We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Molecular Markers and Marker-Assisted Breeding in Plants

Guo-Liang Jiang

Additional information is available at the end of the chapter

1. Introduction

Molecular breeding (MB) may be defined in a broad-sense as the use of genetic manipulation performed at DNA molecular levels to improve characters of interest in plants and animals, including genetic engineering or gene manipulation, molecular marker-assisted selection, genomic selection, etc. More often, however, molecular breeding implies molecular marker-assisted breeding (MAB) and is defined as the application of molecular biotechnologies, specifically molecular markers, in combination with linkage maps and genomics, to alter and improve plant or animal traits on the basis of genotypic assays. This term is used to describe several modern breeding strategies, including marker-assisted selection (MAS), marker-assisted backcrossing (MABC), marker-assisted recurrent selection (MARS), and genome-wide selection (GWS) or genomic selection (GS) (Ribaut et al., 2010). In this article, we will address general principles and methodologies of marker-assisted breeding in plants and discuss some issues related to the procedures and applications of this methodology in practical breeding, including marker-assisted selection, marker-based backcrossing, marker-based pyramiding of multiple genes, etc., beginning with a brief introduction to molecular markers as a powerful tool for plant breeding.

2. Genetic markers in plant breeding: Conceptions, types and application

Genetic markers are the biological features that are determined by allelic forms of genes or genetic loci and can be transmitted from one generation to another, and thus they can be used as experimental probes or tags to keep track of an individual, a tissue, a cell, a nucleus, a chromosome or a gene. Genetic markers used in genetics and plant breeding can be classified into two

categories: classical markers and DNA markers (Xu, 2010). Classical markers include morphological markers, cytological markers and biochemical markers. DNA markers have developed into many systems based on different polymorphism-detecting techniques or methods (southern blotting – nuclear acid hybridization, PCR – polymerase chain reaction, and DNA sequencing) (Collard et al., 2005), such as RFLP, AFLP, RAPD, SSR, SNP, etc.

2.1. Classical markers

Morphological markers: Use of markers as an assisting tool to select the plants with desired traits had started in breeding long time ago. During the early history of plant breeding, the markers used mainly included visible traits, such as leaf shape, flower color, pubescence color, pod color, seed color, seed shape, hilum color, awn type and length, fruit shape, rind (exocarp) color and stripe, flesh color, stem length, etc. These morphological markers generally represent genetic polymorphisms which are easily identified and manipulated. Therefore, they are usually used in construction of linkage maps by classical two- and/or three-point tests. Some of these markers are linked with other agronomic traits and thus can be used as indirect selection criteria in practical breeding. In the green revolution, selection of semi-dwarfism in rice and wheat was one of the critical factors that contributed to the success of high-yielding cultivars. This could be considered as an example for successful use of morphological markers to modern breeding. In wheat breeding, the dwarfism governed by gene *Rht10* was introgressed into Taigu nuclear male-sterile wheat by backcrossing, and a tight linkage was generated between *Rht10* and the male-sterility gene *Ta1*. Then the dwarfism was used as the marker for identification and selection of the male-sterile plants in breeding populations (Liu, 1991). This is particularly helpful for implementation of recurrent selection in wheat. However, morphological markers available are limited, and many of these markers are not associated with important economic traits (e.g. yield and quality) and even have undesirable effects on the development and growth of plants.

Cytological markers: In cytology, the structural features of chromosomes can be shown by chromosome karyotype and bands. The banding patterns, displayed in color, width, order and position, reveal the difference in distributions of euchromatin and heterochromatin. For instance, Q bands are produced by quinacrine hydrochloride, G bands are produced by Giemsa stain, and R bands are the reversed G bands. These chromosome landmarks are used not only for characterization of normal chromosomes and detection of chromosome mutation, but also widely used in physical mapping and linkage group identification. The physical maps based on morphological and cytological markers lay a foundation for genetic linkage mapping with the aid of molecular techniques. However, direct use of cytological markers has been very limited in genetic mapping and plant breeding.

Biochemical/protein markers: Protein markers may also be categorized into molecular markers though the latter are more referred to DNA markers. Isozymes are alternative forms or structural variants of an enzyme that have different molecular weights and electrophoretic mobility but have the same catalytic activity or function. Isozymes reflect the products of different alleles rather than different genes because the difference in electrophoretic mobility is caused by point mutation as a result of amino acid substitution (Xu, 2010). Therefore, iso-

zyme markers can be genetically mapped onto chromosomes and then used as genetic markers to map other genes. They are also used in seed purity test and occasionally in plant breeding. There are only a small number of isozymes in most crop species and some of them can be identified only with a specific strain. Therefore, the use of enzyme markers is limited.

Another example of biochemical markers used in plant breeding is high molecular weight glutenin subunit (HMW-GS) in wheat. Payne et al. (1987) discovered a correlation between the presence of certain HMW-GS and gluten strength, measured by the SDS-sedimentation volume test. On this basis, they designed a numeric scale to evaluate bread-making quality as a function of the described subunits (*Glu-1* quality score) (Payne et al., 1987; Rogers et al., 1989). Assuming the effect of the alleles to be additive, the Bread-making quality was predicted by adding the scores of the alleles present in the particular line. It was established that the allelic variation at the *Glu-D1* locus have a greater influence on bread-making quality than the variation at the others *Glu-1* loci. Subunit combination 5+10 for locus *Glu-D1* (*Glu-D1* 5+10) renders stronger dough than *Glu-D1* 2+12, largely due to the presence of an extra cysteine residue in the Dx-5 subunit compared to the Dx-2 subunit, which would promote the formation of polymers with larger size distribution. Therefore, breeders may enhance the bread-making quality in wheat by selecting subunit combination Glu-D1 5+10 instead of Glu-D1 2+12. Of course, the variation of bread-making quality among different varieties cannot be explained only by the variation in HMW-GS composition, because the low molecular weight glutinen subunit (LMW-GS) (as well as the gliadins in a smaller proportion) and their interactions with the HMW-GS also play an important role in the gluten strength and bread-making quality.

2.2. DNA markers

DNA markers are defined as a fragment of DNA revealing mutations/variations, which can be used to detect polymorphism between different genotypes or alleles of a gene for a particular sequence of DNA in a population or gene pool. Such fragments are associated with a certain location within the genome and may be detected by means of certain molecular technology. Simply speaking, DNA marker is a small region of DNA sequence showing polymorphism (base deletion, insertion and substitution) between different individuals. There are two basic methods to detect the polymorphism: Southern blotting, a nuclear acid hybridization technique (Southern 1975), and PCR, a polymerase chain reaction technique (Mullis, 1990). Using PCR and/or molecular hybridization followed by electrophoresis (e.g. PAGE – polyacrylamide gel electrophoresis, AGE – agarose gel electrophoresis, CE – capillary electrophoresis), the variation in DNA samples or polymorphism for a specific region of DNA sequence can be identified based on the product features, such as band size and mobility. In addition to Sothern blotting and PCR, more detection systems have been also developed. For instance, several new array chip techniques use DNA hybridization combined with labeled nucleotides, and new sequencing techniques detect polymorphism by sequencing. DNA markers are also called molecular markers in many cases and play a major role in molecular breeding. Therefore, molecular markers in this article are mainly referred to as DNA markers except specific definitions are given, although isozymes and protein markers are al-

so molecular markers. Depending on application and species involved, ideal DNA markers for efficient use in marker-assisted breeding should meet the following criteria:

- High level of polymorphism

- Even distribution across the whole genome (not clustered in certain regions)

- Co-dominance in expression (so that heterozygotes can be distinguished from homozygotes)

- Clear distinct allelic features (so that the different alleles can be easily identified)

- Single copy and no pleiotropic effect

- Low cost to use (or cost-efficient marker development and genotyping)

- Easy assay/detection and automation

- High availability (un-restricted use) and suitability to be duplicated/multiplexed (so that the data can be accumulated and shared between laboratories)

- Genome-specific in nature (especially with polyploids)

- No detrimental effect on phenotype

Since Botstein et al. (1980) first used DNA restriction fragment length polymorphism (RFLP) in human linkage mapping, substantial progress has been made in development and improvement of molecular techniques that help to easily find markers of interest on a large-scale, resulting in extensive and successful uses of DNA markers in human genetics, animal genetics and breeding, plant genetics and breeding, and germplasm characterization and management. Among the techniques that have been extensively used and are particularly promising for application to plant breeding, are the restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), microsatellites or simple sequence repeat (SSR), and single nucleotide polymorphism (SNP). According to a causal similarity of SNPs with some of these marker systems and fundamental difference with several other marker systems, the molecular markers can also be classified into SNPs (due to sequence variation, e.g. RFLP) and non-SNPs (due to length variation, e.g. SSR) (Gupta et al., 2001). The marker techniques help in selection of multiple desired characters simultaneously using F_2 and back-cross populations, near isogenic lines, doubled haploids and recombinant inbred lines. In view of page limitation, only five marker systems mentioned above are briefly addressed here according to published literatures. The details about the technical methods how to develop DNA markers and the procedures how to detect in practice have been described in the recently published reviews and books in this area (Farooq and Azam, 2002a, 2002b; Gupta et al., 2001; Semagn et al., 2006a; Xu, 2010).

RFLP markers: RFLP markers are the first generation of DNA markers and one of the important tools for plant genome mapping. They are a type of Southern-Boltting-based markers. In living organisms, mutation events (deletion and insertion) may occur at restriction sites or between adjacent restriction sites in the genome. Gain or loss of restriction sites resulting

from base pair changes and insertions or deletions at restriction sites within the restriction fragments may cause differences in size of restriction fragments. These variations may cause alternation or elimination of the recognition sites for restriction enzymes. As a consequence, when homologous chromosomes are subjected to restriction enzyme digestion, different restriction products are produced and can be detected by electrophoresis and DNA probing techniques.

RFLP markers are powerful tools for comparative and synteny mapping. Most RFLP markers are co-dominant and locus-specific. RFLP genotyping is highly reproducible, and the methodology is simple and no special equipment is required. By using an improved RFLP technique, i.e., cleaved amplified polymorphism sequence (CAPS), also known as PCR-RFLP, high-throughput markers can be developed from RFLP probe sequences. Very few CAPS are developed from probe sequences, which are complex to interpret. Most CAPS are developed from SNPs found in other sequences followed by PCR and detection of restriction sites. CAPS technique consists of digesting a PCR-amplified fragment and detecting the polymorphism by the presence/absence of restriction sites (Konieczny and Ausubel, 1993). Another advantage of RFLP is that the sequence used as a probe need not be known. All that a researcher needs is a genomic clone that can be used to detect the polymorphism. Very few RFLPs have been sequenced to determine what sequence variation is responsible for the polymorphism. However, it may be problematic to interpret complex RFLP allelic systems in the absence of sequence information. RFLP analysis requires large amounts of high-quality DNA, has low genotyping throughput, and is very difficult to automate. Radioactive autography involving in genotyping and physical maintenance of RFLP probes limit its use and share between laboratories. RFLP markers were predominantly used in 1980s and 1990s, but since last decade fewer direct uses of RFLP markers in genetic research and plant breeding have been reported. Most plant breeders would think that RFLP is too laborious and demands too much pure DNA to be important for plant breeding. It was and is, however, central for various types of scientific studies.

