crops such as maize where the generation of F1 hybrid seed is a viable proposition. However, as economies of scale and improvements in technology continue to drive down assay price, the penetration of MAS into commercial wheat breeding will surely grow. This growth should progressively allow a widening in the range of possible MAS targets, in particular extending to critical ones such as QTL for yield and its components (mean kernel size, kernel number per ear and number of fertile tillers per unit area). These are already widely exploited in maize breeding and their definition and validation in wheat represent a significant research theme in both the public and private sectors. In the meantime, much MAS use will be directed towards specific purposes such as accelerated selection of a few traits that are difficult to manage by conventional phenotyping, for the maintenance of recessive alleles in backcrossing programmes, for the pyramiding of disease resistance genes and for guiding the choice of parents to be used in crossing programmes.
REFERENCES


CHAPTER 5

Marker-assisted selection for improving quantitative traits of forage crops

Oene Dolstra, Christel Denneboom, Ab L.F. de Vos and E.N. van Loo
Summary

This chapter provides an example of using marker-assisted selection (MAS) for breeding perennial ryegrass (*Lolium perenne*), a pasture species. A mapping study had shown the presence of quantitative trait loci (QTL) for seven component traits of nitrogen use efficiency (NUE). The NUE-related QTL clustered in five chromosomal regions. These QTL were validated through divergent marker selection in an F2 population. The criterion used for plant selection was a summation index based on the number of positive QTL alleles. The evaluation studies showed a strong indirect response of marker selection on NUE. Marker selection using a summation index such as applied here proved to be very effective for difficult and complex quantitative traits such as NUE. The strategy is easily applicable in outbreeding crops to raise the frequency of several desirable alleles simultaneously.
INTRODUCTION
Most agronomical characteristics of forage crops have a quantitative, polygenic and mostly complex nature. For these reasons, genetic improvement of such traits is laborious and time consuming. Improving nitrogen use efficiency (NUE) in perennial ryegrass (*Lolium perenne*, 2n = 14), the major grass species in northern Europe, is in this respect a good example. The high input of nitrogen needed to attain high forage yields for animal husbandry has caused severe water pollution (van Loo et al., 2003), and therefore lowering nitrogen inputs through improving nitrogen use by breeding is of utmost importance.

Selection for NUE, however, is not easily implemented in conventional grass breeding based on field evaluations. Adequate testing requires separate and long-term trials with good control of the N stress, and such experiments tend to be rather inaccurate. To circumvent the disadvantages of field testing, a hydroponics system was used in this study in which the crop situation is simulated with growth-dependent N application (van Loo et al., 1992), the aim being to grow plants having an equal suboptimal N content. The set-up has a capacity to test about 1 600 plants in parallel and enables all plants to experience more or less the same N strain. Criteria used to measure NUE are several plant growth characteristics, such as tillering, and shoot and root growth. Each test usually requires four to five cuts. The trait is vigour-related and complex, and is extremely important in relation to regrowth after cutting. Together, all these aspects make NUE a very attractive trait for MAS.

ANALYSIS OF GENETIC VARIATION
The genetic variation for NUE present in an F₁ plant originating from a cross between two contrasting genotypes for NUE was first analysed by crossing the F₁ with a doubled haploid. The resulting test cross progeny was then used to produce a molecular marker map and analyse the variation. This approach was chosen to avoid inbreeding effects and to be able to use dominant molecular markers. The performance of the mapping population for NUE-related traits was studied on hydroponics with the system set at a moderately low nitrogen deficiency (3.6 percent N of leaf dry weight). The outcome of the mapping study was a genetic map with seven linkage groups.

Putative genes (quantitative trait loci [QTL]) for the components of NUE were found on four linkage groups. The location of the selection markers for QTL is depicted in Figure 1. The map shows five genomic sites with 1-5 QTL. In total, 13 QTL for seven NUE related traits were found. Three sites contain more than one QTL.

The findings of the current study are typical for genetic analyses of quantitative traits in forage crops and also indicative of the problems associated with exploitation of QTL information through marker-assisted breeding. These included uncertainties with respect to effect and location of QTL, the fairly large number of QTL often found in genetic analyses, the cosegregation of QTL and the weighing of the different component traits of NUE and NUE-QTL. Below is a description of how these breeding problems were solved or circumvented in a divergent marker selection study to validate the QTL found in the mapping study.

DIVERGENT MARKER SELECTION
The plant materials used in the validation study were an F₂ generation obtained by selfing of the heterozygous F₁ genotype
used to generate the mapping population mentioned above (van Loo et al., 2003). In total, about 200 genotypes were genotyped for five amplified fragment length polymorphism (AFLP) selection markers using the fluorescent AFLP technique developed by Applied BioSystems (Figure 1). The markers were co-dominantly scored using the heights of the fluorescence peaks relative to those of homozygous fragments as a criterion.

The genotyping data were used subsequently as a basis for a divergent mass selection programme. The selection strategy is outlined in Figure 2. The selection criterion was a genotype-specific selection index, being the summation of all positive QTL alleles (or chromosome segments) over the five QTL sites considered (Figures 1 and 2).

**APPLICATION OF MARKER SELECTION**

The AFLP technique is usually not the marker technology of choice for selection purposes because of its dominant nature and high costs per selection marker. However, co-dominant scoring of the five selection markers was quite adequate. The trimodal frequency distributions allowed proper classification of plants, although some misclassification cannot be fully excluded. The advantages of co-dominant AFLP scoring from a selection point of view are so large that a small number of genotyping errors are acceptable.

The decision to use a summation index as the criterion for selection was made primarily because of the difficulty of weighting the individual NUE related traits and the co-localization of QTL. The designation of the positive QTL alleles (chromosome
fragments) turned out to be straightforward. Figure 3 shows the F2 frequency distribution for the number of “plus alleles”. The population mean is somewhat below the expected number of five owing to the fact that the AFLP marker on LG1 showed a skewed segregation. This is likely due to gametophytic selection in favour of the negative QTL allele, perhaps due to linkage with an incompatibility locus.

The intensities of selection were set at about 25 percent, representing about 50 genotypes per selection (Figure 3). The selection pressure was kept fairly low because of the need to have sufficient seeds for measuring selection responses. In this way, the influence of genetic drift accompanying marker selection was minimized. The cut-off point for the top selection was six positive alleles and three for the opposite selection (Figure 3). The frequency of the plus alleles was on average 0.66 and 0.27, respectively. Selection showed a positive response for all NUE loci. However, the between-selection difference in allele frequency of the loci ranged from 0.18 to 0.77, showing that index selection did not affect all NUE loci to the same degree. The differences were probably mainly due to chance.

**INDIRECT RESPONSE TO MARKER SELECTION**

The selections were then multiplied using a polycross scheme (after vegetative propagation) to obtain sufficient seeds for evaluation on hydroponics and under various field conditions. The marker selections were evaluated for NUE in a replicated trial with two cuts on hydroponics at two N levels, being 2.5 and 5 percent N in leaves (van Loo et al., 2003). The same set of plant characteristics as in the original mapping studies was monitored after each cut. Leaf area expansion rate, leaf length and width, as well as tiller number, were determined one week after cutting. The determination of shoot and root dry weight followed three weeks later. The indirect responses to marker selection are summarized in Figure 4. At low N supply, the NUEplus
selection showed a remarkable 40 percent higher tillering rate and dry matter production than the NUEmin selection. The 40 percent higher tillering rate is associated with a 40 percent higher leaf area increase after defoliation (data not shown). Relative root growth (expressed as the ratio of root to total growth) and leaf length were hardly
changed through marker selection. At high N supply, the performances of NUEplus and NUEmin were fairly similar (Figure 4). The selections also showed striking differences in field trials in Germany, England and the Netherlands. At suboptimal N, the NUEplus selection significantly outperformed its counterpart in yields of dry matter and water soluble carbohydrates, while total N uptake was slightly lower.

CONCLUSIONS
Divergent mass selection has shown that marker selection using a summation index can be very effective for difficult and complex quantitative traits such as NUE. A collateral advantage of such an approach is that it offers a true validation of the putative genes (QTL) for the traits of interest. The associated response to marker selection distinctively indicates the presence of true genes affecting NUE, particularly in the vicinity of markers, which were strongly affected by the selection imposed. The results also indicate that recurrent mass selection to increase the number of positive alleles is worthwhile. The strategy is easily applicable in outbreeding crops.

ACKNOWLEDGEMENTS
The research work was carried out within the framework of the EU-FAIR project NIMGRASS (CT98-4063).

REFERENCES

Targeted introgression of cotton fibre quality quantitative trait loci using molecular markers

Jean-Marc Lacape, Trung-Bieu Nguyen, Bernad Hau and Marc Giband
SUMMARY
Within the framework of a cotton breeding programme, molecular markers are used to improve the efficiency of the introgression of fibre quality traits of *Gossypium barbadense* into *G. hirsutum*. A saturated genetic map was developed based on genotyping data obtained from the BC$_1$ (75 plants) and BC$_2$ (200 plants) generations. Phenotypic measurements conducted over three generations (BC$_1$, BC$_2$ and BC$_2S_1$) allowed 80 quantitative trait loci (QTL) to be detected for fibre length, uniformity, strength, elongation, fineness and colour. Positive QTL, i.e. those for which favourable alleles came from the *G. barbadense* parent, were harboured by 19 QTL-rich regions on 15 “carrier” chromosomes. In subsequent generations (BC$_3$ and BC$_4$), markers framing the QTL-rich regions were used to select about 10 percent of over 400 plants analysed in each generation. Although BC plants selected through the marker-assisted selection (MAS) process show promising fibre quality, only their full field evaluation will allow validation of the procedure.
INTRODUCTION
Among the four species of *Gossypium* that produce seeds with spinnable fibres called cotton, *Gossypium hirsutum* dominates the world’s cotton fibre production, accounting for approximately 90 percent of total world production. The second most cultivated species, *G. barbadense*, includes superior extra long, strong and fine cottons. However, compared with *G. hirsutum*, the marketing advantage of “high quality” *G. barbadense* cottons is offset by their lower productivity and a narrower adaptability to harsh environments. Breeding approaches within these two species have essentially relied on hybridization and selection methods (subsequent to simple or complex crosses, a pedigree system, sometimes combined with recurrent selection, is applied). Although *G. hirsutum* and *G. barbadense* display complementary characteristics, attempts to utilize deliberate interspecific *G. hirsutum*/*G. barbadense* recombinations through conventional breeding have had limited impact on cultivar development.

In the past 10–15 years, DNA markers for analyses of QTL and MAS have received considerable attention by plant and animal breeders (Dekkers and Hospital, 2002). However, following an initial keen interest and promises for molecular-based breeding approaches, the successful application of this technology has been shown to depend on the reliability and accuracy of the QTL analyses, which in turn are strongly affected by both population size and environmental factors (Schön et al., 2004). Examples of applied MAS in breeding programmes are still scarce, particularly when complex traits (yield components, product quality) are under consideration.

In the case of cotton, it is only recently that the results of efforts to gain a better understanding of the genome and the molecular basis of fibre quality have been published. Most of the earlier efforts in cotton molecular breeding concentrated on interspecific hybridization, due to the fact that, intraspecifically, the major species *G. hirsutum* displayed a very low level of molecular variability (Brubaker and Wendel, 2001). Based on studies of interspecific *G. hirsutum* x *G. barbadense* populations, published reports relate (i) to the construction of high-resolution genetic maps (Lacape et al., 2003; Rong et al., 2004); and (ii) to the identification of fibre quality-related QTL (Jiang et al., 1998; Kohel et al., 2001; Paterson et al., 2003; Lacape et al., 2005). In parallel, data have accumulated describing the cotton fibre transcriptome (reviewed by Wilkins and Arpat, 2005). These studies confirmed that key fibre quality properties, such as length, fineness and strength, are controlled quantitatively, thus complicating conventional breeding for fibre improvement.

Within the framework of a marker-assisted backcross introduction scheme aimed at transferring fibre quality traits from a low-productivity line of *G. barbadense* (donor) into a productive line of *G. hirsutum* (recipient), a saturated genetic map of tetraploid cotton was first developed (Lacape et al., 2003). This chapter describes how molecular markers were used in the early BC₁ and BC₂ generations to identify QTL-rich regions involved in determining fibre quality, as recently reported by Lacape et al. (2005), and how MAS was actually implemented in the later BC₃ and BC₄ generations.

METHODOLOGY
The major milestones (Figure 1) in the marker-assisted backcross selection process included the construction of two genetic maps from the BC₁ and BC₂ populations, the detection of fibre quality QTL from
three phenotyping data sets (BC₁, BC₂ and BC₂S₁) and the actual marker-based selection in the BC₃ and BC₄ generations, followed by the analysis of marker-trait associations in the BC₃ and BC₄ generations.

**Plant material**

The initial interspecific cross involved the *G. hirsutum* variety Guazuncho 2 and the *G. barbadense* variety VH8-4602. Guazuncho 2 is a modern pure line *G. hirsutum* variety created in Argentina and was chosen as a recipient in the backcross generations for its good overall agronomic performance. VH8-4602, a *G. barbadense* variety of the Sea Island type, was the donor parent for superior fibre quality, in particular for length (+9 to +12 mm as compared with Guazuncho 2), strength (+12 to +16 g/tex) and fineness (-30 to -50 millitex); conversely its fibre colour indices (reflectance and yellowness) are of lower value.¹

The plant material used in the multigeneration QTL analyses included three populations: BC₁, BC₂ and BC₂S₁ (Lacape et al., 2005). The first backcross generation (BC₁), consisted of 75 plants grown in a greenhouse in Montpellier (France) during the summer of 1999; these served as female parents for the second backcross to Guazuncho 2. Two hundred individual field-grown BC₂ plants that had shown a satisfactory production of BC₃ seeds and originating from 53 different BC₁ plants were used in 2000. Open pollinated seeds harvested from BC₂ plants were grown as 200 BC₂S₁ progenies in 2001 under field conditions in Brazil. Each BC₂S₁ line was

¹ 1 tex = 1 gram/kilometre
planted in two replications, each plot (one row) measuring 5 m. The next BC3 and BC4 generations were grown under field (411 BC3 in 2002) or greenhouse (450 BC4 in 2003) conditions in Montpellier. Every plant in each BC1-4 generation was used for DNA extraction from young fresh leaves using different methods described elsewhere (Lacape et al., 2003; Nguyen et al., 2004). In each BC3 and BC4 generation, an early genotyping was conducted (before flowering of BC3 plants and at the seedling stage for BC4 plants), to reduce the number of plants to be manipulated and raised to flowering for selfing and backcrossing.

From each generation (75 BC1, 200 BC2, 400 BC2S1, 43 selected BC3 and 37 selected BC4), the cotton seed harvest was ginned (separation of the fibre from the seed) on a laboratory roller gin and the fibre was sampled for analyses at the Fibre Technology Laboratory of the French Agricultural Research Centre for International Development (CIRAD).

**Fibre analyses**

All fibre quality measurements (11 traits) were conducted at CIRAD, Montpellier, on a high volume instrument line (Zellweger Uster 900, Uster Technologies, Switzerland). These included length, uniformity, strength, elongation and colour. A FMT3 maturometer (Shirley Dev Ltd., UK) was used to determine micronaire value, maturity and fineness.

**Molecular analyses**

The different types of markers displaying polymorphism between *G. hirsutum* and *G. barbadense* included restriction fragment length polymorphisms (RFLPs) (used only in the BC1 generation), simple sequence repeats (SSRs) and amplified fragment length polymorphisms (AFLPs). Details of the markers and protocols used are provided in Lacape *et al.* (2003) and Nguyen *et al.* (2004). The AFLP markers were all derived from combinations of EcoRI/MseI primer pairs (64 pairs in the BC1, 45 in the BC2 and 30 in the BC3 and BC4 generations). The cotton microsatellites were derived essentially from two public libraries, Brookhaven National Laboratory (BNL) and CIRAD (CIR). In the BC1 generation, the microsatellites used included 188 polymorphic BNL markers out of the 216 available (Lacape *et al.*, 2003) and 204 CIR markers out of 392 developed (Nguyen *et al.*, 2004). From the results of the combined QTL analyses of the BC1/BC2/BC2S1 generations (Lacape *et al.*, 2005), QTL-rich regions were identified on “carrier” chromosomes, and SSR loci present within or in the vicinity of these regions were assembled for constituting groups of three SSRs (one group per region) to be tested as multiplexes, taking into account both annealing temperature and compatibility of sizes of amplified fragments. A subset of 60 SSR (20 region-specific triplexes) was used for early genotyping of all 411 BC3 and 450 BC4 plants (see examples in Figure 2). The individual plants selected from BC3 and BC4 (43 and 37 plants respectively) were further analysed using known AFLPs to provide broad genome coverage. In the context of our marker-assisted introgression programme, the SSR markers target the QTL-rich regions, i.e. those loci of the “foreground genome” expected to have been introgressed, while the AFLP markers essentially serve to cover the rest of the genome, i.e. the “background genome”, aimed at returning to the recipient genome composition.

**Construction of genetic map**

The BC1 (75 individuals) and BC2 (200 individuals) maps were constructed separately using the MapMaker 3.0 software
The MapMaker “group” (using a logarithm of the odds ratio [LOD] of 5.0 and 30 as a maximal recombination frequency), “order” and “sequence” commands were used in each case. After aligning the BC₁ and BC₂ maps using common loci, a consensus framework BC₁/BC₂ map was constructed by simple extrapolation of the positions of the additional BC₂ loci on the BC₁ map used as a backbone map. The allelic constitution throughout the 26 chromosomes of all BC₁₄ individuals was displayed graphically using Graphical GenoTyping software (R. van Berloo, Laboratory of Plant Breeding, Wageningen, Netherlands) and represented along the consensus BC₁ map data.

**QTL analyses**

The combined marker and phenotypic data then served for three (BC₁, BC₂ and BC₂₁) separate QTL analyses of fibre quality components. The association between phenotype and marker genotype was investigated through simple marker analysis (SMA), interval mapping (IM) and composite interval mapping (CIM) using the computer software QTL Cartographer 1.13 (Basten, Weir and Beng, 1999) as described in Lacape et al. (2005). In each data set (trait, generation), permutation-based thresholds were considered at a 5 percent risk at the genome level. Interval methods relied on the positions of the loci on the consensus BC₁ map. Molecular data of further generations (BC₃ and BC₄) were also combined with phenotypic measurements for conducting the SMA option of QTL Cartographer. Cotton fibre properties were considered from a product transformation perspective, meaning that decreases the in fibre fineness and yellow-
ness index, for example, were positively considered.

Details of the plant material used and the types of analyses undertaken during the different steps of the MAS process are given in Figure 1.

RESULTS
Phenotypic variation
The two parents were characterized by their contrasting fibre properties (Table 1) with significant advantages for the *G. barbadense* parent in terms of length (+9.7 mm on average over all data sets), strength (+15.9 g/tex) and fineness (-38 mtex). By contrast, the *G. hirsutum* parent displayed better yellowness index/colour reflectance. For each BC population, it was observed that the data fitted normal distributions, that transgressive segregants were regularly in the lower range of phenotypic values and that, although progeny values rarely reached those of *G. barbadense*, high phenotypic values were observed, including within the most advanced BC4 generation (Table 1).

Genetic mapping
The first step in the programme involved the construction of two genetic maps of tetraploid cotton by combining RFLP, SSR and AFLP markers generated separately from the first two backcross generations (BC1 and BC2). The initial BC1 map comprising 888 loci grouped in 37 linkage groups and spanning 4 400 cM (Lacape et al., 2003), benefited from the development and integration of new additional microsatellite markers (Nguyen et al., 2004). This updated saturated BC1 map spans 5 500 cM and comprises a total of 1 160 loci ordered along 26 chromosomes or linkage groups and spanning 4 400 cM (Lacape et al., 2003), benefited from the development and integration of new additional microsatellite markers (Nguyen et al., 2004). This updated saturated BC1 map spans 5 500 cM and comprises a total of 1 160 loci ordered along 26 chromosomes or linkage groups (Nguyen et al., 2004). On the other hand, the BC2 map constructed using AFLP and SSR markers had 514 loci in total. The two maps agreed perfectly for loci order. They had 373 loci in common (between seven and 26 per chromosome throughout the 26 chromosomes), thus allowing their merger into a combined consensus map. The consensus framework map comprises 1 306 loci and spans 5 597 cM, with an average marker interval of 4.3 cM.

QTL detection
The QTL analyses, conducted through composite interval mapping, used two molecular data sets (BC1 and BC2) and three sets of fibre measurements (per plant

| TABLE 1 | Range of parental (*G. hirsutum* [Gh] and *G. barbadense* [Gb]) values over the five sets of data |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | Gh              | Gb              | BC1 N=75        | BC2 N=200        | BC3 N=200        | BC4 N=43        | BC4 N=37        |
| Length (mm)*   | 27.5–31.8       | 39.2–43.7       | 33.8 (27.8–38.2)| 28.6 (22.9–35.3)| 30.7 (26.9–36.1)| 30.0 (25.0–33.9)| 31.7 (28.1–37.1)|
| Length uniformity | 81.3–85.5      | 83.9–87.1       | 85.0 (82.0–88.2)| 81.3 (73.3–86.7)| 83.3 (80.6–85.1)| 81.9 (77.6–86.4)| 86.1 (82.5–89.5)|
| Strength (g/tex)  | 26.5–32.5       | 41.4–46.7       | 35.7 (29.7–41.6)| 28.3 (17.8–43.7)| 29.0 (23.7–34.5)| 24.5 (16.8–32.8)| 33.8 (29.7–39.1)|
| Fineness (mtex)** | 5.1–6.4        | 5.5–6.0         | 6.3 (5.7–7.4)   | 5.5 (3.9–7.6)   | 6.3 (5.4–7.4)   | 5.7 (4.5–7.0)   | 6.3 (4.9–7.4)   |
| Colour reflectance | 71.2–77.7      | 74.6–75.6       | 74.3 (65.9–81.1)| 72.2 (56.8–81.3)| 74.1 (69.8–77.5)| 71.5 (64.7–76.6)| 75.1 (67.1–82.0)|

* Length is upper half mean length (UHML), ** standard fineness, *** low fibre fineness values in BC3 generation because of poor maturities

Note: Mean values and range (in brackets) observed in each BC1–4 generation (number of plants, N, indicated) of fibre technological parameters.
basis for BC₁ and BC₂ and per-line basis with two replicates for BC₂S₁). The generations BC₁ and BC₂ were conducted with no selection, except for choosing those plants that produced backcrossed seeds. The fibre measurements, which initially included eleven traits, were reduced to six groups after considering the strong correlations that existed between some traits. The fibre characteristics that were retained for measurement included length, length uniformity, strength, elongation, fineness or maturity, and colour.

For the six fibre quality components studied, 50 QTL were identified that met permutation-based LOD thresholds (ranging between 3.2 and 4.0 for most of the traits). Thirty additional suggestive QTL (having a LOD value below threshold but above 2.5) were also taken into consideration after comparing the results between the three populations or between the present results and those reported in the literature (Jiang et al., 1998; Kohel et al., 2001; Paterson et al., 2003; Mei et al., 2004). Table 2 summarizes the data generated from the QTL analyses for the six traits of interest and the phenotypic effects of the detected QTL. In general, the contribution of each QTL, measured as a percentage of explained variation of a given trait, was variable and in most cases fairly low. For example, for traits of economic importance, individual contributions varied from 4.8 to 14.8 percent in the case of fibre length, 4.4 to 21.3 percent for fibre strength and 4.6 to 29.1 percent for colour reflectance.

Overall, it was observed that these 80 QTL partitioned as expected from the phenotypic values of the G. hirsutum and G. barbadense parents: a majority of positive alleles for length (12 of the 15 QTL), strength (8 of the 12 QTL) and fineness (13 of the 21 QTL) derived from the G. barbadense parent, while a majority of positive alleles for fibre colour (13 of the 16 QTL) derived from the G. hirsutum parent (Table 2). Furthermore, the QTL detected for the various traits often co-localized within QTL-rich regions (Lacape et al., 2005). In some cases, QTL detection and mapping were in agreement between generations (BC₁ and BC₂) and, very interestingly, in 26 cases (33 percent of the 80 QTL) they confirmed the results reported in the literature, both for the position of a QTL and for the sign of its phenotypic effect. The most prominent cases of QTL consistently detected in this study as well as in those of Paterson et al. (2003) and Kohel et al. (2001), i.e. in different crosses/populations, were found

### Table 2

<table>
<thead>
<tr>
<th>Trait</th>
<th>Number of QTL</th>
<th>QTL range</th>
<th>Phenotypic effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length (mm)*</td>
<td>12</td>
<td>+0.7 to +2.1</td>
<td>3</td>
</tr>
<tr>
<td>Length uniformity</td>
<td>3</td>
<td>+0.5 to +1.5</td>
<td>3</td>
</tr>
<tr>
<td>Strength (g/tex)</td>
<td>8</td>
<td>+0.8 to +2.8</td>
<td>4</td>
</tr>
<tr>
<td>Elongation</td>
<td>6</td>
<td>+0.2 to +0.5</td>
<td>4</td>
</tr>
<tr>
<td>Fineness (mtex)**</td>
<td>13</td>
<td>–10 to –20</td>
<td>8</td>
</tr>
<tr>
<td>Colour reflectance</td>
<td>3</td>
<td>+1.8 to +2.5</td>
<td>13</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>45</strong></td>
<td></td>
<td><strong>35</strong></td>
</tr>
</tbody>
</table>

* Length is upper half mean length (UHML), ** standard fineness.
along chromosome 3 for QTL for fibre strength and fineness, and chromosome 23 for QTL for fibre strength and length.

