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

Babu, R.C., Nguyen, B.D., Chamarerk, V., Shanmugasundaram, P., Chezhian, P., Jeyaprakash, P., Ganesh, S.K., Palchamy, A., Sadasivam, S., Sarkarung, S., Wade, L.J. & Nguyen, H.T. 2003. Genetic analysis of drought resistance in rice by molecular markers: association between secondary traits and field performance. *Crop Sci.* 43: 1457–1469.

Butcher, P.A. 2003. Molecular breeding of tropical trees. *In* A. Rimbawanto & M. Susanta, eds. *Proc. Int. Conf. Adv. in Genetic Improvement of Tropical Tree Species*, 1–3 October 2002. Centre for Forest Biotechnology and Tree Improvement, Yogyakarta, Indonesia.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* 82: E313–E328 (available at http://jas.fass.org/cgi/reprint/82/13_suppl/E313)

Dekkers, J.C.M. & Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Revs. Genet.* 3: 22–32.

FAO. 2001. *Agricultural biotechnology for developing countries – results of an electronic forum,* by J. Ruane & M. Zimmermann. FAO Research and Technology Paper No. 8. Rome (also available at www.fao.org/docrep/004/Y2729E/Y2729E00.htm).

Koebner, R. 2003. MAS in cereals: green for maize, amber for rice, still red for wheat and barley. In *Proc. Int. Workshop on Marker-Assisted Selection: A Fast Track to Increase Genetic Gain in Plant and Animal Breeding?* (available at www.fao.org/biotech/docs/Koebner.pdf).

Morris, M., Dreher, K., Ribaut, J-M. & Khairallah, M. 2003. Money matters (II): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. *Mol. Breed.*11: 235–247.

Nichols, K.M., Young, W.P., Danzmann, R.G., Robison, B.D., Rexroad, C., Noakes, M., Phillips, B., Bentzen, P., Spies, I., Knudsen, K., Allendorf, F.W., Cunningham, B.M., Brunelli, J., Zhang, H., Ristow, S., Drew, R., Brown, K.H., Wheeler, P.A. & Thorgaard, G.H. 2003. A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Anim. Genet.* 34: 102–115.

Paterson A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E. & Tanksley, S.D. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721–726.

Sonesson, A.K. 2003. Possibilities for marker-assisted selection in fish breeding schemes. In *Proc. Int. Workshop on Marker-Assisted Selection: A Fast Track to Increase Genetic Gain in Plant and Animal Breeding?* (available at www.fao.org/biotech/docs/Sonesson.pdf).

Tartarini, S. 2003. Marker-assisted selection in pome fruit breeding. In *Proc. Int. Workshop on Marker-Assisted Selection: A Fast Track to Increase Genetic Gain in Plant and Animal Breeding?* (available at www.fao.org/biotech/docs/Tartarini.pdf).

Yanchuk, A. 2002. The role and implications of biotechnology in forestry. *Forest Genetic Resources* 30: 18–22 (available at www.fao.org/docrep/005/Y4341E/Y4341E06.htm).

Young, N.D. 1999. A cautiously optimistic vision for marker-assisted breeding. *Mol. Breed.* 5: 505–510.

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

TABLE 1
Number of research initiatives utilizing genetic markers in the crop and forestry sectors sorted by type of markers

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

TABLE 2
Number of research initiatives utilizing genetic markers in the crop and forestry sectors according to the development stage of the technique or product

Phase	Crop	Forestry	Total
Experimental phase	344	179	523
Field tests	107	8	115
Commercial phase	4	1	5
Unspecified	34	0	34
Total	**489**	**188**	**677**

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 3
Number of research initiatives utilizing genetic markers in the crop and forestry sectors according to geographical origin

Region	Crop	Forestry	Total
Africa	52	17	69
Asia and Pacific	98	103	201
Europe (transition countries)	42	13	55
Latin America and Caribbean	249	55	304
Near East and North Africa	48	0	48
Total	**489**	**188**	**677**

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

TABLE 4
Number of research initiatives utilizing genetic markers according to the crop of application

Crop group	Number of projects
Cereals and pseudo-cereals	134
Pulses	54
Root and tubers	51
Fruit trees	53
Vegetables	29
Industrial crops	74
Fodder crops	16
Aromatics	5
Other or not specified	73
Total	**489**

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

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

Country	TC	MAS	IC	DH	MC	GE
Algeria	X[1]	X	X	X	N[2]	N
Angola	X	N	N	N	N	N
Armenia	X	N	X	X	X	N
Cameroon	X	X	N	N	N	N
Costa Rica	X	N	X	N	X	X
Dominican Republic	N	N	N	N	N	N
Ethiopia	X	X	N	X	N	N
Georgia	X	X	X	N	X	N
Ghana	X	X	N	N	N	N
Mali	X	N	N	N	N	N
Kenya	X	X	X	X	N	X
Malawi	X	N	N	N	N	N
Moldova	X	N	N	N	N	N
Mozambique	N	N	N	N	N	N
Nicaragua	X	X	N	N	X	N
Niger	X	X	N	N	N	N
Nigeria	X	X	X	X	X	N
Senegal	X	N	X	X	N	N
Sierra Leone	X	N	N	N	N	N
Sri Lanka	N	N	N	N	N	N
Sudan	X	N	N	N	N	N
Tunisia	X	X	N	X	N	N
Uzbekistan	X	N	N	X	X	N
Zambia	X	N	N	N	N	N
Zimbabwe	X	X	N	N	N	X

[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

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 6
Use of molecular markers reported in country reports on AnGR management

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

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). 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).

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.

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.

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

FAO region	Number of countries
Africa	13
Asia and the Pacific	19
Europe	37
Latin America and the Caribbean	10
Near East	9
North America	2
Total	**93**

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 8
Number of projects and countries in which samples were collected according to animal species and FAO regions

Species	Number of projects	Number of countries	FAO region
Buffalo	3	9	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean
Cattle	24	40	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean
Goat	11	28	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean
Sheep	19	56	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean
Pig	6	19	Africa, Asia and the Pacific, Europe
Ass	1	1	Europe
Horse	5	25	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean, North America
Bactrian camel	1	2	Asia and the Pacific
Dromedary	2	7	Africa, Near East
Alpaca	3	2	Near East, Latin America and the Caribbean
Rabbit	1	19	Africa, Asia and the Pacific, Europe
Chicken	8	34	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean, Near East
Yak	2	8	Asia and the Pacific, Europe, Near East

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

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 instable, 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

Baumung, R., Simianer, H. & Hoffmann, I. 2004. Genetic diversity studies in farm animals – a survey. *J. Anim. Breed. Genet*. 121: 361-373.

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

FAO. 1998a. *Secondary guidelines for development of National Farm Animal Genetic Resources Management Plans. Measurement of domestic animal diversity (MoDAD): original Working Group report*. Rome.

FAO. 1998b. *Secondary guidelines for development of National Farm Animal Genetic Resources Management Plans. Measurement of domestic animal diversity (MoDAD): recommended microsatellite markers*. Rome.

FAO. 2001. *Preparation of the first report on the state of the world's animal genetic resources*. Guidelines for the preparation of CRs. Rome.

FAO. 2004. *Measurement of domestic animal diversity – a review of recent diversity studies.* Document prepared for the third session of the Intergovernmental Working Group on Animal Genetic Resources. Rome (available at www.fao.org/ag/againfo/programmes/en/genetics/documents/ITWG3_Inf3.pdf).

FAO. 2005. *Status of research and application of crop biotechnologies in developing countries: preliminary assessment,* by Z. Dhlamini, C. Spillane, J.P. Moss, J. Ruane, N. Urquia & A. Sonnino. Rome (available at www.fao.org/docrep/008/y5800e/y5800e00.htm).

Guimarães, E.P., Kueneman, E. & Carena, M.J. 2006. Assessment of national plant breeding and biotechnology capacity in Africa and recommendations for future capacity building. *Hort. Sci.* 41: 50–52.

Hallauer, A.R. 1999. Temperate maize and heterosis. *In* J.G. Coors & S. Pandey, eds. *The genetics and exploitation of heterosis in crops,* pp. 353-361. Proc. Internat. Symp. of Heterosis in Crops. Mexico City, 18–22 August 1997. ASA, CSSA and SSSA, Madison, WI, USA.