RAPD markers: RAPD is a PCR-based marker system. In this system, the total genomic DNA of an individual is amplified by PCR using a single, short (usually about ten nucleotides/ bases) and random primer. The primer which binds to many different loci is used to amplify random sequences from a complex DNA template that is complementary to it (maybe including a limited number of mismatches). Amplification can take place during the PCR, if two hybridization sites are similar to one another (at least 3000 bp) and in opposite directions. The amplified fragments generated by PCR depend on the length and size of both the primer and the target genome. The PCR products (up to 3 kb) are separated by agarose gel electrophoresis and imaged by ethidium bromide (EB) staining. Polymorphisms resulted from mutations or rearrangements either at or between the primer-binding sites are visible in the electrophoresis as the presence or absence of a particular RAPD band.

RAPD predominantly provides dominant markers. This system yields high levels of polymorphism and is simple and easy to be conducted. First, neither DNA probes nor sequence information is required for the design of specific primers. Second, the procedure does not involve blotting or hybridization steps, and thus it is a quick, simple and efficient technique.

Third, relatively small amounts of DNA (about 10 ng per reaction) are required and the procedure can be automated, and higher levels of polymorphism also can be detected compared with RFLP. Fourth, no marker development is required, and the primers are non-species specific and can be universal. Fifth, the RAPD products of interest can be cloned, sequenced and then converted into or used to develop other types of PCR-based markers, such as sequence characterized amplified region (SCAR), single nucleotide polymorphism (SNP), etc. However, RAPD also has some limitations/disadvantages, such as low reproducibility and incapability to detect allelic differences in heterozygotes.

AFLP markers: AFLPs are PCR-based markers, simply RFLPs visualized by selective PCR amplification of DNA restriction fragments. Technically, AFLP is based on the selective PCR amplification of restriction fragments from a total double-digest of genomic DNA under high stringency conditions, i.e., the combination of polymorphism at restriction sites and hybridization of arbitrary primers. Because of this AFLP is also called selective restriction fragment amplification (SRFA). An AFLP primer (17-21 nucleotides in length) consists of a synthetic adaptor sequence, the restriction endonuclease recognition sequence and an arbitrary, non-degenerate 'selective' sequence (1-3 nucleotides). The primers used in this technique are capable of annealing perfectly to their target sequences (the adapter and restriction sites) as well as a small number of nucleotides adjacent to the restriction sites. The first step in AFLP involves restriction digestion of genomic DNA (about 500 ng) with two restriction enzymes, a rare cutter (6-bp recognition site, *EcoR*I, *Pts*I or *Hind*III) and a frequent cutter (4-bp recognition site, *Mse*I or *Taq*I). The adaptors are then ligated to both ends of the fragments to provide known sequences for PCR amplification. The double-stranded oligonucleotide adaptors are designed in such a way that the initial restriction site is not restored after ligation. Therefore, only the fragments which have been cut by the frequent cutter and rare cutter will be amplified. This property of AFLP makes it very reliable, robust and immune to small variations in PCR amplification parameters (e.g., thermal cycles, template concentration), and it also can produce a high marker density. The AFLP products can be separated in high-resolution electrophoresis systems. The fragments in gel-based or capillary DNA sequencers can be detected by dye-labeling primers radioactively or fluorescently. The number of bands produced can be manipulated by the number of selective nucleotides and the nucleotide motifs used.

A typical AFLP fingerprint (restriction fragment patterns generated by the technique) contains 50-100 amplified fragments, of which up to 80% may serve as genetic markers. In general, AFLP assays can be conducted using relatively small DNA samples (1-100 ng per individual). AFLP has a very high multiplex ratio and genotyping throughput, and is relatively reproducible across laboratories. Another advantage is that it does not require sequence information or probe collection prior to generating the fingerprints, and a set of primers can be used for different species. This is especially useful when DNA markers are rare. However, AFLP assays have some limitations also. For instance, polymorphic information content for bi-allelic markers is low (the maximum is 0.5). High quality DNA is required for complete restriction enzyme digestion. AFLP markers usually cluster densely in centromeric regions in some species with large genomes (e.g., barley and sunflower). In ad-

dition, marker development is complicated and not cost-efficient, especially for locus-specific markers. The applications of AFLP markers include biodiversity studies, analysis of germplasm collections, genotyping of individuals, identification of closely linked DNA markers, construction of genetic DNA marker maps, construction of physical maps, gene mapping, and transcript profiling.

SSR markers: SSRs, also called microsatellites, short tandem repeats (STRs) or sequence-tagged microsatellite sites (STMS), are PCR-based markers. They are randomly tandem repeats of short nucleotide motifs (2-6 bp/nucleotides long). Di-, tri- and tetra-nucleotide repeats, e.g. (GT)n, (AAT)n and (GATA)n, are widely distributed throughout the genomes of plants and animals. The copy number of these repeats varies among individuals and is a source of polymorphism in plants. Because the DNA sequences flanking microsatellite regions are usually conserved, primers specific for these regions are designed for use in the PCR reaction. One of the most important attributes of microsatellite loci is their high level of allelic variation, thus making them valuable genetic markers. The unique sequences bordering the SSR motifs provide templates for specific primers to amplify the SSR alleles via PCR. SSR loci are individually amplified by PCR using pairs of oligonucleotide primers specific to unique DNA sequences flanking the SSR sequence. The PCR-amplified products can be separated in high-resolution electrophoresis systems (e.g. AGE and PAGE) and the bands can be visually recorded by fluorescent labeling or silver-staining.

SSR markers are characterized by their hyper-variability, reproducibility, co-dominant nature, locus-specificity, and random genome-wide distribution in most cases. The advantages of SSR markers include that they can be readily analyzed by PCR and easily detected by PAGE or AGE. SSR markers can be multiplexed, have high throughput genotyping and can be automated. SSR assays require only very small DNA samples (~100 ng per individual) and low start-up costs for manual assay methods. However, SSR technique requires nucleotide information for primer design, labor-intensive marker development process and high start-up costs for automated detections. Since the 1990s SSR markers have been extensively used in constructing genetic linkage maps, QTL mapping, marker-assisted selection and germplasm analysis in plants. In many species, plenty of breeder-friendly SSR markers have been developed and are available for breeders. For instance, there are over 35,000 SSR markers developed and mapped onto all 20 linkage groups in soybean, and this information is available for the public (Song et al., 2010).

SNP markers: An SNP is a single nucleotide base difference between two DNA sequences or individuals. SNPs can be categorized according to nucleotide substitutions either as transitions (C/T or G/A) or transversions (C/G, A/T, C/A or T/G). In practice, single base variants in cDNA (mRNA) are considered to be SNPs as are single base insertions and deletions (indels) in the genome. SNPs provide the ultimate/simplest form of molecular markers as a single nucleotide base is the smallest unit of inheritance, and thus they can provide maximum markers. SNPs occur very commonly in animals and plants. Typically, SNP frequencies are in a range of one SNP every 100-300 bp in plants (Edwards et al., 2007; Xu, 2010). SNPs may present within coding sequences of genes, non-coding regions of genes or in the intergenic regions between genes at different frequencies in different chromosome regions.

Based on various methods of allelic discrimination and detection platforms, many SNP genotyping methods have been developed. A convenient method for detecting SNPs is RFLP (SNP-RFLP) or by using the CAPS marker technique. If one allele contains a recognition site for a restriction enzyme while the other does not, digestion of the two alleles will produce different fragments in length. A simple procedure is to analyze the sequence data stored in the major databases and identify SNPs. Four alleles can be identified when the complete base sequence of a segment of DNA is considered and these are represented by A, T, G and C at each SNP locus in that segment. There are several SNP genotyping assays, such as allele-specific hybridization, primer extension, oligonucleotide ligation and invasive cleavage based on the molecular mechanisms (Sobrino et al., 2005), and different detection methods to analyze the products of each type of allelic discrimination reaction, such as gel electrophoresis, mass spectrophotometry, chromatography, fluorescence polarization, arrays or chips, etc. At the present, SNPs are also widely detected by sequencing. Detailed procedures are described in the review by Gupta at el. (2001) and the book Molecular Plant Breeding by Xu (2010).

SNPs are co-dominant markers, often linked to genes and present in the simplest/ultimate form for polymorphism, and thus they have become very attractive and potential genetic markers in genetic study and breeding. Moreover, SNPs can be very easily automated and quickly detected, with a high efficiency for detection of polymorphism. Therefore, it can be expected that SNPs will be increasingly used for various purposes, particularly as whole DNA sequences become available for more and more species (e.g., rice, soybean, maize, etc.). However, high costs for start-up or marker development, high-quality DNA required and high technical/equipment demands limit, to some extent, the application of SNPs in some laboratories and practical breeding programs.

The features of the widely used DNA markers discussed above are compared in Table 1. The advantages or disadvantages of a marker system are relevant largely to the purposes of research, available genetic resources or databases, equipment and facilities, funding and personnel resources, etc. The choice and use of DNA markers in research and breeding is still a challenge for plant breeders. A number of factors need to be considered when a breeder chooses one or more molecular marker types (Semagn et al., 2006a). A breeder should make an appropriate choice that best meets the requirements according to the conditions and resources available for the breeding program.