The chromosome regions carrying co-localized QTL (corresponding to a single or to several traits measured on a single or on several populations) whose positive alleles derived from the *G. barbadense* donor genome, were reduced to 19 QTL-rich regions that were carried by 15 different “carrier” chromosomes (Table 3). Altogether, the confidence intervals (one LOD) of the involved QTL-rich regions delimited a total length of 636 cM (20 percent of the carrier genome), or 11.5 percent of the total genome (Table 3). Eleven non-carrier chromosomes were devoid of positive QTL, or harboured negative (positive alleles derived from the *G. hirsutum* alleles) QTL.

### Table 3

<table>
<thead>
<tr>
<th>Carrier chromosome</th>
<th>Chromosome length (cM)</th>
<th>Target interval (cM)</th>
<th>Target size (cM)</th>
<th>Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>c14</td>
<td>197</td>
<td>28–57</td>
<td>29</td>
<td>Length</td>
</tr>
<tr>
<td>c3</td>
<td>153</td>
<td>32–67</td>
<td>35</td>
<td>Length, fineness</td>
</tr>
<tr>
<td>c4</td>
<td>190</td>
<td>102–118</td>
<td>16</td>
<td>Fineness</td>
</tr>
<tr>
<td>c22</td>
<td>139</td>
<td>112–139</td>
<td>27</td>
<td>Fineness</td>
</tr>
<tr>
<td>c5</td>
<td>360</td>
<td>78–101</td>
<td>23</td>
<td>Strength</td>
</tr>
<tr>
<td>c6</td>
<td>296</td>
<td>137–144</td>
<td>7</td>
<td>Length, fineness</td>
</tr>
<tr>
<td>c25</td>
<td>183</td>
<td>44–73</td>
<td>29</td>
<td>Length, strength</td>
</tr>
<tr>
<td>c16</td>
<td>168</td>
<td>65–117</td>
<td>52</td>
<td>Strength, fineness, colour</td>
</tr>
<tr>
<td>c23</td>
<td>173</td>
<td>45–66</td>
<td>21</td>
<td>Strength (elongation –, colour –)</td>
</tr>
<tr>
<td>c10</td>
<td>192</td>
<td>0–21</td>
<td>21</td>
<td>Fineness</td>
</tr>
<tr>
<td>c20</td>
<td>268</td>
<td>88–161</td>
<td>73</td>
<td>Elongation, fineness</td>
</tr>
<tr>
<td>c26</td>
<td>195</td>
<td>67–143</td>
<td>76</td>
<td>Length (colour –)</td>
</tr>
<tr>
<td>A01</td>
<td>233</td>
<td>16–54</td>
<td>38</td>
<td>Length</td>
</tr>
<tr>
<td>c18</td>
<td>158</td>
<td>32–46</td>
<td>14</td>
<td>Fineness</td>
</tr>
<tr>
<td>A03</td>
<td>271</td>
<td>209–234</td>
<td>25</td>
<td>Strength, uniformity</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3176</strong></td>
<td><strong>Total</strong></td>
<td><strong>636</strong></td>
<td></td>
</tr>
</tbody>
</table>

Note: All targeted QTL show a positive contribution from the *G. barbadense* allele, except for a few negative cases indicated in brackets. The target region is defined as situated between the two loci flanking the QTL peak LOD value at a one LOD confidence interval.

#### MAS in the BC3 and BC4 generations and allelic transmission throughout generations

The early selection of BC3 and BC4 plants using SSR markers that framed the 19 targeted regions of interest made it possible to choose those plants that showed an allelic constitution with as many introgressed loci within the targeted regions as possible. In total, 43 BC3 plants out of 411 (11.4 percent) and 37 BC4 plants out of 450 (8.2 percent) were retained based upon the information provided by the markers, i.e. without any phenotypic selection at this stage. These plants were backcrossed to the recurrent parent (and self-pollinated in the case of the BC4 plants).

The allelic transmission observed in the four groups of BC4 derived from four different BC1 plants is given in Table 4.
TABLE 4
Percentage of introgressed loci (i%), at the heterozygous state, of 37 BC4 plants that were derived from four different BC1 plants (Nos. 3, 11, 16, and 27)

<table>
<thead>
<tr>
<th>BC1 Plant Number</th>
<th>No. of Loci</th>
<th>i% of Loci</th>
<th>No. of Plants</th>
<th>i% of Plants</th>
<th>No. of Loci</th>
<th>i% of Loci</th>
<th>No. of Plants</th>
<th>i% of Plants</th>
<th>No. of Loci</th>
<th>i% of Loci</th>
<th>No. of Plants</th>
<th>i% of Plants</th>
<th>Global i% of Loci</th>
<th>Target/Non-Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 3</td>
<td>646</td>
<td>55</td>
<td>1</td>
<td>479</td>
<td>14</td>
<td>1</td>
<td>467</td>
<td>8</td>
<td>9</td>
<td>456</td>
<td>5</td>
<td>10/4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. 11</td>
<td>654</td>
<td>64</td>
<td>1</td>
<td>446</td>
<td>28</td>
<td>1</td>
<td>403</td>
<td>13</td>
<td>1</td>
<td>408</td>
<td>10</td>
<td>29/6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. 16</td>
<td>681</td>
<td>67</td>
<td>2</td>
<td>464</td>
<td>31</td>
<td>3</td>
<td>420</td>
<td>15</td>
<td>21</td>
<td>428</td>
<td>9</td>
<td>25/6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. 27</td>
<td>668</td>
<td>63</td>
<td>1</td>
<td>471</td>
<td>26</td>
<td>1</td>
<td>433</td>
<td>15</td>
<td>6</td>
<td>439</td>
<td>10</td>
<td>25/6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>62</td>
<td>62</td>
<td>14</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21/5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The number of plants and of loci analysed at each generation are given. At the BC4 generation, the percentage of introgression is also differentiated between target and non-target (as defined in Table 3) regions.

Moderate deviations were observed from theoretical transmission values (62, 26, 14 and 8 percent compared with 50, 25, 12.5 and 6.25 percent at the BC1, BC2, BC3 and BC4 stages, respectively), with a bias in favour of a higher rate of G. barbadense allele transmission. This bias was probably due to the selection pressure imposed at least in the BC3 and BC4 generations. Throughout the BC1 and BC2 generations that have undergone no deliberate selection, the introgression of G. barbadense alleles (at the heterozygous state) covered the complete genome fairly well, i.e. introgressed segments were found on all of the 26 chromosomes (not shown). This result contradicts the findings of Jiang et al. (2000) who detected important deficiencies in donor (G. barbadense) allele transmission in a population of 3662 BC3 plants originating from 21 BC1 plants.

After combining the SSR and AFLP marker data, it was observed that the introgression rate differed between target and non-target regions. When averaged over the 37 BC4 plants, the percentage of introgressed loci (8 percent genome-wide) was much lower in the non-target regions (5 percent) than that reached within target regions (21 percent) (Table 4). The different BC4 plants introgressed between three and six QTL-rich target regions in different combinations. As an illustration of the selection pressure applied through the use of molecular markers, Figure 3 shows the graphical genotype of two BC4 individuals as well as that of the BC1 plant (No. 16) from which these individuals were derived. The two BC4 plants had a common BC1 ancestor but originated from two different BC2 plants. In this particular example, starting from a common BC1 plant (No. 16) which harboured 13 out of 19 possible QTL-rich regions, the two BC4 plants (Nos. 104 and 419) derived from it partly or completely retained respectively five (c16, c23top, c23bot, c25 and A03) and four (c6, c25, c26 and A01bot) genomic regions carrying favourable alleles. The other regions carrying QTL on c3, c4, c23, c20, A01 and A03, which had been introgressed and were heterozygous in the BC1 plant, had returned to the homozygous G. hirsutum/G. hirsutum state. The percentages of introgressed loci in target and non-target regions in these two examples were 29 and 10 percent, and of 29 and 5 percent in the two BC4 plants (Nos. 104 and 419) respectively.

This example shows that, at least in some cases, the process used was efficient in selecting for chromosomal regions of interest (foreground selection), while allowing the rest of the genome to return towards that of the recurrent parent.
Chapter 6 – Targeted introgression of cotton fibre quality quantitative trait loci using molecular markers

Owing to the limited number of individuals and the unbalanced frequencies of genotypic classes in the BC3 and BC4 material, significant marker-trait associations were less frequent than observed from the BC1 and BC2 data. For example, markers mapped along five, nine and six chromosome regions contributed (P=0.01), respectively, to length, strength or fineness variation using BC4 marker-trait data, as compared with 15, 12 and 21 from the BC1 and BC2 data (Table 2). However, the majority of significant associations, particularly those determined in the BC4 generation, were observed within previously detected regions (not shown). Using fibre strength as an example, out of the eight strength QTL-harbouring regions on chromosomes c3bot, c5, c16, c23sup, c23bot, c25, A01 and A03 identified from the combined BC1 and BC2 data (Table 3), the BC4 data confirmed significant marker-trait associations in five of these regions, i.e. for markers mapped on chromosomes c3bot,

**Figure 3**

Graphical genotypes (26 chromosomes) of a BC1 plant (No. BC1/16) (upper panel) and of two selected BC4 plant (Nos. BC4/104 and BC4/419) (lower panel) derived from it.

The two possible allelic forms, homozygous Gh/Gh and heterozygous Gh/Gb, are denoted in dark grey and black respectively. Regions in black are introgressed with G. barbadense alleles. Light grey areas indicate portions of unknown allelic composition. Boxed areas represent the localization of QTL-rich regions localized on 15 carrier chromosomes shown to the left (11 non-carrier chromosomes are shown to the right). Arrows indicate the regions totally or partially introgressed.

**Fibre characteristics of BC3 and BC4 generation plants**

Owing to the limited number of individuals and the unbalanced frequencies of genotypic classes in the BC3 and BC4 material, significant marker-trait associations were less frequent than observed from the BC1 and BC2 data. For example, markers mapped along five, nine and six chromosome regions contributed (P=0.01), respectively, to length, strength or fineness variation using BC4 marker-trait data, as compared with 15, 12 and 21 from the BC1 and BC2 data (Table 2). However, the majority of significant associations, particularly those determined in the BC4 generation, were observed within previously detected regions (not shown). Using fibre strength as an example, out of the eight strength QTL-harbouring regions on chromosomes c3bot, c5, c16, c23sup, c23bot, c25, A01 and A03 identified from the combined BC1 and BC2 data (Table 3), the BC4 data confirmed significant marker-trait associations in five of these regions, i.e. for markers mapped on chromosomes c3bot,
c16, c23bot, c25 and A03. Furthermore, it is worth noting that the BC₄ plant No. 104 presented in Figure 3, which had introgressed all these five regions, also displayed the highest fibre strength value of its generation (39.1 g/tex, compared with 33.1 g/tex for the Guazuncho 2 parent). The concomitant introgression of *G. barbadense* alleles displaying positive marker-trait associations for other fibre properties such as length or fineness was also observed. This translated into the development of different highly valuable BC progenies. These preliminary results suggest that the improvement of *G. hirsutum* fibre properties through the introgression of *G. barbadense* fibre QTL appears feasible.

**DISCUSSION**

In an attempt to overcome the limitations of conventional breeding for improving cotton fibre quality through the use of interspecific hybridization, molecular markers were used in a MAS scheme to improve the efficiency of introgressing fibre quality traits. The advanced backcross-QTL (AB-QTL) strategy ( Tanksley and Nelson, 1996) was used as this allowed concomitant development of a genetic map of the cotton genome and analysis of fibre quality QTL, and attempts to introgress favourable alleles in an adequate recipient genetic background (Figure 1).

In contrast to monogenic characteristics such as disease and insect resistance, many important traits including yield and quality show continuous phenotypic variation and are governed by a number of QTL. Cotton fibre quality is a complex concept that involves a number of traits or characteristics. Each of these is under the influence of numerous QTL, indicating a complex genetic determinism. Indeed, from the present results, at least six QTL govern fibre uniformity and up to 21 QTL influence fibre fineness. When considering six traits that can account for fibre quality, a total of 80 QTL were detected (Table 2). This figure falls within the same range as that found by Paterson *et al.* (2003). As some of these QTL co-localized within the same chromosome region, by choosing those QTL whose positive allele derived from the donor parent and had the strongest effect on economically important fibre characteristics, the number of target regions to be introgressed was reduced to 19 (Table 3). Nevertheless, this number of QTL remains too high to identify a single plant that would carry them all. Indeed, in the authors’ experience, at the BC₃ stage, single plants carried a maximum of five regions of interest (eight if considering regions only partially introgressed), while at the BC₄ stage, this number was reduced to four (seven if considering regions only partially introgressed).

At this stage of the MAS process, two routes are under way (Figure 1). The first involves identifying the best BC₄ plants, i.e. those showing the highest amount of favourable QTL introgression, and then fixing the favourable allele by self-pollination. Such BC₄S₁ plants have been crossed with other BC₄S₁ plants of different ascent in order to pyramid as many QTL as possible (each contributing to different traits) within the same genome. Similarly, BC₄S₁ plants were used to pyramid various QTL responsible for a given trait (“selective pyramiding”). This latter strategy could especially apply to traits of commercial importance, such as fibre strength or fineness. The second avenue involves repeating the backcrossing process until near isogenic lines differing only at a given QTL (QTL-NILs) are developed. Such plant material could prove useful not only to study the
effect of a single given QTL on the phenotypic value of a plant harbouring it, but also in case the introgressed QTL is proven to contribute significantly to the improvement of a given trait (Bernacchi et al., 1998). Also, QTL-NILs could be used as donor material for QTL pyramiding (Peleman and van der Voort, 2003). Finally, an introgression library, i.e. a collection of NILs, will typically serve as primary plant material for QTL fine mapping and eventual QTL cloning (Salvi and Tuberosa, 2005).

However successful marker-aided introgression of genomic regions of interest may be, only phenotypic analysis of plant material stemming from the MAS process, including the assessment of its adaptability to any given set of local agronomic and ecological conditions, will allow validation of this procedure.

REFERENCES


CHAPTER 7

Marker-assisted selection in common beans and cassava

Mathew W. Blair, Martin A. Fregene, Steve E. Beebe and Hernán Ceballos
SUMMARY

Marker-assisted selection (MAS) in common beans (Phaseolus vulgaris L.) and cassava (Manihot esculenta) is reviewed in relation to the breeding system of each crop and the breeding goals of International Agricultural Research Centres (IARCs) and National Agricultural Research Systems (NARS). The importance of each crop is highlighted and examples of successful use of molecular markers within selection cycles and breeding programmes are given for each. For common beans, examples are given of gene tagging for several traits that are important for bean breeding for tropical environments and aspects considered that contribute to successful application of MAS. Simple traits that are tagged with easy-to-use markers are discussed first as they were the first traits prioritized for breeding at the International Center for Tropical Agriculture (CIAT) and with NARS partners in Central America, Colombia and eastern Africa. The specific genes for MAS selection were the bgm-1 gene for bean golden yellow mosaic virus (BGMYV) resistance and the bc-3 gene for bean common mosaic virus (BCMV) resistance. MAS was efficient for reducing breeding costs under both circumstances as land and labour savings resulted from eliminating susceptible individuals. The use of markers for other simply inherited traits in marker-assisted backcrossing and introgression across Andean and Mesoamerican gene pools is suggested. The possibility of using MAS for quantitative traits such as low soil phosphorus adaptation is also discussed as are the advantages and disadvantages of MAS in a breeding programme. For cassava, the use of multiple flanking markers for selection of a dominant gene, CMD2 for cassava mosaic virus (CMV) resistance at CIAT and the International Institute of Tropical Agriculture (IITA) as well as with NARS partners in the United Republic of Tanzania using a participatory plant breeding scheme are reviewed. MAS for the same gene is important during introgression of cassava green mite (CGM) and cassava brown streak (CBS) resistance from a wild relative, M. esculenta sub spp. flabellifolia. The use of advanced backcrossing with additional wild relatives is proposed as a way to discover genes for high protein content, waxy starch, delayed post-harvest physiological deterioration, and resistance to whiteflies and hornworm. Other potential targets of MAS such as beta carotene and dry matter content as well as lower cyanogenic potential are given. In addition, suggestions are made for the use of molecular markers to estimate average heterozygosity during inbreeding of cassava and for the delineation of heterotic groups within the species. A final section describes the similarities and differences between the MAS schemes presented for the two crops. Differences between the species can be ascribed partially to the breeding and propagation systems of common beans (seed propagated, self-pollinating) and cassava (clonally propagated, cross-pollinating). In addition, differences in growth cycles, breeding methods, availability of genetic markers, access to selection environments and the accompanying opportunities for phenotypic selection influence the decisions in both crops of when and how to apply MAS. Recommendations are made for applying MAS in breeding of both crops including careful prioritization of traits, marker systems, genetic stocks, scaling up, planning of crosses and the balance between MAS and phenotypic selection.
COMMON BEANS: IMPORTANCE AND GENETICS

Common beans (Phaseolus vulgaris L.) are the most important grain legume for direct human consumption, especially in Latin America and eastern and southern Africa. They are seed-propagated, true diploids (2n = 22) and have a relatively small genome (650 Mb) (Broughton et al., 2003). Originating in the Neotropics, common beans were domesticated in at least two major centres in Mesoamerica and the Andes (Gepts, 1988) and possibly in a third minor centre in the northern Andes (Islam et al., 2002). Wide DNA polymorphism is expressed between the two major gene pools. Mesoamerican beans typically have small to medium size seeds and can be classed into four races that are distinguished by randomly amplified polymorphic DNA (RAPD) polymorphisms (Beebe et al., 2000). Andean beans usually have medium to large seeds, and landraces have been classed into three races based on plant morphology and agro-ecological adaptation (Singh, Gepts and Debouck, 1991). These can be differentiated by microsatellites (M. Blair, unpublished data) but the genetic distance among Andean races is narrower than that among Mesoamerican races (Beebe et al., 2001). A large number of gene tagging studies have been conducted in common beans, predominantly with RAPD markers, some of which have been converted subsequently to sequence characterized amplified regions (SCARs; reviewed most recently by Miklas et al., 2006).

Beans display a wide range of growth habits (Van Schoonhoven and Pastor-Corrales, 1987), from determinate bush types, to indeterminate upright or viny bush types, to vigorous climbers. Bush types are the most widely grown, and are a relatively short season crop, maturing in as little as 60 days from seeding in a tropical climate and yielding from 700 to 2 000 kg/ha on average. On the other hand, in small-holder agriculture where land is scarce, labour-intensive, high-yielding climbing beans enjoy continuing or even expanding popularity. Climbing beans can mature in 100 to 120 days at mid-elevations, but can delay as long as ten months at higher elevations and can produce the highest yields for the crop, up to 5 000 kg/ha. These features have significant implications for breeding programmes. In bush types it is possible to obtain up to three cycles per year in the field, or even four cycles in greenhouse conditions. Breeding bush beans is thus quite agile with regard to advance of generations, although seed harvest of individual plants is sometimes limited. With climbing beans, on the other hand, at best it is possible to obtain two cycles per year with field grown plants, while managing climbing beans in the greenhouse is logistically difficult. However, while bush beans produce on average 20 to 50 seeds/plant, individual plants of climbing beans often produce enough seeds to plant several rows (100 to 150 seeds).

Beans are self-pollinating and thus breeding methods for autogamous crops are employed. Pedigree selection or some adaptation thereof is most common, and both recurrent (Muñoz et al., 2004) and advanced (or inbred) backcrossing (Sullivan and Bliss, 1983; Buendia et al., 2003; Blair, Iriarte and Beebe, 2003) have been used. Recurrent selection has also been employed (Kelly and Adams, 1987; Beaver et al., 2003) but seldom in a formal sense with a defined population structure. Singh et al. (1998) suggested a system that they called gamete selection in which individual F1 plants of multiple parent crosses give rise
Marker-assisted selection – Current status and future perspectives in crops, livestock, forestry and fish

84

to families. This system takes advantage of the variability among F1 plants that is created between segregating parental plants. The choice of breeding method and its adaptation to specific circumstances, the growth cycle of the crop in relation to different planting seasons, the access to selection environments and the accompanying opportunities for phenotypic selection and the ease of implementing the specific markers to be used will all influence the decisions about where and how MAS will be most cost effective and used to best advantage.

**MAS in bean breeding: experiences of CIAT and NARS**

Molecular markers have been sought for both simple and complex traits in beans, with an eye to eventual application in MAS. Tagging of genes and QTL in common bean and their application to MAS have been reviewed previously (Kelly et al., 2003; Miklas et al., 2006). In the present chapter, some of the aspects that contribute to the successful use of MAS are considered in greater detail, referring to examples taken from bean breeding in the tropics at CIAT and within NARS. Simple and complex traits are discussed separately, as they represent two contrasting sorts of experience.

**Simple traits**

*Bean golden yellow mosaic virus resistance*

Bean golden yellow mosaic virus (BGYMV) is a white fly-transmitted Gemini virus, and a major production limitation of beans in the mid-to-low altitude areas of Central America, Mexico and the Caribbean. Host resistance to the virus is the most practical means of control, and any new variety in these production areas must carry resistance. Studies on inheritance of resistance revealed a major gene denominated *bgm-1* in breeding line A429 (Blair and Beaver, 1993) that originates in the Mexican (Durango race) accession “Garrapato” or G2402. Minor genes (Miklas et al., 2000c) as well as additional recessive and dominant resistance genes exist for the virus (Miklas et al., 2006). In most production areas where BGYMV exists, it is necessary to pyramid genes for adequate disease control. Although lines developed in CIAT target these areas, BGYMV does not exist at levels that would permit selection under field conditions in Palmira, Colombia, at CIAT headquarters. Therefore, MAS was desirable to assure recovery of at least the most important resistance genes. MAS has also been employed in the Panamerican School in Zamorano, Honduras, as a complement to field screening, to extend selection to sites and seasons with less disease pressure (J.C. Rosas, personal communication).

A co-dominant RAPD marker was identified for the *bgm-1* gene (Urrea et al., 1996) that was subsequently converted to a SCAR marker named SR2 (CIAT, 1997). The DNA fragment associated with *bgm-1* gene has only been observed in one genotype other than G2402 and its derivatives, and thus the polymorphism has been very useful for recognizing the presence of the gene in different genetic backgrounds. This SCAR was evaluated on as many as 7,000 plants in a single sowing (CIAT, 2001; 2003). The uniqueness of the marker’s polymorphism and its reliability over laboratories, seasons and genetic backgrounds have facilitated its wide use. More recently, a second SCAR (SW12.700) was developed from the W12.700 RAPD for a QTL located on linkage group b04 (Miklas et al., 2000c), and this has also been incorporated into the breeding programme of CIAT. The combination of *bgm-1* and the QTL is expected to offer an
intermediate level of resistance, while other minor genes must be recovered through conventional phenotypic selection to assure higher resistance.

Scaling up of MAS required the development of simple operational procedures in both the field (tagging, tissue collection) and the laboratory (DNA extraction, marker evaluation). For gamete selection strategies in the field, individual, evenly-spaced plants from segregating populations were marked with numbered tags that were coated with paraffin to protect them until seed harvest. Leaf disks were sampled from young vegetative tissue with a paper hole puncher and placed directly into pre-numbered cells of microtitre 96-well plates stored on ice, ready for grinding and extraction in the laboratory. The implementation of MAS for \textit{bgm}-1 and subsequently for SW12.700 in the laboratory required substantial adaptation of standard protocols to establish high-throughput procedures. Grinding of samples in microtitre plates was accomplished with a block of 96 pegs that fit into each well. Alkaline DNA extraction (Klimyuk \textit{et al.}, 1993) was employed with success for both markers, and eventually it was possible to multiplex the markers in both the amplification and gel phases using multiple primer PCR and multiple loading per gel wells (Figure 1A). With experience and improved procedures, efficiency more than doubled over a two-year period. MAS was often carried out before flowering to decide on a plant’s status as a carrier of the resistant allele for further use in crossing.