Pilling, D., Cardellino, R., Zjalic, M., Rischkowsky, B., Tempelman, K.A. & Hoffmann, I. 2007. The use of reproductive and molecular biotechnology in animal genetic resources management – a global overview. *Anim. Genet. Resources Information*. FAO, Rome (in press).

Toenniessen G.P., O'Toole, J.C. & DeVries, J. 2003. Advances in plant biotechnology and its adoption in developing countries. *Curr. Opin. in Plant Biol*. 6: 191–198.

SECTION II

Marker-assisted selection in crops – case studies

CHAPTER 3

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 *Pst*I. 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 polyacylamide 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 (Zietkietkiewicz, 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

TABLE 1
SNP technologies

Allele discrimination	Detection methods
• Hybridization	• Gel separation
• Primer extension	• Arrays
• Ligation	• Mass spectrometry
• Invasive cleavage	• Plate readers

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.

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 labelling 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.

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

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 TEs 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 mole-cular 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).

TABLE 2
Key features of common molecular marker technologies

Marker type	PCR-based	Uses restriction enzymes	Poly-morphism	Abun-dance	Co-dominant	Auto-mation	Loci per assay	Specialized equipment
RFLP	no	yes	moderate	moderate	yes	no	1 to few	Radioactive isotope
RAPD	yes	no	moderate	moderate	no	yes	many	Agarose gels
AFLP	yes	no	moderate	moderate	no	yes	many	Polyacrylamide gels/capillary
ISSR	yes	no	moderate	moderate	no	yes	many	Agarose/polyacrylamide gels
DArT	yes	yes	moderate	moderate	no	yes	many	Microarray
CAPS	yes	yes	variable	moderate	yes	yes	single	Agarose gels
SCAR	yes	no	low	moderate	yes	yes	single	Agarose gels
SSR	yes	no	low	moderate	yes	yes	1 to about 20	Polyacrylamide gels/capillary
TE-Anchor	yes	no	variable	variable	yes	yes	single	Agarose gels
SNP	yes	no	variable	highest	yes	yes	1 to thousands	Variable

CHOOSING A MOLECULAR MARKER TECHNOLOGY

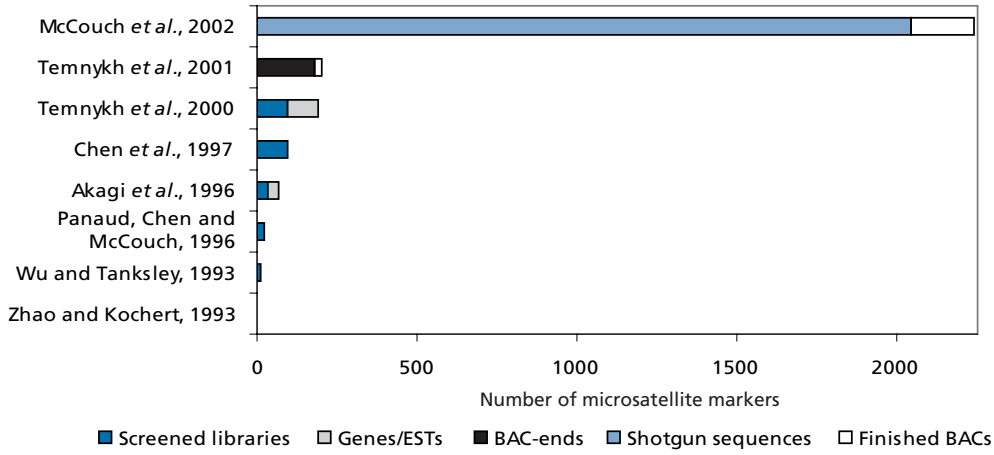
Clearly there is no single best choice of molecular marker for all situations. Factors influencing the decision may include the objectives of the study, availability of organism specific sequences, equipment and technical resources, and biological features of the species. Several important advantages/disadvantages 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

FIGURE 1
Progress of microsatellite marker development in rice



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 though 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

TABLE 3
Examples of SSR markers available across different plant species

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

From: Thomson, Septiningsih and Sutrisno, 2003 (reprinted with permission of author)

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

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 funnelled 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

Akagi, H., Yokozeki, Y., Inagaki, A. & Fujimura, T.Y. 1996. Microsatellite DNA markers for rice chromosomes. *Theor. Appl. Genet*. 93: 1071–1077.

Aparicio, S., Chapman, J., Stupka, E., Putman, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., Gelpke, M.D.S., Roach, J., Oh, T., Ho, I.Y., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S.F., Clark, M.S., Edwards, Y.J.K., Doggett, N., Zharkikh, A., Tavtigian, S.V., Pruss, D., Barnstead, M., Evans, C., Baden, H., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y.H., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, D. & Brenner, S. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297: 1301–1310.

Aranzana, M.J., Pineda, A., Cosson, P., Dirlewanger, E., Ascasibar, J., Cipriani, G., Ryder, C.D., Testolin, R., Abbott, A., King, G.J., Iezzoni, A.F. & Arus, P. 2003. A set of simple-sequence repeat (SSR) markers covering the *Prunus* genome. *Theor. Appl. Genet.* 106: 819–825.

Barrett, M.T., Scheffer, A., Ben-Dor, A., Sampas, N., Lipson, D., Kincaid, R., Tsang, P., Curry, B., Baird, K., Meltzer, P.S., Yakhini, Z., Bruhn, L. & Laderman, S. 2004. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc. Nat. Acad. Sci. USA* 101: 17765–17770.

Blair, M.W. & McCouch, S.R. 1997. Microsatellite and sequence-tagged site markers diagnostic for the rice bacterial leaf blight resistance gene *xa-5*. *Theor. Appl. Genet.* 95: 174-184.

Blair, M.W., Panaud O. & McCouch, S.R. 1999. Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 98: 780–792.

Borevitz, J. O., Liang, D., Plouffe, D., Chang, H.S., Zhu, T., Weigel, D., Berry, C.C., Winzeler, E. & Chory, J. 2003. Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res*. 13: 513–523.

Brondani, C., Rangel, N., Brondani V. & Ferreira, E. 2002. QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. *Theor. Appl. Genet.* 104: 1192–1203.

Bruskiewich, R.M., Cosico, A.B., Eusebio, W., Portugal, A.M., Ramos, L.M., Reyes, M.T., Sallan, M.A., Ulat, V.J., Wang, X., McNally, K.L., Sackville Hamilton, R. & McLaren, C.G. 2003. Linking genotype to phenotype: the International Rice Information System (IRIS). *Bioinformatics 19 Suppl 1*: i63–i65.

Bryan, G.T., Wu, K.S., Farrall, L., Jia, Y., Hershey, H.P., McAdams, S.A., Faulk, K.N., Donaldson, G.K., Tarchini, R. & Valent, B. 2000. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene Pi-ta. *The Plant Cell* 12: 2033–2046.

Casa, A.M., Brouwer, C., Nagel, A., Wang, L., Zhang Q., Kresovich, S. & Wessler, S.R. 2000. Inaugural article: the MITE family heartbreaker (Hbr): molecular markers in maize. *Proc. Nat. Acad. Sci. USA* 97: 10083–10089.

Champoux, M.C., Sarkarung, G.W.S., Mackill, D.J., O'Toole, J.C., Huang, N. & McCouch, S.R. 1995. Locating genes associated with root morphology and drought avoidance via linkage to molecular markers. *Theor. Appl. Genet.* 90: 969–981.

Chen, X., Cho Y.G. & McCouch, S.R. 2002. Sequence divergence of rice microsatellites in *Oryza* and other plant species. *Mol. Genet. Genomics* 268: 331–343.

Chen, X., Temnykh, S., Xu, Y., Cho, X.G. & McCouch, S.R. 1997. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theor. Appl. Genet*. 95: 553–567.

Comai, L., Young, K., Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R. & Henikoff, S. 2004. Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J.* 37: 778–786.