Feature and description	RFLP	RAPD	AFLP	SSR	SNP
Genomic abundance	High	High	High	Moderate to high	Very high
Genomic coverage	Low copy coding region	Whole genome	Whole genome	Whole genome	Whole genome
Expression/inheritance	Co-dominant	Dominant	Dominant / co-dominant	Co-dominant	Co-dominant

Feature and description	RFLP	RAPD	AFLP	SSR	SNP
Number of loci	Small (<1,000)	Small (<1,000)	Moderate (1,000s)	High (1,000s – 10,000s)	Very high (>100,000)
Level of polymorphism	Moderate	High	High	High	High
Type of polymorphism	Single base changes, indels	Single base changes, indels	Single base changes, indels	Changes in length of repeats	Single base changes, indels
Type of probes/primers	Low copy DNA or cDNA clones	10 bp random nucleotides	Specific sequence	Specific sequence	Allele-specific PCR primers
Cloning and/or sequencing	Yes	No	No	Yes	Yes
PCR-based	Usually no	Yes	Yes	Yes	Yes
Radioactive detection	Usually yes	No	Yes or no	Usually no	No
Reproducibility/ reliability	High	Low	High	High	High
Effective multiplex ratio	Low	Moderate	High	High	Moderate to high
Marker index	Low	Moderate	Moderate to high	High	Moderate
Genotyping throughput	Low	Low	High	High	High
Amount of DNA required	Large (5 – 50 µg)	Small (0.01 – 0.1 µg)	Moderate (0.5 – 1.0 µg)	Small (0.05 – 0.12 µg)	Small (≥ 0.05 µg)
Quality of DNA required	High	Moderate	High	Moderate to high	High
Technically demanding	Moderate	Low	Moderate	Low	High
Time demanding	High	Low	Moderate	Low	Low
Ease of use	Not easy	Easy	Moderate	Easy	Easy
Ease of automation	Low	Moderate	Moderate to high	High	High
Development/start-up cost	Moderate to high	Low	Moderate	Moderate to high	High
Cost per analysis	High	Low	Moderate	Low	Low
Number of polymorphic loci per analysis	1.0 – 3.0	1.5 – 5.0	20 – 100	1.0 – 3.0	1.0
Primary application	Genetics	Diversity	Diversity and genetics	All purposes	All purposes

Table 1. Comparison of most widely used DNA marker systems in plants; Adapted from Collard et al. (2005), Semagn et al. (2006a), Xu (2010), and others.

3. Pre-requisites and general activities of marker-assisted breeding

3.1. Prerequisites for an efficient marker-assisted breeding program

Compared with conventional breeding approaches, molecular breeding, mainly referred to as DNA marker-assisted breeding, needs more complicated equipment and facilities. In general, the pre-requisites listed below are essential for marker-assisted breeding (MAB) in plants.

a. Appropriate marker system and reliable markers: For a plant species or crop, a suitable marker system and reliable markers available are critically important to initiate a marker-assisted breeding program. As discussed above, suitable markers should have following attributes:

- Ease and low-cost of use and analysis;

- Small amount of DNA required;

- Co-dominance;

- Repeatability/reproducibility of results;

- High levels of polymorphism; and

- Occurrence and even distribution genome wide

In addition, another important desirable attribute for the markers to be used is close association with the target gene(s). If the markers are located in close proximity to the target gene or present within the gene, selection of the markers will ensure the success in selection of the gene. Although they can also be used in plant breeding programs, the number of classical markers possessing these features is very small. DNA markers for polymorphism are available throughout the genome, and their presence or absence is not affected by environments and usually do not directly affect the phenotype. DNA markers can be detected at any stage of plant growth, but the detection of classical markers is usually limited to certain growth stages. Therefore, DNA markers are the predominant types of genetic markers for MAB. Each type of markers has advantages and disadvantages for specific purposes. Relatively speaking, SSRs have most of the desirable features and thus are the current marker of choice for many crops. SNPs require more detailed knowledge of the specific, single nucleotide DNA changes responsible for genetic variation among individuals. However, more and more SNPs have become available in many species, and thus they are also considered an important type for marker-assisted breeding.

b. Quick DNA extraction and high throughput marker detection: For most plant breeding programs, hundreds to thousands of plants/individuals are usually screened for desired marker patterns. In addition, the breeders need the results instantly to make selections in a timely manner. Therefore, a quick DNA extraction technique and a high throughput marker detection system are essentially required to handle a large number of tissue samples and a large-scale screening of multiple markers in breeding programs. Extract-

ing DNA from small tissue samples in 96- or 384-well plates and streamlined operations are adopted in many labs and programs. High throughput PAGE and AGE systems are commonly used for marker detection. Some labs also provide marker detection services using automated detection systems, e.g. SNP chips based on thousands to ten thousands of markers.

c. Genetic maps: Linkage maps provide a framework for detecting marker-trait associations and for choosing markers to use in marker-assisted breeding. Therefore, a genetic linkage map, particularly high-density linkage map is very important for MAB. To use markers and select a desired trait present in a specific germplasm line, a proper population of segregation for the trait is required to construct a linkage map. Once a marker or a few markers are found to be associated with the trait in a given population, a dense molecular marker map in a standard reference population will help identify makers that are close to (or flank) the target gene. If a region is found associated with the desired traits of interest, fine mapping also can be done with additional markers to identify the marker(s) tightly linked to the gene controlling the trait. A favorable genetic map should have an adequate number of evenly-spaced polymorphic markers to accurately locate desired QTLs/genes (Babu et al., 2004).

d. Knowledge of marker-trait association: The most crucial factor for marker-assisted breeding is the knowledge of the associations between markers and the traits of interest. Only those markers that are closely associated with the target traits or tightly linked to the genes can provide sufficient guarantee for the success in practical breeding. The more closely the markers are associated with the traits, the higher the possibility of success and efficiency of use will be. This information can be obtained in various ways, such as gene mapping, QTL analysis, association mapping, classical mutant analysis, linkage or recombination analysis, bulked segregant analysis, etc. In addition, it is also critical to know the linkage situation, i.e. the markers are linked in cis/trans (coupling or repulsion) with the desired allele of the trait. Even if some markers have been reported to be tightly linked with a QTL, a plant breeder still needs to determine the association of alleles in his own breeding material. This makes QTL information difficult to directly transfer between different materials.

e. Quick and efficient data processing and management: In addition to above-mentioned pre-requisites, quick and efficient data process and management may provide timely and useful reports for breeders. In a marker-assisted breeding program, not only are large numbers of samples handled, but multiple markers for each sample also need to be screened at the same time. This situation requires an efficient and quick system for labeling, storing, retrieving, processing and analyzing large data sets, and even integrating data sets available from other programs. The development of bioinformatics and statistical software packages provides a useful tool for this purpose.

3.2. Activities of marker-assisted breeding

Marker-assisted breeding involves the following activities provided the prerequisites are well equipped or available:

a. Planting the breeding populations with potential segregation for traits of interest or polymorphism for the markers used.

b. Sampling plant tissues, usually at early stages of growth, e.g. emergence to young seedling stage.

c. Extracting DNA from tissue sample of each individual or family in the populations, and preparing DNA samples for PCR and marker screening.

d. Running PCR or other amplifying operation for the molecular markers associated with or linked to the trait of interest.

e. Separating and scoring PCR/amplified products, by means of appropriate separation and detection techniques, e.g. PAGE, AGE, etc.

f. Identifying individuals/families carrying the desired marker alleles.

g. Selecting the best individuals/families with both desired marker alleles for target traits and desirable performance/phenotypes of other traits, by jointly using marker results and other selection criteria.

h. Repeating the above activities for several generations, depending upon the association between the markers and the traits as well as the status of marker alleles (homozygous or heterozygous), and advancing the individuals selected in breeding program until stable superior or elite lines that have improved traits are developed.

4. Marker-assisted selection

4.1. MAS procedure and theoretical and practical considerations

Marker-assisted selection (MAS) refers to such a breeding procedure in which DNA marker detection and selection are integrated into a traditional breeding program. Taking a single cross as an example, the general procedure can be described as follow:

a. Select parents and make the cross, at least one (or both) possesses the DNA marker allele(s) for the desired trait of interest.

b. Plant F_1 population and detect the presence of the marker alleles to eliminate false hybrids.

c. Plant segregating F_2 population, screen individuals for the marker(s), and harvest the individuals carrying the desired marker allele(s).

d. Plant $F_{2:3}$ plant rows, and screen individual plants with the marker(s). A bulk of F_3 individuals within a plant row may be used for the marker screening for further confirmation in case needed if the preceding F_2 plant is homozygous for the markers. Select and harvest the individuals with required marker alleles and other desirable traits.

e. In the subsequent generations (F_4 and F_5), conduct marker screening and make selection similarly as for $F_{2:3}$s, but more attention is given to superior individuals within homozygous lines/rows of markers.

f. In $F_{5:6}$ or $F_{4:5}$ generations, bulk the best lines according to the phenotypic evaluation of target trait and the performance of other traits, in addition to marker data.

g. Plant yield trials and comprehensively evaluate the selected lines for yield, quality, resistance and other characters of interest.

A frequently asked question about marker-assisted selection is that "how many QTLs should be selected for MAS?" Theoretically, all the QTLs contributing to the trait of interest could be taken into account. For a quantitatively-inherited character like yield, numerous QTLs or genes are usually involved. It is almost impossible to select all QTLs or genes simultaneously so that the selected individuals incorporate all the desired QTLs due to the limitation of resources and facilities. The number of individuals in the population increases exponentially with the increase of target loci involved. The relative efficiency of MAS decreases as the number of QTLs increases and their heritability decreases (Moreau et al., 1998). In other words, MAS will be less effective for a highly complex character governed by many genes than for a simply inherited character controlled by a few genes. The number of genes/QTLs not only impacts the efficiency of MAS, but also the breeding design and implement scheme (detail will be discussed below). Typically no more than three QTLs are regarded as an appropriate and feasible choice (Ribaut and Betran, 1999), although five QTLs were used in improvement of fruit quality traits in tomato via marker-assisted introgression (Lecomte et al., 2004). With development of SNP markers (especially rapid automated detection and genotyping technologies), selection of more QTLs at the same time might be preferred and practicable (Kumpatla et al., 2012).

For MAS for multiple genes/QTLs, it was suggested to limit the number of genes undergoing selection to three to four if they are QTLs selected on the basis of linked markers, and to five to six if they are known loci selected directly (Hospital, 2003). Only the multi-environmentally verified QTLs that possess medium to large effects are selected. The first priority should be given to the major QTLs that can explain greatest proportion of phenotypic variation and/or can be consistently detected across a range of environments and different populations. In addition, an index for selection that weights markers differently could be constructed, depending on their relative importance to the breeding objectives. Flint-Garcia et al. (2003) presented an example of such an index used to select for QTLs with different effect magnitudes.

Another question that is commonly asked also is that "how many markers should be used in MAS?" The more markers associated with a QTL are used, the greater opportunity of success in selecting the QTL of interest will be ensured. However, efficiency is also important for a breeding program, especially when the resources and facilities are limited. From the point of both effectiveness and efficiency, for a single QTL it is usually suggested to use two markers (i.e. flanking markers) that are tightly linked to the QTL of interest. The markers to be used should be close enough to the gene/QTL of interest (<5cM) in order to ensure that

only a minor proportion of the selected individuals will be recombinants. If a marker (e.g. the peak marker) is found to be located within the region of gene sequence of interest or in such a close proximity to the QTL/gene that no recombination occurs between the marker and the QTL/gene, such a marker only should be preferable. However, if a marker is not tightly linked to a gene of interest, recombination between the marker and gene may reduce the efficiency of MAS because a single crossover may alternate the linkage association and leads to selection errors. The efficiency of MAS decreases as the recombination frequency (genetic distance) between the marker and gene increases. Use of two flanking markers rather than one may decrease the chance of such errors due to homologous recombination and increase the efficiency of MAS. In this case, only a double crossover (i.e. two single crossovers occurring simultaneously on both sides of the gene/QTL in the region) may result in selection errors, but the frequency of a double crossover is considerably rare. For instance, if two flanking markers with an interval of 20cM or so between them are used, there will be higher probability (99%) for recovery of the target gene than only one marker used.