Two small red seeded lines developed in the Panamerican School using MAS have reached the stage of validation in Honduras (J.C. Rosas, personal communication) and shown resistance to the BGYMV strains prevalent there. Resistance to BGYMV of drought tolerant lines selected at CIAT was maintained using MAS for one or more genes, followed by field selection in Central America. Similarly, red mottled lines developed in CIAT with the aid of MAS showed field resistance in the Caribbean and one of these lines from the red mottled advanced line for the Caribbean (RMC) series has been released (Blair \textit{et al.}, 2006). MAS has also been an important element of maintaining BGYMV viral resistance in CIAT’s programme as other breeding objectives such as nutritional value have been assumed, necessitating the inclusion of susceptible parents in crosses with resistant
lines. MAS for this trait has also been practised at the University of Puerto Rico and at the Biotechnology Institute of Cuba.

**Bean common mosaic virus and bean common mosaic necrotic virus**

Bean common mosaic virus (BCMV) and the related necrotic strains (bean common mosaic necrotic virus [BCMNV]) are aphid-transmitted potyviruses that are found worldwide and are seed-borne from season to season. BCMNV resistance is very important in Africa where necrotic strains are prevalent and has become a renewed priority for parts of the Caribbean where necrotic strains have been discovered. BCMV is also endemic in the Andean region where it persists in farmer-saved seed and long-season climbing beans. Climbing beans are grown in both intensive (trellised/staked monoculture) and extensive (intercropping with maize) farming systems. In both systems the need to protect the crop from easily transmitted viral diseases such as BCMV or BCMNV is great; however, very few climbing beans have been bred for resistance to BCMV. A number of BCMV/BCMNV resistance genes have been tagged including the dominant I gene (with which the necrotic strains interact to produce necrosis) and the recessive bc-3, bc-2 and bc-1\(^2\) genes (Haley, Afanador and Kelly, 1994; Melotto, Afanador and Kelly, 1996; Johnson et al., 1997; Miklas et al., 2000a). The genes can be distinguished by inoculation with different viral isolates, and a range of molecular marker tags are available for each gene (reviewed in Kelly et al., 2003; Miklas et al., 2006). The dominant I gene was incorporated into a wide range of small seeded bush beans at CIAT, while resistant bush beans of the bush bean resistant to black root (BRB) series carrying recessive genes were developed in the 1990s and have been widely distributed as breeding parents. The need to reselect the recessive genes with confidence from segregating populations makes MAS a priority.

CIAT started a collaborative project with the Colombian national bean programme based at the Colombian Agricultural Research Corporation (CORPOICA) in 2002 to introgress BCMV resistance genes from BRB lines into local landraces and improved genotypes of Andean climbing beans (CIAT, 2002, 2003, 2004; Santana et al., 2004). During the breeding programme for BCMV and over the course of four years, MAS was used extensively based primarily on the SCAR marker ROC11 developed for the bc-3 gene (Johnson et al., 1997) and the SCAR marker SW13 for the I gene (Melotto, Afanador and Kelly, 1996) along with virus screening to confirm the selection of resistant progeny. The programme was successful in moving bc-3 resistance into a background of cream mottled and red mottled seed types for both highland areas (known as Cargamanto commercial class) as well as mid-altitude areas through triple-, double- and backcrosses. Although virus resistance was also screened phenotypically, the frequency of escape, the complex interaction of multiple genes and the recessive nature of most of these made MAS the best option for breeding resistant varieties rapidly. In addition, as climbing bean breeding is a more time-consuming and expensive endeavour than bush bean breeding due to the longer season, wider plant spacing and need for staking material, MAS was also found to be a very effective measure to reduce breeding costs and save on breeding nursery space.

The implementation of MAS for BCMV was based on a combination of the previously developed SCAR markers previously mentioned and techniques developed at
CIAT for the selection of BGYMV resistance as discussed previously. Although most BCMV and BCMNV resistance genes had been tagged with SCAR markers, implementation required efforts to validate and scale up the use of the markers in applied breeding programmes. Genotyping for the ROC11 marker was carried out on advanced lines given that this marker is dominant and in repulsion with the resistance allele. In other words, the absence of a band was indicative of the presence of the recessive bc-3 allele and therefore it was more appropriate to evaluate after fixation of the alleles to homozygosity through mass or pedigree selection with single plant selections in the F4 and F5 generation when single plant rows were evaluated for the resistance gene marker. To determine whether the advanced line continued to segregate for the gene, alkaline DNA extraction was conducted on leaf discs collected from four leaflets from four individual plants per line using a hole-puncher rather than from a single plant per family or advanced line. The presence or absence of polymerase chain reaction (PCR) products was evaluated for each genotype based on scanned photographs or gel capture imagery of multiplexed gels (Figure 1B) to predict if the genotype contained the resistance or the susceptible allele.

Once optimized for parental genotypes, MAS was conducted on a large number of progeny rows. For example in 2003, more than 4,000 advanced lines were evaluated for the ROC11 marker for genotypes grown at three sites within Colombia (CIAT-Darien, CIAT headquarters and CORPOICA-Rionegro). DNA was collected at all three sites and shipped successfully to the laboratory in 96-well plate format as discussed above. Both the ROC11 and SW13 markers were single copy SCARs that did not produce extra bands and therefore were easy to multiplex. To facilitate the evaluation of markers on a large number of advanced lines, usually within two to three weeks, and increase the efficiency of MAS, several innovations were implemented: loading of agarose gels (first with two and then three loadings), increasing numbers of wells per comb (first 30-well and then 42-well combs were used), use of 384-well PCR plates and multipipetor loading of gels. The resulting savings decreased the time to PCR amplify and load a gel by approximately 50 percent and increased the number of genotypes run per gel by 225 percent.

The rapid increase in efficiency obtained during the application of the ROC11 marker shows the advantages of testing new markers in practical breeding programmes. The use and advantages of these molecular markers has been presented at an Organization of American States-sponsored course in Colombia given in 2002 and a Rockefeller Foundation-sponsored course in Uganda given in 2003. Based on this programme and the training courses, MAS for BCMV genes was initiated as part of a recently approved Association for Strengthening Agricultural Research in Eastern and Central Africa (ASARECA) project for three countries in eastern Africa and training of researchers from the Andean region has allowed more breeding lines from Peru to be screened (CIAT, 2004).

Other examples of MAS for simply inherited traits
Several pathogens, especially fungal pathogens, have co-evolved with the bean host, and present a population structure (Andean/MesoAmerican) that mimics the major gene pools of bean (Pastor-Corrales, Jara and Singh, 1998). This is the case with Phaeoisariopsis griseola, the
causal agent of angular leaf spot (ALS), and *Colletotrichum lindemutheanum*, which induces anthracnose. In both cases, pathogen isolates tend to be more virulent on host genotypes of the same gene pool (Andean or Mesoamerican) and less so on host genotypes from the contrary gene pool. Resistance genes of utility to one host gene pool thus tend to originate in the other gene pool and require introgression from one gene pool to the other. MAS has great potential for introgression as DNA polymorphisms are maximized in wide crosses across gene pools, and markers are available for this purpose for both ALS (Carvalho et al., 1998; Sartorato et al., 1999; Nietsche et al., 2000; Ferreira et al., 2000; Mahuku et al., 2004) and anthracnose (Young et al., 1998; Awale and Kelly 2001; Vallejo and Kelly, 2001).

Other cases of wide crosses in which MAS can be of use include those for the selection of genes for resistance to a storage insect, the Mexican bean weevil (*Zabrotes subfasciatus* [Boheman]) derived from wild bean accessions from Mexico. Selection for resistance has also been achieved by analysis for the active resistance agent, a seed protein called arcelin, by either antibody reaction or electrophoresis, but MAS is simpler and more efficient than either of these analyses that require protein extraction. Even wider crosses of common bean with *Phaseolus acutifolius* have recovered resistance to common bacterial blight (caused by *Xanthomonas axonopodis* pv. *phaseoli*) (Muñoz et al., 2004) and markers have also been developed for these resistance genes (Jung et al., 1997; Miklas et al., 2000b; Park et al., 1999; CIAT, unpublished data). In these cases also, the fact of deploying genes from relatively wide crosses favours maintaining a state of DNA polymorphism in relation to the target genotypes.

**Complex multigenic traits**

In addition to the studies previously discussed, several attempts have been carried out to tag quantitative trait loci (QTL) for abiotic stress tolerance or insect resistance in common bean, although most of these traits might better be described as oligogenic, as results usually suggest that a limited number of loci (from three to six) are involved in their genetic control.

One example is tolerance to low soil phosphorus that was investigated in the landrace G21212. Linkage group b08 proved to be especially important to yield under low phosphorus, with as many as three important and loosely linked QTL (Beebe, Velasco and Pedraza, 1999; Miklas et al., 2006). Interestingly, these same QTL were linked to QTL for resistance to *Thrips palmi* Karny derived from the same source (Frei et al., 2005). This is a promising candidate for applying MAS in the short term for abiotic stress tolerance, although another notable attempt was also made for drought tolerance breeding with MAS through a joint programme between Michigan State University and the National Institute for Forestry, Agriculture and Livestock Research (INIFAP) in Mexico (Schneider, Brothers and Kelly, 1997).

In theory, a breeder would prefer markers for low heritability quantitative traits that are difficult to select through phenotypic selection. However, in general, markers for polygenic or oligogenic traits have not moved into the application phase. The same problems that make phenotypic selection difficult apply in some degree to MAS. Multiple minor genes that are often associated with poor heritability also imply that it is difficult to identify QTL with highly significant effects and that merit the investment of MAS. Furthermore, good genome coverage is usually necessary to
detect the QTL that explain the highest amount of genetic variability, and this has been difficult to achieve in intragene pool crosses in common beans.

However, genetic analysis by markers has been very useful for revealing the inheritance of quantitative traits, especially physiological traits, even when the markers involved did not result in application in MAS. Analysis of QTL was applied to root traits of bean as they relate to absorption of phosphorus from soil (Liao et al., 2004; Yan et al., 2005; Beebe et al., 2006). This permitted associating different physiological traits to P uptake and estimating their importance in nutrient acquisition. Once traits are better understood, then an appropriate selection strategy can be devised, be it phenotypic or MAS. Thus, markers can be useful to a breeding programme by elucidating basic plant mechanisms even if they are not applied directly in selection.

**Breeding schemes: adaptation to include MAS**

The eventual application of MAS requires careful prioritization of traits and even specific genes for which markers are to be sought, in light of the importance of the trait and genes, and options for phenotypic selection. One should never assume that MAS is necessarily superior to phenotypic selection, which for some traits may be as effective and efficient as the use of molecular markers. However, if a gene is sufficiently important in a breeding programme to demand that advanced lines have such a gene (as in the case of the bgm-1 gene for virus resistance in Central America), there is probably some point in the selection process at which MAS would be useful. Also, it is not necessary to select many genes by MAS for it to be of great value. For example, if a single gene is segregating and 50 percent of plants lack the gene in advanced generations, an effective selection would eliminate half the population and increase the subsequent efficiency of the breeding programme by a factor of 2.

Once markers are available, a key issue is determining the range of parental genotypes within which a marker is polymorphic and therefore useful for selection. Markers of genes that originate from wider crosses (e.g. from different races, gene pools or species) will have a progressively greater chance of being polymorphic among a range of parents (Figure 2) and therefore diagnostic for the gene of interest. The example of bgm-1 is again a good case in point as the resistance allele and the SR2 marker are both unique to the Durango gene pool and polymorphic in combinations across other Mesoamerican races as well as the Andean gene pool. In contrast, the ROC11 marker for the bc-3 gene is only polymorphic across gene pools and therefore not diagnostic for the resistant allele.

If a breeder has several potential parents among which to choose and these are comparable with regard to other traits, it might be preferable to eliminate those that carry a band that would be confused with the linked marker and would result in false positives. Conversely, if more than one marker is available for a given gene, one might focus on those linked markers that maintain polymorphism in the greater number of combinations. In some combinations it might be informative to use both linked markers simultaneously, both to discern recombinants and to confirm markers.

Several possible schemes for the introduction of MAS to different breeding schemes are represented in Figure 3. A breeder must consider at what generation in the breeding programme selection
by MAS will give the greatest cost/benefit ratio. This would probably be early in the breeding programme for the pedigree method or for gamete selection while it would be later in the programme for bulk method or mass selection (Figure 3). In the case of early generation selection, elimination of plants without the gene(s) will avoid unproductive investment.

**Advantages and disadvantages of MAS**

MAS provides real advantages where the conditions are not favourable for phenotypic selection, for example, in the case of BGYMV, which does not exist at epiphytotic levels in CIAT. Indeed, bgm-1 behaves as a recessive gene, so phenotypic selection in early generations would be inefficient in recovering the gene in the heterozygous state.

The same principle would apply to the recessive bc-3 gene, although the lack of a marker linked in coupling to this gene has been a serious drawback and has limited the effectiveness of MAS to advanced generations when the gene is fixed by inbreeding. In this case, early generation selection with MAS would be limited to negative selection against homozygous dominant and heterozygous plants, and this eliminates potentially useful allele-carrying genotypes. Indeed, MAS is impossible in generations such as the F₁ or BC₁F₁ to the susceptible parent when no homozygous recessive plants exist at all.
In other cases where phenotypic selection methods are available, the advantage of MAS resides in its simplicity. This is the case in the selection of arcelin, which can be achieved through protein extraction followed by antibody detection or electrophoresis, but both of these are laborious while MAS can be applied more rapidly and with much greater throughput. Similarly, markers for common blight resistance and anthracnose have the advantage of obviating the need for field inoculations that are sometimes ineffective if environmental conditions are not favourable. The advantage of MAS is much greater if a single DNA extraction can serve for the evaluation of several markers, as in the multiplexing of bgm-1 and SW12.700 markers.

In spite of attempts to apply MAS to complex traits, examples of successful application are still limited to relatively simple traits. This is contrary to some previous expectations that markers would benefit mostly traits of low heritability. However, experience has shown that the ability to manipulate even one important gene with confidence can make a breeding programme more efficient, if that gene is highly desirable and valuable for advanced materials.

Meanwhile the disadvantages of MAS compared with phenotypic selection are based on effectiveness and cost considerations. The effectiveness of MAS is relative to the ease of applying a given marker, its reliability and its level of linkage with the gene of interest. Although molecular markers theoretically have a heritability of 1.0, variability among laboratories or among runs within a laboratory make markers less than 100 percent reliable. This is especially true for RAPD markers for which band amplification is dependent on DNA concentration and quality, annealing temperature and thermocycling conditions, Taq polymerase concentration and the relative proportion of various other ingredients to the PCR cocktail. In comparison, SCAR markers are much more reliable and repeatable and therefore have higher heritability than RAPD markers. Linkage distance between a marker for a
gene of interest and the actual locus itself also affects the reliability of a marker. In turn, the type of cross (wide versus narrow) and parents involved (closely or distantly related) affect the frequency of recombination around introgressed genes as well as the level of polymorphism of the cross and whether the marker will present distinct alleles for the desirable and undesirable character states. In this regard, there is a tradeoff as MAS is most effective when there is high polymorphism in the crosses being evaluated (Figure 2). However, this is precisely the breeding situation in which gene introgression is most difficult, time-consuming and plagued by linkage drag, as is the case for interspecific or interspecific-derived crosses, hybridizations with wild or wild-derived genotypes and crosses between the Andean and Mesoamerican gene pools. This issue is being addressed in beans with the development and mapping of microsatellite markers (Blair et al., 2003) that are much more polymorphic and useful for diagnosing the inheritance of genomic segments in narrow crosses. The first application of microsatellite markers for MAS in common beans was the selection of arcelin based bruchid resistance using gene-derived simple sequence repeats that are diagnostic for the introgression of alleles for resistance from wild beans into cultivated backgrounds (CIAT, 2004), but others should also show promise.

In terms of cost considerations, the relative costs of MAS versus phenotypic selection are relative to each trait and situation. The widely held perception that MAS is expensive is often due to the ingredients and time used to prepare DNA extractions and PCR reactions, although these costs have been reduced by innovations such as the alkaline DNA extraction technique (Klimyuk et al., 1993) that obviates the need for organic solvents or expensive enzymes involved in other mini-preparation techniques (Afanador and Hadley, 1993). While experienced labour was previously required for DNA extraction at CIAT or in NARS breeding programmes, the alkaline extraction method allows most laboratory steps to be carried out even by untrained personnel. Furthermore, MAS costs can be reduced by miniaturization, especially in the PCR reaction (for example, use of 384-well PCR plates and small reaction volumes) and re-use of ingredients (for example plasticware including pipette tips and microtitre plates as well as agarose from used gels). As previously mentioned, multiplexing adds to the efficiency and therefore reduces the datapoint costs of MAS.

Currently, MAS with SCAR markers and alkaline extraction at CIAT cost less than US$0.25 per datapoint. Therefore the expense of MAS is now not as important an issue as previously. In this regard, MAS sometimes has the advantage of being implemented in any generation and under both field or greenhouse conditions, while phenotypic selection often requires a separate planting and specialized labour for inoculation, agronomic management and evaluations or scoring. However, in the final analysis, the most efficient and cost effective breeding programme will probably be one that combines MAS and phenotypic selection in some optimal combination. It is precisely the challenge of the breeder to define that optimal combination.

One last disadvantage of relying on MAS is that it commits a breeder to a unique gene(s) for a given trait. For example, there might be multiple genes or gene combinations for resistance to a disease, or for a physiological trait such as root structure. To the extent that a breeder relies on MAS for selection, this excludes other
possible genes and the use of other potentially useful parents that do not share the DNA polymorphism that is used in MAS. On the other hand, phenotypic selection would permit recognizing different genetic options for a desired phenotype. Thus, MAS is most useful when it is applied to truly unique genes.

**CASSAVA: IMPORTANCE AND GENETICS**

Cassava is a perennial shrub but it is generally harvested as an annual crop at 10–11 months of age. Basically every part of the plant can be utilized. The starchy roots are a valuable source of energy and can be boiled or processed in different ways for human consumption and different industrial purposes such as starches, animal feed or alcohol (Ceballos et al., 2006). Cassava storage roots are not tubers and therefore cannot be used for reproductive purposes; stems are the common planting materials. Cassava foliage is not widely exploited in spite of its high nutritive value (Buitrago, 1990; Babu and Chatterjee, 1999). Foliage consumption by humans is relatively common in certain countries of Africa, Asia and Latin America. The use of foliage for animal feeding is generating increased interest in Asia.

Cassava can be propagated by either stem cuttings or botanical seed. However, the former is the practice most widely used by farmers for multiplication and planting purposes. Propagation from true seed occurs under natural conditions and is common in breeding programmes. Occasionally botanical seed is also used in commercial propagation schemes (Rajendran et al., 2000).

Cassava is monoecious and allogamous, with female flowers opening 10–14 days before the male ones on the same branch. Pollination can be done manually in a controlled way to produce full-sib families or else in polycross nurseries where open pollination takes place and, therefore, half-sib families are produced. Self-pollination is feasible when using male and female flowers on different branches or on different plants of the same genotypes (Jennings and Iglesias, 2002). Some clones flower relatively early at four or five months after planting whereas others only do so at eight to ten months after planting. As a result, the time required for the seed to mature, the growing cycle of the crop and the need to plant with the arrival of the rains take about two years between a given cross being planned and the respective seed becoming available. On average, between one and two seeds (out of the three possible in the trilocular fruit) per pollination are obtained (Kawano, 1980; Jennings and Iglesias, 2002).

**Breeding objectives**

Productivity plays a major role in industrial uses of cassava, whereas stability of production is fundamental in the many regions where cassava is the main subsistence crop. Industrial uses of cassava require high dry matter content as the main quality trait for the roots, whereas for human consumption the emphasis is on cooking quality, frequently even over productivity, as the determining trait. Stability of production is associated with resistance or tolerance to major biotic and abiotic stresses, with the emphasis varying with the target environment. Genetic resistance to the most important diseases and pests and the prevalent abiotic stresses can be found in cassava germplasm (Hillocks and Wydra, 2002; Bellotti et al., 2002; Belloti, 2002; Ceballos et al., 2004). Although cyanogenic glucosides are found in every tissue except
the cassava seed, most processing methods allow a rapid release and elimination of the cyanide. Depending on the end use, high or low cyanide clones are preferred. Other relevant traits for the roots are dry matter, protein and carotenoid content (Chávez et al., 2005).

**Breeding schemes**

Genetic improvement of clonally propagated non-inbred crops such as cassava is made possible by the fact that a superior genotype can be fixed at any stage in the breeding scheme, even after a single cross, the equivalent of an F1 in commercial hybrids such as maize. Therefore, non-additive gene actions including dominance and epistasis become important components of the genetic variance to be manipulated by the breeder (Jaramillo et al., 2005; Calle et al., 2005; Perez et al., 2005a). Large effective breeding population sizes are required to retain favourable dominant alleles and epistatic loci combination.

As in most crop breeding activities, cassava genetic improvement starts with the production of new recombinant genotypes derived from selected elite clones. Scientific cassava breeding began only a few decades ago, and the divergence between landraces and improved germplasm is not as wide as in other crops. Therefore, accessions for germplasm bank collections from different research institutions play a more relevant role in cassava than in other crops that have been scientifically bred for longer periods of time. Parental lines are selected based mainly on their performance per se and little progress has been made to use general combining ability (Hallauer and Miranda Fo, 1988) as a criterion for parental selection. Sexual seeds obtained by the different crossing schemes are germinated to initiate a new cycle of selection. The multiplication rate of cassava planting material is low as five to ten cuttings can be obtained from one plant. This implies a lengthy selection process, and in fact it takes about six years from the time the botanical seed is germinated until enough planting material is available for multilocation replicated trials.

Table 1 illustrates a typical selection cycle in cassava. It begins with the crossing of elite clones and finishes when the few clones surviving the selection process reach the stage of regional trials across several locations. It should be emphasized that there is some variation among the few cassava-breeding programmes in the world with respect to the number of genotypes and types of evaluation trials.

### Table 1

<table>
<thead>
<tr>
<th>Year</th>
<th>Activity</th>
<th>Number</th>
<th>Plants per genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>Crosses among elite clones planned, nurseries planted and pollinations made</td>
<td>Up to 100 000</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>F1: Evaluation of seedlings from botanical seeds. Strong selection for African cassava mosaic virus (ACMV) in Africa.</td>
<td>100 000; 50 0000; 50 000</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Clonal evaluation trial (CET)</td>
<td>20 000–30 000; 50 000; 700</td>
<td>6–8 (1 rep, 1 location)</td>
</tr>
<tr>
<td>5</td>
<td>Preliminary yield trial (PYT)</td>
<td>100; 300; 80</td>
<td>20–60 (3 reps, 1 location)</td>
</tr>
<tr>
<td>6</td>
<td>Advanced yield trial (AYT)</td>
<td>25; 100; 20–25</td>
<td>100–500 (3 reps, 2–3 location)</td>
</tr>
<tr>
<td>7-9</td>
<td>Regional trials (RT)</td>
<td>5–30; 500-4 000 (3 reps, 3–4 locations)</td>
<td></td>
</tr>
</tbody>
</table>

Figures for cassava breeding at a IITA (Ibadan, Nigeria); b CIAT (Cali, Colombia) and c CIAT and Rayong Field Research Station from Department of Agriculture (Thailand).

Source: adapted from Jennings and Iglesias, 2002.
and plants representing them through the different stages. Table 1 also provides an idea of the selection pressures generally applied.

Strong emphasis on highly heritable traits (plant type, branching habits and reaction to diseases, harvest index and dry matter content) is applied during the early phases of selection (F1 and CET), (Hahn, Howland and Terry, 1980; Hahn, Terry and Leuschner, 1980; Hershey, 1984; Kawano, 2003; Ceballos et al., 2004). As the number of plants representing each genotype increases, the weight of selection criteria shifts towards low heritability traits such as root yield. The clones that show outstanding performance in the regional trials are released as new varieties and, eventually, incorporated as parents in the crossing nurseries. With that the selection cycle is finished and a new one begins. The whole process has the following characteristics (Ceballos et al., 2004):

- the process is indeed phenotypic selection because no family data are involved;
- no data are collected in the early stages of selection. Therefore, data regarding general combining ability effects (∼breeding value) are not available for a better selection of parental materials;
- there is no proper separation between general (GCA ∼ additive) and specific (SCA ∼ heterotic) combining ability effects. The outstanding performance of selected materials is likely to depend on positive heterotic effects that cannot be transferred to the progenies that are sexually derived from them;
- no inbreeding is incorporated purposely in the selection process. Therefore, large genetic loads are likely to remain hidden in cassava populations and useful recessive traits are difficult to detect;
- several stages of selection are based on unreplicated trials. A large proportion of genotypes is eliminated without proper evaluation.