Conaway-Bormans, C.A., Marchetti, M.A., Johnson, C.W., McClung, A.M. & Park, W.D. 2003. Molecular markers linked to the blast resistance gene Pi-z in rice for use in marker-assisted selection. *Theor. Appl. Genet*. 107: 1014–1020.

Davierwala, A.P., Reddy, A.P., Lagu, M.D., Ranjekar, P.K. & Gupta, V.S. 2001. Marker assisted selection of bacterial blight resistance genes in rice. *Biochem. Genet*. 39: 261–278.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* 82 E-Suppl: E313–E328.

Doyle, J.J., Morgante, M., Tingey S.V. & Powell, W. 1998. Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). *Mol. Biol. Evol.* 15: 215–218.

Edwards, J.D., Lee, V.M. & McCouch, S.R. 2004. Sources and predictors of resolvable indel polymorphism assessed using rice as a model. *Mol. Genet. Genomics* 271: 298–307.

Fazio, G., Staub, J.E. & Chung, S.M. 2002. Development and characterization of PCR markers in cucumber. *J. Am. Soc. Hortic. Sci.* 127: 545–557.

Ferguson, M.E., Burow, M.D., Schulze, S.R., Bramel, P.J., Paterson, A.H., Kresovich, S. & Mitchell, S. 2004. Microsatellite identification and characterization in peanut (*A. hypogaea* L.). *Theor. Appl. Genet.* 108: 1064–1070.

Feschotte, C., Jiang N. & Wessler, S.R. 2002. Plant transposable elements: where genetics meets genomics. *Nature Revs. Genet.* 3: 329–341.

Flint-Garcia, S.A., Thornsberry, J.M. & Buckler, E.S.T. 2003. Structure of linkage disequilibrium in plants. *Ann. Rev. Plant Biol.* 54: 357–374.

Flowers, T.J., Koyama, M.L., Flowers, S.A., Sudhakar, C., Singh, K.P. & Yeo, A.R. 2000. QTL: their place in engineering tolerance of rice to salinity. *J. Exp. Bot.* 51: 99–106.

Garris A., McCouch, S. & Kresovich, S. 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). *Genetics* 165: 759–769.

Ghislain, M., Spooner, D.M., Rodriguez, F., Villamon, F., Nunez, J., Vasquez, C., Waugh, R. & Bonierbale, M. 2004. Selection of highly informative and user friendly microsatellites (SSRs) for genotyping of cultivated potato. *Theor. Appl. Genet.* 108: 881–890.

Goodwin, I.D., Aitken, E.A.B. & Smith, L.W. 1997. Application of inter-simple sequence repeat (ISSR) markers to plant genetics. *Electrophoresis* 18: 1524–1528.

Gregorio, G.B. 2002. Progress in breeding for trace minerals in staple crops. *J. Nutr*. 132: 500S–502S.

Gupta, M., Chyi, Y.S., Romero-Severson, J. & Owen, J.L. 1994. Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. *Theor. Appl. Genet.* 89: 998–1006.

Gupta, P.K., Rustgi, S., Sharma, S., Singh, R., Kumar, N. & Balyan, H.S. 2003. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol. Genet. Genomics* 270: 315–323.

Guryev, V., Koudijs, M.J., Berezikov, E., Johnson, S.L., Plasterk, R.H., van Eeden, F.J. & Cuppen, E. 2006. Genetic variation in the zebrafish. *Genome Res.* **16:** 491–497.

Gut, I.G. 2001. Automation in genotyping of single nucleotide polymorphisms. *Hum. Mutat.* 17: 475–492.

Hayashi, K., Hashimoto, N., Daigen, M. & Ashikawa, I. 2004. Development of PCR-based SNP markers for rice blast resistance genes at the Piz locus. *Theor. Appl. Genet*. 108: 1212–1220.

Hittalmani, S., Foolad, M.R., Mew, T., Rodriguez, R.L. & Huang, N. 1995. Development of a PCR-based marker to identify rice blast resistance gene, Pi-2(T), in a segregating population. *Theor. Appl. Genet.* 91: 9–14.

Hittalmani, S., Huang, N., Courtois, B., Venuprasad, R., Shashidhar, H.E., Zhuang, J.Y., Zheng, K.L., Liu, G.F., Wang, G.C., Sidhu, J.S., Srivantaneeyakul, S., Singh, V.P., Bagali, P.G., Prasanna, H.C., McLaren, G. & Khush, G.S. 2003. Identification of QTL for growth- and grain yield-related traits in rice across nine locations of Asia. *Theor. Appl. Genet*. 107: 679–690.

Innan, H., Terauchi, R., Kahl, G. & Tajima, F. 1999. A method for estimating nucleotide diversity from AFLP data. *Genetics* 151: 1157-1164.

Jaccoud, D., Peng, K., Feinstein, D. & Kilian, A. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucl. Acids Res*. 29: E25.

James, C. 2003. Global review of commercialized transgenic crops: 2002 Feature: *Bt* Maize. *ISAAA Briefs.*

Jones, M.P., Dingkuhn, M. Aluko, G.K. & Semon, M. 1997. Interspecific *O. sativa* L.x *O. glaberrima* Steud. Progenies in upland rice improvement. *Euphytica* 92: 237–246.

Kamoshita, A., Wade, J., Ali, L., Pathan, S., Zhang, J., Sarkarung, S. & Nguyen, T. 2002. Mapping QTLs for root morphology of a rice population adapted to rain-fed lowland conditions. *Theor. Appl. Genet.* 104: 880–893.

Kochert, G., Jena, K. & Zhao, X.P. 1990. Genome characterization and introgression. In *4th Ann. Mtg. Rockefeller Foundation's Internat. Program on Rice Biotech.* Los Baños, Philippines, IRRI.

Konieczny, A. & Ausubel, F.M. 1993. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J*. 4: 403–410.

Kwok, P.Y. 2001. Methods for genotyping single nucleotide polymorphisms. *Ann. Rev. Genomics Hum. Genet*. 2: 235–258.

La Rota, M., Kantety, R.V., Yu, J.K. & Sorrells, M.E. 2005. Non-random distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* 6: 23.

Li, Z.K., Yu, S.B., Lafitte, H.R., Huang, N., Courtois, B., Hittalmani, S., Vijayakumar, C.H., Liu, G.F., Wang, G.C., Shashidhar, H.E., Zhuang, J.Y., Zheng, K.L., Singh, V.P., Sidhu, J.S., Srivantaneeyakul, S. & Khush, G.S. 2003. QTL x environment interactions in rice. I. Heading date and plant height. *Theor. Appl. Genet.* 108: 141–153.

Lorieux, M., Reversat, G., Garcia Diaz, S.X., Denance, C., Jouvenet, N., Orieux, Y., Bourger, N., Pando-Bahuon, A. & Ghesquiere, A. 2003. Linkage mapping of Hsa-1(Og), a resistance gene of African rice to the cyst nematode, *Heterodera sacchari*. *Theor. Appl. Genet*. 107: 691–696.

McCouch, S.R., Kochert, G., Yu, Z.H., Wang, Z.Y., Khush, G.S., Coffman, W.R. & Tanksley, S.D. 1988. Molecular mapping of rice chromosomes. *Theor. Appl. Genet.* 76: 815–829.

McCouch, S.R., Chen, X., Panaud, O., Temnykh, S., Xu, Y., Cho, Y.G., Huang, N., Ishii, T. & Blair, M. 1997. Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Mol. Biol.* 35**:** 89–99.

McCouch, S.R., Teytelman, L., Xu, Y., Lobos, K.B., Clare, K., Walton, M., Fu, B., Maghirang, R., Li, Z., Xing, Y., Zhang, Q., Kono, I., Yano, M., Fjellstrom, R., DeClerck, G., Schneider, D., Cartinhour, S., Ware, D. & Stein, L. 2002. Development and mapping of 2 240 new SSR markers for rice (*Oryza sativa* L.) (supplement). *DNA Res*. 9: 257–279.

Monna, L., Lin, X., Kojima, S., Sasaki, T. & Yano, M. 2002. Genetic dissection of a genomic region for a quantitative trait locus, Hd3, into two loci, Hd3a and Hd3b, controlling heading date in rice. *Theor. Appl. Genet.* 104**:** 772–778.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51 Pt 1**:** 263–273.