In practical MAS, a breeder is also concerned about how the markers should be detected, how many generations of MAS have to be conducted, and how large size of the population is needed. In general, detection of marker polymorphism is performed at early stages of plant growth. This is true especially for marker-assisted backcrossing and marker-assisted recurrent selection, because only the individuals that carry preferred marker alleles are expected to be used in backcrossing to the recurrent parent and/or inter-mating between selected individuals/progenies. The generations of MAS required vary with the number of markers used, the degree of association between the markers and the QTLs/genes of interest, and the status of marker alleles. In many cases, marker screening is performed for two to four consecutive generations in a segregating population. If fewer markers are used and the markers are in close proximity to the QTL or gene of interest, fewer generations are needed. If homozygous status of marker alleles of interest is detected in two consecutive generations, marker screening may not be performed in their progenies. Bonnett et al. (2005) discussed the strategies for efficient implementation of MAS involving several issues, e.g. breeding systems or schemes, population sizes, number of target loci, etc. Their strategies include F_2 enrichment, backcrossing, and inbreeding.

In MAS, phenotypic evaluation and selection is still very helpful if conditions permit to do so, and even necessary in cases when the QTLs selected for MAS are not so stable across environments and the association between the selected markers and QTLs is not so close. Moreover, one should also take the impact of genetic background into consideration. The presence of a QTL or marker does not necessarily guarantee the expression of the desired trait. QTL data derived from multiple environments and different populations help a better understanding of the interactions of QTL x environment and QTL x QTL or QTL x genetic background, and thus help a better use of MAS. In addition to genotypic (markers) and phenotypic data for the trait of interest, a breeder often pays considerable attention to other important traits, unless the trait of interest is the only objective of breeding.

There are several indications for adoption of molecular markers in the selection for the traits of interest in practical breeding. The situations favorable for MAS include:

- The selected character is expressed late in plant development, like fruit and flower features or adult characters with a juvenile period (so that it is not necessary to wait for the plant to become fully developed before propagation occurs or can be arranged)

- The target gene is recessive (so that individuals which are heterozygous positive for the recessive allele can be selected and/or crossed to produce some homozygous offspring with the desired trait)

- Special conditions are required in order to invoke expression of the target gene(s), as in the case of breeding for disease and pest resistance (where inoculation with the disease or subjection to pests would otherwise be required), or the expression of target genes is highly variable with the environments.

- The phenotype of a trait is conditioned by two or more unlinked genes. For example, selection for multiple genes or gene pyramiding may be required to develop enhanced or durable resistance against diseases or insect pests.

4.2. MAS for major genes or improvement of qualitative traits

In crop plants, many economically important characteristics are controlled by major genes/ QTLs. Such characteristics include resistance to diseases/pests, male sterility, self-incompatibility and others related to shape, color and architecture of whole plants and/or plant parts. These traits are often of mono- or oligogenic inheritance in nature. Even for some quality traits, one or a few major QTLs or genes can account for a very high proportion of the phenotypic variation of the trait (Bilyeu et al., 2006; Pham et al., 2012). Transfer of such a gene to a specific line can lead to tremendous improvement of the trait in the cultivar under development. The marker loci which are tightly linked to major genes can be used for selection and are sometimes more efficient than direct selection for the target genes. In some cases, such advantages in efficiency may be due to higher expression of the marker mRNA in such cases that the marker is actually within a gene. Alternatively, in such cases that the target gene of interest differs between two alleles by a difficult-to-detect SNP, an external marker of which polymorphism is easier to detect, may present as the most realistic option.

Soybean cyst nematode (SCN) (*Heterodera glycines* Inchinoe) may be taken as an example of MAS for major genes. This pathogen is the most economically significant soybean pest. The principal strategy to reduce or eliminate damage from this pest is the use of resistant cultivars (Cregan et al., 1999). However, identifying resistant segregants in breeding populations is a difficult and expensive process. A widely used phenotypic assay takes five weeks, requires a large greenhouse space, and about 5 to 10 h of labor for every 100 plant samples processed (Young, 1999). Fortunately, the SSR marker Satt309 has been identified to be located only 1–2 cM away from the resistance gene *rhg1* (Cregan et al., 1999), which forms the basis of many public and commercial breeding efforts. In a direct comparison, genotypic selection with Satt309 was 99% accurate in predicting lines that were susceptible in subsequent greenhouse assays for two test populations, and 80% accurate in a third population, each with a different source of SCN resistance (Young, 1999). In soybean, Shi et al. (2009) reported that using molecular markers in a cross J05 x V94-5152, they developed five $F_{4:5}$

lines that were homozygous for all eight marker alleles linked to the genes/loci of resistance to soybean mosaic virus (SMV). These lines exhibited resistance to SMV strains G1 and G7 and presumably carried all three resistance genes (*Rsv1*, *Rsv3* and *Rsv4*) that would potentially provide broad and durable resistance to SMV.

4.3. MAS for improvement of quantitative traits

Most of the important agronomic traits are polygenic or controlled by multiple QTLs. MAS for the improvement of such traits is a complex and difficult task because it is related to many genes or QTLs involved, QTL x E interaction and epistasis. Usually, each of these genes has a small effect on the phenotypic expression of the trait and expression is affected by environmental conditions. Phenotyping of quantitative traits becomes a complex endeavor consequently, and determining marker-phenotype association becomes difficult as well. Therefore, repeated field tests are required to accurately characterize the effects of the QTLs and to evaluate the stability across environments. The QTL x E interaction reduces the efficiency of MAS and epistasis can result in a skewed QTL effect on the trait.

Despite a tremendous amount of QTL mapping experiments over the past decade, application and utilization of the QTL mapping information in plant breeding has been constrained by a number of factors (Collard and Mackill, 2008):

1. Strong QTL-environmental interaction which make phenotyping difficult since expression may vary from one location/year to another;

2. Lack of universally valid QTL-marker associations applicable across populations. The notion that QTL mapping to identify new QTL markers whenever a new germplasm is used, puts some people off and they lose interest in MAS;

3. Deficiencies in QTL statistical analysis which lead to either overestimation or underestimation of the number of QTLs involved and their effect on the trait;

4. Often times, there are no QTLs with major effects on the trait and this means a large number of QTLs have to be identified and in many cases this becomes a tough goal to achieve and further complicates identification of marker-QTL association.

In order to improve the efficiency of MAS for quantitative traits, appropriate field experimental designs and approaches have to be employed. Attention should be given to replications both over time and space, consistency in experimental techniques, samplings and evaluations, robust data processing and statistical analysis. For example, composite interval mapping (CIM) allows the integration of data from different locations for joint analysis to estimate QTL-environment interaction so that stable QTLs across environments can be identified. A saturated linkage map enables accurate identification of both targeted QTLs as well as linked QTLs in coupling and repulsion linkage phases. In practical breeding for improvement of a quantitative trait, usually not many minor QTLs are considered but only a few major QTLs are used in MAS. In case many QTLs especially minor-effect QTLs are involved, a breeder would prefer to consider the strategy of gene pyramiding (see the later section).

Fusarium head blight (FHB) caused by *Fusarum* species is one of the most destructive diseases in wheat and barley worldwide. To combat this disease, a great effort from multiple fields, including plant breeding and genetics, molecular genetics and genomics, plant pathology, and integrated management, has been dedicated since 1990s. Resistance to HFB in both wheat and barley is quantitatively inherited, and many QTLs have been identified from different resources of germplasm (Buerstmayr et al., 2009). Use of MAS to improve the resistance has become a choice for many breeding programs. In wheat, a major QTL designated as *Fhb1* was consistently detected across multiple environments and populations, and explained 20-40% of phenotypic variation in most cases (Buerstmayr et al., 2009; Jiang et al., 2007a, 2007b). Thus wheat breeders would especially prefer to use this major QTL to develop new cultivars with FHB resistance. Pumphrey et al. (2007) compared 19 pairs of NIL for *Fhb1* derived from an ongoing breeding program and found that the average reduction in disease severity between NIL pairs was 23% for disease severity and 27% for kernel infection. Later investigation from the group also demonstrated successful implementation of MAS for this QTL (Anderson et al. 2007).

In addition, researchers also tried to incorporate multiple QTLs by MAS. Miedaner et al. (2006) demonstrated that MAS for three FHB resistance QTLs simultaneously was highly effective in enhancing FHB resistance in German spring wheat. FHB resistance was the highest in recombinant lines with multiple QTLs combined, especially 3B plus 5A. Jiang et al. (2007a) made a comparison of multiple-locus combinations in a RIL population derived from the cross "Veery x CJ 9306". For three loci, the average levels of resistance from low to high in genotypes were: no favorable allele – one favorable allele – two favorable alleles – three favorable alleles, except for the non-reciprocal comparisons. When four or five loci carrying favorable alleles from the resistant parent CJ 9306 were considered simultaneously, the coefficients of determination between the accumulated effects of alleles for different combinations and the averages of number or percentage of diseased spikelets for the corresponding RILs were 0.33-0.41 (P<0.01) (Jiang et al., 2007a). Therefore, the authors concluded that the effects of FHB resistance QTLs could be accumulated and the resistance could be feasibly enhanced by selection of favorable marker alleles for multiple loci in breeding programs.

In the U.S., the Coordinated Agricultural Projects (CAPs) with aims to encourage collaborative efforts in applied plant genomics and molecular research have been implemented in several crops, such as rice, wheat, barley, beans, potato, tomato, etc. An important strategy CAPs take is applying marker-assisted selection to plant breeding and efficiently using genetic resources and facilities available, including thousands and ten thousands of DNA markers and plant introductions, to facilitate development of crop cultivars with improved yield, resistance and quality.