For the above-mentioned reasons, cassava breeding is difficult, expensive and to a certain degree inefficient (Perez et al., 2005a; Cach et al., 2005a, b). Kawano et al. (1998) mention that, during a 14-year period about 372 000 genotypes derived from 4 130 crosses were evaluated at the CIAT-Rayong Field Crop Research Center. Only three genotypes emerged from the selection process to be released as official varieties. Similar experiences have been observed at the International Institute of Tropical Agriculture (IITA), CIAT-Colombia and Brazil. Therefore, the development and adaptation of molecular tools for cassava genetic improvement offer important advantages to make the process more efficient and effective.

**MAS in cassava breeding**

Cassava genetic improvement can be made more efficient through the use of easily assayable molecular genetic or DNA markers (MAS) that enable the precise identification of genotype without the confounding effect of the environment, thereby increasing heritability. MAS can also contribute to the efficient reduction of large breeding populations at the seedling stage based upon “minimum selection criteria”. This is particularly important given the length of the growing cycle of cassava and the expense involved in the evaluation process. Therefore, a pre-selection at the F1 phase (see Table 1) could greatly enhance the efficiency of the CET experiments. The selection of progenies based on genetic values derived from molecular marker data substantially increases the rate of genetic gain, especially if the number of cycles of evaluation or generations can be reduced (Meuwissen, Hayes and Goddard, 2001).
Another application of MAS in cassava breeding is reducing the length of time required for the introgression of traits from wild relatives. Wild relatives are important sources of genes for pest and disease resistance in cassava (Hahn, Howland and Terry, 1980; Hahn, Terry and Leuschner, 1980; Chavarriaga et al., 2004), but the need to reduce or eliminate undesirable donor genome content and linkage drag can lengthen the process, making it unrealistic for most breeders. Simulations by Stam and Zeven (1981) indicate that markers could reduce linkage drag and would reduce the number of generations required in the backcross scheme. Hospital, Chevalet and Mulsant (1992) corroborated this in achieving a reduction of two backcross generations with the use of molecular marker selection. Frisch, Bohn and Melchinger (1999), through a simulation study, found that use of molecular markers for the introgression of a single target allele saved two to four backcross generations. They inferred that MAS had the potential to reach the same level of recurrent parent genome in generation BC3 as reached in BC7 without molecular markers.

The decision to employ DNA-based markers in cassava breeding is primarily based on the heritability of a trait and the amount of genotypic variance explained by the marker. There are many instances in cassava breeding where $h^2$ is low or zero. Some examples are:

- plant health traits where the pathogen or pest pressure is absent or low, such as cassava mosaic disease (CMD) in the New World tropics or cassava green mite (CGM) during the wet season;
- variable or erratic pest pressure, e.g. the CGM or diseases such as the cassava frog skin disease (FSD);
- evaluation based upon a single plant;
- variable experimental fields and/or poor management resulting in large experimental errors;
- traits that are affected by the stage of plant growth or the part of the organ used for tissue analysis, e.g. cyanogenic potential.

In the above-mentioned instances, having a marker(s) that explains a large proportion of the genetic variance can accelerate progress in breeding. Even where $h^2$ is moderate or high, selection by markers can be advantageous:

- where different sources of genes exist for the trait that are indistinguishable by phenotype alone and pyramiding is difficult and time consuming, e.g. for different sources of resistance to a disease or pest;
- where molecular tags that can be used inexpensively and rapidly to identify desirable genotypes early in the breeding cycle exist, thereby eliminating the need to evaluate large numbers of plants phenotypically, and obviating the confounding effects of the environment. Markers may permit the efficient elimination of undesirable genotypes at the seedling stage. For example, the number of genotypes at the seedling stage can be reduced by 50 percent if a trait is controlled by a single gene, or by 87.5 percent if controlled by three genes;
- for the introgression of useful genes from exotic germplasm into adapted gene pools. MAS can be used to identify genotypes that carry minimal amounts of flanking donor parent genome around the gene of interest for faster backcrossing;
- for definition of heterotic pools in a group of germplasm accessions for more directed crosses.
• for definition of average heterozygosity in the selection of partially inbred lines for tolerance to inbreeding;
• for identification of the male parent in elite germplasm derived from polycrosses by fingerprinting. This tool is also useful for checking the identity of different genotypes to eliminate duplication in germplasm collections.

Best results are achieved when MAS is combined with phenotypic data as compared with either approach independently (Hospital, Chevalet and Mulsant, 1992). Phenotypic data would reduce the cost of genotyping especially if phenotypic evaluation is conducted on early generations (Gimelfarb and Lande, 1994). This not only reduces the cost of MAS but also increases its efficiency. Some examples of MAS in cassava breeding conducted at an international centre and national programmes are described below.

**Molecular MAS for CMD resistance at an IARC**

An ideal target for MAS is breeding for disease resistance in the absence of the pathogen. This is the case of CMD in the Americas, where the disease does not occur. CMD is a viral disease first reported by Warburg in 1894 in eastern Africa (quoted by Storey and Nichols, 1938). Several variants of the disease (East Africa cassava mosaic virus [EACMV], South Africa cassava mosaic virus [SACMV], Indian cassava mosaic virus [ICMV]) have been reported (Swanson and Harrison, 1994) and are endemic in all cassava growing regions of Africa and southern India, where it is the most severe production constraint. The white fly vector of CMD, *Bemisia tabaci* biotype A, does not colonize cassava in the New World but recently a new biotype of *B. tabaci*, biotype B (also referred to as *B. argentifolia*), has become widespread in the Americas and has a wide host range including cassava (Polston and Anderson, 1997), increasing the possibility that CMD, EACMV, SACMV, ICMV or a native American gemini virus will become established on cassava in the neo-tropics. This is a frightening prospect for cassava production in Latin America, considering that most Latin American cassava germplasm is very susceptible to CMD (Okogbenin et al., 1998). The susceptibility of neo-tropical germplasm to CMD also limits the utilization of germplasm from the crop’s centre of diversity in the neo-tropics for these key cassava production regions. Breeding for resistance to CMD in Latin America, where the disease does not exist and is unlikely to be introduced due to very strict quarantine controls, requires the tools of MAS.

Evaluations at IITA identified an excellent source of resistance to CMD in some Nigerian landraces (A.G.O. Dixon 1989, unpublished data), namely TME3, TME7, TME5, TME8, TME14 and TME28. This resistance is effective against all known strains of the virus, including the virulent Ugandan variant (UgV) (Akano et al., 2002; CIAT, 2001). CIAT, in collaboration with IITA in Ibadan, Nigeria, and with support from the Rockefeller Foundation, developed several molecular markers for this source of CMD resistance, revealed to be controlled by a single dominant gene designated as *CMD2* (Akano et al., 2002). At least five markers tightly associated to *CMD2* have been developed, the closest being RME1 and NS158 at distances of four and seven cM respectively. The dominant nature of *CMD2* and its effectiveness against a wide spectrum of viral strains makes its deployment very appealing for protecting cassava against the actual or potential ravages of CMD.
in both Africa and Latin America. CIAT and IITA undertook a project to verify the utility of these markers for MAS in breeding CMD resistance by developing crosses between the sources of TME3 and susceptible varieties. A total of six families, ranging in size from 36–840 genotypes, and a total of 2,490 genotypes were used. The crosses were genotyped with two markers and also evaluated for CMD resistance in a high CMD pressure area in Nigeria. Results of the marker analysis and phenotypic evaluation of CMD resistance in the field revealed that the markers RME1 and NS158 SSR were excellent prediction tools for CMD resistance in some crosses (a prediction accuracy of 70–80 percent). In a few families, however, the markers were not polymorphic between the resistant and susceptible parent and, therefore, were not useful. This highlights the need to develop many markers around a gene of interest in a MAS programme and then to use those markers to evaluate the parents and identify the best markers for the different cross combinations.

Eighteen progenies from TME3 carrying the CMD2 marker were established from embryo axes and imported to CIAT from IITA. They were crossed extensively to elite parents. Seeds harvested from the crosses were germinated in vitro from embryo axes according to standard protocols for cassava (Fregene et al., 1997, CIAT, 2002) to allow sharing the CMD resistant genotypes with collaborators in Africa and India. Each plantlet was multiplied after three to four weeks of growth to obtain three to five plants. After another four weeks, leaves of all plants were removed for molecular analysis and the plants multiplied again to obtain 10–20 plantlets. DNA isolation was by a rapid mini preparation method developed for rice (Nobuyuki et al., 2000). The DNA obtained is sufficient for 100 reactions and can be held in the Costar plates for two months at −20 °C without any degradation. PCR amplification, polyacrylamide gel electrophoresis (PAGE) or agarose gel analysis of SSR markers NS158 and RME1 were as described by Mba et al. (2001). The versatility of spreadsheets makes them the appropriate software to handle the diverse information generated by MAS. Gel images from the marker analysis were entered directly into a spreadsheet that contains information on the parents, tissue culture and greenhouse records, and subsequent phenotypic evaluation of the progenies. After molecular analysis, genotypes that carry the marker allele associated with CMD2 were further multiplied to obtain at least 30 plants. Ten plants were sent to the greenhouse for hardening and later transferred to the breeding programme for evaluation. Five plants were kept in vitro, while 15 plants were shipped to partners in India and Africa as shown in the flow chart for MAS (Figure 4).

To date, more than 50,000 progeny have been evaluated with CMD linked markers and resistant lines shared with national programmes in India or Africa, and also incorporated into the breeding scheme at CIAT. The cost of a single marker data point is US$0.30 and 32,000 samples can be processed in a year.

**MAS for CMD resistance at a NARS**

Although evaluation for CMD resistance in sub-Saharan Africa is relatively easy and most areas have sufficient disease pressure to permit moderate to high heritability
of resistance, overlapping outbreaks of CGM, cassava bacterial blight (CBB), and CMD are common (Legg and Ogwal, 1998) and the need for modest-sized breeding populations make MAS for CMD resistance a powerful tool to accelerate cassava improvement even in Africa. A MAS and participatory plant breeding (PPB) project was initiated in 2003 with funding from the Rockefeller Foundation to improve the resistance of local cassava varieties in the United Republic of Tanzania to CMD and CGM and also to provide proof of concept for the use of MAS to accelerate cassava improvement. The United Republic of Tanzania is the fourth largest producer of cassava in Africa with average yields of about 8 tonnes/ha (FAO, 2001). This is below the continent’s average of 10 tonnes/ha, and well below the average yield of 14 tonnes/ha of Africa’s (and the world’s) largest producer, Nigeria.

The low yield in the United Republic of Tanzania is caused by many factors, including the susceptibility of commonly grown varieties to major diseases and pests such as CMD and the cassava brown streak disease (CBSD). The project crosses farmer-preferred germplasm, by agro-ecology, to improved introductions that are resistant to CMD and to CGM. Markers associated with resistance to CMD are used to reduce
the population size and a small set of genotypes with the “minimum criteria” for successful cassava production are evaluated in a single season in the corresponding agro-ecology and then evaluated over two cycles in collaboration with end-users (rural communities and cassava processors). Figure 5 describes the scheme of the United Republic of Tanzanian MAS and PPB project. CMD resistant F1 generated by MAS at CIAT were crossed to BC1 derivatives of M. esculenta sub spp. flabellifolia, showing good resistance to CGM, to produce progenies that combine some CMD and CGM resistance (Kullaya et al., 2004). The progenies were established from embryo axes as in vitro plants to aid shipment to Africa. Molecular markers associated with resistance to CMD and phenotypic evaluation for CGM resistance were used to screen and select progenies that combine resistance to CMD and CGM. Resistant plants (300 genotypes and ten plants per genotype), were shipped to the United Republic of Tanzania as in vitro plantlets for use as improved parents. A selection based on harvest index, a highly heritable trait, and total biomass was made and 80 genotypes selected. These were planted in the second year in a controlled crossing block together with 54 local germplasm from the eastern and southern zones of the country. Emphasis was placed on local varieties with, or tolerance to, CBSD, which is a major disease of cassava in coastal east Africa and Mozambique. Over 40,000 crosses were made between the improved genotypes and the local varieties producing more than 60,000 seeds.

Sexual seeds obtained from crossing improved and local genotypes were planted in the screen house and transferred to the field 40 days after planting. Parental lines were also planted in the screen house from woody stakes. DNA was isolated from parental lines using the rapid mini-preparation method and evaluated with the five markers associated with the

---

**FIGURE 5**

MAS scheme to improve local varieties of cassava in the United Republic of Tanzania using improved disease and pest resistant introductions from Latin America

<table>
<thead>
<tr>
<th>Year</th>
<th>Improved introductions (≈90)</th>
<th>Crossing block (Polycross design)</th>
<th>Local varieties (≈60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Crossing block</td>
<td></td>
<td>MAS</td>
</tr>
<tr>
<td>3</td>
<td>Seedling trial (60,000 seedlings)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Combining ability studies</td>
<td>Single row trial (≈10,000 genotypes)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Farmer participatory trials</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The scheme is now in its second year.
CMD2 mediated resistance to CMD. Polymorphism in pair-wise combinations of the parental lines was observed with at least one of the five markers and will be used on the progeny. The phenotype of the progeny will be evaluated at three and six months after planting for resistance to CMD, CBSD and CGM. Markers are currently being tested for CGM resistance and are being developed for resistance to CBSD; when their utility is confirmed, they will also be used on progenies.

Using published broad sense heritability of 0.6 for CMD resistance (Hahn, Terry and Leuschner, 1980), it is expected that 24 000 symptomless genotypes will be analysed with markers associated with resistance to CMD. The gain of MAS will be the elimination of at least 38 400 (4 800 x 8 plants) that would have been carried to the single row trial stage (eight plant-rows per genotype), considering that breeders traditionally select 20 percent at the seedling trial stage. This represents a reduction of about 4 ha at the CET. If markers can be used to select for resistance to CGM and CBSD, then an additional number of genotypes can be eliminated from the CET leading to even greater savings. Using MAS for CMD alone would reduce the size of field trials by 50 percent. If additional second and third traits were included, reductions could be as high as 75 and 87.5 percent, respectively. Perhaps the most important advantage, however, comes from the increased genetic gain arising from higher heritabilities in these field evaluations with fewer genotypes.

**MAS for transferring useful traits from wild relatives of cassava into the cultivated gene pool**

Wild *Manihot* germplasm offers a wealth of useful genes for the cultivated *M. esculenta* species, but its use in regular breeding programmes is restricted by linkage drag and the long reproductive breeding cycle. For example, several accessions of *M. esculenta* sub spp. *flabellifolia*, *M. peruviana* and *M. tristis* have high levels of proteins (Nichols, 1947; Asiedu *et al.*, 1992; CIAT, 2004). Low amylose content starch (3–5 percent) or waxy starch of relevance to the cassava starch industry has also been identified in two wild relatives of cassava, namely *M. crassisepala* and *M. chlorostricta*. The only source of dramatically delayed post-harvest physiological deterioration (PPD) has been identified in an interspecific hybrid between cassava and *M. walkerae*. The *M. walkerae* parent was collected in Mexico and held at the Washington University, St. Louis, United States of America (Bertram, 1993). It was brought to CIAT in 1998 in an attempt to use it in improving PPD. Furthermore, the only source of resistance to the cassava hornworm and the most widely deployed source of resistance to CMD were identified in fourth backcross generation progenies of *M. glaziovii* (Jennings, 1976; Chavarriaga *et al.*, 2004). Moderate to high levels of resistance to CGM, whiteflies and the cassava mealybug have been found in interspecific hybrids of *M. esculenta* sub spp. *flabellifolia*. The delayed PPD trait and resistance to the pests were successfully transferred to F₁ interspecific hybrids suggesting dominant or additive gene action of the gene(s) involved (CIAT, unpublished data).

The long reproductive cycle and lengthy time required to develop new cassava varieties (10–15 years) often discourages the use of wild species in most conventional cassava breeding programmes. However, the use of molecular markers to introgress a single target region of the genome can
save between two to four backcross generations (Frisch et al., 1999). Indeed, it has been shown in several crops that the “tremendous genetic potential” locked up in wild relatives can be released more efficiently through the aid of new tools of molecular genetic maps and the advanced backcross QTL mapping scheme (ABC-QTL) ( Tanksley and McCouch, 1997).

For several years now molecular marker tools and a modified ABC-QTL scheme have been tested in cassava at CIAT for the introgression of useful genes from wild relatives. The scheme entails generating BC₁ crosses and carrying out QTL mapping followed by selection of genotypes carrying the genome region of interest with minimum segments of the donor genome (Figure 6). The modified ABC-QTL is currently being used at CIAT to introgress genes for high protein content, waxy starch, delayed PPD, and resistance to whiteflies and the hornworm. The most advanced of these MAS projects is the introgression of high protein content from close wild relatives of cassava. Two BC₁ families of between 250 and 300 progenies were developed from two accessions of M. esculenta sub spp. flabellifolia OW284-1 and OW231-3, and the improved cassava variety from Thailand Rayong 60 (MTAI 8 in the germplasm collection). The BC₁ families were planted
in a CET for evaluation of root protein content at ten months. The grand parental lines of the BC\textsubscript{1} population were genotyped with over 800 simple sequence repeat (SSR) markers available for cassava and about 300 polymorphic markers were identified. The polymorphic markers are being assayed in the progenies after which QTL analysis will be conducted using the phenotypic protein and molecular marker data. Genotypes that have QTL for protein and a minimum of the donor parent genome will be selected and used for producing the BC\textsubscript{2} generation.

For introgression of naturally occurring mutant granule-bound starch synthase (GBSSI) for waxy starch in wild relatives, a more targeted approach was taken. Sequencing of the glycosyl transferase region of the GBSSI gene from the wild relatives and two cassava accessions identified four single nucleotide polymorphisms (SNPs) that differentiated the wild accessions from cassava. Allele-specific molecular markers unique to these SNPs were developed for selection of these alleles in a breeding scheme.

Genetic crosses were made between *M. chlorosticta* accession CW14-11 and MTAI8, and the resulting F\textsubscript{1} was backcrossed to MTAI8. The allele specific marker will be used together with other agronomic traits, particularly performance, to select for BC\textsubscript{1} that carry the mutant GBSS alleles for self-pollination to recover the waxy trait. The identification of natural mutants in a key gene and development of markers represent an innovative molecular tool to accelerate the introgression of favourable alleles from wild relatives into cassava. Backcross derivatives have also been developed from *M. walkerae* (MWal 001) for delayed post-harvest physiological deterioration; from MNG11 (a BC\textsubscript{4} derivative of *M. glaziovii*) for resistance to hornworm; and from *M. esculenta* sub spp. *flabellifolia* (FLA447-1) for resistance to whiteflies. Phenotypic and genetic mapping of these backcross populations are in progress to be followed by identification of QTL and selection of progenies to generate the next generation. MAS will later be used to combine these genes into progenitors for use as parents in breeding which, together with low cost marker technologies, will be distributed extensively to national programmes in Africa, Asia and Latin America to produce improved varieties.

**Marker-assisted estimation of average heterozygosity during inbreeding of cassava**

A principal use of molecular markers by private sector breeding companies is to accelerate the development of inbred lines. Cassava genotypes are heterozygous and very little inbreeding has been practised to date. However, inbred lines are better as parents as they do not have the confounding effect of dominance and carry lower levels of genetic load (undesirable alleles). Speed of inbreeding depends upon the average heterozygosity of the original parental lines, the homozygosity level of the selected genotypes at the end of the self-pollinating phase and the process of selection of progenies to be self-pollinated (Scotti et al., 2000). Basically in the inbreeding process two events go together: phenotypically there is a decrease in vigour, which is correlated with the increased levels of homozygosity. While the aim is to select vigorous plants (tolerant to inbreeding), in the process plants may be selected that are less homozygous than the expected average for their generation. It is expected that the first few cycles of self-pollination will result in a marked reduction of vigour (inbreeding depression associated with the genetic load of the parental lines); therefore, selection for
tolerance to inbreeding depression must be exerted. However, such selection is biased by the differences in homozygosity levels of segregating partially inbred genotypes. This highlights the need for a method to measure the level of heterozygosity in these partially inbred individuals and to use this in a co-variance correction in the selection of phenotypically vigorous genotypes. Molecular markers can be used to estimate the level of homozygosity of a given plant, enabling selection of plants with true tolerance to inbreeding.

Molecular markers can identify regions in the genome that are particularly related to the expression of heterosis and for measuring genetic distances among inbred lines to direct crosses with higher probabilities of high heterosis. Co-dominant SSR markers on a genome-wide basis are suitable for this purpose. The effect of self-pollination on vigour and heterozygosity was analysed in nine S1 families, heterozygosity being estimated in the S1 families by 100 mapped SSR markers that cover over 80 percent of the cassava genome and plant vigour by dry root yield and plant biomass. Results will assist in selecting the best performing and least heterozygous plants during inbreeding by identifying superior partially inbred parental lines. Molecular markers could also be used to delineate heterotic groups in cassava. Genetic resources of cassava have been characterized at the regional (Fregene et al., 2003) and global (Hurtado et al., 2005) levels. Highly differentiated groups of accessions were observed particularly among groups of materials from Guatemala and Africa and they may represent heterotic pools. These groupings are being tested based on molecular markers by genetic crossing between and within the groups as a first step to define heterotic patterns for a more systematic improvement of combining ability via recurrent reciprocal selection.

**Other potential MAS targets**

Several other traits for which MAS can be applied to increase efficiency of breeding include:

**Beta-carotene**

CIAT and a number of partners are involved in a project to produce cassava varieties with higher levels of β-carotene in yellow roots. This is one way of combating the deficiency of this key micronutrient in areas where cassava is a major staple. The experimental approach to increasing cassava β-carotene content includes conventional breeding and genetic transformation. The discovery of a wide segregation pattern of root colour in two S1 families from the Colombian landrace MCOL 72 (cross code AM 273) and MTAI 8 (AM 320) was the basis for molecular genetic analysis of β-carotene content in cassava. Three markers, SSRY251, NS980 and SSRY330, were found to be associated with β-carotene content. These are in the same region of the genome and together explain >80 percent of phenotypic variation for β-carotene content in the population used for this study. The homozygous state of certain alleles of these markers translates into higher β-carotene content, suggesting that breeding for this trait can benefit from molecular markers to assist in combining favourable alleles in breeding populations. The work is continuing with the search for additional favourable alleles in yellow-rooted germplasm to give the best possible phenotypic expression of the trait.

**Cyanogenic potential**

A collaborative project between the Swedish University of Agricultural Sciences (SLU), Uppsala, the Medical Biotechnology
Laboratories (MBL), Kampala, and CIAT, is aimed at the genetic mapping of CNP in cassava. An S1 family-AM 320, derived from the bitter variety MTAI 8 is the basis for the study. This family has been evaluated for cyanogenic glucoside content and has been genotyped with more than 200 diversity array technology (DarT) markers at CAMBIA, Australia, and 150 SSR markers at CIAT. The discovery of molecular markers for CNP will provide a tool to select efficiently for low cyanogenic potential in cassava. Also ongoing is the genetic mapping of the two cytochrome P450 genes CYP79D1 and D2 that catalyse the rate-limiting step of the biosynthesis of the cyanogenic glucosides, linamarin in the S1 family AM 320. The group is also looking for an association with QTL for CNP. It is expected that markers associated with CNP will be identified at the end of the study.

Dry matter content
Few key traits in cassava hold greater potential for increasing cost-effectiveness via MAS than root dry matter content (DMC). This trait is usually measured at the end of the growth cycle. A number of genetic and environmental effects influence DMC. It is usually highest before the onset of rains, but drops after the rains begin as the plant mobilizes starch from the roots for re-growth of leaves (Byrne, 1984). Defoliation from pest and disease attacks can lower DMC. Breeding programmes have been quite successful in improving DMC, especially for industrial markets. The entry point for developing markers associated with DMC was recent diallel experiments (Jaramillo et al., 2005; Calle et al., 2005; Pérez et al., 2005a, b; Cach et al., 2005b). Diallels, in this case made up of 90 families, are an ideal method to identify genes controlling DMC that are useful in many genetic backgrounds. Estimates of general and specific combining ability (SCA and GCA, respectively) for many traits of agronomic interest were calculated, with emphasis on DMC. Based on GCA estimates, parents were selected to generate larger-sized progenies for DMC mapping. Sizes of families in the original diallel experiment were about 30 progenies, which is rather small for genetic mapping. Parallel to the development of mapping populations was the search for markers associated with DMC using two F1 families, GM 312 and GM 313, selected from the diallel experiment having parents with high GCA for DMC.