Ndjiondjop, M.N., Albar, L., Fargette, D., Fauquet, C. & Ghesquiere, A. 1999. The genetic basis of high resistance to rice yellow mottle virus (RYMV) in cultivars of two cultivated rice species. *Plant Disease* 83: 931–935.

Nei, M. & Miller, J.C. 1990. A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics* 125: 873–879.

Nguyen, B.D., Brar, D.S., Bui, B.C., Nguyen, T.V., Pham, L.N. & Nguyen, H.T. 2002. Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., into indica rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 106: 583–593.

Nielsen, R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154: 931–942.

Nuffield Council on Bioethics. 2004. *The use of genetically modified crops in developing countries*. London, UK.

Panaud, O., Chen, X. & McCouch, S.R. 1996. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol. Gen. Genet*. 252: 597–607.

Paran, I. & Michelmore, R.W. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Genetics* 85: 985–993.

Poke, F.S., Vaillancourt, R.E., Potts, B.M. & Reid, J.B. 2005. Genomic research in *Eucalyptus*. *Genetica* 125: 79–101.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. & Rafalski, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2: 225–238.

Price, A.H., Townend, J., Jones, M.P., Audebert, A. & Courtois, B. 2002. Mapping QTLs associated with drought avoidance in upland rice grown in the Philippines and West Africa. *Plant Mol. Biol*. 48: 683–695.

Remington, D.L., Ungerer, M.C. & Purugganan, M.D. 2001. Map-based cloning of quantitative trait loci: progress and prospects. *Genet. Res.* 78: 213–218.

Ronald, P.C., Salmeron, J.M., Carland, F.M. & Staskawicz, B.J. 1992. The cloned avirulence gene 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 sequence tags. *Theor. Appl. Genet*. 109: 1283–1294.

Septiningsih, E.M., Prasetiyono, J., Lubis, E., Tai, T.H., Tjubaryat, T., Moeljopawiro, S. & 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 *O. rufipogon*. *Theor. Appl. Genet*. 107**:** 1419–1432.

Septiningsih, E.M., Trijatmiko, K.R., Moeljopawiro, S. & McCouch, S.R. 2003b. Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theor. Appl. Genet.* 107: 1433–1441.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H., Fauquet, C. & Ronald, P. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270: 1804–1806.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol*. 98: 503–510.

Su, C.C., Cheng, X.N., Zhai, H.Q. & Wan, J.M. 2002. Detection and analysis of QTL for resistance to the brown planthopper, *Nilaparvata lugens* (Stal), in rice (*Oryza sativa* L.), using backcross inbred lines]. *Yi Chuan Xue Bao* 29: 332–338.

Sun, X., Cao, Y., Yang, Z., Xu, C., Li, X., Wang, S. & Zhang, Q. 2004. Xa26, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes an LRR receptor kinase-like protein. *Plant J.* 37: 517–527

Takeuchi, Y., Lin, S.Y., Sasaki, T. & Yano, M. 2003. Fine linkage mapping enables dissection of closely linked quantitative trait loci for seed dormancy and heading in rice. *Theor. Appl. Genet.* 107: 1174–1180.

Tanksley, S.D. & McCouch, S.R. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277: 1063–1066.

Taylor, G., 2002. *Populus*: *arabidopsis* for forestry. Do we need a model tree? *Ann. Bot. (London)* 90: 681–689.

Temnykh, S.P., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., Cho,Y-G., Ishii, T. & McCouch, S.R. 2000. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa L.*). *Theor. Appl. Genet*. 100: 697–712.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S. & McCouch, S. 2001. 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.

Testolin, R., Huang, W.G., Lain, O., Messina, R., Vecchione, A. & Cipriani, G. 2001. A kiwifruit (*Actinidia* spp.) linkage map based on microsatellites and integrated with AFLP markers. *Theor. Appl. Genet.* 103**:** 30–36.

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

Thomson, M.J., Tai, T.H., McClung, A.M., Lai, X.H., Hinga, M.E., Lobos, K.B., Xu, Y., Martinez, C.P. & McCouch, S.R. 2003. Mapping quantitative trait loci for yield, yield components and

morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theor. Appl. Genet.* 107: 479–493.

Upholt, W.B. 1977. Estimation of DNA sequence divergence from comparison of restriction endonuclease digests. *Nucl. Acids Res*. 4: 1257–1265.

Valent, B., Bryan, G.T., Jia, Y., Farrall, L., McAdams, S.A., Faulk, K.N. & Levy, M. 2001. Enhancing deployment of rice blast resistance genes: opportunities from cloning a resistance gene/avirulence gene pair. *In* G.J. Khush, D.S. Brar & B. Hardy, eds. *Rice genetics IV,* pp. 309-322. Enfield, NH, USA, Science Publishers.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J. & Kuiper, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucl. Acids Res.* 23: 4407–4414.

Wang Z.X., Yano, M., Yamanouchi, U., Iwamoto, M.I., Monna, L., Hayasaka, H., Katayose, Y. & Sasaki, T. 1999. The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. *Plant J.* 19: 55–64.

Weber, J.L. & May, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet*. 44: 388–396.

Welsh, J. & McClelland, M. 1990. Finger-printing genomes using PCR with arbitrary primers. *Nucl. Acids Res*. 18: 7213–7218.

Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Kleinhofs, A. & Kilian, A. 2004. Diversity arrays technology (DArT) for whole-genome profiling of barley. *Proc Natl Acad Sci USA* 101: 9915–9920.

Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski J.A. & Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* 18: 6531–6535.

Wu, K.S. & Tanksley, S.D. 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. *Mol. Gen. Genet.* 241: 225–235.

Xu, Y., Ishii, T. & McCouch, S.R. 2003. Marker-assisted evaluation of germplasm resources for plant breeding, *In* T.W. Mew, D.S. Brar, S. Peng, D. Dawe & B. Hardy, eds. *Rice science: innovations and impact for livelihoods. Proc. Internat. Rice Res. Conf.,* pp. 213-230. Beijing, China, IRRI, Chinese Academy of Engineering and Chinese Academy of Agricultural Sciences.

Yamamoto, T., Lin, H., Sasaki, T. & Yano, M. 2000. Identification of heading date quantitative trait locus Hd6 and characterization of its epistatic interactions with Hd2 in rice using advanced backcross progeny. *Genetics* 154: 885–891.

Ye, X., Al-Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P. & Potrykus, I. 2000. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287: 303–305.

Yoshimura, S., Yoshimura, A., Saito, A., Kishimoto, N., Kawase, M., Yano, M., Nakagahara, M., Ogawa, T. & Iwata, N. 1992. RFLP analysis of introgressed chromosomal segments in three near-isogenic lines of rice for bacterial blight resistance genes, Xa-1, Xa-3 and Xa-4. *Jpn. J. Genet*. 67: 29–37.

Yoshimura. S., Iwata, N.A.Y., McCouch, S.R., Abenes, M.L., Baraoidan, M.R., Mew, T.W. & Nelson, R.J. 1995. Tagging and combining bacterial blight resistance genes using RAPD and RFLP markers. *Mol. Breed.* 1: 375–387.

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 bacterial inoculation. *Proc. Nat. Acad. Sci. USA* 95: 1663–1668.

Zane, L., Bargelloni, L. & Atarnello, T.P. 2002. Strategies for microsatellite isolation: a review. *Mol. Ecol.* 11: 1–16.

Zhao, X. & Kochert, G. 1993. Phylogenetic distribution and genetic mapping of a (GGC)n microsatellite from rice (*Oryza sativa* L.). *Plant Mol. Biol.* 21: 607–614.

Zhang, L.Y., Bernard, M., Leroy, P., Feuillet, C. & Sourdille, P. 2005. High transferability of bread wheat EST-derived SSRs to other cereals. *Theor. Appl. Genet.* 111: 677–687.

Zheng, H.G., Babu, R.C., Pathan, M.S., Ali, L., Huang, N., Courtois, B. & Nguyen, H.T. 2000. Quantitative trait loci for root-penetration ability and root thickness in rice: comparison of genetic backgrounds. *Genome* 43: 53–61.