5. Marker-assisted backcrossing

5.1. MABC procedure and theoretical and practical considerations

Marker-assisted or marker-based backcrossing (MABC) is regarded as the simplest form of marker-assisted selection, and at the present it is the most widely and successfully used

method in practical molecular breeding. MABC aims to transfer one or a few genes/QTLs of interest from one genetic source (serving as the donor parent and maybe inferior agronomically or not good enough in comprehensive performance in many cases) into a superior cultivar or elite breeding line (serving as the recurrent parent) to improve the targeted trait. Unlike traditional backcrossing, MABC is based on the alleles of markers associated with or linked to gene(s)/QTL(s) of interest instead of phenotypic performance of target trait. The general procedure of MABC is as follow, regardless of dominant or recessive nature of the target trait in inheritance:

a. Select parents and make the cross, one parent is superior in comprehensive performance and serves as recurrent parent (RP), and the other one used as donor parent (DP) should possess the desired trait and the DNA markers allele(s) associated with or linked to the gene for the trait.

b. Plant F_1 population and detect the presence of the marker allele(s) at early stages of growth to eliminate false hybrids, and cross the true F_1 plants back to the RP.

c. Plant BCF_1 population, screen individuals for the marker(s) at early growth stages, and cross the individuals carrying the desired marker allele(s) (in heterozygous status) back to the RP. Repeat this step in subsequent seasons for two to four generations, depending upon the practical requirements and operation situations as discussed below.

d. Plant the final backcrossing population (e.g. BC_4F_1), and screen individual plants with the marker(s) for the target trait and discard the individuals carrying homozygous markers alleles from the RP. Have the individuals with required marker allele(s) selfed and harvest them.

e. Plant the progenies of backcrossing-selfing (e.g. BC_4F_2), detect the markers and harvest individuals carrying homozygous DP marker allele(s) of target trait for further evaluation and release.

Theoretically, the proportion of the RP genome after n generations of backcrossing is given by $1 - (1/2)^{n+1}$ for a single locus and $[1 - (1/2)^{n+1}]^k$ for k loci, respectively, for a population large enough in size (or with adequate individuals) and no selection being made during backcrossing (i.e. "blind" backcrossing only). The percentage of the RP genome is the average of the population, with some individuals possessing more of the RP genome than others. To fully recover the genome of the RP, 6-8 generations of backcrossing is needed typically in case no selection is made for the RP. However, this process is usually slower than expected for the target gene-carrier chromosome, i.e. linkage drag, especially in case a linkage exists between the target gene and other undesirable traits. On the other hand, the process of introgression of QTLs/genes and recovery of the RP genome may be accelerated by selection using markers flanking QTLs and evenly spaced markers from other chromosomes (i.e. unlinked to QTLs) of the RP (Collard et al., 2005) or selection for the performance of the RP conducted simultaneously. For MABC program, therefore, there are two types of selection recognized: Foreground selection and background selection (Hospital, 2003).

In foreground selection, the selection is made only for the marker allele(s) of donor parent at the target locus to maintain the target locus in heterozygous state until the final backcrossing is completed. Then the selected plants are selfed and the progeny plants with homozygous DP allele(s) of selected markers are harvested for further evaluation and release. As described above, this is the general procedure of MABC. The effectiveness of foreground selection depends on the number of genes/loci involved in the selection, the marker-gene/QTL association or linkage distance and the undesirable linkage to the target gene/QTL.

In background selection, the selection is made for the marker alleles of recurrent parent in all genomic regions of desirable traits except the target locus, or selection against the undesirable genome of donor parent. The objective is to hasten the restoration of the RP genome and eliminate undesirable genes introduced from the DP. The progress in recovery of the RP genome depends on the number of markers used in background selection. The more markers evenly located on all the chromosomes are selected for the RP alleles, the faster recovery of the RP genome will be achieved but larger population size and more genotyping will be required as well. In addition, the linkage drag also can be efficiently addressed by background selection using DNA markers, although it is difficult to overcome in a traditional backcrossing program.

Foreground selection and background selection are two respective aspects of MABC with different foci of selection. In practice, however, both foreground and background selection are usually conducted in the same program, either simultaneously or successively. In many cases, they can be performed alternatively even in the same generation. The individuals that have the desired marker alleles for target trait are selected first (foreground selection). Then the selected individuals are screened for other marker alleles again for the RP genome (background selection). It is understandable to do so because selection of the target gene/QTL is the essential and only critical point for backcrossing program, and the individuals that do not have the allele of target gene will be discarded and thus it is not necessary to genotype them for other traits.

The efficiency of MABC depends upon several factors, such as the population size for each generation of backcrossing, marker-gene association or the distance of markers from the target locus, number of markers used for target trait and RP background, and undesirable linkage drag. Based on simulations of 1000 replicates, Hospital (2003) presented the expected results of a typical MABC program, in which heterozygotes were selected at the target locus in each generation, and RP alleles were selected for two flanking markers on target chromosome each located 2 cM apart from the target locus and for three markers on non-target chromosomes. As shown in Table 2, a faster recovery of the RP genome could be achieved by MABC with combined foreground and background selection, compared to traditional backcrossing. Therefore, using markers can lead to considerable time savings compared to conventional backcrossing (Frisch et al., 1999; Collard et al., 2005).

Backcross generation	Number of individuals	% homozygosity of recurrent parent alleles at selected markers		% recurrent parent genome	
		Chromosome with target locus	All other chromosomes	Marker-assisted backcross	Conventional backcross
BC_1	70	38.4	60.6	79.0	75.0
BC_2	100	73.6	87.4	92.2	87.5
BC_3	150	93.0	98.8	98.0	93.7
BC_4	300	100.0	100.0	99.0	96.9

Table 2. Expected results of a MABC program with combined foreground and background selection used; Adapted from Hospital (2003).

In a MABC program, the population to be analyzed should contain at least one genotype that has all favorable alleles for a particular QTL. Later, the number of QTLs may be increased progressively, but not beyond six QTLs in most cases because of prohibitive difficulty in handling all QTLs (Hospital, 2003). In addition, the more QTLs/genes are transferred, the larger the proportion of unwanted genes would be due to linkage drag. In general, most of the unwanted genes are located on non-target chromosomes in early BC generations, and are rapidly removed in subsequent BC generations. On the contrary, the quantity of DP genes on the target chromosome decreases much more slowly, and even after generation BC_6 many of the unwanted donor genes are still located on the target chromosome in segregating state (Newbury, 2003). Given a total genome length is 3000 cM, 1% donor DNA fragments after six backcrosses represents a 30 cM chromosomal segment or region, which may host many unwanted genes, especially if the DP is a wild genetic resource. Young and Tanksley (1989) genotyped a collection of tomato varieties in which the resistance gene was previously transferred at the *Tm*-2 locus with RLFP markers. Their data indicated that the size of chromosomal segment retained around the *Tm*-2 locus during backcross breeding was very variable, with one line exhibiting a donor segment of 50 cM after 11 backcrosses and other one possessing 36 cM donor segment after 21 backcrosses. This clearly demonstrates the need for background selection.

As discussed above, linkage drag can be reduced by performing background selection. Typically, two markers flanking the target gene are used, and the individuals (or double recombinants) that are heterozygous at the target locus and homozygous for the recipient (RP) alleles at both flanking markers are selected. Use of closer flanking markers leads to more effective and faster reduction of linkage drag compared to distant markers. However, less distance between two flanking markers implies less probability of double recombination, and thus larger populations and more genotyping are needed. In order to optimize genotyping effort (i.e. the cost of the program), therefore, it is important to determine the minimal population sizes necessary to ensure the desired genotypes can be obtained. Hospital and Decoux (2002) developed a statistical software for determining the minimum population size required in BC program to identify at least one individual that is double-recombinant with heterozygosity at target locus and homozygosity for recurrent parent alleles at flanking

marker loci. In addition, for closely-linked flanking markers, it is unlikely to obtain double recombinant genotypes through only one generation of backcrossing. Therefore, additional backcrossing should be conducted. For instance, in one BC generation (e.g. BC_1) single recombination on one side of the target gene is selection, and single recombination on the other side may be selected in another BC generation (e.g. BC_2) (Young and Tanksley 1989). In this way, individuals with desired RP alleles at two flanking markers and donor allele at target locus can be finally obtained.

To accelerate the recovery of RP genome on non-target chromosomes, scientists suggested using markers in backcrossing and discussed how many makers should be used (Tanksley et al., 1989; Hospital et al. 1992; Visscher et al. 1996). In background selection, the approaches involve selecting individuals that are of homozygous recipient type at a collection of markers located on non-carrier chromosomes. From a point of both effectiveness and efficiency, it is important to determine an appropriate number of markers to be used. More markers do not necessarily mean better benefits in practice. Generally, several markers are involved and MABC should be performed over two or more generations. It is unlikely that the selection objective can be realized in a single BC generation.

Dense marker coverage of non-target chromosomes is not mandatory to increase the overall proportion of recurrent parent genome, unless fine-mapping of specific chromosome regions is highly important. An appropriate number of markers and optimal position on chromosomes are important. Computer simulation suggested that for a chromosome of 100 cM, two to four markers are sufficient, and selection based on markers would be most efficient if the markers are optimally positioned along the chromosomes (Servin and Hospital, 2002). In practice, at least two or three markers per chromosome are needed, and every chromosome should be involved. In such a MABC scheme, three to four generations of backcrossing is generally enough to achieve more than 99% of the recurrent parent genome. With respect to the time necessary to release new varieties, the gain due to background selection can be economically valuable. In addition, background selection is more efficient in late BC generations than in early BC generations. For example, if a BC breeding scheme is conducted over three successive BC generations and yet the preference is to genotype individuals only once, then it is more efficient to genotype and select the individuals in BC_3 generation rather than in the BC_1 generation (Hospital et al. 1992, Ribaut et al. 2002).

5.2. Application of MABC

Success in integrating MABC as a breeding approach lies in identifying situations in which markers offer noticeable advantages over conventional backcrossing or valuable complements to conventional breeding effort. MABC is essential and advantageous when:

1. Phenotyping is difficult and/or expensive or impossible;

2. Heritability of the target trait is low;

3. The trait is expressed in late stages of plant development and growth, such as flowers, fruits, seeds, etc.;

4. The traits are controlled by genes that require special conditions to express;

5. The traits are controlled by recessive genes; and

6. Gene pyramiding is needed for one or more traits.

Among the molecular breeding methods, MABC has been most widely and successfully used in plant breeding up to date. It has been applied to different types of traits (e.g. disease/pest resistance, drought tolerance and quality) in many species, e.g. rice, wheat, maize, barley, pear millet, soybean, tomato, etc. (Collard et al., 2005; Dwivedi et al., 2007; Xu, 2010). In maize, for example, *Bacillus thuringiens* is a bacterium that produces insecticidal toxins, which can kill corn borer larvae when they ingest the toxins in corn cells (Ragot et al. 1995). The integration of the *Bt* transgene into various corn genetic backgrounds has been achieved by using MABC. Aroma in rice is controlled by a recessive gene which is due to an eight base-pair deletion and three single nucleotide polymorphism in a gene that codes for betaine aldehyde dehydrogenase 2 (Bradbury et al., 2005a). This discovery allows identification of the aromatic and non-aromatic rice varieties and discriminates homozygous recessive and dominant as well as heterozygous individuals in segregating population for the trait. MABC has been used to select for aroma in rice (Bradbury et al. 2005b). High lysine *opaque2* gene in corn was incorporated using MABC (Babu et al. 2005). However, the rate of success decreases when large numbers of QTL$_s$ are targeted for introgression. Sebolt et al. (2000) used MABC for two QTL for seed protein content in soybeans. However, only one QTL was confirmed in BC$_3$F$_{4:5}$. When that QTL was introduced in three different genetic backgrounds, it had no effect in one background. In tomato, Tanksley and Nelson (1996) proposed a MABC strategy, called advanced backcross-QTL (AB-QTL), to transfer resistance genes from wild relative/unadapted genotype into elite germplasm. The strategy has proven effective for various agronomically important traits in tomato, including fruit quality and black mold resistance (Tanksley and Nelson, 1996; Bernacchi et al., 1998; Fulton et al., 2002). In addition, AB-QTL has been used in other crop species, such as rice, barley, wheat, maize, cotton and soybean, collectively demonstrating that this strategy is effective in transferring favorable alleles from the wild/unadapted germplasm to elite germplasm (Wang and Chee, 2010; Concibido et al., 2003).