Initial marker analysis using bulked segregant analysis led to the discovery of two molecular genetic markers, SSRY160 and SSRY150, which explain about 30 and 18 percent, respectively, of phenotypic variance for DMC. These markers are being analysed on approximately 700 genotypes derived from 23 crosses with parents having high GCA for DMC in order to confirm their utility across genetic backgrounds. Parallel to this, larger families are being developed from selected parents for QTL mapping of DMC.

Disadvantages of MAS
Perhaps the greatest disadvantage of MAS is the time and financial investment required to develop markers that are widely applicable for traits of agronomic importance. Often a marker developed in one or a few related genotypes will not work for other genotypes in a breeding scheme due to allelic effects. Furthermore, development of markers, particularly for QTL, is complicated by epistatic interactions and the critical need for good quality phenotypic data. Several ways around this
problem have been proposed, such as the use of candidate genes involved in the traits directly as selectable markers without the need for laborious gene tagging experiments. However, unravelling the genetics and the development of markers for such traits is still many years down the road. New methods of association mapping and linkage disequilibrium mapping that rely upon non-random association of candidate genes or markers on a high resolution map with a phenotype of interest in a non-structured collection of genotypes have been used extensively in human medicine to identify genes involved in disease (Cardon and Bell, 2001). Given the enormous difficulties of quantitative mapping in humans and the success of association mapping, these methods have also been proposed as ways around the problems in developing markers for low heritability traits in plants (Gaut and Long, 2003). The development of (partially) inbred cassava genetic stocks will certainly accelerate the application of MAS for the genetic improvement of the crop.

CONCLUSIONS
Given limited resources, further prioritization of traits is needed for the development of markers if they do not already exist. Top priority should be given to MAS for the most important pests and diseases prevalent in the region for which durable sources of resistance genes exist. Priority should also be given to DMC as this is another trait that, although having a high narrow sense heritability at the time of evaluation (usually after the onset of the rains to permit planting immediately thereafter), is significantly affected by non-genetic factors and is not as highly heritable. There are several initiatives to assist national programmes acquire new molecular tools to increase the cost-effectiveness of breeding. Prominent among these are the “molecular breeding communities of practice” project of the Generation Challenge Programme (GCP, www.generationcp.org) and the Rockefeller Foundation-funded African Molecular Marker Network (AMMANET, www.africancrops.net/ammanet). Both have training programmes on molecular breeding that are open to national programme scientists. The CIAT cassava project has also developed a Web-based database resource including protocols, populations, and markers for MAS in cassava that can easily be accessed by national programmes (www.ciat.cgiar.org/mascas).

Cassava and common beans: contrasts
Cassava and beans are similar with respect to the modest level of research input they have enjoyed over the past three to five decades. Both have been part of the research agenda of CIAT and of the CGIAR for nearly thirty years, and especially beans have benefited from inputs from laboratories and programmes in the United States of America and, to a lesser degree, Europe. However, research investments for high-scale genomics through marker development in these crops has been far less than for the “super crops” like maize, rice or soybean that enjoy participation by the private sector, but are more than minor orphan crops with local usage in the tropics.

Yet biologically, these two crops are widely contrasting. Cassava is a perennial versus beans, which are short-season annuals, although climbing beans at high altitudes can be similar to cassava in growth cycle. Beans are an autogamous seed crop while cassava is an allogamous crop with vegetative propagation. Accompanying this latter dichotomy are differences in gene action. Beans present largely additive gene
action, while cassava expresses important components of dominance and epistatic action. Finally, cassava as a clonal crop can fix heterotic combinations, while a lack of genetic male sterility or apomixis systems in common bean have curtailed the development of a hybrid industry for this seed crop even though heterosis is observed.

In spite of their biological and other differences, the results of several years experience with MAS in beans and cassava are surprisingly similar. In both crops, MAS is being employed principally to bolster phenotypic selection for disease resistance genes. Disease resistance is often governed by relatively few genes, and phenotypic data are obtained more easily. On the other hand, MAS for more complex traits has yet to find ready application. While there are candidates for such traits in both crops (root bulking in cassava; low phosphorus or drought tolerance in beans), the complexity of these traits has made the identification of reliable markers more difficult and has delayed application. Obtaining reliable phenotypic data for complex traits is especially difficult and is often the biggest bottleneck to eventual application of MAS. In the case of cassava, no inbred parents have been used to date for the development of molecular markers, making the genetic analysis more difficult.

However, some differences in the application of MAS for the two crops may be noted, arising from the form of reproduction of each crop. The time frame to select cassava clones through multilocalional trials is about six to seven years. During this period and with each step the number of genotypes is reduced as a result of the selection exerted, but the genotype of each individual clone remains stable. In the case of beans from the F₁ until stabilization of pure lines there is an intense segregation process in the early generations which tapers off in later generations. In both crops MAS can be used in the early stages of the selection process but with different objectives. In cassava, MAS can help to select early on the clone that will ultimately be released, whereas in beans MAS is used to “direct” the segregation process in the more desirable direction. Although maps with significant saturation are available for both crops, these have been constructed over several years, employing genotypes (in the case of beans) from different gene pools with wide polymorphism. A small proportion of these markers (often 20–30 percent) is polymorphic in other hybrid combinations among the genotypes within the same gene pool or race that have been created to tag a specific trait. Thus, genome coverage is often still not optimal for the high quality QTL analysis that is usually needed for complex traits.

**RECOMMENDATIONS**

**Careful prioritization of traits, marker system and genetic stocks for MAS**

The limited resources available for cassava or bean research require a judicious allocation of efforts. In the past 10–20 years there has been increased investment in molecular marker research in both crops. However, a considerable proportion of that research was directed at demonstrating the usefulness of different techniques, e.g. RAPD, restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), etc. Over this period there has been an ever-changing set of technologies but relatively little actual benefit derived from their application. There is a trade-off between being on the cutting edge with the newest technologies and “sticking it out” with an “outdated” technology
until some benefit is extracted from it. On the other hand, far too much effort has been expended in the identification of markers for traits without carrying these through to application. Often gene tagging is a component of a short-term project, and does not receive the necessary follow up in implementation. In each case, the essential question is: what are the key genes for each crop? And once defined, which genes merit the investment to develop molecular markers? For investments in molecular marker development to yield results, it is important that traits are chosen for which molecular breeding has both a clear advantage over field-based selection and is feasible in the short to medium term. It is also important that emphasis be given to selecting the relevant crosses, pedigrees and populations in which to practise MAS, and to have in place appropriate phenotyping strategies for the confirmation of MAS results. In this regard, the use of parental surveys of many of the genotypes involved in a given breeding programme is an important first step in implementing MAS.

**Short- and long-term research related to MAS**
The present research structure that is normally based on short-term projects, usually of three years’ duration, can seldom be expected to deliver results of usable markers for complex traits. Such short-term projects that seek to establish the basis for MAS or to implement selection should limit their objectives to simply inherited traits. On the other hand, longer-term funding either of a programmatic or successive project funding nature, must be obtained to address more complex traits governed by QTL as these would normally require at least two phases of three-year projects. The first phase might be expected to reveal the inheritance of a given trait, establishing the location and numbers of QTL, while a second phase would be required to validate these over more environments and to find markers that are polymorphic over a wide number of genotypes and therefore widely useful for breeding, as well as adapted to rapid laboratory techniques. A medium-to long-term investment likewise implies careful prioritization of such traits, with regard to potential impact and the eventual need for MAS. These reflections are based upon presently available laboratory techniques, but as techniques for more detailed and widespread evaluation of loci and genotypes are developed (e.g. gene chips for analysis of multiple loci), conclusions could change significantly.

**Scaling-up technologies**
After the development of molecular markers for a trait and their initial implementation, a period of scaling-up in use of the specific markers is necessary. Sometimes this involves changes to MAS protocols, in the marker detection technique or in the markers themselves. Marker re-design has been a common element of scaling-up exercises and can involve something as simple as changing a PCR fragment size to implementing a SNP assay for the actual sequence differences between alleles. Technologies that speed up the implementation process and lower the costs associated with scaling-up are crucial to the success of MAS and are often neglected.

**Development of markers that are useful in a large number of crosses**
Often a marker developed for a particular trait in one or a few related genotypes will not work for other genotypes with high value of the trait due to differences in gene or allelic effects. Unravelling the genetics of
major traits of agronomic interest even in a subset of elite parents used for breeding is beyond the resources available for bean and cassava research. Association mapping and linkage disequilibrium mapping, which rely upon non-random association of candidate genes or markers on a high resolution map with a phenotype of interest in a non-structured collection of genotypes, have been proposed as a way around this problem. Association mapping can be used to discover new marker-trait associations or to validate associations that were found through conventional genetic mapping. The GCP is facilitating association mapping of traits of agronomic importance in cassava and beans with the goal of discovering more useful markers for a wider range of genotypes.

The need to strike a balance between MAS and field-based selection
Occasionally the question is raised: which is better, MAS or conventional selection? This very question betrays a false dichotomy that hinders progress. By itself, MAS is seldom an adequate selection tool and therefore must be combined with conventional phenotypic selection. The objective should be to develop the optimal balance between conventional and molecular breeding, and the “best” balance will be unique to each situation, crop, selection scheme, environment and opportunities for different selection methods. More emphasis is needed on combined selection systems, rather than viewing MAS as a replacement for phenotypic or field selection.

REFERENCES


Chapter 7 – Marker-assisted selection in common beans and cassava


Marker-assisted selection in maize: current status, potential, limitations and perspectives from the private and public sectors

Michel Ragot and Michael Lee
**SUMMARY**

More than twenty-five years after the advent of DNA markers, marker-assisted selection (MAS) has become a routine component of some private maize breeding programmes. Line conversion has been one of the most productive applications of MAS in maize breeding, reducing time to market and resulting in countless numbers of commercial products. Recently, applications of MAS for forward breeding have been shown to increase significantly the rate of genetic gain when compared with conventional breeding. Costs associated with MAS are still very high. Further improvements in marker technologies, data handling and analysis, phenotyping and nursery operations are needed to realize the full benefits of MAS for private maize breeding programmes and to allow the transfer of proven approaches and protocols to public breeding programmes in developing countries.
INTRODUCTION
The ability to identify genetic components of traits, particularly quantitative traits, in Mendelian factors, and to monitor or direct their changes during breeding through the use of DNA-based markers has created much enthusiasm. Claims were sometimes made that marker-assisted selection (MAS) would rapidly replace phenotypic selection and dramatically reduce the time required to develop commercial varieties (Mazur, 1995). At the turn of this century, phenotypic selection was still the approach on which maize breeding programmes mostly relied to develop new and improved cultivars while MAS had contributed to advances in introgression, or backcross breeding (Ragot et al., 1995; Hå, McCouch and Smith, 2002; Ribaut, Jiang and Hoisington, 2002; Morris et al., 2003). Overly optimistic statements and exaggerated promises about the power of MAS to improve complex traits created excessively high and largely unfulfilled hopes and prompted a wave of cautious and sometimes pessimistic views (Melchinger, Utz and Schön, 1998; Young, 1999; Goodman and Carson, 2000; Bernardo, 2001).

Recently, multinational corporations with large maize breeding programmes reported the routine and successful use of MAS (Johnson, 2004; Niebur et al., 2004; Eathington, 2005; Crosbie et al., 2006). Rates of genetic gain twice as high as those achieved through conventional breeding were reported for MAS in maize. Accounts were also given of a number of MAS-derived single-cross (i.e. simple) hybrids being currently on the market. Although too little is known about the methods (e.g. breeding schemes, mathematical algorithms) and tools (e.g. marker technologies, computer programs, databases) used to develop these hybrids, these results have raised confidence in the ability of MAS to increase the rate of genetic gain over what can be achieved through conventional breeding. As technologies evolve and marker genotypes become less expensive, MAS becomes increasingly within the reach of developing countries. Whenever necessary, transfer of methods or tools from private companies to developing countries should be made possible while preserving the commercial interests of the companies concerned, thereby contributing to increasing the rate of genetic gain where it is most needed.

Much has happened in maize breeding since Stuber and Moll (1972) first reported that selection for grain yield in maize had resulted in changes in allele frequencies at several isozyme loci throughout the genome. In so doing, they essentially laid the grounds for MAS in maize. Indeed, if phenotypic selection could produce a change in marker allele frequencies, then why could deliberately altering marker allele frequencies at specific loci not produce predictable phenotypic changes for one or several traits?

The objectives of this chapter are to provide the scientific community and decision-makers with information on the current status of MAS in maize breeding programmes, including the major steps that led to it, and to provide suggestions to developing countries for deploying the technology and methods involved in an efficient, cost-effective and realistic manner.

HOW HAS MAS BEEN USED BY THE PRIVATE SECTOR TO IMPROVE THE MAIZE CROP?
Applications of DNA markers in private maize breeding programmes started in the 1980s with the identification of DNA clones used to detect restriction fragment length polymorphisms (RFLPs)
in the nuclear genome. As described below, the methods used to detect RFLPs were incompatible with the magnitude, speed and efficiency of all but a few aspects of selection in maize breeding programmes. Gradually, however, the methods used to detect DNA polymorphisms and to create meaningful information from DNA marker and phenotypic data sets have evolved to the point where they are routine components of some maize breeding programmes in the private sector.

Selection occurs at various stages in maize breeding programmes. The first opportunity arises when choosing inbred lines to mate as parents of new populations. In some programmes, all such inbreds are genotyped systematically at DNA marker loci (Smith and Smith, 1992). If the marker loci are sufficiently close on genetic or physical maps then reasonably good inferences may be made about the inbred’s haplotype. Such information is used to establish identity, resolve disagreements related to germplasm ownership and acquisition, enforce laws intended to encourage genetic diversity of the hybrids and avoid using inbreds that contain transgenes which may violate regulatory considerations and restrictions. These selection practices, while admittedly not conventional MAS, have led to improvements in the maize crop by enabling more informed stewardship and deployment of genetic resources and by providing a degree of protection of intellectual property and related investments in maize breeding.

Unquestionably, the most pervasive and direct use of MAS in maize by the private sector has been with backcrossing of transgenes into elite inbred lines, the direct parents of the commercial hybrids (Ragot et al., 1995; Crosbie et al., 2006). Currently, the most widely deployed transgenes and combinations thereof (i.e. gene stacks) are for resistance to herbicides or insects (e.g. Ostrinia and Diabrotica). As the commercial maize crop of any region, maturity zone, market or country is not yet uniform or homogeneous for any transgene, maize breeders have elected to develop near-isogenic versions (transgenic and non-transgenic) of elite inbreds and commercial hybrids in order to satisfy combinations of licensing agreements, agronomic practices, regulatory requirements, market demands and product development schemes. This has required companies to have two parallel maize breeding programmes, transgenic and non-transgenic. In this manner, marker-assisted backcrossing (MABC) of transgenes, and to a lesser degree, of native genes and quantitative trait loci (QTL) for other traits, has expedited the development of commercial hybrids.

More recently, marker-assisted recurrent selection (MARS) schemes and infrastructure have been developed for “forward breeding” of native genes and QTL for relatively complex traits such as disease resistance, abiotic stress tolerance and grain yield (Ribaut and Betrán, 1999; Ragot et al., 2000; Ribaut, Jiang and Hoisington, 2000; Eathington, 2005; Crosbie et al., 2006). Simulation studies suggested that MAS could be effective for such traits under certain conditions (Edwards and Page, 1994; Gimelfarb and Lande, 1994), but the initial empirical attempts at such selection were not successful (Stromberg, Dudley and Rufener, 1994; Openshaw and Frascaroli, 1997; Holland, 2004; Moreau, Charcosset and Gallais, 2004) except in the special case of sweetcorn (Edwards and Johnson, 1994; Yousef and Juvik, 2001). The success reported for sweetcorn is due to the fact that the genetic base of sweetcorn is extremely narrow relative to dent or flint maize; thus
predicted gains and extrapolations across populations are more reliable. Also, phenotypic analyses of many traits in a sweetcorn breeding programme are extremely expensive because they involve processing large volumes of grain; therefore, MAS would be relatively inexpensive and effective under such circumstances. However, subsequent developments in technology, refinements in analytical methods and improvements in experimental designs have been assembled into a process that has shown promise for some reference populations of dent maize (Ragot et al., 2000; Johnson, 2004; Crosbie et al., 2006) as improvement in grain yield from MAS often exceeded that from non-MAS approaches. Presumably, such results will lead to the development of new and superior inbred lines and commercial hybrids in a cost-effective manner. While the impact of such MAS has not yet been fully realized in the maize crop, the methods have been employed to various degrees by programmes in the private sector that have the necessary infrastructure.

The potential for MAS to contribute to improvements in the maize crop should increase in parallel with our understanding of the relationships among genomes, the environment and phenotypes. Candidate transgenes will be developed on a regular basis and their contributions to maize improvement will be realized in the most efficient manner with MAS. Likewise, the identification of candidate native genes and their gene products and functions, and of other DNA sequences (e.g. miRNA, matrix attachment and regulatory regions), will improve the power of methods such as association mapping and genome scans to assess their genotypic value in the context of defined reference populations of significance to maize breeding (Thornsberry et al., 2001; Rafalski, 2002; Niebur et al., 2004; Varshney, Graner and Sorrels, 2005). Beyond its use in MARS schemes, this information might make it reasonable to reconsider ideas such as methods for predicting hybrid performance that may have been limited by the amount and type of information and by the design of the experiment when they were initially evaluated (Bernardo, 1994).

**METHODODOLOGY AND DESIGN OF BREEDING PROGRAMMES SUPPORTED BY MAS**

As expected, private sector maize programmes focus entirely on inbred-hybrid breeding schemes intended to develop elite inbred lines that enable the profitable production of commercial F$_1$ hybrids. To a large extent, MAS breeding programmes use the same designs and methods known to maize breeders for decades and generic descriptions of these have been published (Hallauer and Miranda, 1981; Sprague and Dudley, 1988; Bernardo, 2002). When MAS is included in the breeding programme, the significant differences are, of course, the availability of genotypic data at different stages of selection and some knowledge of the relationships between the genotypic and phenotypic data sets for the reference population(s) in the target environment(s).

In contrast to conventional breeding schemes, the methods and design of infrastructure needed to support MAS have been the areas of greatest change. In order to utilize MAS, companies had to make significant investments to assemble or modify various aspects of infrastructure such as methods to detect DNA polymorphism, manage information, or analyse and track samples, software to relate genotype with phenotype, and off-season or continuous nurseries. These components had to be integrated with each
other and with breeding activities, which meant that scientists needed to learn how and when MAS provided a comparative advantage over other methods.

**MAS: enabling methods, tools and infrastructure**

Perhaps the component of infrastructure in greatest need of development was related to the acquisition of genotypic data (i.e. DNA markers). Although the concept of associating markers with quantitative traits was not new (Sax, 1923), the discovery reported by Stuber and Moll (1972) was very significant. Stuber and Moll (1972) described for the first time associations between molecular markers and quantitative traits while previous associations had been based on morphological markers (Sax, 1923). The advantages of molecular over morphological markers soon became obvious and detailed descriptions of these advantages were published by Tanksley et al. (1989) and Stuber (1992).

Two of these advantages are of particular importance. First, molecular marker genotypes can usually be obtained from any plant tissue, even from young seedlings or kernels, while morphological markers frequently require the observation of whole, mature plants. Selection can therefore occur earlier in the plant's cycle when using molecular markers than when using morphological markers. The ability to conduct early selection, possibly before flowering, can have a tremendous impact on the rate of genetic gain of a breeding programme and therefore constitutes a very significant advantage of molecular over morphological markers.

Second, molecular markers are neutral markers. They are not affected by environmental or growing conditions. They are not affected by the genetic background either, nor do they affect phenotypes. The expression of morphological traits, by contrast, can be dependent on environmental or growing conditions. In addition, epistatic interactions are often observed among morphological marker loci or between morphological marker loci and the genetic background. These epistatic interactions prevent distinguishing all genotypes associated with morphological markers and further limit the number of morphological markers that can be studied simultaneously.

Although isozyme markers had many advantages over morphological markers, the lack of a sufficient number of polymorphic loci limited their use for MAS (Goodman et al., 1980). Nevertheless, isozyme markers are still used for quality control during seed production.

RFLPs (Botstein et al., 1980) are based on DNA polymorphisms detected through restriction nuclease digestions followed by DNA blot hybridizations. The abundance and high level of polymorphism of RFLPs, especially in maize, allowed the construction of extensive maize genetic maps (Helentjaris et al., 1986; Burr et al., 1988; Hoisington, 1989; Coe et al., 1995; Davis et al., 1999) as well as the identification and mapping of many QTL.

Being robust, reproducible and co-dominant, RFLPs are perfectly suited for genetic studies as well as for MAS applications. Their two main disadvantages are the large quantities of DNA required, and the difficulty to miniaturize and automate. Nevertheless, RFLPs were quickly adopted and represented the marker system of choice for many plant species including maize throughout the 1980s and during much of the 1990s.

The development of the polymerase chain reaction (PCR) (Saiki et al., 1988) turned out to be a major breakthrough in
molecular marker technology. PCR-based markers require little DNA, allowing sampling of young seedlings and very early selection and thereby optimization of breeding schemes. PCR-based marker protocols are very amenable to automation and miniaturization and improvements to protocols resulted in considerable reductions in both cost and time required to produce data points. The first two PCR-based marker systems were random amplified polymorphic DNA (RAPDs), and amplified fragment length polymorphisms (AFLPs). Detailed descriptions and critical assessments of these two systems can be found in Welsh and McClelland (1990), Williams et al. (1990), Penner et al. (1993), Ragot and Hoisington (1993), Skroch and Nienhuis (1995) and Jones et al. (1997) for RAPDs, and in Vos et al. (1995), Jones et al. (1997) and Castiglioni et al. (1999) for AFLPs. They are also described in other chapters of this book.

Simple sequence repeats (SSRs) or microsatellites rapidly became the marker of choice in maize, almost entirely displacing RFLPs and previously developed PCR-based marker systems. Polymorphism of SSRs is due to variable numbers of short tandem repeats, often two or three base pairs in length and usually flanked by unique regions (Tautz, 1989). SSRs are very reproducible (Jones et al., 1997) and co-dominant (Shattuck-Eidens et al., 1990; Senior and Heun, 1993; Senior et al., 1996) and are therefore very suitable for maize MAS applications.

Many additional variations of PCR-based marker systems have been developed and a thorough review can be found in Mohan et al. (1997).

All the DNA-based marker systems described to date are gel-based systems, a major constraint for automation. Single nucleotide polymorphisms (SNPs) (Lindblad-Toh et al., 2000) can be revealed in many ways including allele-specific PCR, primer extension approaches, or DNA chips, all of which are not gel-based. SNPs can generally be scored as co-dominant markers, except in the case of insertion-deletion polymorphisms. Although allelic diversity at SNPs is usually limited to two alleles, this limitation can be offset by the abundance of SNPs and the analysis of haplotypes, combinations of genotypes at several neighbouring SNPs. Haplotypes analyses increase informativeness (Ching et al., 2002), although at some expense because two to four SNPs have to be genotyped where one SSR sufficed. SNP genotyping can be highly miniaturized and automated, thereby reducing the cost and allowing the production of very large numbers of data points. With genetic maps containing several thousand mapped SNPs, these have become the marker of choice for private maize MAS programmes.

DNA marker technology has been a dynamic and often expensive component of the infrastructure needed for MAS. For example, one corporation indicated having spent tens of millions of United States dollars to develop an automated system for detecting RAPDs, a technology that was never suited for MAS in a large maize breeding programme. Later, another corporation spent an even greater amount of money to acquire technology for matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) analysis of amplified DNA fragments. These technologies were either rapidly replaced or never used. Such decisions would have bankrupted most national maize programmes or a couple of centres belonging to the Consultative Group on International Agricultural Research (CGIAR). Fortunately, this area of infrastructure has matured somewhat.
and become more stable so that start-up and operating costs, while still high for some programmes, are more predictable.

Statistical methods and related software have also been areas of significant development, especially for the detection and description of putative QTL. QTL, which are nothing more than associations between markers and traits, were first described using simple association tests between trait values and marker genotypes (Stuber and Moll, 1972). These tests consider each marker locus independently and neither require nor take advantage of the existence of genetic maps. Statistical methods have been developed that take advantage of the existence of genetic maps (see review by Manly and Olson, 1999). These statistical methods, simple interval mapping (Lander and Botstein, 1989) and composite interval mapping (Jansen, 1993; Zeng, 1993, 1994), test the existence of associations between hypothetical marker genotypes and trait values at several points in intervals between pairs of adjacent marker loci on the genetic map, allowing the positioning of QTL on these genetic maps. All of the previous methods are based on single QTL models. Other statistical methods have been developed that simultaneously test the presence of several QTL in the genome (Kao, Zeng and Teasdale, 1999).