Zietkietkiewicz, E.A.R., Rafalski, A. & Labuda, D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20: 176–183.

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 F_1 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 F_1 hybrids to the variety pool. Most varietal development programmes are therefore based on versions of the long-established pedigree breeding system, where large F_2 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 F_1 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 tricothecin 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 F_1 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 F_1 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

Anderson, J.A., Stack, RW., Liu, S., Waldron B.L., Fjeld, A.D., Coyne, C., Moreno-Sevilla, B., Mitchell Fetch, J., Song, Q.J., Cregan, P.B. & Frohberg, R.C. 2001. DNA markers for *Fusarium* head blight resistance QTLs in two wheat populations. *Theor. Appl. Genet.* 102: 1164–1168.

Ayala, L., Henry, M., Gonzalez-de-Leon, D., van Ginkel, M., Mujeeb-Kazi, A., Keller, B. & Khairallah, M.A. 2001a. Diagnostic molecular marker allowing the study of *Th. intermedium*-derived resistance to BYDV in bread wheat segregating populations. *Theor. Appl. Genet.* 102: 942–949.

Ayala, L., van Ginkel, M., Khairallah, M., Keller, B. & Henry, M. 2001b. Expression of *Thinopyrum intermedium*-derived barley yellow dwarf virus resistance in elite bread wheat backgrounds. *Phytopath.* 91: 55–62.

Buerstmayr, H., Lemmens, M., Hartl, L., Doldi, L., Steiner, B., Stierschneider, M. & Ruckenbauer, P. 2002. Molecular mapping of QTLs for *Fusarium* head blight resistance in spring wheat. I. Resistance to fungal spread (Type II resistance). *Theor. Appl. Genet*. 104: 84–91.

Buerstmayr, H., Steiner, B., Hartl, L., Griesser, M., Angerer, N., Lengauer, D., Miedaner, T., Schneider, B. & Lemmens, M. 2003. Molecular mapping of QTLs for *Fusarium* head blight resistance in spring wheat. II. Resistance to fungal penetration and spread. *Theor. Appl. Genet*. 107: 503–508.

Chartrain, L., Brading, P.A. & Brown, J.K.M. 2004. Presence of the *Stb6* gene for resistance to *Septoria tritici* blotch (*Mycosphaerella graminicola*) in cultivars used in wheat-breeding programmes worldwide. *Plant Pathol*. 54: 134–143.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. & Weissenbach, J. 1996. A comprehensive genetic map of the human genome based on 5 264 microsatellites. *Nature* 380: 152–154.

Law, C.N. & Worland, A.J. 1997. The control of adult plant resistance to yellow rust by the translocated chromosome 5BS-7BS of bread wheat. *Plant Breed*. 116: 59–63.

Liu, W.H., Nie, H., Wang, S.B., Li, X., He, Z.T., Han, C.G., Wang, J.R., Chen, X.L., Li, L.H. & Yu, J.L. 2005. Mapping a resistance gene in wheat cultivar Yangfu 9311 to yellow mosaic virus, using microsatellite markers. *Theor. Appl. Genet.* 111: 651–657.

Mallard, S., Gaudet, D., Aldeia, A., Abelard, C., Besnard, A.L., Sourdille, P. & Dedryver, F. 2005. Genetic analysis of durable resistance to yellow rust in bread wheat. *Theor. Appl. Genet.* 110: 1401–1409.

Thomas, J., Fineberg, N., Penner, G., McCartney, C., Aung, T., Wise, I. & McCallum, B. 2005. Chromosome location and markers of *Sm1*: a gene of wheat that conditions antibiotic resistance to orange wheat blossom midge. *Mol. Breed.* 15: 183–192.

Waldron, B.L., Moreno-Sevilla, B., Anderson, J.A., Stack, R.W. & Frohberg, R.C. 1999. RFLP mapping of QTLs for *Fusarium* head blight resistance in wheat. *Crop Sci.* 39: 805–811.

Young, N.D. 1999. A cautiously optimistic vision for marker-assisted selection. *Mol. Breed.* 5: 505–510.

Zhang, Z.Y., Xu, J.S., Xu, Q.J., Larkin, P. & Xin, Z.Y. 2004. Development of novel PCR markers linked to the BYDV resistance gene *Bdv2* useful in wheat for marker-assisted selection. *Theor. Appl. Genet.* 109: 433–439.

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 F_2 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 F_1 plant originating from a cross between two contrasting genotypes for NUE was first analysed by crossing the F_1 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.

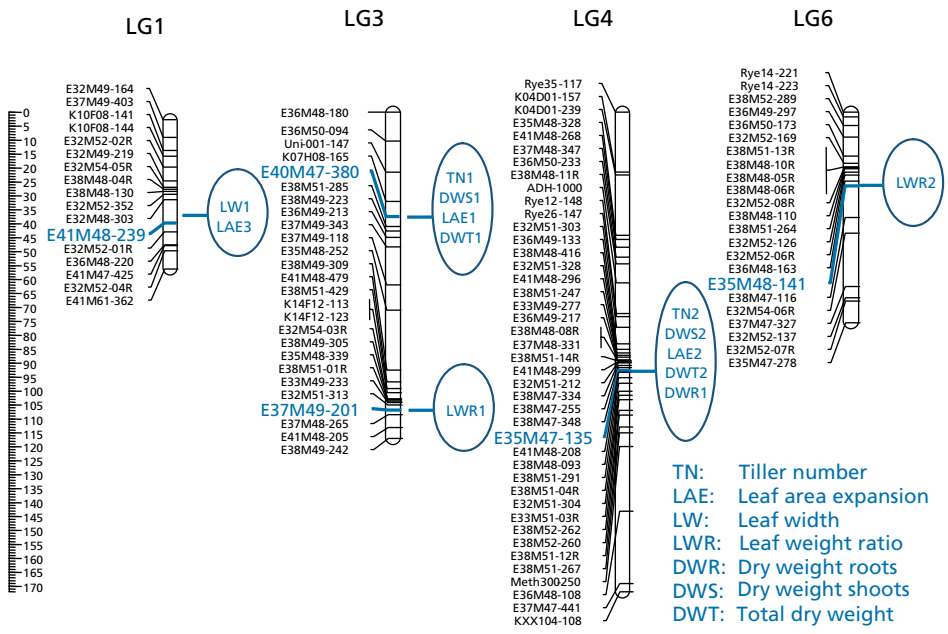
Putative genes (quantitative trait loci [QTL]) for the components of NUE were found on four linkage groups. The location of the selection markers for QTL is depicted in Figure 1. The map shows five genomic sites with 1-5 QTL. In total, 13 QTL for seven NUE related traits were found. Three sites contain 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 co-segregation 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 F_2 generation obtained by selfing of the heterozygous F_1 genotype

FIGURE 1
Map position of the selection markers (in blue) for QTL for seven components of nitrogen use efficiency



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.

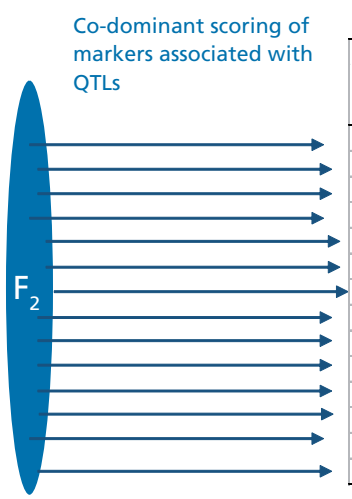
The genotyping data were used subsequently as a basis for a divergent mass selection programme. The selection strategy is outlined in Figure 2. The selection criterion was a genotype-specific selection index, being the summation of all positive QTL alleles (or chromosome segments) over the five QTL sites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

FIGURE 2
Strategy applied for marker selection



Genotype	Number of plus alleles					
	QTL1	QTL2	QLT3	QTL4	QTL5	Sum
1	2	1	0	0	2	5
2	1	1	0	0	2	4
3	0	2	2	1	1	6
4	2	2	2	1	1	8
5	2	2	1	1	1	7
.						
.						
.						
195	1	1	2	2	2	8
196	1	1	2	0	2	6
197	0	2	0	1	2	5
198	0	2	0	1	1	4
199	2	2	1	1	1	7
200	2	1	2	2	1	8