In barley, a marker linked (0.7 cM) to the *Yd2* gene for resistance to barley yellow dwarf virus was successfully used to select for resistance in a backcrossing scheme (Jefferies et al., 2003). Compared to lines without the marker, the BC$_2$F$_2$-derived lines carrying the linked marker had lighter leaf symptoms and higher yield when infected by the virus. In maize, marker-facilitated backcrossing was also successfully employed to improve complex traits such as grain yield. Using MABC, six chromosomal segments each in two elite lines, Tx303 and Oh43, were transferred into two widely used inbred lines, B73 and Mo17, through three generations of backcrossing followed by two selfing generations. Then the enhanced lines with better performance were selected based on initial evaluations of testcross hybrids. The single-cross hybrids of enhanced B73 x enhanced Mo17 out-yielded the check hybrids by 12-15% (Stuber et al., 1999). Zhao et al. (2012) reported that a major quantitative trait locus (named *qHSR1*) for resistance to head smut in maize was successfully integrated into ten high-yielding inbred lines (susceptible to head smut). Each of the ten high-yielding lines was crossed with a donor parent Ji 1037 that contains *qHSR1* and is completely resistant to head smut, followed by five gen-
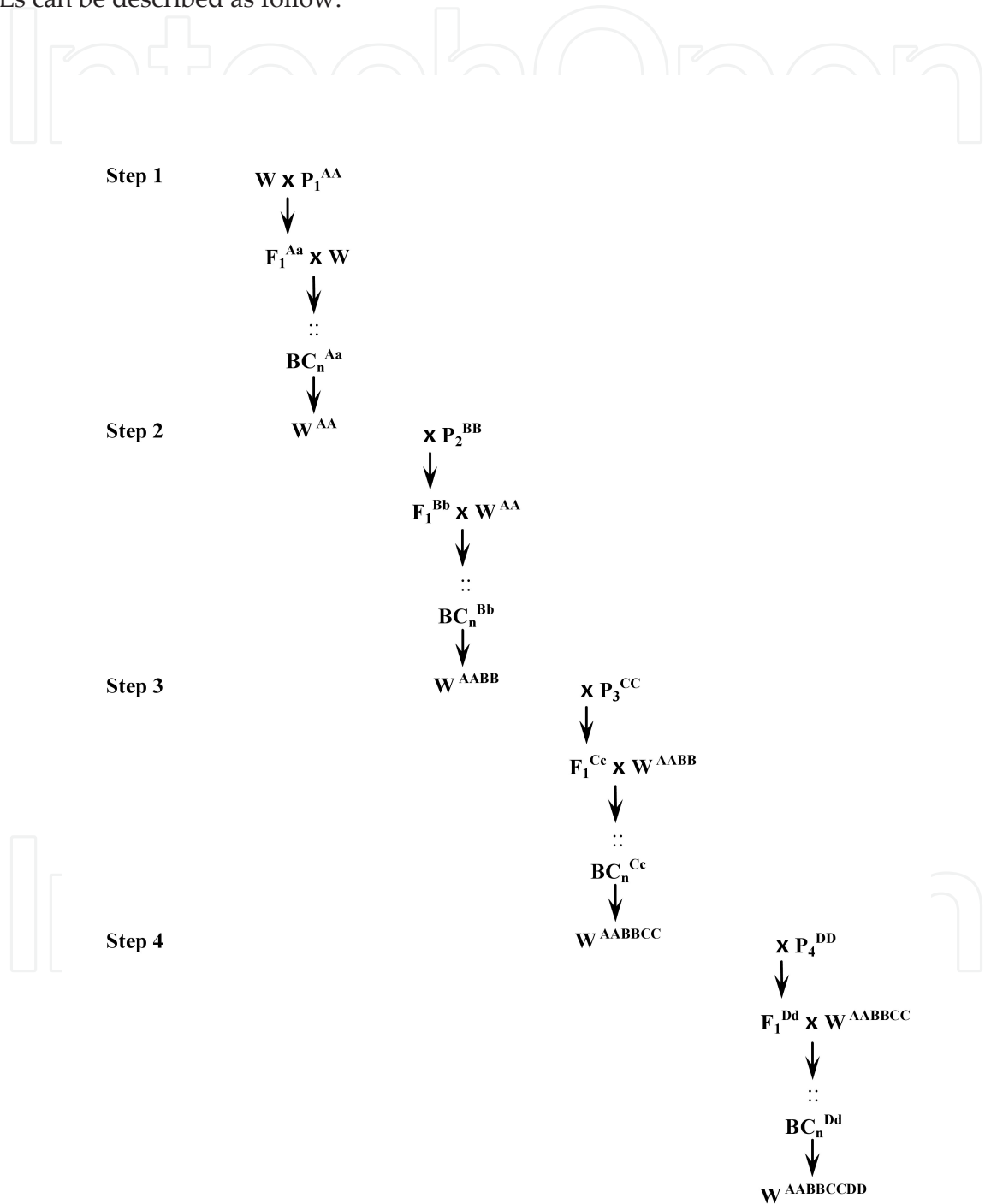
erations of backcrossing to the respective recurrent parents. In BC_1 through BC_3 only phenotypic selection was conducted to identify highly resistant individuals after artificial inoculation. In BC_4 phenotypic selection, foreground selection and recombinant selection were conducted to screen for resistant individuals with the shortest *qHSR1* donor regions. In BC_5, phenotypic selection, foreground selection and background selection were performed to identify resistant individuals with the highest proportion of the recurrent parent genome, followed by one generation of self-pollination to obtain homozygous genotypes at the *qHSR1* locus. The ten improved inbred lines all showed substantial resistance to head smut, and the hybrids derived from these lines also showed a significant increase in the resistance. Semagn et al. (2006b) provided a detail review on the progress and prospects of MABC in crop breeding.

Currently, a cooperative marker-based backcrossing project for high-oleic acid in soybean has been initiated among multiple U.S. land-grant universities and USDA-ARS. Backcrossing and selection will be performed using the markers tightly linked to the high-oleic genes/ loci. Hopefully, the high-oleic (80% or higher) traits will be successfully transferred from mutant lines or derived lines into other locally superior cultivars/lines, or combined with other unique traits like low linolenic acid (Pham et al., 2012).
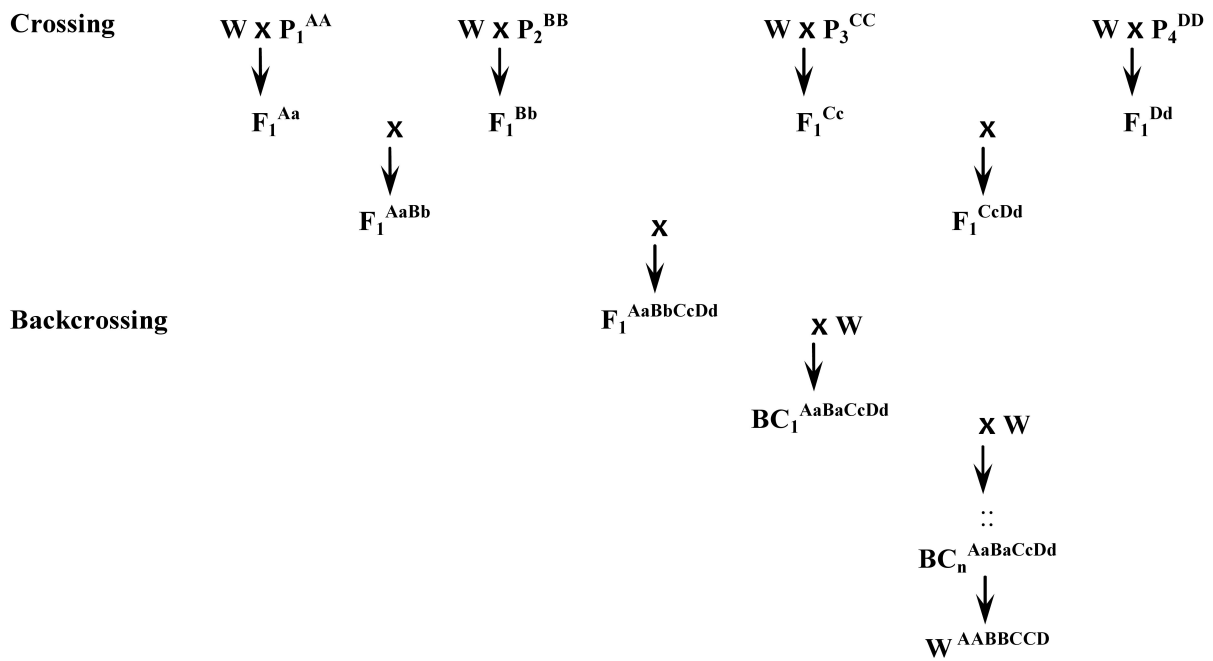
6. Marker-assisted gene pyramiding and marker-assisted recurrent selection