Many software packages are available for QTL mapping and based on one or several of the statistical methods developed to date. No two packages are exactly alike and all have specific strengths and weaknesses with respect to particular situations, making it sometimes beneficial to use more than one package to perform QTL mapping analyses. The software packages most commonly used for QTL mapping in maize include QTL Cartographer (Basten, Weir and Zeng, 1994), MapQTL (van Ooijen and Maliepaard, 1996), and PLABQTL (Utz and Melchinger, 1996). All of these only handle bi-allelic populations, while MCQTL (Jourjon et al., 2005) also performs QTL mapping in multi-allelic situations, including bi-parental populations made from segregating parents, or sets of bi-parental, bi-allelic populations.

More recently, methods based on Bayesian analysis (Jansen, Jannink and Beavis, 2003; Gelman et al., 2004) and association (Varshney, Graner and Sorrels, 2005) or in silico mapping (Parisseaux and Bernardo, 2004) have been proposed as more powerful and refined approaches to assess the relationships between genotype and phenotype that are needed for MAS. Methods of Bayesian analysis should be less affected by the uncertainties of QTL effects and locations and produce better estimates of those parameters in MAS. Association mapping approaches are particularly useful to validate the relevance of genes and alleles in specific germplasm such as that used by maize breeders. In silico mapping takes advantage of the pedigree relationships among individuals to structure the population used to establish marker-trait associations. This approach, which is highly complex due to the population structure resulting from pedigree breeding, is particularly appropriate for maize where data across many years and environments are available for large sets of related individuals. Certainly, as the annotation of genomes gradually improves, such methods will be common components of breeding programmes. Currently, the applications of methods such as association mapping for MAS are hindered by the fact that a very low percentage of the genes in crop plants have a function assigned to them on the basis of direct experimentation. However, this impoverished situation
is being enriched through a variety of projects on functional genomics.

In sharp contrast to the many methods and software packages developed for QTL identification and mapping, little has been published for MAS. This paucity of information on MAS tools most likely reflects both the low level of activity in the public sector and the fully proprietary nature of developments in the private sector.

In parallel with advancements in DNA technology and statistical methods, private sector programmes have enhanced the capabilities and capacities of their continuous nurseries. Such nurseries have been used for decades by programmes in both the private and public sectors. In order to conduct MAS to its greatest advantage, continuous nurseries had to be managed, equipped and staffed in new ways so that the plants complete their life cycle as quickly as possible and that the genotypic data (and sometimes some phenotypic data) needed for MAS may be collected at each sexual generation. Three to four sexual generations per year may be completed at such nurseries.

These activities and the continuous collection of both genotypic and phenotypic data in the target environment and their integrated analyses create huge data sets that must be analysed quickly and related to other large extant data sets. Data management and bioinformatics for breeders have therefore become critical components of the infrastructure needed to use MAS. Prior to the advent of MAS, some large private breeding programmes had established a group of dedicated data managers to assist with research and marketing, and with the arrival of genomics and MAS the need for such dedicated specialists has increased greatly.

Once the basic infrastructure had been established to complement the activities of maize breeders, programmes were ready to implement several basic aspects of MAS; many of which are derived from well established methods and principles of maize breeding.

**MAS-based breeding**

Selection occurs at various stages in maize breeding programmes. The first opportunity for selection is the choice of inbred lines to mate as parents of new populations. Prior to the advent of DNA marker data, the selection of such parents would be based on a combination of phenotypic assessments, pedigree information, breeding records and chance (Hallauer and Miranda, 1981; Sprague and Dudley, 1988). In some programmes today, all such inbreds are genotyped systematically at DNA marker loci. Depending on the resources and objectives, the degree of genotyping may range from a low density of marker loci (e.g. SNPs in candidate genes) to higher density whole genome scans (Varshney, Graner and Sorrels, 2005). These genotypic data, alone or integrated with phenotypic information, may reveal novel aspects of maize gene pools, heterotic groups, haplotype evolution, gene content and parents used in MAS for specific target environments (Fu and Dooner, 2002; Niebur *et al.*, 2004; Crosbie *et al.*, 2006). When properly integrated with phenotypic information and functional genomics, genotypic data of inbred lines should allow breeders to choose parents that, when mated, should provide populations or gene pools enriched for the more desirable combinations of favourable alleles. Such a starting point is a huge advantage in plant breeding because it increases the probability of selecting progeny that are superior to the parents and that approximate a predicted optimum genotype.
MABC is certainly the form of MAS with the most immediate and obvious benefits for maize breeding. MABC is used for three main purposes: selection of transgenes (or of native DNA sequences of the maize genome, whether genes or QTL), elimination of unwanted regions of the donor-parent genome linked to the transgene and selection of unlinked regions of the recurrent-parent genome. With the exception of DNA markers and transgenes, these have been the same goals of backcross breeding since the inception of that method decades ago (Fehr, 1987). Of course, DNA markers enable breeders to identify progeny that contain the desired recombinant chromosomes and donor-parent genome in a more direct manner. Also, MABC facilitates the process of combining more than one transgene in a given inbred line (e.g. “gene or trait stacking or pyramiding”). This reduces the number of generations needed to reach certain stages of a breeding programme and reduces the time needed to produce commercial hybrids for the market.

Generic MABC schemes suitable for maize breeding programmes have been described in detail for single genes (Hospital, Chevalet and Mulsant, 1992; Ragot et al., 1995; Frisch, Bohn and Melchinger, 1999a, 1999b; Frisch and Melchinger, 2001a; Hospital, 2001; Ribaut, Jiang and Hoisington, 2002), for QTL (Hospital and Charcosset, 1997; Bouchez et al., 2002) and for gene stacks (Frisch and Melchinger, 2001b). Versions of such schemes have been used in maize breeding programmes in the private sector, often at their continuous nurseries (Ragot et al., 1995). Most recently, MABC has also been adopted as a tool to develop sets of near-isogenic lines (NILs) for genomics research (Peleman and van der Voort, 2003).

Theoretical and simulation studies have been conducted to identify the most efficient MABC protocols. Parameters most commonly studied include the number of individuals genotyped at each generation, the number of markers used, relative selection pressure for recombination around the target locus or global recovery of recurrent parent genome and the number of individuals selected at any generation. Optimal values for each of the above depend on the objective of the MABC approach in terms of quality (required level of recurrent parent genome recovery), speed (fastest possible conversion or set number of generations) and resources (unlimited or limited). While the fastest and highest quality MABC approaches have the most expensive protocols, less intensive approaches can result in significant time savings and quality improvements when compared with conventional backcrossing approaches and at a fraction of the cost of the most expensive MABC protocols.

Frisch, Bohn and Melchinger (1999b) showed that to minimize linkage drag around the target locus (loci), selection of recombination events close to the target locus (loci) should be conducted in the early backcross generations. Frisch and Melchinger (2001a) and Ribaut, Jiang and Hoisington (2002) further demonstrated that minimizing linkage drag around the target locus requires very large numbers of individuals (possibly hundreds) to be genotyped. Hospital and Charcosset (1997) proposed a selection scheme based on selecting a single individual to be backcrossed. By contrast, Frisch and Melchinger (2001b) proposed selecting several individuals and determining the family size of their backcross progeny based on the individuals’ genotypes. By using varying rather than constant numbers of individuals or markers at the different backcross generations, it was shown that the number...
of marker data points required could be reduced and thus the efficiency of MABC improved (Hospital, Chevalet and Mulsant, 1992; Frisch, Bohn and Melchinger, 1999b). Several studies also showed that using a limited number of markers on non-carrier chromosomes was sufficient to recover more that 95 percent of the recurrent parent genome in three or fewer backcross generations (Hospital, Chevalet and Mulsant, 1992; Visscher, Haley and Thompson, 1996; Servin and Hospital, 2002).

One of the most important lessons from the various theoretical and simulation studies of MABC is that the effects of the different MABC parameters are not independent of each other. With maize, large backcross populations can be generated from a single plant when that plant is used as the male and recurrent parent plants are used as females. Marker systems in maize are also such that very large amounts of marker data can be generated on plants before flowering. Potential MABC protocols are almost endless in maize and identifying the most efficient is only possible on a case-by-case basis. For example, while achieving almost complete recovery of the recurrent parent’s genome is necessary for registering backcross-derived lines and hybrids in many European countries, partial recovery might be sufficient to improve the agronomic performance of varieties in developing countries. The optimal MABC protocols for these two strikingly different objectives will be very different. Protocols for the first objective will involve background selection and the use of background markers very close to the target locus (loci). Protocols for the second objective might involve markers for the target locus (loci) only, while relying on successive backcross generations to recover an adequate amount of recurrent parent genome.

Successful examples of MABC in maize include backcrossing of transgenes (Ragot et al., 1995), and QTL for insect resistance (Willcox et al., 2002), flowering maturity (Ragot et al., 2000; Bouchez et al., 2002) and grain yield (Ho, McCouch and Smith, 2002).

Methods of “forward breeding” with DNA markers have also been proposed and implemented by maize breeding programmes. As with the pedigree-based methods of maize breeding favoured by the private sector, many of the “new” methods that utilize genetic data from DNA markers integrated with phenotypic data are essentially a form of recurrent selection, a method that has been in use for several decades (Hallauer and Miranda, 1981). The key advantages of the new versions of recurrent selection are, of course, the availability of genetic data for all progeny at each generation of selection, the integration of genotypic and phenotypic data, and the rapid cycling of generations of selection and information-directed matings at continuous nurseries.

At least two distinct forms of forward breeding with MAS have been proposed: single large-scale MAS (SLS-MAS) (Ribaut and Betrán, 1999) and MARS (Edwards and Johnson, 1994; Lee, 1995; Stam, 1995). A key difference between the methods is that SLS-MAS employs DNA markers at only one generation and attempts to retain genetic variation in regions of the genome unlinked to the DNA markers, while MARS uses markers at each generation, exhausting genetic variation in most regions of the genome. Versions of both SLS-MAS and MARS have been used by breeding programmes in the private sector (Johnson, 2004; Eathington, 2005; Crosbie et al., 2006).

SLS-MAS is of particular interest in pedigree breeding as it consists of screening
and selecting individuals at a few loci at early generations, usually F2 or F3, (Eathington, Dudley and Rufener, 1997), using large populations (Ribaut and Betrán, 1999). Individuals displaying homozygous favourable genotypes at the loci of interest are selected and self-pollinated while others are discarded. Self-pollinated progeny of the selected plants then proceed normally through subsequent steps of pedigree breeding. Screening large populations is necessary to ensure that genetic diversity is maintained at regions not under genotypic selection, thereby allowing further phenotypic selection to be conducted. Loci at which marker selection operates can be QTL as described by Ribaut and Betrán (1999). SLS-MAS is thus limited by issues such as the precision of the QTL parameters (position, effect), and relevance of the QTL across environments or gene pools. SLS-MAS can also be conducted for genes, eliminating many of the limitations pertaining to QTL. Although a powerful approach adopted in several species (barley, soybean, sunflower, wheat) to enrich breeding populations at a few loci (Crosbie et al., 2006), SLS-MAS does not appear to have been widely implemented in maize breeding programmes.

MARS targets all traits of importance in a breeding programme and for which genetic information can be obtained. Genetic information is usually obtained from QTL analyses performed on experimental populations and comes in the form of maps of QTL with their corresponding effects. If the QTL mapping analysis is conducted based on a bi-parental population, the sign of the effect at each QTL indicates which of the two parents carried the favourable allele at that QTL. As both parents often contribute favourable alleles, the ideal genotype is a mosaic of chromosomal segments from the two parents. This assumes that the goal is to obtain individuals with as many accumulated favourable alleles as possible, a different goal from that of marker-assisted population improvement as studied elsewhere (Lande and Thompson, 1990; Gimelfarb and Lande, 1994; Gallais, Dillmann and Hospital, 1997; Hospital, Chevalet and Mulsant, 1997; Knapp, 1998; Moreau et al., 1998; Xie and Xu, 1998). Population improvement schemes are generally based on the random mating of selected individuals while the scheme proposed here is based on directed recombination between specific individuals. As reported by Stam (1995), the ideal genotype, defined as the mosaic of favourable chromosomal segments from two parents, will usually never occur in any Fn population of realistic size. It is, however, possible to design a breeding scheme to produce or approach this ideal genotype based on individuals of the experimental population. This breeding scheme could involve several successive generations of crossing individuals (Stam, 1995; Peleman and van der Voort, 2003) and would therefore constitute what is referred to as MARS or genotype construction. This idea can be extended to situations where favourable alleles come from more than two parents (Stam, 1995; Peleman and van der Voort, 2003).

Van Berloo and Stam (1998, 2001) and Charmet et al. (1999) developed computer simulations around this idea and assessed the relative merits of marker-assisted genotype construction over phenotypic selection. MARS was simulated in an experimental population where QTL had been mapped. Index (genetic) values were computed for each individual based on its genotypes at QTL-flanking markers (van Berloo and Stam, 1998, 2001). All simulation studies of MARS found that
it was generally superior to phenotypic selection in accumulating favourable alleles in one individual (van Berloo and Stam, 1998, 2001; Charmet et al., 1999). MARS appeared to take better advantage of the genetic diversity present in the populations to which it was applied than phenotypic selection. Simulation research conducted by van Berloo and Stam (2001) showed that MARS was between 3 and almost 20 percent more efficient than phenotypic selection. The advantage of MARS over phenotypic selection was greater when the population under selection was larger or more heterozygous (BC1s or F2s vs. RILs, recombinant inbred lines, or DHs, doubled haploids). Although van Berloo and Stam (2001) limited their simulations to populations of up to 200 individuals, their results seem to indicate that the relative advantage of marker-assisted over phenotypic selection would keep increasing as population size increased. The same simulation studies showed that the advantage of marker-assisted over phenotypic selection was larger when dominant QTL were involved in the selection index, or when trait heritability was low in the case of selection for a single trait (van Berloo and Stam, 1998, 2001). These latter observations are of little relevance to most commercial maize breeding programmes, the goal of which is generally the development of inbred lines improved for several traits that will be later combined into superior hybrid varieties. They should, however, increase the appeal of MARS approaches for breeding programmes aimed at developing open-pollinated varieties.

Simulation have also addressed the impact of the amount and quality of QTL information on selection efficiency. Simulation and empirical studies (Beavis, 1994, 1999) showed that QTL mapping experiments based on segregating populations of less than 500 individuals generally revealed only a subset of all QTL affecting the complex traits segregating in these populations. Quantitative trait loci information used in subsequent MARS was therefore necessarily incomplete. Van Berloo and Stam (2001) showed that the relative advantage of MARS over phenotypic selection decreased rapidly when the fraction of the total genotypic variance explained by the QTL included in the selection index decreased. By contrast (van Berloo and Stam, 1998; Charmet et al., 1999), the efficiency of MARS seems to be rather robust to the well-documented (Lee, 1995) uncertainty of QTL genetic locations. The use of genotypic information at markers flanking the QTL possibly explains this observation.

The cost efficiency of MARS was also investigated through simulation (Moreau et al., 2000; Xie and Xu, 1998). When simulating selection for a single trait, Moreau et al. (2000) found that, irrespective of the heritability of the trait, MARS was always more cost efficient than phenotypic selection if the cost of genotyping was less than that of evaluating one individual in one plot. When simulating simultaneous selection for multiple traits, Xie and Xu (1998) found that MARS was more cost efficient than phenotypic selection if the cost of genotyping was less than that of phenotyping one individual for all traits. These studies were based on a single generation of MARS. Also, they did not take into consideration any factors besides genotyping and phenotyping costs, although factors influencing the length of a selection cycle or the number of cycles that can be completed in a year can obviously affect the relative economic merits of marker-assisted and phenotypic selection.
In contrast to the abundance of QTL mapping reports, very few accounts of MARS experiments are found in the literature. Moreau, Charcosset and Gallais (2004) compared phenotypic, marker-only, and combined recurrent selection for grain yield and grain moisture at harvest over several cycles and years in maize. Combined selection was based both on phenotypic and marker information while marker-only selection was based on marker information only. Both the marker-only and the combined selection methods constitute MARS approaches. Several combinations of these three methods of selection were applied to the segregating population that served to map the QTL used in marker-based selection indices. Over the six years of the experiment, two cycles of phenotypic selection, two cycles of combined selection, one cycle of combined selection followed by two cycles of marker-only selection, and one cycle of marker-only selection were conducted in parallel. A reassessment of the positions and effects of QTL was conducted after the first cycle for the three schemes containing multiple cycles. All MARS methods were more efficient than phenotypic selection to increase the frequency of favourable alleles at QTL. Nevertheless, Moreau, Charcosset and Gallais (2004) reported no significant difference between marker-assisted and phenotypic selection on the multitrait performance index, although all MARS methods resulted in genetic gain for both grain yield and grain moisture while phenotypic selection resulted in genetic gain for grain yield but an unfavourable evolution of grain moisture. This disappointing result was tentatively explained by the high heritability of the traits, favourable to phenotypic selection, while the percentage of total phenotypic variance explained by the QTL detected for both traits was only about 50 percent. One very encouraging result of this experiment, although Moreau, Charcosset and Gallais (2004) failed to present it as such, was that the first cycle of marker-only selection was as efficient as phenotypic or combined selection in delivering genetic gain. Two conclusions can be drawn from this observation. First, the QTL identified in the initial experimental population were in general not artefacts. Second, selection pressure applied at these QTL, and aimed at fixing alleles identified as favourable, resulted in a change in performance of the selected population in the desired direction when compared with the initial population.

A similar experiment, although based solely on marker-only recurrent selection, was reported by Openshaw and Frascaroli (1997). They conducted MARS in maize simultaneously for four traits, for each of which about ten QTL had been identified. They showed that genetic gain had been achieved in the first cycle of MARS, but that later cycles did not result in any gain. Possible explanations given for these results included uncertainties about QTL parameters (location and effect), interaction effects (epistasis, genetic x environment interaction), and the fact that selection was based on single markers rather than chromosomal segments (Openshaw and Frascaroli, 1997).

Recent communications from several private MARS research programmes (Ragot et al., 2000; Eathington, 2005; Crosbie et al., 2006) revealed large-scale successful applications in maize. Accounts were given of commercial maize hybrids for which at least one of the parental lines was derived through MARS. Eathington (2005) and Crosbie et al. (2006) reported that the rates of genetic gain achieved through MARS were about twice those
of phenotypic selection in some reference populations. Marker-only recurrent selection schemes have been implemented for a variety of traits including grain yield and grain moisture (Eathington, 2005), or abiotic stress tolerance (Ragot et al., 2000), and multiple traits are being targeted simultaneously. Selection indices were apparently based on 10 to probably more than 50 loci, these being either QTL identified in the experimental population where MARS was being initiated, QTL identified in other populations, or genes. Marker genotypes are generated for all markers flanking QTL included in the selection indices (Ragot et al., 2000). Plants are genotyped at each cycle and specific combinations of plants are selected for crossing, as proposed by van Berloo and Stam (1998). Several, probably three to four, cycles of MARS are conducted per year using continuous nurseries. In maize, early versions of such schemes have been tested and implemented (Johnson, 2004; Crosbie et al., 2006).

Results reported in these recent communications about private MARS experiments (Ragot et al., 2000; Eathington, 2005) are in sharp contrast to those in earlier publications (Openshaw and Frascaroli, 1997; Moreau, Charcosset and Gallais, 2004). Several factors can explain these discrepancies:

- **Size of the populations submitted to selection at each cycle.** Given reports that increasing population size should result in higher genetic gain through MARS (van Berloo and Stam, 2001,) it is likely that populations submitted to selection in private programmes are rather large, larger than the 160 and 300 individuals reported respectively by Openshaw and Frascaroli (1997) and Moreau, Charcosset and Gallais (2004).

- **Use of flanking versus single markers.** The use of flanking markers for QTL under selection allows better prediction of the genotype at the QTL than when using single markers. When single markers are used, recombination events that occur between the marker and the QTL lead to loss of linkage between the marker and the QTL much faster than when flanking markers are used, thereby rapidly reducing the predictive power of the single marker.

- **Early selection, pre-flowering.** The ability to select plants before flowering ensures optimal mating schemes as the genotypes of plants being selfed or intercrossed are fully known. However, this is not the case when selection cannot take place before flowering and involves intercrossing selfed progenies of selected plants, the genotypes of which might have drifted significantly from those of their genotyped parents.

- **Number of generations per year.** To the authors’ knowledge, none of the simulation or experimental studies of MARS has assessed the effects of cycle length on its efficiency despite its direct relationship to the rate of genetic gain. In maize, cycle length can be reduced three- to six-fold when using marker-only recurrent selection compared with phenotypic recurrent selection. Consequently, marker-only recurrent selection will be superior to phenotypic selection as soon as the genetic gain achieved through one cycle of MARS is, respectively, more than a third or a sixth of that achieved through one cycle of phenotypic selection. Private maize breeding programmes have access to off-season nurseries. Furthermore, they have often established efficient continuous nurseries where three to four generations of maize can be grown per year. The use of such nurseries allows them to carry MARS continuously, i.e.
with up to four cycles per year, whereas phenotypic recurrent selection is limited to one cycle per year at most. The impact on the rate of genetic gain of such an implementation of MARS might be very positive even if MARS did not present any advantage over phenotypic selection on a per-cycle basis.

- **Cost of marker data points.** Large private companies have made considerable efforts to reduce both the cost of marker data points and the cost of experimental field plots. The ratio of cost of marker data point to cost of experimental field plot is most likely lower in large private breeding programmes than in most public research laboratories or small private programmes, potentially leading to different views on the economic efficiency of MARS.

  Marker-based and phenotypic selection can be mobilized in many different ways, with respect to each other, in marker-assisted breeding schemes. Marker and phenotypic information can be used either simultaneously or sequentially. Selection of parents for breeding populations can be made using marker information alone, phenotypic information alone, or a combination of each. Selection of individuals in a backcross programme can be made on the sole basis of either marker or phenotypic information, or using both. Advancement of individuals in a line development programme can also be made at each generation on the basis of either marker information only, phenotypic information only, or a combination of each. In order to maximize the rate of genetic gain it is likely that MAS breeding schemes such as MABC and MARS will involve generations of marker-only selection conducted at continuous nurseries. The advent of improved methods of producing doubled haploids will certainly further influence the way marker-based and phenotypic selection are mobilized with respect to each other.

  In spite of the development of marker-only selection and regardless of the underlying technology and breeding scheme, high-quality phenotyping remains vital and without substitute at several stages; but it may become more focused. Phenotypic evaluation remains the ultimate screen before any cultivar is released. MAS-derived lines and hybrids that meet phenotypic requirements are selected for further evaluation and selection on the basis of their phenotypic value, while those that do not are discarded. Phenotypic evaluation is also critical to establish marker-trait associations or perform the candidate gene validations required to conduct MAS. Here, high quality phenotyping is necessary. Phenotyping protocols will therefore likely be different from those commonly used for phenotypic selection. Experiments may be conducted that involve side-by-side comparisons of different treatments such as water stress or nitrogen fertilization levels to dissect complex traits into their components and facilitate the elucidation of their genetic basis.

  Enhancements of such approaches to maize breeding will be based on the incorporation of improved methods of producing doubled haploid inbred lines, information from functional genomics and by learning how to incorporate favourable native genetic variation systematically after MAS has reduced the genetic variation in the original reference populations to unacceptable levels.

**ADVANTAGES AND LIMITATIONS OF MAS IN MAIZE BREEDING PROGRAMMES**

**Advantages of MAS**

For private breeding programmes, MAS has offered several attractive features, most
of which are related to time and resource allocations.

MABC clearly provides the information needed to reduce the number of generations of backcrossing, to combine (i.e. “stack”) transgenes, “native” genes or QTL into one inbred or hybrid quickly, and to maximize the recovery of the recurrent parent’s genome in the backcross-derived progeny. In several private breeding programmes, MABC has enabled the number of backcrossing generations needed to recover 99 percent of the recurrent parent genome to be reduced from six to three, reducing the time needed to develop a converted variety by one year (Crosbie et al., 2006; Ragot et al., 1995). As a line derived by MABC can be made to be very similar to the original non-converted line, most of its attributes, including agronomic performance, can be assumed to be equal or similar to those of the original line. Only limited phenotyping is therefore necessary to verify these assumptions, compared with the extensive multiyear phenotyping required when backcrossing is conducted without markers. One or two years can be saved with MABC during post-conversion phenotyping when compared with conventional backcrossing, resulting in an overall time advantage of MABC over conventional backcrossing of up to three years.

In many situations, the greatest advantages and profits are realized by those who are first to the market with their products. Also, for reasons related to the practices of seed production or legal aspects of crop registration procedures, it may be quite important to be able to produce near-isogenic versions of inbreds and hybrids; MABC provides such ability at a higher probability.