Selection criterion

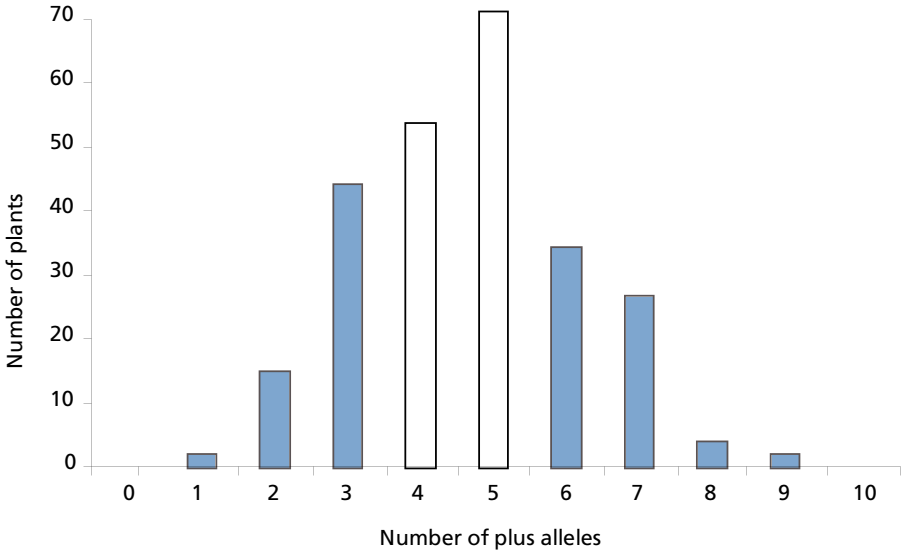
fragments) turned out to be straightforward. Figure 3 shows the F_2 frequency distribution for the number of "plus alleles". The population mean is somewhat below the expected number of five owing to the fact that the AFLP marker on LG1 showed a skewed segregation. This is likely due to gametophytic 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

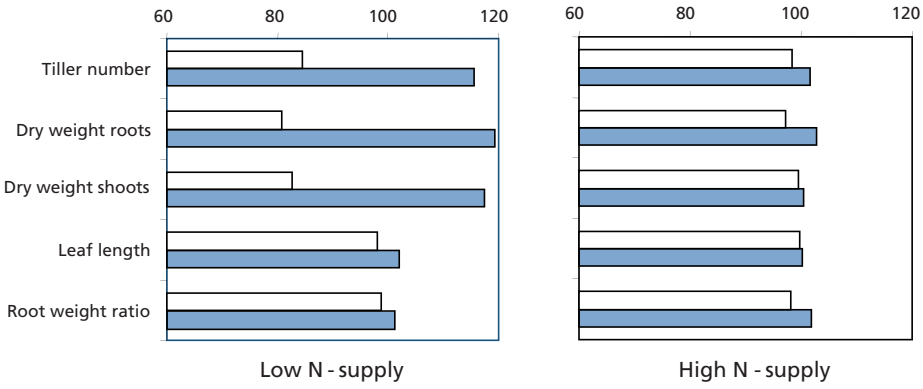
The selections were then multiplied using a polycross scheme (after vegetative propagation) to obtain sufficient seeds for evaluation on hydroponics and under various field conditions. The marker selections were evaluated for NUE in a replicated trial with two cuts on hydroponics at two N levels, being 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

FIGURE 3
Frequency distribution for number of plus alleles found for five QTLs



The two contrasting marker selections, i.e. NUEplus (right) and NUEmin (left), are marked in blue.

FIGURE 4
Performance of the marker selections, NUEmin (white) and NUEplus (blue) on hydroponics at low N-supply (2.5%N in leaves) and high N-supply (5%N in leaves). The performance is expressed as a percentage of the mean of the two selections



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

van Loo, E.N., Schapendonk, A.H.C.M. & de Vos, A.L.F. 1992. Effects of nitrogen supply on tillering dynamics and regrowth of perennial ryegrass populations. *Netherlands J. Agric. Sci.* 40: 401–419.

van Loo, E.N., Dolstra, O., Humphreys, M.O., Wolters, L., Luessink, W., de Riek, W. & Bark N. 2003. Lower nitrogen losses through marker assisted selection for nitrogen use efficiency and feeding value (NIMGRASS). *Vorträge Pflanzenzüchtung* 59: 270–279.

CHAPTER 6

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.

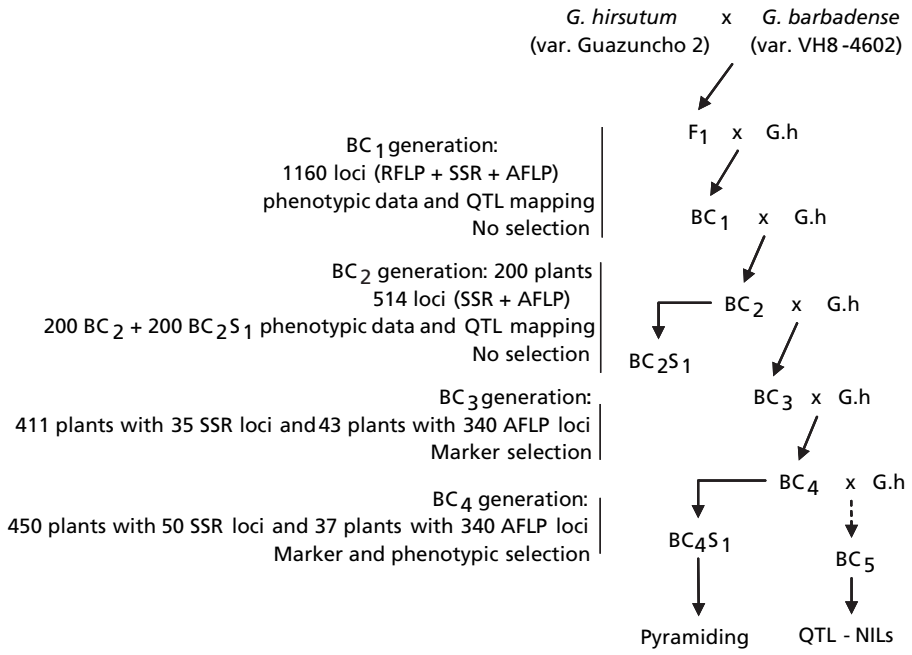
In the case of cotton, it is only recently that the results of efforts to gain a better understanding of the genome and the molecular basis of fibre quality have been published. Most of the earlier efforts in cotton molecular breeding concentrated on interspecific hybridization, due to the fact that, 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 introgression 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 BC_1 and BC_2 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 BC_3 and BC_4 generations.

METHODOLOGY

The major milestones (Figure 1) in the marker-assisted backcross selection process included the construction of two genetic maps from the BC_1 and BC_2 populations, the detection of fibre quality QTL from

FIGURE 1
Schematic representation of the marker-assisted backcross selection scheme



three phenotyping data sets (BC_1, BC_2 and BC_2S_1) and the actual marker-based selection in the BC_3 and BC_4 generations, followed by the analysis of marker-trait associations in the BC_3 and BC_4 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 multi-generation QTL analyses included three populations: BC_1, BC_2 and BC_2S_1 (Lacape *et al.*, 2005). The first backcross generation (BC_1), 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 BC_2 plants that had shown a satisfactory production of BC_3 seeds and originating from 53 different BC_1 plants were used in 2000. Open pollinated seeds harvested from BC_2 plants were grown as 200 BC_2S_1 progenies in 2001 under field conditions in Brazil. Each BC_2S_1 line was

[1] 1 tex = 1 gram/kilometre

planted in two replications, each plot (one row) measuring 5 m. The next BC_3 and BC_4 generations were grown under field (411 BC_3 in 2002) or greenhouse (450 BC_4 in 2003) conditions in Montpellier. Every plant in each $BC_{1\text{-}4}$ 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 BC_3 and BC_4 generation, an early genotyping was conducted (before flowering of BC_3 plants and at the seedling stage for BC_4 plants), to reduce the number of plants to be manipulated and raised to flowering for selfing and backcrossing.