Marker-assisted gene pyramiding (MAGP) is one of the most important applications of DNA markers to plant breeding. Gene pyramiding has been proposed and applied to enhance resistance to disease and insects by selecting for two or more than two genes at a time. For example in rice such pyramids have been developed against bacterial blight and blast (Huang et al., 1997; Singh et al., 2001; Luo et al., 2012). Castro et al. (2003) reported a success in pyramiding qualitative gene and QTLs for resistance to stripe rust in barley. The advantage of using markers in this case allows selecting for QTL-allele-linked markers that have the same phenotypic effect. To enhance or improve a quantitatively inherited trait in plant breeding, pyramiding of multiple genes or QTLs is recommended as a potential strategy (Richardson et al., 2006). The cumulative effects of multiple-QTL pyramiding have been proven in crop species like wheat, barley and soybean (Richardson et al., 2006; Jiang et al., 2007a, 2007b; Li et al., 2010; Wang et al., 2012). Pyramiding of multiple genes/QTLs may be achieved through different approaches: multiple-parent crossing or complex crossing, backcrossing, and recurrent selection. A suitable breeding scheme for MAGP depends on the number of genes/QTLs required for improvement of traits, the number of parents that contain the required genes/QTLs, the heritability of traits of interest, and other factors (e.g. marker-gene association, expected duration to complete the plan and relative cost). Assuming three or four desired genes/QTLs exist separately in three or four lines, pyramiding of them can be realized by three-way, four-way or double crossing. They may also be integrated by convergent backcrossing or stepwise backcrossing. However, if there are more than four genes/QTLs to be pyramided, complex or multiple crossing and/or recurrent selection may be often preferred.
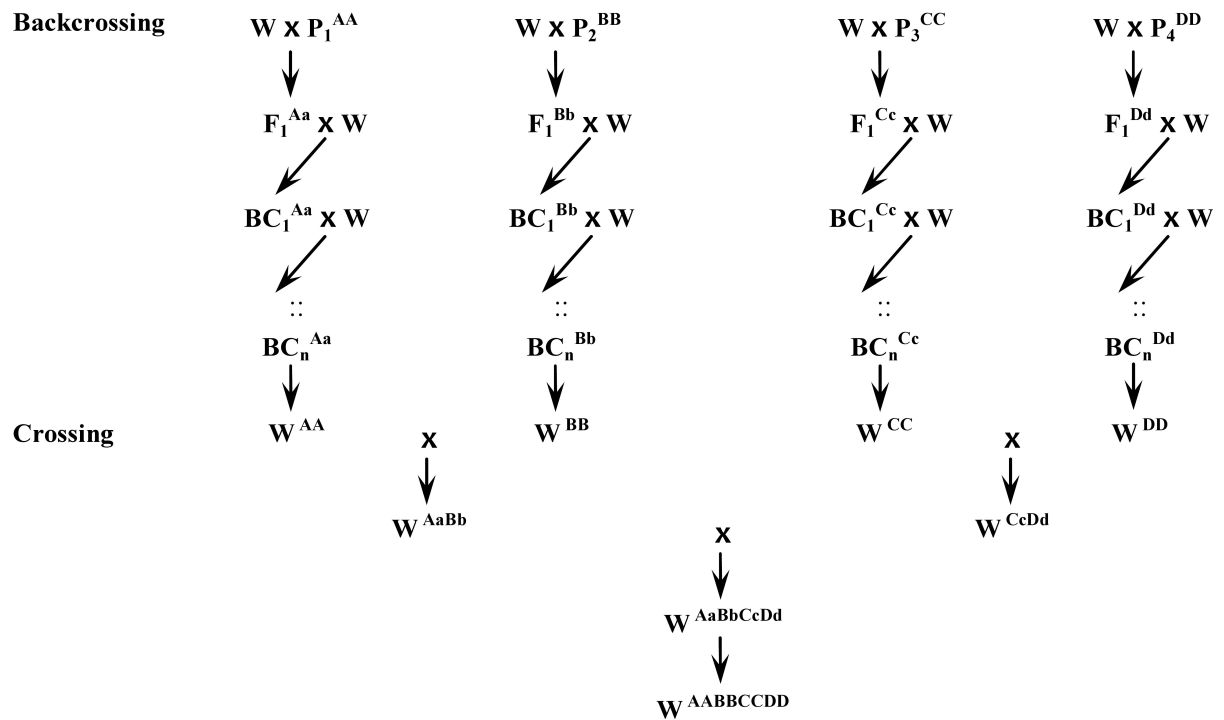
For MABC-based gene pyramiding, in general, there may be three strategies or breeding schemes: stepwise, simultaneous/synchronized and convergent backcrossing or transfer. Supposing one cultivar W is superior in comprehensive performance but lack of a trait of interest, and four different genes/QTLs contributing to the trait have been identified in four germplasm lines (e.g. P1, P2, P3 and P4). Three MABC schemes for pyramiding the genes/ QTLs can be described as follow.

Step 1 $W \times P_1^{AA}$
$\downarrow$
$F_1^{Aa} \times W$
$\downarrow$
$::$
BC_n^{Aa}
$\downarrow$
Step 2 W^{AA} $\times P_2^{BB}$
$\downarrow$
$F_1^{Bb} \times W^{AA}$
$\downarrow$
$::$
BC_n^{Bb}
$\downarrow$
Step 3 W^{AABB} $\times P_3^{CC}$
$\downarrow$
$F_1^{Cc} \times W^{AABB}$
$\downarrow$
$::$
BC_n^{Cc}
$\downarrow$
Step 4 W^{AABBCC} $\times P_4^{DD}$
$\downarrow$
$F_1^{Dd} \times W^{AABBCC}$
$\downarrow$
$::$
BC_n^{Dd}
$\downarrow$
$W^{AABBCCDD}$

Scheme 1. Stepwise Backcrossing

Crossing

$W \times P_1{}^{AA}$ $W \times P_2{}^{BB}$ $W \times P_3{}^{CC}$ $W \times P_4{}^{DD}$

$F_1{}^{Aa}$ **x** $F_1{}^{Bb}$ $F_1{}^{Cc}$ **x** $F_1{}^{Dd}$

$F_1{}^{AaBb}$ $F_1{}^{CcDd}$

x

$F_1{}^{AaBbCcDd}$

Backcrossing

x W

$BC_1{}^{AaBaCcDd}$

x W

::

$BC_n{}^{AaBaCcDd}$

$W^{AABBCCD}$

Scheme 2. Simultaneous/Synchronized Backcrossing

Backcrossing

$W \times P_1{}^{AA}$ $W \times P_2{}^{BB}$ $W \times P_3{}^{CC}$ $W \times P_4{}^{DD}$

$F_1{}^{Aa}$ **x W** $F_1{}^{Bb}$ **x W** $F_1{}^{Cc}$ **x W** $F_1{}^{Dd}$ **x W**

$BC_1{}^{Aa}$ **x W** $BC_1{}^{Bb}$ **x W** $BC_1{}^{Cc}$ **x W** $BC_1{}^{Dd}$ **x W**

::   ::   ::   ::

$BC_n{}^{Aa}$ $BC_n{}^{Bb}$ $BC_n{}^{Cc}$ $BC_n{}^{Dd}$

Crossing

W^{AA} W^{BB} W^{CC} W^{DD}

x   **x**

W^{AaBb} W^{CcDd}

x

$W^{AaBbCcDd}$

$W^{AABBCCDD}$

Scheme 3. Convergent Backcrossing

In the stepwise backcrossing, four target genes/QTLs are transferred into the recurrent parent W in order. In one step of backcrossing, one gene/QTL is targeted and selected, followed by next step of backcrossing for another gene/QTL, until all target genes/QTLs have been introgressed into the RP. The advantage is that gene pyramiding is more precise and easier to implement as it involves only one gene/QTL at one time and thus the population size and genotyping amount will be small. The improved recurrent parent may be released before the final step as long as the integrated genes/QTLs (e.g. two or three) meet the requirement at that time. The disadvantage is that it takes a longer time to complete. In the simultaneous or synchronized backcrossing, the recurrent parent W is first crossed to each of four donor parents to produce four single-cross F_1s. Two of the four single-cross F_1s are crossed with each other to produce two double-cross F_1s, and these two double-cross F_1s are crossed again to produce a hybrid integrating all four target genes/QTLs in heterozygous state. The hybrid and/or progeny with heterozygous markers for all four target genes/QTLs is subsequently crossed back to the RP W until a satisfactory recovery of the RP genome, and finalized by one generation of selfing. The advantage of this method is that it takes the shortest time to complete. However, in the backcrossing all target genes/QTLs are involved at the same time and thus it requires a large population and more genotyping. Convergent backcrossing is a strategy combining the advantages of stepwise and synchronized backcrossing. First the four target gene/QTLs are transferred separately from the donors into the recurrent parent W by single crossing followed by backcrossing based on markers linked to the target genes/QTLs, to produce four improved lines (W^{AA}, W^{BB}, W^{CC}, and W^{DD}). Two of the improved lines are crossed with each other and the two hybrids are then intercrossed to integrate all four genes/QTLs together and develop the final improved line with all four genes/QTLs pyramided (i.g. $W^{AABBCCDD}$). Relatively speaking, convergent backcrossing is more acceptable because in this scheme not only is time reduced (compared to stepwise transfer) but gene fixation and/or pyramiding is also more easily assured (compared to simultaneous transfer).

Theoretical issues and efficiency of MABC for gene pyramiding have been investigated through computer simulations (Ribaut et al., 2002; Servin et al., 2004; Ye and Smith, 2008). Practical application of MABC to gene pyramiding has been reported in many crops, including rice, wheat, barley, cotton, soybean, common bean and pea, especially for developing durable resistance to stresses in crops. However, there is very limited information available about the release of commercial cultivars resulted from this strategy. Somers et al. (2005) implemented a molecular breeding strategy to introduce multiple pest resistance genes into Canadian wheat. They used high throughput SSR genotyping and half-seed analysis to process backcrossing and selection for six FHB resistance QTLs, plus orange blossom wheat midge resistance gene *Sm1* and leaf rust resistance gene *Lr21*. They also used 45-76 SSR markers to perform background selection in backcrossing populations to accelerate the restoration of the RP genetic background. This strategy resulted in 87% fixation of the elite genetic background at the BC_2F_1 on average and successfully introduced all (up to 4) of the chromosome segments containing FHB, *Sm1* and *Lr21* resistance genes in four separate crosses(Somers et al., 2005). Joshi and Nayak (2010) and Xu (2010) recently reviewed the techniques and practical cases in marker-based gene pyramiding.

Similar to the simultaneous/synchronized backcrossing scheme, marker-assisted complex or convergent crossing (MACC) can be undertaken to pyramid multiple genes/QTLs. In particular, MACC is a proper option of breeding schemes for gene pyramiding if all the parents are improved cultivars or lines with good comprehensive performance and have different or complementary genes or favorable alleles for the traits of interest. The difference from simultaneous backcrossing is that selfing hybrid and progenies replaces backcrossing hybrid to the recurrent parent. In MACC, the hybrid of convergent crossing is subsequently self-pollinated and marker-based selection for target traits is performed for several consecutive generations until genetically stable lines with desired marker alleles and traits have been developed. In order to reduce population size and to avoid loss of most important genes/QTLs, different markers may be used and selected in different generations, depending on their relative importance. The markers for the most important genes/QTLs can be detected and selected first in early generations and less important markers later. Once homozygous alleles of the markers for a gene/locus are detected, they may not be necessarily detected again in the subsequent generations. Instead, phenotypic evaluation should be conducted if conditions permit.