By contrast with MABC, SLS-MAS and MARS do not necessarily decrease the time needed to develop inbred lines. The use of MARS might actually increase it. The advantage of SLS-MAS and MARS resides in their ability to increase the rate of genetic gain (Eathington, 2005), which potentially results in higher performing lines and hybrids than can be developed through phenotypic selection only. Both SLS-MAS and MARS increase the frequency of favourable alleles in the population of selected individuals. The difference between the two approaches is that SLS-MAS operates on few loci while MARS operates on many. When SLS-MAS or MARS are used, the effective size of the population on which selection operates is increased, either directly for SLS-MAS or indirectly through several consecutive generations for MARS when compared with phenotypic pedigree selection. This increase in effective population size permits the application of a greater selection intensity and hence produces a higher genetic gain. SLS-MAS and MARS can also be seen as pre-selection steps if conducted prior to phenotypic selection and therefore improve the chances of evaluating genotypes with a higher frequency of favourable alleles phenotypically because the truly undesirable portion of the population may have been eliminated prior to phenotyping. Phenotypic selection can therefore be conducted with higher selection intensity than would be possible if no pre-selection had taken place, resulting potentially in additional genetic gain.

Alternatively, the resources used for phenotyping can be allocated differently based on whether individuals have been pre-selected or not with MAS. MAS schemes for forward breeding should enable breeding programmes to reallocate or focus resources for phenotypic evaluation in the target environment. For example, if DNA markers are linked to genes for resistance
to a disease or insect then it should be possible initially to select resistant progeny in the absence of the disease or insect by using the DNA data at continuous nurseries. The selected progeny could then be evaluated using relatively more expensive bioassays with the pest(s) in the target environment. This shift in resources is inherent to MARS schemes for complex traits (Edwards and Johnson, 1994; Johnson, 2004; Crosbie et al., 2006). By enriching populations through rapid cycles of MARS at continuous nurseries, breeders should derive a higher frequency of progeny with favourable alleles and haplotypes that are then evaluated in the target environment. Without MARS, resources for evaluation in the target environment would be diluted by the inclusion of too many progeny with an undesirable genetic constitution.

Concerns about reduced genetic diversity among commercial maize hybrids and depletion of genetic diversity in gene pools used in breeding may be partially alleviated by successful implementations of MAS. MABC may revive interest in using essentially untapped maize exotic germplasm as a source of favourable alleles for improvement of elite varieties. Very small and targeted chromosomal segments of exotic origin can be introgressed into elite inbred lines with limited risk of carrying along undesirable characteristics. Such an approach could be beneficial in maize although no accounts of its implementation have been reported despite the many years as reports of its successful use in tomato (Tanksley et al., 1996; Bernacchi et al., 1998a, b; Robert et al., 2001), rice (Xiao et al., 1998), and soybean (Concibio et al., 2003). MARS, in turn, may also contribute to increasing genetic diversity among commercial maize hybrids because, by focusing on selecting specific recombination events, it will result in the development of genuinely new genomic rearrangements. As QTL identified in any experiment represent only a fraction of the loci responsible for the phenotypes of complex traits, one can assume that breeding programmes in different private companies will conduct MARS based on their different genetic models and select for different genomic rearrangements. As a result, hybrids of similar and high performance might be developed that are based on different sets of favourable alleles at different loci, representing distinct “genetic solutions” and contributing to increased genetic diversity in farmers’ fields.

An indirect but important advantage of MAS and its underlying information and technology relates to intellectual property. Some maize breeding programmes have created a form of wealth through their collection and knowledge of maize germplasm. Significant investments have been made in maize breeding as exemplified by the billions of United States dollars that were used to purchase a few private programmes between 1995 and 2005. Protecting and maximizing returns on such investments have always been important but are now of greater concern. Information from MAS should be advantageous for addressing issues concerning ownership and derivation of germplasm, relatedness among germplasm and for the formation of some claims in patents and similar documents.

Perhaps one of the greatest advantages of MAS is that, for the first time, maize breeders have the means of learning some of the genetic details about germplasm and the response to selection. Some maize programmes in the private sector have started this process (Niebur et al., 2004). As real functions become associated with the many candidate genes and other DNA sequences,
the opportunities for learning about and understanding the response to selection will increase dramatically. It may then be possible to ameliorate some of the limitations of MAS and truly breed by design.

**Limitations of MAS**

While not truly an inherent limitation of the methods involved, one unavoidable limitation of MAS is the cost of assembling and integrating the necessary infrastructure and personnel. These can be substantial and beyond the means of many programmes. For such programmes, implementation of MAS could lead to a delusional or unbalanced reallocation of resources from vital activities such as high-quality phenotypic evaluation and selection in the target environment. Currently, only the largest maize breeding programmes in a given market or region have the scale of sales and diversity of products that can justify and support MAS and withstand some of the financial burdens of establishing and replacing components of the system (e.g. changes in the methods and platforms for detecting DNA polymorphisms).

Some inherent limitations to MAS are related to the estimates of QTL position and genetic effects and the rates of false positives and negatives. Confidence intervals for QTL are typically 10–15 cM; a genetic region that should not be a major barrier for implementing MAS although it could become a limitation to achieving genetic gain by preventing the selection of desired recombination events. The advent of association mapping and a growing pool of candidate genes should provide some resources needed to minimize problems related to the estimation of QTL position. The genetic effects of QTL are overestimated for many reasons, some of which are linked to experimental designs for phenotyping or population development while others are inherent to the process of QTL detection (Lee, 1995; Beavis, 1998; Melchinger, Utz and Schön, 1998; Holland, 2004). In addition, genetic effects related to epistasis are either poorly estimated or ignored by programmes in the private sector (Holland, 2001; Crosbie et al., 2006). Such assessments of genetic effects will inflate predictions of genetic gain. The relative merit of MAS will depend on the nature of predictions, actual results and costs of alternative methods.

A possible limitation of MAS with maize is the structure and content of various gene pools. Examples of maize gene pools would include European flint and dent germplasm, United States dents and various heterotic groups within each of these and other larger pools. Surveys with DNA markers have established differences among such groups of germplasm (Smith and Smith, 1992; Niebur et al., 2004). The relatively allele-rich maize gene pools coupled with genetic heterogeneity for many traits will hinder the ability to extrapolate information about genotype-phenotype relationships across gene pools. Such transfer of information is expected to be more successful in relatively homogeneous and less diverse maize gene pools (e.g. sweetcorn or popcorn) and with self-pollinated plant species (Lee, 1995). There have been undocumented reports of a few alleles at QTL that have relatively universal genetic effects across a relatively broad range of maize populations and target environments, but details of such genetic factors have not been publicly disclosed (Crosbie et al., 2006). More resources will need to be devoted to discovering where genetic information cannot be easily extrapolated across gene pools or even populations within a gene pool compared with situations where
it could. Although this should not impact the economic efficiency of MABC or forward breeding, it could affect the overall cost efficiency of MAS.

Finally, the efficacy for MAS in relatively complex populations such as synthetics and open-pollinated varieties (OPVs) has not been investigated. Compared with the bi-allelic populations used in the private sector, such populations are likely to have more than two alleles at a given locus. Also, unlike the simple bi-allelic populations, allele frequency should be an important component of predictions with such populations. Therefore, there should be more genetic effects and interactions to consider when making predictions based on MAS with OPVs and synthetics.

In the future, successful implementation of MAS in maize may lead to more frequent problems related to limited genetic variation. The emphasis of aggressive private sector maize breeding programmes on crosses between elite, related inbred lines to create segregating source populations has led to concerns about the depletion of genetic diversity in such gene pools and the ability to enhance such gene pools with high quality genetic variation (Niebur et al., 2004). Such concerns, which existed prior to the deployment of DNA markers and MAS in maize, are likely to increase as MAS becomes more prevalent. If MAS in forward breeding schemes is as effective as reported, then alleles and haplotypes may approach fixation more rapidly (Crosbie et al., 2006). At that point, breeding programmes will need to repeat the process of calibrating genotype-phenotype relationships in a slightly different array of reference populations to start the next metacycle of MAS (Johnson, 2004).

There is much anticipation for the future of MAS as genic sequences become the marker loci, functional information is discovered for the many candidate genes and gene products are assessed for their potential as useful sources of information in breeding programmes (Varshney, Graner and Sorrels, 2005; Lee, 2006). Certainly, these huge sets of raw data will contribute to progress. Eventually, other sources of genetic variation unrelated to the primary DNA sequence such as DNA methylation will be evaluated for their influence on genotype-phenotype relationships. Currently, epigenetic variation is mostly ignored from that assessment although it is well known that much of the maize genome may be methylated (Kaeppler, 2004) and may be more dynamic than predicted by current genetic models and mechanisms (Fu and Dooner, 2002). Also, the influences of non-coding sequences such as small interfering RNA (siRNA), matrix attachment regions and long-distance regulatory sequences have yet to be considered for their effects on genetic variation and estimates of genetic values used in MAS (Lee, 2006).

Most of the early limitations of MAS, due to the availability or cost of genotypic data, have been overcome. However, the availability or cost of high-quality phenotypic information is becoming one of the major limitations of MAS. During the past 20 years, development of new technologies and automation and miniaturization of laboratory procedures have contributed to reducing the cost of marker data points as well as the time needed to produce them. Large-scale marker laboratories produce marker data points at less than a tenth of the cost of 20 years ago. By contrast, neither cost nor the time required to produce phenotypic data has changed much, if at all, in the same timeframe. As the establishment of marker-trait associations and ultimately
the success of MAS depends on access to high-quality phenotypic data, means will have to be found to decrease the cost of phenotypic information while maintaining or increasing its quality. Alternatively, a greater proportion of budgets needs to be devoted to collecting phenotypic information.

**ACHIEVEMENTS OF MAIZE BREEDING PROGRAMMES WITH MAS**

In some important ways, maize breeding has gradually changed since the mid 1990s with the advent of genomics. Genetic principles were always an important component of modern maize breeding and now genetic information of various types is seeping into breeding schemes. MAS is the connection between the growing pool of genetic information and actual plant breeding. Establishing and enhancing this connection have been important achievements.

For the simplest breeding scenario, programmes in the private sector have demonstrated that MABC is an effective and routine method to backcross one or more transgenes into established elite inbred lines, the direct parents of commercial hybrids. Hybrids with effective combinations of transgenes have been very successful in the market. Consequently, MAS has accelerated the delivery of some products to the market; an important achievement in competitive economies.

Programmes in the private sector have also demonstrated a sufficient degree of efficacy of MAS methods to secure protection of intellectual property in patents. Methods, ideas and linkage relationships have been included in claims of patents or patent applications related to MAS (e.g. US5 746 023P 1998; US6 368 806B1 2002; US6 399 855B1 2002). Given the magnitude of the investments made in maize breeding by the private sector, receiving such a legal position may be a valuable achievement for the owner of the patent.

The efficacy of MAS for forward breeding of complex traits has yet to be firmly established. Positive results from calibration studies have been reported, but although accounts of MAS-derived commercial varieties have been made (Eathington, 2005), the impact on actual breeding and the development of new commercial hybrids has not been disclosed to a significant extent (Johnson, 2004; Niebur et al., 2004; Crosbie et al., 2006). At this point in time, it is therefore too early to make a definitive and database assessment of this aspect of MAS.

The history and cost of the genetic gain achieved through MAS will certainly vary among target environments. In some regions of the world, such as the central United States, maize breeding achieved steady genetic gains in grain yield for several consecutive decades prior to the advent of MAS (Duvick, Smith and Cooper, 2004). Nevertheless, the cost per unit gain has increased as more resources are needed for phenotypic evaluation in more environments (Smith et al., 1999). However, the advent of applied genomics and the discovery of many genes and gene functions, coupled with MAS, could reduce the dependence on costly phenotypic information for breeding. In regions where biotic and abiotic stress factors are more important than in the central United States, MAS may be very effective. Ultimately, the value and achievements of MAS will depend on the ecological and socio-economic context of the target environment.
COLLABORATION BETWEEN THE PRIVATE AND PUBLIC SECTORS IN MAS AND MAIZE IMPROVEMENT

The increased investments in maize breeding, expected returns on investment and concerns regarding intellectual property by the private sector have made it more difficult for corporations to collaborate with external parties of any kind. Such factors hinder the exchange of information and material that is common in collaborative projects. Nevertheless, around the world, the private and public sectors still manage to collaborate through various mechanisms and at different levels in the pursuit of maize improvement. Such collaboration involves interactions among multinational corporations, philanthropic foundations, national and subnational governments, universities and individuals. Major categories of collaboration include social programmes and institutions, research and development, and education.

In many regions of the world, private sector maize breeding would not have grown without some critical social programmes and institutions. For example, legislation related to intellectual property, transfer of capital and material, and regulatory approval of biotechnical innovations in maize improvement have been important components of legal systems that have encouraged financial investment in maize breeding. The stability of these systems and the rule of law have contributed to the long-term gains in selection. Also, long-term crop subsidy programmes in some regions have provided an element of security for investments in maize research and development by the private sector (Troyer, 2004; Crosbie et al., 2006). In those same regions, MAS has been deployed initially and on the largest scale for maize breeding.

With respect to research and development, there is a long history of effective collaboration between the public and private sectors in maize breeding. While such interaction continues in the era of MAS, the nature of the collaboration has changed with the growth and development of the breeding programmes in the private sector. Initially, collaboration was absolutely vital for the private sector because breeding programmes in the public sector were important, or the sole, sources of the inbred lines used directly by the private sector to produce commercial hybrids or to source populations from which elite inbreds were derived. Also, the inbred lines from the public sector were usually provided on an unrestricted basis and without payments of royalties or licensing fees. Public breeding programmes continue to develop elite inbred lines, occasionally in collaboration with the private sector (e.g. the Germplasm Enhancement of Maize programme in the United States; Pollak, 2003). However, the direct impact of contemporary public germplasm varies greatly among regions and gradually, in many regions of the world, the private sector has become the primary source of elite maize inbred lines and commercial hybrids.

In addition to germplasm, most or all of the critical concepts, methods and basic technologies have their origins in the public sector (Niebur et al., 2004; Troyer, 2004; Crosbie et al., 2006). The private sector, with its unique ability to concentrate capital through various mechanisms (e.g. profits from products or licence fees, venture capital and financial markets), is in the best position to allocate resources quickly to assess, modify and apply new developments in MAS and ancillary areas of maize improvement across large geographical and political regions of a market zone. As
described in previous sections, cost-effective MAS requires several components of an integrated infrastructure, some features of which have had a relatively high rate of renovation and replacement (e.g. methods of detecting DNA polymorphism), and therefore required substantial financial resources. Competing corporations and the potential for profit provide the necessary motivation for such investments (Troyer, 2004; Crosbie et al., 2006). To the authors’ knowledge, such financial mechanisms either do not exist or are limited in the public sector.

Collaboration between the public and private sectors in MAS for maize may be strongest in basic genetics and genome annotation. In order for MAS to reach its full potential, it may be necessary to acquire a much better understanding of gene function and products. For any plant species, only a small percentage of genes and other DNA sequences have a function defined through direct experimentation (Lee, 2006). Discoveries in plant gene function will occur in many laboratories around the world and, ultimately, the development groups in the private sector will have the necessary concentration of resources and sense of purpose to assemble the relatively raw basic information into tools and products from MAS. The maize nuclear genome, with tens of thousands of genes and many other important DNA sequences, is mostly a “black box” with respect to understanding the role of these in mediating phenotypes in response to environmental cues. Such understanding, a potential key to MAS and maize improvement, can only occur through informal and formal collaboration between the public and private sectors investigating a broad array of plant species.

Examples of collaborative research between the public and private sectors relevant to MAS in maize include attempts to select for hybrid yield (Stromberg, Dudley and Rufener, 1994), QTL mapping and selecting of hybrid yield (Eathington, Dudley and Rufener, 1997) and grain quality (Laurie et al., 2004), the development of the IBM population of recombinant inbred lines, and mapping genomic regions that include the vgt1 locus in maize (Lee et al., 2002; Salvi et al., 2002). National collaborative research programmes such as Génoplante in France and GABI in Germany, as well as several projects within the European Commission-sponsored framework programmes, are additional examples of such collaboration. Certainly, other collaborative projects between the public and private sectors have been conducted in maize MAS but their proprietary nature prevents public disclosure.

Future collaborative research activities in maize MAS could assume many forms. In most regions of the world, the private sector has the obvious superiority in terms of infrastructure needed for genotyping, phenotyping and data analysis. These resources are mostly devoted to the direct pursuit of products and profits. That pursuit may also be the greatest disadvantage of the private sector because such a focus limits the attention devoted to many interesting yet seemingly ancillary observations of genotype-phenotype relations in MAS. Some components of that infrastructure could possibly be made accessible to the public sector as “in-kind” contributions to collaborative or service-related projects in regions that are unlikely to emerge as important markets for the private sector or for phenotypes and germplasm that are not of direct interest to the private sector.

Education and training are also important areas in which the public and
private sectors should collaborate. With the advent of MAS, there has been an obvious need for maize breeders in the private sector to become familiar with all aspects of the process, and the public sector has developed several new short courses and training sessions in MAS-related concepts (Niebur et al., 2004; Crosbie et al., 2006).

Such knowledge is now considered a standard component of recent graduate training. However, while new students may have an adequate grasp of the theoretical aspects of MAS, their lack of exposure to the private sector’s advanced infrastructure represents a gap in their education. This situation is similar to that of students with a new degree in engineering who join advanced engineering and design groups in other industries: the private sector’s capacity to concentrate and focus capital often leads to advanced infrastructure that does not exist in the public sector. In such situations, new students have to navigate a rather steep learning curve before they become productive members of their new group. To reduce the slope of the learning curve, the private sector could provide internships to graduate students or to professors who teach plant breeding courses. It is unlikely that the public sector will have the resources to duplicate or exceed some features of the infrastructure that has been developed for maize MAS in the private sector. Therefore, for some aspects of education, it will be to everyone’s benefit to find ways to work together.

**PRIVATE SECTOR PERSPECTIVES ON MAS FOR MAIZE IMPROVEMENT**

The development of molecular markers in the 1980s provided the first tools to dissect the genetic basis of traits and select individuals based on their predicted genetic value. Back in these early days, the availability of genetic information was a limiting factor. Today’s landscape is very different as advances in applied genomics and laboratory technology have provided the tools to generate genetic information for all traits of interest. Gene similarities and synteny across genomes mean that much of the information generated on any plant species has relevance to other plant species. The speed at which genetic information becomes available never ceases to increase. Rather than its availability, it is the ability to handle and utilize genetic information that is becoming the limiting factor for MAS. New and improved information technology and bioinformatics capabilities therefore need to be developed that connect the growing wealth of genetic information with maize breeding programmes where knowledge about the genetic basis of traits and allelic variation at these loci is translated into varieties.

QTL and gene mapping will remain key for the generation and use of genetic information. As sequencing of cereal genomes including maize progresses, physical mapping of cloned genes will become a powerful alternative to statistical approaches. Characterization of allelic diversity at loci of interest can proceed from analyses of bi-parental populations or association studies. An effective alternative is the use of sets of NILs, or introgression line (IL) libraries (Peleman and van der Voort, 2003). As NILs developed around a specific locus differ only by the allele at this locus, and because most traits of agronomic interest in maize are quantitative, phenotypic differences among such NILs are expected to be rather small. High precision phenotyping will not only be required but will be critical for the evaluation of such material (Peleman and van der Voort, 2003). Private corporations have realized the need
for such high precision phenotyping as can be seen from their active recruiting of trait-specific phenotyping scientists often located in targeted areas where the trait of interest can be more easily measured (e.g. positions dedicated to drought tolerance and located in arid regions of the world).

In order to further the implementation of MAS in breeding, increased numbers of marker data points will be required. Private corporations have established or are developing the capacity to produce hundreds of millions of data points per year in service laboratories, distinct from research units. Besides, smaller “biotech” companies are developing technologies that could reduce the cost of each marker data point to a mere few United States cents. Moving to marker systems that are not based on gels is permitting the automation of most laboratory steps. Data points are being produced around the clock with laboratory technicians working in shifts. Here again, private companies are actively recruiting highly qualified technology specialists as well as laboratory managers whose role is more to optimize the running of production plants than dwell on the science. Beyond laboratories, plant handling is becoming a bottleneck to high-throughput protocols. High-throughput facilities have to be established and equipped at continuous nursery sites potentially to handle millions of plants per year.

There is little doubt that the largest private maize breeding programmes are investing very heavily in the implementation of MAS. Unless regulatory issues change dramatically, MABC will remain the preferred means of delivering transgenes to the market. Faster MABC protocols will always represent a potential commercial advantage in an area where competition is fierce and a one-year advantage may mean much on the market. Most recent investments have been directed at implementing MARS in breeding. The size of the investment in this approach seems to suggest that private corporations have more insight into its benefits compared with conventional breeding than has been reported publicly. Genotype-driven breeding should also allow faster development of specialized varieties as the maize market becomes more and more fragmented based on end-use of the harvest: animal feed (silage or grain), ethanol, dry or wet milling. Favourable alleles for traits of interest are likely to be spread across more than two lines therefore requiring the assembly of alleles from many different sources in a single inbred line. Proposals have been made to achieve such goals (Peleman and van der Voort 2003), although software tools to determine the optimal breeding schemes are not yet available to generate these “ideal” genotypes.

Maize breeding is likely to change more in the coming 10 or 20 years than it has over the past 50. Developing new hybrids efficiently now requires integrating data from many sources, sometimes beyond maize, generating high-quality genotypic and phenotypic data needed for the construction of “ideal” genotypes, and finally selecting phenotypically the best individuals from populations of marker-assisted-derived materials. Many stakeholders beyond maize breeders now take an active part in the development of new varieties and therefore breeding will increasingly become the responsibility of groups of individuals with complementary skills than stand-alone breeders. Training of all to understand and challenge the contribution of others will be critical to operating multidisciplinary breeding teams efficiently.
MAS FOR MAIZE IMPROVEMENT IN DEVELOPING COUNTRIES

A rapid analysis of the implementation of MAS in private maize breeding programmes points to three elements as being of particular importance: availability of high-quality phenotypic data, access to low-cost molecular marker data points and access to reliable continuous nurseries. The importance of high-quality phenotypic analyses has been clearly recognized by groups in the private sector (Niebur et al., 2004; Crosbie et al., 2006). Implementation of MAS in maize breeding requires large amounts of marker data points to be generated. Private groups have spent much effort developing technologies and platforms to achieve cost-efficient genotyping. Simultaneously, highly efficient continuous nurseries have been established in tropical environments or local greenhouses.

By contrast, maize breeding for developing countries is rather fragmented. National agricultural research institutions and international centres of the CGIAR such as the International Maize and Wheat Improvement Center (CIMMYT) focus much of their efforts on poor farmers and underserved regions. Private maize breeding programmes are also established in a number of developing countries. Due to the large up-front costs of assembling infrastructure and personnel for genotyping, it is unlikely that individual national marker laboratories could produce data points in a cost-efficient manner. However, regional facilities serving the needs of several national programmes and supported by local laboratories that could process samples (processing samples could be as easy as taking and air-drying them) and provide information in a timely manner, would probably be very sustainable alternatives. Such a regional molecular service laboratory has been established recently in Nairobi, Kenya, in a joint effort by two CGIAR centres, CIMMYT and the International Livestock Research Institute (ILRI) and Kenya’s Agricultural Research Institute (KARI), under the Canadian International Development Agency (CIDA)-funded Biosciences eastern and central Africa (BecA) platform, to provide technical access and training for African maize breeders (Delmer, 2005). Such a facility could be an excellent component of a comprehensive maize breeding effort if it is possible to establish and maintain high-quality personnel and facilities for all of the other aspects of maize breeding in key target environments. However, without high-quality capabilities in phenotypic evaluation and selection, molecular laboratories will be worthless. Research projects involving large-scale (transnational) phenotypic evaluations of key genetic material and focused on specific traits (tolerance to biotic or abiotic stresses) should provide genetic information that is both locally relevant and broadly applicable (geographically and in terms of germplasm). Such projects would also spread the cost of phenotyping across all participants but would only be successful with effective transnational coordination.

Private companies running MAS in maize could contribute to its implementation in developing countries in several ways. First, they could make some of their genetic information available, thereby adding to that already available in the public domain. Much information is being generated in the private sector on traits of importance to developing countries such as disease resistance (e.g. grey leaf spot, northern corn leaf blight, *Fusarium* stalk and ear rots), drought tolerance and nitrogen use efficiency. After validation of its relevance
to the germplasm and environments of target areas, this genetic information could be used to select efficiently for specific traits through MAS. Second, private companies could provide access to some of their genotyping or nursery platforms. Genotyping samples for MAS projects in developing countries would not substantially disrupt private companies’ own research if conducted in periods of lower activity, and would provide these MAS projects with marker data points for as low a cost as possible. Third and probably most critically, private companies could train scientists from developing countries on the principles, mechanics and logistics of applying and implementing MAS in maize. Scientists in private maize breeding groups have already identified many of the pitfalls and overcome many of the hurdles linked with the implementation of MAS. Transfer of this knowledge to scientists from developing countries would help them immensely to design marker-based breeding schemes adapted to their sets of constraints.