From each generation (75 BC_1, 200 BC_2, 400 BC_2S_1, 43 selected BC_3 and 37 selected BC_4), 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 maturimeter (Shirley Dev Ltd., UK) was used to determine micronaire value, maturity and fineness.

Molecular analyses

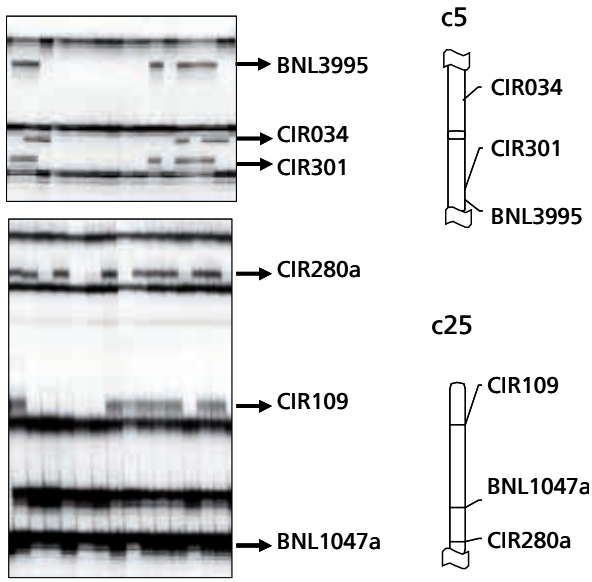
The different types of markers displaying polymorphism between *G. hirsutum* and *G. barbadense* included restriction fragment length polymorphisms (RFLPs) (used only in the BC_1 generation), simple sequence repeats (SSRs) and amplified fragment length polymorphisms (AFLPs). Details of the markers and 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 BC_1, 45 in the BC_2 and 30 in the BC_3 and BC_4 generations). The cotton microsatellites were derived essentially from two public libraries, Brookhaven National Laboratory (BNL) and CIRAD (CIR). In the BC_1 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 BC_1/BC_2/BC_2S_1 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 BC_3 and 450 BC_4 plants (see examples in Figure 2). The individual plants selected from BC_3 and BC_4 (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 BC_1 (75 individuals) and BC_2 (200 individuals) maps were constructed separately using the MapMaker 3.0 software

FIGURE 2

Examples of autoradiograms showing the segregation of 2 SSR triplexes observed among BC_4 plants and used for the targeted genotyping of specific regions along chromosomes c5 and c25 (a subset of around 15 BC_4 plants from the 450 plants analysed is shown)



(Lander *et al.*, 1987). 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 BC_1 and BC_2 maps using common loci, a consensus framework BC_1/BC_2 map was constructed by simple extrapolation of the positions of the additional BC_2 loci on the BC_1 map used as a backbone map. The allelic constitution throughout the 26 chromosomes of all $BC_{1\text{-}4}$ individuals was displayed graphically using Graphical GenoTyping software (R. van Berloo, Laboratory of Plant Breeding, Wageningen, Netherlands) and represented along the consensus BC_1 map data.

QTL analyses

The combined marker and phenotypic data then served for three (BC_1, BC_2 and BC_2S_1) 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 BC_1 map. Molecular data of further generations (BC_3 and BC_4) 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-

TABLE 1
Range of parental (*G. hirsutum [Gh]* and *G. barbadense [Gb]*) values over the five sets of data

	Gh	*Gb*	BC_1 N=75	BC_2 N=200	BC_2S_1 N=200	BC_3 N=43	BC_4 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)
Elongation	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)
Fineness (mtex)**	207–243	178–191	218 (177–308)	224 (165–379)	225 (176–285)	128*** (117–148)	243 (192–283)
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 BC_3 generation because of poor maturities
Note: Mean values and range (in brackets) observed in each BC_{1-4} generation (number of plants, N, indicated) of fibre technological parameters.

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 BC_4 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 (BC_1 and BC_2). The initial BC_1 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 (Nguyen *et al.*, 2004). On the other hand, the BC_2 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 (BC_1 and BC_2) and three sets of fibre measurements (per plant

TABLE 2
Number of QTL for each trait and range of observed phenotypic effects conferred by the *G. barbadense* alleles (either positive, "*Gb* +", or negative, "*Gb* –") detected over the three populations (BC_1, BC_2 and BC_2S_1)

	QTL *Gb* +	Range phenotypic effects	QTL *Gb* –	Range phenotypic effects
Length (mm)*	12	+0.7 to +2.1	3	–1.6 to –1.8
Length uniformity	3	+0.5 to +1.5	3	–1.1 to –3.3
Strength (g/tex)	8	+0.8 to +2.8	4	–0.9 to –3.4
Elongation	6	+0.2 to +0.5	4	–0.3 to –0.6
Fineness (mtex)**	13	–10 to –20	8	+9 to +40
Colour reflectance	3	+1.8 to +2.5	13	–0.9 to –3.5
Total	**45**		**35**	

* Length is upper half mean length (UHML), ** standard fineness.

basis for BC_1 and BC_2 and per-line basis with two replicates for BC_2S_1). The generations BC_1 and BC_2 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 (BC_1 and BC_2) 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 3
Identification of the 19 targeted regions mapped on 15 different chromosomes and comprising one or several co–localized fibre quality QTL from *G. barbadense* for introgression into a *G. hirsutum* genetic background

Carrier chromosome	Chromosome length (cM)	Target interval (cM)	Target size (cM)	Trait
c14	197	28–57	29	Length
c3	153	32–67	35	Length, fineness
		90–138	48	Length, strength, fineness
c4	190	102–118	16	Fineness
c22	139	112–139	27	Fineness
c5	360	78–101	23	Strength
c6	296	137–144	7	Length, fineness
c25	183	44–73	29	Length, strength
c16	168	65–117	52	Strength, fineness, colour
c23	173	45–66	21	Strength (elongation –, colour –)
		113–135	22	Length, strength
c10	192	0–21	21	Fineness
		78–120	42	Length, fineness, colour
c20	268	88–161	73	Elongation, fineness
c26	195	67–143	76	Length (colour –)
A01	233	16–54	38	Length
		171–209	38	Strength
c18	158	32–46	14	Fineness
A03	271	209–234	25	Strength, uniformity
Total	**3176**	**Total**	**636**	

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.

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.

MAS in the BC_3 and BC_4 generations and allelic transmission throughout generations

The early selection of BC_3 and BC_4 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 BC_3 plants out of 411 (11.4 percent) and 37 BC_4 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 BC_4 plants).

The allelic transmission observed in the four groups of BC_4 derived from four different BC_1 plants is given in Table 4.

TABLE 4
Percentage of introgressed loci (I%), at the heterozygous state, of 37 BC_4 plants that were derived from four different BC_1 plants (Nos. 3, 11, 16, and 27)

BC_1			BC_2			BC_3			BC_4			
Plant number	Number of loci	I%	Number of plants	Number of loci	I%	Number of plants	Number of loci	I%	Number of plants	Number of loci	I% global	I% target/non-target
No. 3	646	55	1	479	14	1	467	8	9	456	5	10/4
No. 11	654	64	1	446	28	1	403	13	1	408	10	29/6
No. 16	681	67	2	464	31	3	420	15	21	428	9	25/6
No. 27	668	63	1	471	26	1	433	15	6	439	10	25/6
Mean		62			26			14			8	21/5

Note: The number of plants and of loci analysed at each generation are given. At the BC_4 generation, the percentage of introgression is also differentiated between target and non-target (as defined in Table 3) regions.