Beyond their contribution to the implementation of MAS in maize in developing countries, private companies could, in very similar ways, contribute to MAS programmes in other species of importance to developing countries but remote from their core interests. Synteny and gene conservation across species should allow some of the maize genetic information to be transferable to other species. Technology platforms and breeding approaches developed for MAS in maize should be good models for other crops and some might be directly usable. Mechanisms or organizations need to be put in place for these transfers of knowledge and technologies to occur from private maize MAS programmes to other crops in developing countries. Private programmes will likely not drive these transfers but might be very willing to contribute or be directly involved in specific projects provided adequate frameworks exist.

Public–private partnerships will need to be established to manage intellectual property issues related to the transfers of information, material or technologies from private companies to developing countries (Naylor et al., 2004). The African Agricultural Technology Foundation (AATF) is one initiative that has been established to deal with such issues. Several private corporations with major investments in MAS in maize have agreed to provide access to germplasm and knowledge for African countries (Naylor et al., 2004; Delmer, 2005).

As with the private sector in Europe and North America, it will be necessary to provide regular and easy access to education and training in maize MAS as the phenotypes and population structures are likely to differ from those encountered by programmes in the private sector in relatively high-input production environments. Also, and in common with the changes in the private sector, some reorganization or restructuring of public sector programmes may be warranted with the advent of more specialized roles for some personnel.

Understanding the genetic basis of traits and cloning and sequencing the underlying genes will not have an impact on poor farmers unless translated into varieties through breeding. Implementing MAS requires significant investments in both people and infrastructures. Some of the most promising marker-based breeding schemes (e.g. MARS), take about as long as conventional breeding schemes to develop improved varieties and therefore require long-term funding commitments. Funding of practical crop improvement has declined for several years, particularly
in the international public sector (Knight, 2003), and as a result investments have favoured research at the expense of practical applications (Naylor et al., 2004). Whether current funding mechanisms based on short-term (two to five years) grants are adequate to allow maize or any other breeding programmes in developing countries to benefit from the much needed advantages of MAS is questionable.

REFERENCES


Senior, M.L. & Heun, M. 1993. Mapping maize microsatellites and polymerase chain reaction confir-
mation of the target repeats using a CT primer. *Genome* 36: 884–889.


Molecular marker-assisted selection for resistance to pathogens in tomato

Amalia Barone and Luigi Frusciante
**SUMMARY**

Since the 1980s, the use of molecular markers has been suggested to improve the efficiency of releasing resistant varieties, thus overcoming difficulties met with classical breeding. For tomato, a high-density molecular map is available in which more than 40 resistance genes are localized. Markers linked to these genes can be used to speed up gene transfer and pyramiding. Suitable PCR markers targeting resistance genes were constructed directly on the sequences of resistance genes or on restriction fragment length polymorphisms (RFLPs) tightly linked to them, and used to select resistant genotypes in backcross schemes. In some cases, the BC$_5$ generation was reached, and genotypes that cumulated two homozygous resistant genes were also obtained. These results supported the feasibility of using marker-assisted selection (MAS) in tomato and reinforcing the potential of this approach for other genes, which is today also driven by the development of new techniques and increasing knowledge about the tomato genome.
INTRODUCTION

Tomato (Solanum lycopersicum, formerly Lycopersicon esculentum) is one of the most widely grown vegetable crops in the world. It is used as a fresh vegetable and can also be processed and canned as a paste, juice sauce, powder or as a whole. World volume has increased approximately 10 percent since 1985, reflecting a substantial increase in dietary use of the tomato. Nutritionally, tomato is a significant source of vitamins A and C. Furthermore, recent studies have shown the importance of lycopene, a major component of red tomatoes, which has antioxidant properties and may help to protect against cancer and heart disease (Rao and Agarwal, 2000).

One of the main constraints to tomato cultivation is damage caused by pathogens, including viruses, bacteria, nematodes and fungi, which cause severe losses in production. The control of pathogen spread mainly involves three strategies: husbandry practices, application of agrochemicals and use of resistant varieties. Husbandry techniques generally help to restrict the spread of pathogens and their vectors as well as to keep plants healthy, thus allowing pathogen attack to be limited. Chemical control gives good results for some pathogens, but poor results against others, such as bacteria, and has practically no effect on viruses. Moreover, reducing chemical treatments lowers the health risks to farmers and consumers. Therefore, in order to achieve sustainable agriculture and obtain high-quality, safe and healthy products, the use of resistant varieties is one of the principal tools to reduce pathogen damage.

Since the early part of the twentieth century, breeding for disease resistance has been a major method for controlling plant disease. Varieties that are resistant or tolerant to one or a number of specific pathogens are already available for many crops, and hybrids with multiple resistance to several pathogens exist and are currently used in vegetable production. In tomato, genetic control of pathogens is a very useful practice with most resistance being monogenic and dominant. Various sources of resistance have been used in traditional breeding programmes, and resistant breeding lines, varieties and F₁ hybrids have been developed with varying stability and levels of expression (Table 1) (Laterrot, 1996; Gardner and Shoemaker, 1999; Scott, 2005).

<table>
<thead>
<tr>
<th>Table 1</th>
<th>List of pathogen resistances present in tomato breeding lines, varieties and F₁ hybrids obtained through conventional breeding</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Virus</strong></td>
<td>Beet curly top virus (BCTV)                                                          Tobacco mosaic virus (TMV)                                                          Tomato mosaic virus (ToMV)                                                          Tomato yellow leaf curl virus (TYLCV)                                                          Tomato spotted wilt virus (TSWV)</td>
</tr>
<tr>
<td><strong>Bacteria</strong></td>
<td>Corynebacterium michiganense                                                          Pseudomonas solanacearum                                                          Pseudomonas syringae pv. tomato</td>
</tr>
<tr>
<td><strong>Nematodes</strong></td>
<td>Meloidogyne spp.</td>
</tr>
<tr>
<td><strong>Fungi</strong></td>
<td>Alternaria alternata f. sp. lycopersici                                                          Alternaria solani                                                          Cladosporium fulvum                                                          Fusarium oxysporum f. sp. lycopersici                                                          Fusarium oxysporum f. sp. radicis-lycopersici                                                          Phytophthora infestans                                                          Pyrenochaeta lycopersici                                                          Stemphylium solani                                                          Verticillium dahliae</td>
</tr>
</tbody>
</table>

Modified from Laterrot (1996) and updated as reported in the text.

MARKER-ASSISTED BREEDING FOR PATHOGEN RESISTANCE

Although conventional plant breeding has had a significant impact on improving tomato for resistance to important diseases, the time-consuming process of
### TABLE 2
**Resistance genes mapped on the tomato molecular map**

<table>
<thead>
<tr>
<th>Pathogen</th>
<th>Gene1</th>
<th>Chromosomal location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Virus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alfalfa mosaic virus (AMV)</td>
<td>Am</td>
<td>6</td>
<td>Parrella et al., 2004</td>
</tr>
<tr>
<td>Cucumber mosaic virus (CMV)</td>
<td>Cmr</td>
<td>12</td>
<td>Stamova and Chetelat, 2000</td>
</tr>
<tr>
<td>Potato virus Y (PVY)</td>
<td>pot-1</td>
<td>3</td>
<td>Parrella et al., 2002</td>
</tr>
<tr>
<td>Tomato mosaic virus (ToMoV)</td>
<td></td>
<td>2 genes</td>
<td>Griffiths and Scott, 2001</td>
</tr>
<tr>
<td>Tobacco mosaic virus (TMV)</td>
<td>Tm-1, Tm2a</td>
<td>2, 9</td>
<td>Young and Tanksley, 1988; Levesque et al., 1990</td>
</tr>
<tr>
<td>Tomato spotted wilt virus (TSWV)</td>
<td>Sw5</td>
<td>9</td>
<td>Stevens, Lamb &amp; Rhoads, 1995</td>
</tr>
<tr>
<td>Tomato yellow leaf curl virus (TYLCV)</td>
<td>Ty-1 (Q), Ty-2</td>
<td>6, 11</td>
<td>Zamir et al., 1994; Chagué et al., 1997; Hanson et al., 2000</td>
</tr>
<tr>
<td><strong>Bacteria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clavibacter michiganensis</td>
<td>Cm1.1- Cm 10.1 (Q)</td>
<td>1, 6, 7, 8, 9, 10</td>
<td>Sandbrink et al., 1995</td>
</tr>
<tr>
<td></td>
<td>QTLs</td>
<td>5, 7, 9</td>
<td>van Heusden et al., 1999</td>
</tr>
<tr>
<td></td>
<td>Rcm2.0 (Q), Rcm5.1 (Q)</td>
<td>2, 5</td>
<td>Kabelka, Franchino &amp; Francio, 2002</td>
</tr>
<tr>
<td>Pseudomonas syringae pv. tomato</td>
<td>Prf</td>
<td>6</td>
<td>Salmeron et al., 1996</td>
</tr>
<tr>
<td></td>
<td>Pto</td>
<td>6</td>
<td>Martin et al., 1993</td>
</tr>
<tr>
<td>Ralstonia solanacearum</td>
<td>Bw 1, Bw 3, Bw 4, Bw 5 (Q)</td>
<td>6, 10, 4, 6</td>
<td>Danesh et al., 2004; Thoquet et al., 1996</td>
</tr>
<tr>
<td>Xanthomonas campestris pv vesicatoria</td>
<td>Bs4</td>
<td>5</td>
<td>Ballvora et al., 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rx-1, rx-2, rx-3</td>
<td>1</td>
<td>Yu et al., 1995</td>
<td></td>
</tr>
<tr>
<td><strong>Nematodes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Globodera rostochiensis</td>
<td>Hero</td>
<td>4</td>
<td>Ganal et al., 1995</td>
</tr>
<tr>
<td>Meloidogyne spp.</td>
<td>Mi, Mi-3, Mi-9</td>
<td>6, 12, 6</td>
<td>Williamson et al., 1994; Yaghoobi et al., 1995; Ammiraju et al., 2003</td>
</tr>
<tr>
<td><strong>Fungi</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alternaria alternata f. sp. lycopersici</td>
<td>Asc</td>
<td>3</td>
<td>van der Biezen, Glagotlkaya &amp; Overduin, 1995</td>
</tr>
<tr>
<td></td>
<td>QTLs 2a, 2c, 3, 9, 12</td>
<td>2, 3, 9, 12</td>
<td>Robert et al., 2001</td>
</tr>
<tr>
<td>Alternaria solani</td>
<td>EBR-QTLs</td>
<td>2, 3, 9, 12</td>
<td>Foolad et al., 2002; Zhang et al., 2003</td>
</tr>
<tr>
<td>Cladosporium fulvum</td>
<td>Cf-1, Cf-2, Cf-4, Cf-5, Cf-9</td>
<td>1, 6, 1, 6, 1</td>
<td>Balint-Kurti et al., 1994; Jones et al., 1993</td>
</tr>
<tr>
<td>Fusarium oxysporum f. sp. radicis- lycopersici</td>
<td>Frl</td>
<td>9</td>
<td>Vakalounakis et al., 1997</td>
</tr>
<tr>
<td>Fusarium oxysporum f. sp. lycopersici</td>
<td>I1, I2, I3</td>
<td>7, 11, 7</td>
<td>Bournival, Vallejos and Scott, 1990; Sarfatti et al., 1991; Tanksley and Costello, 1991; Ori et al., 1997</td>
</tr>
<tr>
<td>Leveillula taurica</td>
<td>Lv</td>
<td>12</td>
<td>Chunwongse et al., 1994</td>
</tr>
<tr>
<td>Oidium lycopersici</td>
<td>Oi-1, Oi-2, Oi-3, Oi-4</td>
<td>6, 4, 6, 6</td>
<td>Huang et al., 2000; Bai et al., 2004; De Giovanni et al., 2004</td>
</tr>
<tr>
<td>Phytophthora infestans</td>
<td>lb1-lb12 (Q)</td>
<td>6, 12</td>
<td>Bai et al., 2003</td>
</tr>
<tr>
<td></td>
<td>Ph-1, Ph-2, Ph-3</td>
<td>7, 10, 9</td>
<td>Moreau et al., 1998; Chunwongse et al., 2002</td>
</tr>
<tr>
<td>Pyrenochaeta lycopersici</td>
<td>py-1</td>
<td>3</td>
<td>Doganlar et al., 1998</td>
</tr>
<tr>
<td>Stemphylium spp.</td>
<td>Sm</td>
<td>11</td>
<td>Behare et al., 1991</td>
</tr>
<tr>
<td>Verticillium dahliae</td>
<td>Ve1, Ve2</td>
<td>9</td>
<td>Diwan et al., 1999; Kawchuck et al., 2001</td>
</tr>
</tbody>
</table>

1 Q in parenthesis, QTL and qtl indicate quantitative trait loci. Recessive resistance genes are reported with small letters.
making crosses and backcrosses, and the selection of the desired resistant progeny, make it difficult to respond adequately to the evolution of new virulent pathogens. Moreover, several interesting resistances are difficult to use because the diagnostic tests often cannot be developed due to the challenge posed by inoculum production and maintenance. In addition, where symptoms are detectable only on adult plants and/or fruits, diagnostic tests can be particularly expensive and difficult to perform.

Since the 1980s, the use of molecular markers has been suggested as a tool for breeding many crops, including tomato. In the last two decades, molecular markers have been employed to map and tag major genes and quantitative trait loci (QTL) involved in monogenic and polygenic resistance control, known respectively as vertical and horizontal resistance. To date, more than 40 genes (including many single genes and QTL) that confer resistance to all major classes of plant pathogens have been mapped on the tomato molecular map (Table 2) and/or cloned from Solanaceous species, as reported by Grube, Radwanski and Jahn (2000). Since then, other resistance genes together with resistance gene analogues (RGAs), which are structurally related sequences based on the protein domain shared among cloned R genes (Leister et al., 1996), have been added to the map. A molecular linkage map of tomato based on RGAs has also been constructed in which 29 RGAs were located on nine of the 12 tomato chromosomes (Foolad et al., 2002; Zhang et al., 2002). Several RGA loci were found in clusters and their locations coincided with those of several known tomato R genes or QTL. This map provides a basis for further identifying and mapping genes and QTL for disease resistance and will be useful for MAS.

In fact, independently of the type of marker used for selection, by making it possible to follow the gene under selection through generations rather than waiting for phenotypic expression of the resistance gene, markers tightly linked to resistance genes can greatly aid disease resistance programmes. In particular, genetic mapping of disease resistance genes has greatly improved the efficiency of plant breeding and also led to a better understanding of the molecular basis of resistance.

DNA marker technology has been used in commercial plant breeding programmes since the early 1990s, and has proved helpful for the rapid and efficient transfer of useful traits into agronomically desirable varieties and hybrids ( Tanksley et al., 1989; Lefebvre and Chèvre, 1995). Markers linked to disease resistance loci can now be used for MAS programmes, thus also allowing several resistance genes to be cumulated in the same genotype (“pyramiding” of resistance genes), and they may be also useful for cloning and sequencing the genes. In tomato, several resistance genes have been sequenced to date, among them *Cf-2, Cf-4, Cf-5, Cf-9, Pto, Mi, I2*, and *Sw5*. These cloned R genes now provide new tools for tomato breeders to improve the efficiency of breeding strategies, via MAS. Although MAS is still not used routinely for improving disease resistance in many important crops (Michelmore, 2003), it is being used by seed companies for improving simple traits in tomato (Foolad and Sharma, 2005). Furthermore, while the deep knowledge of the tomato genome and the availability of a high-density molecular map for this species ( Pillen et al., 1996) should provide further opportunities to accelerate breeding through MAS, the time-consuming and expensive process of developing markers associated with genes of inter-
est and the high cost of genotyping large populations has and will continue to limit the use of MAS in most tomato breeding programmes.

The potential of MAS to speed up the breeding of tomato using molecular markers linked to various resistance genes has been examined in the authors’ laboratory. The two main goals of the research were to find the most suitable markers, and to test the feasibility of MAS for pyramiding resistance genes in tomato “elite” lines selected for their good processing qualities.

**STRATEGIES FOR GENE TRANSFER AND PYRAMIDING**

Six tomato genotypes carrying various resistance genes (Table 3) were crossed with tomato “elite” lines previously selected for yield and quality but lacking resistance traits. Each resistant genotype was crossed initially with each “elite” tomato line and various backcross schemes were then carried out starting from different F₁ hybrids. At each backcross generation the screening of resistant genotypes was performed using molecular markers linked to the resistance genes and DNA extracted from young leaves at seedling stage. Only resistant plants were then transplanted and grown in the greenhouse. At flowering, crosses were made with the recurrent parent to obtain the subsequent generations.

As the efficiency of MAS depends on the availability of polymerase chain reaction (PCR)-based markers highly linked to the resistance gene to be selected, for each resistance gene the most suitable marker system was investigated. For this purpose, three different strategies were undertaken. The first involved searching PCR markers already available in the literature and verifying their usefulness on the genetic material used. The second consisted of designing PCR primers from the sequence of cloned genes reported in the GeneBank database of the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/Genbank), while the third involved designing PCR primers from RFLP markers tightly linked to resistance genes. This last strategy was made possible by the availability of sequences of various mapped tomato RFLPs in the SolGenes database (www.sgn.cornell.edu).

In most cases, the results were obtained using cleaved amplified polymorphic sequences (CAPS; Konieczyn and Ausubel, 1993), which require one PCR reaction followed by restriction digest of the amplified fragment. In three cases (markers linked to genes Mi, Sw5 and Tm2a), the primers and enzymes used were those reported in the literature (Williamson et al., 1994; Folkertsma et al., 1999; Sobir et al., 2000). In the case of gene py-1, the procedure reported in the literature (Doganlar et al., 1998) was

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Resistance gene</th>
<th>Pathogen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Momor</td>
<td>Frl, Tm2a, Ve</td>
<td>Fusarium oxysporum f. sp. radicis-lycopersici, TMV, Verticillium dahliae</td>
</tr>
<tr>
<td>Motelle</td>
<td>I2, Mi, Ve</td>
<td>Fusarium oxysporum f. sp lycopersici, Meloidogyne spp., Verticillium dahliae</td>
</tr>
<tr>
<td>Okitzu</td>
<td>I2, Tm2a</td>
<td>Fusarium oxysporum f. sp lycopersici, TMV</td>
</tr>
<tr>
<td>Ontario</td>
<td>Pto</td>
<td>Pseudomonas syringae</td>
</tr>
<tr>
<td>Pyrella</td>
<td>py-1</td>
<td>Pyrenochaeta lycopersici</td>
</tr>
<tr>
<td>Stevens</td>
<td>Sw5</td>
<td>TSWV</td>
</tr>
</tbody>
</table>
simplified, enabling a faster and cheaper marker system, i.e. a sequence characterized amplified region (SCAR; Kawchuk, Hachey and Lynch, 1998) marker, which only requires one PCR reaction to detect polymorphism between the resistant and the susceptible genotypes, to be set up. (Barone et al., 2004).

The second strategy was followed to design primers and enzymes suitable for targeting three resistance genes (I2, Pto and Ve2). This strategy allowed gene-assisted selection to be achieved through the simple PCR procedure. Finally, the third strategy was applied in the case of one CAPS marker targeting the resistance gene Frl; it was derived from one RFLP tomato marker (TG101) linked to the gene (Fazio, Stevens and Scott, 1999).

The markers found were used to select resistant genotypes in backcross breeding schemes, while the process itself allowed three generations to be screened annually. At present, for some cross combinations, the BC5 generation has been reached, for others the BC2-BC3 (Figure 1). Where a BC5 generation was already available, the breeding programme continued by selfing BC5 resistant genotypes. In all other cases the backcross programme will continue up to the fifth backcross generation. At the end of each backcross scheme, the resistant BC5F3 genotypes, selected through molecular marker analysis, will also be tested directly for resistance by inoculating the pathogen and monitoring signs of disease. This will allow verification that no linkage breakage and loss of resistance gene occurred.

This procedure was already adopted in the case of one backcross scheme aimed at transferring a resistance gene to tomato spotted wilt virus (TSWV) to the susceptible genotype AD17 (Langella et al., 2004). The in vivo test performed on F1BC5, F2BC5 and, F3BC5 generations confirmed the introgression of the resistance trait and revealed that the resistance gene Sw5 was
fixed at the homozygous stage at the F3BC5 generation.

Finally, besides the transfer of one resistance gene to each susceptible genotype, a crossing scheme was undertaken to accumulate two or three resistance genes in the same genotype. In this case, the decision was made to stop the backcross scheme at the BC3 or BC4 generation as both parental lines were cultivated varieties and hence genetically very similar, and therefore the recovery of the recurrent genome could be satisfactory. F1BC4 hybrids carrying the same genetic background in the recurrent parent have been intercrossed, following the breeding scheme shown in Figure 2. At the end of each F1BC4 x F1BC4 cross and after selecting the genotypes carrying all the resistant alleles at the heterozygous level, one or two selfing generations will be carried out to fix all the resistant genes at the homozygous level.

This strategy has already started in some cases and the first homozygous multiresistant genotypes have been obtained. Also available are two F2 genotypes out of 52 analysed plants, obtained by intercrossing the F1BC4 progeny from PI15 x Stevens with the F1BC4 progeny from PI15 x Ontario (Table 4). This F2 generation exhibited two genotypes carrying both resistant genes Sw5 and Pto at the homozygous level as well as 29 genotypes carrying both genes at the heterozygous level.

The work reported here on transferring resistance genes among tomato genotypes demonstrates the usefulness of MAS for improving traditional breeding strategies. The contribution of molecular markers linked to resistance genes was very efficient in reducing the time and space necessary for selection, enabling both early screening for resistance and reduced numbers of genotypes to be transplanted. The most challenging work was the search for suitable markers, which often required both considerable time and financial resources. Different strategies were used successfully to find the most suitable markers to perform MAS for transferring eight resistance genes into superior tomato genotypes; such strategies could be repeated in tomato for many other genes due to advanced molecular knowledge of the genome of this species.

**PERSPECTIVES**

The availability of PCR-based markers for many resistance genes allows MAS for biotic resistance in tomato to be applied successfully in any laboratory without the need for highly sophisticated techniques.
Indeed, once a marker has been set up, its use on large populations for resistance screening is then routine. Technical facilities are today available for screening many samples simultaneously and also costs for equipment are decreasing. In addition, the rapid development of new molecular techniques, combined with the ever-increasing knowledge about the structure and function of resistance genes (Hulbert et al., 2001), will help to identify new molecular markers for MAS, such as single nucleotide polymorphisms (SNPs). Moreover, thanks to the International Solanaceae Genome Project (SOL), sequencing of the tomato genome is in progress, and in a few years this will enhance information on resistance genes. This in turn will facilitate the development of molecular markers from transcribed regions of the genome, thereby allowing large-scale gene-assisted selection (GAS) to be achieved.

Over the coming years, selection for pathogen resistance in tomato will be underpinned by research aimed at: mapping other resistance genes for new pathogens; developing PCR-based functional markers (Andersen and Luddenstedt, 2003); designing the most suitable breeding schemes (Peleman and van der Voort, 2003), especially for transferring QTL resistances; large-scale screening through automation; allele-specific diagnostics (Yang et al., 2004); and DNA microarrays (Borevitz et al., 2003). In effect, the combination of new knowledge and new tools will lead to changes in the strategies used for breeding by exploiting the potential of integrating “omics” disciplines with plant physiology and conventional plant breeding, a process that will drive the evolution of MAS into genomics-assisted breeding (Morgante and Salamini, 2003; Varshney, Graner and Sorrells, 2005).

ACKNOWLEDGEMENTS
The authors wish to thank Mr Mark Walters for editing the manuscript and Angela Cozzolino for her technical assistance. This research was partially financed by the Ministry of Agriculture and Forestry of Italy in the framework of project PROMAR (Plant Protection by the Use of Molecular Markers). Contribution no. 119 from the Department of Soil, Plant and Environmental Sciences.

REFERENCES


Salmeron, J.M., Oldroyd, G.E.D., Rommens, C.M.T., Scofield, S.R., Kim, H.S., Lavello, D.T., Dahlbeck, D. & Staskawicz, B.J. 1996. Tomato Prf is a member of the leucine-rich repeat class of
plant disease resistance genes and lies embedded within the tomato Pto kinase gene cluster. *Cell* 86: 123–133.