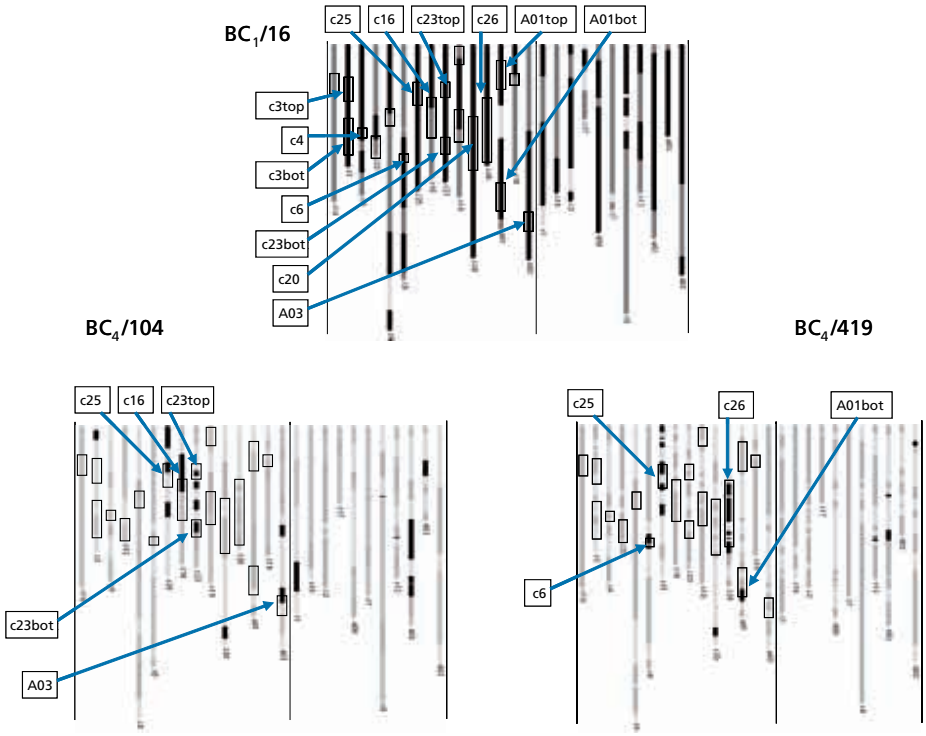
Moderate deviations were observed from theoretical transmission values (62, 26, 14 and 8 percent compared with 50, 25, 12.5 and 6.25 percent at the BC_1, BC_2, BC_3 and BC_4 stages, respectively), with a bias in favour of a higher rate of *G. barbadense* allele transmission. This bias was probably due to the selection pressure imposed at least in the BC_3 and BC_4 generations. Throughout the BC_1 and BC_2 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 3 662 BC_3 plants originating from 21 BC_1 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 BC_4 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 BC_4 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 BC_4 individuals as well as that of the BC_1 plant (No. 16) from which these individuals were derived. The two BC_4 plants had a common BC_1 ancestor but originated from two different BC_2 plants. In this particular example, starting from a common BC_1 plant (No. 16) which harboured 13 out of 19 possible QTL-rich regions, the two BC_4 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 BC_1 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 BC_4 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.

FIGURE 3

Graphical genotypes (26 chromosomes) of a BC_1 plant (No. BC1/16) (upper panel) and of two selected BC_4 plant (Nos. BC_4/104 and BC_4/419) (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 BC_3 and BC_4 generation plants

Owing to the limited number of individuals and the unbalanced frequencies of genotypic classes in the BC_3 and BC_4 material, significant marker-trait associations were less frequent than observed from the BC_1 and BC_2 data. For example, markers mapped along five, nine and six chromosome regions contributed (P=0.01), respectively, to length, strength or fineness variation using BC_4 marker-trait data, as compared with 15, 12 and 21 from the BC_1 and BC_2 data (Table 2). However, the majority of significant associations, particularly those determined in the BC_4 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 BC_1 and BC_2 data (Table 3), the BC_4 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 BC_4 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 BC_3 stage, single plants carried a maximum of five regions of interest (eight if considering regions only partially introgressed), while at the BC_4 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 BC_4 plants, i.e. those showing the highest amount of favourable QTL introgression, and then fixing the favourable allele by self-pollination. Such BC_4S_1 plants have been crossed with other BC_4S_1 plants of different ascent in order to pyramid as many QTL as possible (each contributing to different traits) within the same genome. Similarly, BC_4S_1 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

Basten, C., Weir, B. & Beng, Z-B. 1999. *QTL cartographer, version 1.13.* Dept. of Statistics, North Carolina State Univ., Raleigh, NC, USA.

Bernacchi, D., Beck-Bunn, T., Emmatty, D., Eshed, Y., Inai, S., Petiard, V., Sayama, H., Uhlig, J., Zamir, D. & Tanksley, S.D. 1998. Advanced backcross QTL analysis in tomato. II. Evaluation of near isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derives from *Lycopersicon hirsutum* and *L. pimpinellifolium*. *Theor. Appl. Genet.* 97: 170–180.

Brubaker, C.L. & Wendel, J.F. 2001. RFLP diversity in cotton. *In* J.N. Jenkins & S. Saha, eds. *Genetic improvement of cotton, emerging technologies,* pp 81-101. Enfield, NH, USA, Science Publisher Inc.

Dekkers, J.C.M. & Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Genet.* 3: 22–32.

Jiang, C., Chee, P., Draye, X., Morrell, P., Smith, C.W. & Paterson, A.H. 2000. Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). *Evolution* 54: 798–814.

Jiang, C., Wright, R., El-Zik, K.M. & Paterson, A.H. 1998. Polyploid formation created unique venues for response to selection in *Gossypium* (cotton). *Proc. Nat. Acad. Sc. USA* 95: 4419–4424.

Kohel, R.J., Yu, J., Park, Y.-H. & Lazo, G. 2001. Molecular mapping and characterization of traits controlling fiber quality in cotton. *Euphytica* 121: 163–172

Lacape, J.-M., Nguyen, T.-B., Thibivilliers, S., Courtois, B., Bojinov, B.M., Cantrell, R.G., Burr, B. & Hau, B. 2003. A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. *Genome* 46: 612–626.

Lacape, J.-M., Nguyen, T.-B., Courtois, B., Belot, J.-L., Giband, M., Gourlot, J.-P., Gawryziak, G., Roques, S. & Hau, B. 2005. QTL analysis of cotton fiber quality using multiple *G. hirsutum* x *G. barbadense* backcross generations. *Crop Sci.* 45: 123–140.

Lander, E., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S. & Newburg, L. 1987. MapMaker: an interactive computer package for constructing primary linkage maps of experimental and natural populations. *Genomics* 1:174–181.

Mei, M., Syed, N.H., Gao, W., Thaxton, P.M., Smith, C.W., Stelly, D.M. & Chen, Z.J. 2004. Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). *Theor. Appl. Genet.* 108: 280–291.

Nguyen, T.-B., Giband, M., Brottier, P., Risterucci, A.-M. & Lacape, J.-M. 2004. Wide coverage of tetraploid cotton genome using newly developed microsatellite markers. *Theor. Appl. Genet.* 109: 167–175.

Paterson, A.H., Saranga, Y., Menz, M., Jiang, C. & Wright, R.J. 2003. QTL analysis of genotype x environment interactions affecting cotton fiber quality. *Theor. Appl. Genet.* 106: 384–396.

Peleman, J.D. & van der Voort, J.R. 2003. Breeding by design. *Trends Plant Sci.* 7: 330–334.

Rong, J., Abbey, C., Bowers, J.E., Brubaker, C.L., Chang, C., Chee, P., Delmonte, T.A., Ding, X., Garza, J.J., Marler, B.S., Park, C.-H., Pierce, G.J., Rainey, K.M., Rastogi, V.K., Schulze, S.R., Trolinder, N., Wendel, J.F., Wilkins, T.A., Williams-Coplin, T.D., Wing, R.A., Wright, R.J., Zhao, X., Zhu, L. & Paterson, A.H. 2004. A 3347 locus genetic recombination map of sequence tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166: 389–417.

Salvi, S. & Tuberosa, R. 2005. To clone or not to clone plant QTLs: present and future challenges. *Trends in Plant Sci.* 10: 297–304.

Schön, C., Utz, H.F., Groh, S., Truberg, B., Openshaw, S. & Melchinger, A.E. 2004. Quantitative trait locus mapping based on resampling in a vast maize test cross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167: 485–498.

Tanksley, S.D. & Nelson, J.C. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor. Appl. Genet.* 92: 191–203.

Wilkins, T.A. & Arpat, A. 2005. The cotton fiber transcriptome. *Physiologia plantarum* 124: 295–300.