Using markers to select or pyramid for multiple genes/QTLs is more complex and less proven. Recurrent selection is widely regarded as an effective strategy for the improvement of polygenic traits. However, the effectiveness and efficiency of selection are not so satisfactory in some cases because phenotypic selection is highly dependent upon environments and genotypic selection takes a longer time (2-3 crop seasons at least for one cycle of selection). Marker-assisted recurrent selection (MARS) is a scheme which allows performing genotypic selection and intercrossing in the same crop season for one cycle of selection (Fig. 1). Therefore, MARS could enhance the efficiency of recurrent selection and accelerate the progress of the procedure (Jiang et al., 2007a), particularly helps in integrating multiple favorable genes/QTLs from different sources through recurrent selection based on a multiple-parental population.

Figure 1. General procedure of marker-assisted recurrent selection (MARS)

For complex traits such as grain yield, biotic and abiotic resistance, MARS has been proposed for "forward breeding" of native genes and pyramiding multiple QTLs (Ragot et al., 2000; Ribaut et al., 2000, 2010; Eathington, 2005; Crosbie et al., 2006). As defined by Ribaut et al. (2010), MARS is a recurrent selection scheme using molecular markers for the identification and selection of multiple genomic regions involved in the expression of complex traits to assemble the best-performing genotype within a single or across related populations. Johnson (2004) presented an example to demonstrate the efficiency of MARS for quantitative traits. In their maize MARS programs, a large-scale use of markers in bi-parental populations, first for QTL detection and then for MARS on yield (i.e. rapid cycles of recombination and selection based on associated markers for yield), could allow increased efficiency of long-term selection by increasing the frequency of favorable alleles (Johnson, 2004). Eathington (2005) and Crosbie et al. (2006) also indicated that the genetic gain achieved through MARS in maize was about twice that of phenotypic selection (PS) in some reference populations. In upland cotton, Yi et al. (2004) reported significant effectiveness of MARS for resistance to *Helicoverpa armigera*. The mean levels of resistance in improved populations after recurrent selection were significantly higher than those of preceding populations.

7. Genomic selection

Genomic selection (GS) or genome-wide selection (GWS) is a form of marker-based selection, which was defined by Meuwissen (2007) as the simultaneous selection for many (tens or hundreds of thousands of) markers, which cover the entire genome in a dense manner so that all genes are expected to be in linkage disequilibrium with at least some of the markers. In GS genotypic data (genetic markers) across the whole genome are used to predict complex traits with accuracy sufficient to allow selection on that prediction alone. Selection of desirable individuals is based on genomic estimated breeding value (GEBV) (Nakaya and Isobe, 2012), which is a predicted breeding value calculated using an innovative method based on genome-wide dense DNA markers (Meuwissen et al., 2001). GS does not need significant testing and identifying a subset of markers associated with the trait (Meuwissen et al., 2001). In other words, QTL mapping with populations derived from specific crosses can be avoided in GS. However, it does first need to develop GS models, i.e. the formulae for GEBV prediction (Nakaya and Isobe, 2012). In this process (training phase), phenotypes and genome-wide genotypes are investigated in the training population (a subset of a population) to predict significant relationships between phenotypes and genotypes using statistical approaches. Subsequently, GEBVs are used for the selection of desirable individuals in the breeding phase, instead of the genotypes of markers used in traditional MAS. For accuracy of GEBV and GS, genome-wide genotype data is necessary and require high marker density in which all quantitative trait loci (QTLs) are in linkage disequilibrium with at least one marker.

GS can be possible only when high-throughput marker technologies, high-performance computing and appropriate new statistical methods become available. This approach has be-

come feasible due to the discovery and development of large number of single nucleotide polymorphisms (SNPs) by genome sequencing and new methods to efficiently genotype large number of SNP markers. As suggested by Goddard and Hayes (2007), the ideal method to estimate the breeding value from genomic data is to calculate the conditional mean of the breeding value given the genotype at each QTL. This conditional mean can only be calculated by using a prior distribution of QTL effects, and thus this should be part of the research to implement GS. In practice, this method of estimating breeding values is approximated by using the marker genotypes instead of the QTL genotypes, but the ideal method is likely to be approached more closely as more sequence and SNP data are obtained (Goddard and Hayes, 2007).

Since the application of GS was proposed by Meuwissen et al. (2001) to breeding populations, theoretical, simulation and empirical studies have been conducted, mostly in animals (Goddard and Hayes, 2007; Jannink et al., 2010). Relatively speaking, GS in plants was less studied and large-scale empirical studies are not available in public sectors for plant breeding (Jannink et al., 2010), but it has attracted more and more attention in recent years (Bernardo, 2010; Bernardo and Yu, 2007; Guo et al., 2011; Heffner et al., 2010, 2011; Lorenzana and Bernardo, 2009; Wong and Bernardo, 2008; Zhong et al., 2009). Studies indicated that in all cases, accuracies provided by GS were greater than might be achieved on the basis of pedigree information alone (Jannink et al., 2010). In oil palm, for a realistic yet relatively small population, GS was superior to MARS and PS in terms of gain per unit cost and time (Wong and Bernardo, 2008). The studies have demonstrated the advantages of GS, suggesting that GS would be a potential method for plant breeding and it could be performed with realistic sizes of populations and markers when the populations used are carefully chosen (Nakaya and Isobe, 2012).

GS has been highlighted as a new approach for MAS in recent years and is regarded as a powerful, attractive and valuable tool for plant breeding. However, GS has not become a popular methodology in plant breeding, and there might be a far way to go before the extensive use of GS in plant breeding programs. The major reason might be the unavailability of sufficient knowledge of GS for practical use (Nakaya and Isobe, 2012). Statistics and simulation discussed in terms of formulae in GS studies are most likely too specific and hard for plant breeders to understand and to use in practical breeding programs. From a plant breeder's point of view, GS can be practicable for a few breeding populations with a specific purpose, but may be impractical for a whole breeding program dealing with hundreds and thousands of crosses/populations at the same time. Therefore, GS must shift from theory to practice, and its accuracy and cost effectiveness must be evaluated in practical breeding programs to provide convincing empirical evidence and warrant a practicable addition of GS to a plant breeder's toolbox (Heffner et al., 2009). Development of easily understandable formulae for GEBVs and user-friendly software packages for GS analysis is helpful in facilitating and enhancing the application of GS in plant breeding. Kumpatla et al. (2012) recently presented an overall review on the GS for plant breeding.

8. Marker-based breeding and conventional breeding: Challenges and perspectives

Marker-assisted breeding became a new member in the family of plant breeding as various types of molecular markers in crop plants were developed during the 1980s and 1990s. The extensive use of molecular markers in various fields of plant science, e.g. germplasm evaluation, genetic mapping, map-based gene discovery, characterization of traits and crop improvement, has proven that molecular technology is a powerful and reliable tool in genetic manipulation of agronomically important traits in crop plants. Compared with conventional breeding methods, MAB has significant advantages:

a. MAB can allow selection for all kinds of traits to be carried out at seedling stage and thus reduce the time required before the phenotype of an individual plant is known. For the traits that are expressed at later developmental stages, undesirable genotypes can be quickly eliminated by MAS. This feature is particularly important and useful for some breeding schemes such as backcrossing and recurrent selection, in which crossing with or between selected individuals is required.

b. MAB can be not affected by environment, thus allowing the selection to be performed under any environmental conditions (e.g. greenhouse and off-season nurseries). This is very helpful for improvement of some traits (e.g. disease/pest resistance and stress tolerance) that are expressed only when favorable environmental conditions present. For low-heritability traits that are easily affected by environments, MAS based on reliable markers tightly linked to the QTLs for traits of interest can be more effective and produce greater progress than phenotypic selection.

c. MAB using co-dominance markers (e.g. SSR and SNP) can allow effective selection of recessive alleles of desired traits in the heterozygous status. No selfing or test crossing is needed to detect the traits controlled by recessive alleles, thus saving time and accelerating breeding progress.

d. For the traits controlled by multiple genes/QTLs, individual genes/QTLs can be identified and selected in MAB at the same time and in the same individuals, and thus MAB is particularly suitable for gene pyramiding. In traditional phenotypic selection, however, to distinguish individual genes/loci is problematic as one gene may mask the effect of additional genes.

e. Genotypic assays based on molecular markers may be faster, cheaper and more accurate than conventional phenotypic assays, depending on the traits and conditions, and thus MAB may result in higher effectiveness and higher efficiency in terms of time, resources and efforts saved.

The research and use of MAB in plants has continued to increase in the public and private sectors, particularly since 2000s. However, MAS and MABC were and are primarily constrained to simply-inherited traits, such as monogenic or oligogenic resistance to diseases/pests, although quantitative traits were also involved (Collard and Mackill, 2008; Segmagn

et al., 2006; Wang and Chee, 2010). The application of molecular markers in plant breeding has not achieved the results as expected previously in terms of extent and success (e.g. release of commercial cultivars). Collard and Mackill (2008) listed ten reasons for the low impact of MAS and MAB in general. Improvement of most agronomic traits that are of complicated inheritance and economic importance like yield and quality is still a great challenge for MAB including the newly developed GS. From the viewpoint of a plant breeder, MAB is not universally or necessarily advantageous. The application of molecular technologies to plant breeding is still facing the following drawbacks and/or challenges:

a. Not all markers are breeder-friendly. This problem may be solved by converting of non-breeder-friendly markers to other types of breeder-friendly markers (e.g. RFLP to STS, sequence tagged site, and RAPD to SCAR, sequence characterized amplified region).

b. Not all markers can be applicable across populations due to lack of marker polymorphism or reliable marker-trait association. Multiple mapping populations are helpful in understanding marker allelic diversity and genetic background effects. In addition, QTL positions and effects also need to be validated and re-estimated by breeders in their specific germplasm (Heffner et al., 2009).

c. False selection may occur due to recombination between the markers and the genes/QTLs of interest. Use of flanking markers or more markers for the target gene/QTL can help.

d. Imprecise estimates of QTL locations and effects result in slower progress than expected. The efficiency of QTL detection is attributed to multiple factors, such as algorithms, mapping methods, number of polymorphic markers, and population type and size (Wang et al., 2012). High marker density fine mapping with large populations and well-designed phenotyping across multiple environments may provide more accurate estimates of QTL location and effects.

e. A large number of breeding programs have not been equipped with adequate facilities and conditions for a large-scale adoption of MAB in practice.

f. The methods and schemes of MAB must be easily understandable, acceptable and implementable for plant breeders, unless they are not designed for a large scale use in practical breeding programs.

g. Higher startup expenses and labor costs.

With a long history of development, especially since the fundamental principles of inheritance were established in the late 19th and early 20th centuries, plant breeding has become an important component of agricultural science, which has features of both science and arts. Conventional breeding methodologies have extensively proven successful in development of cultivars and germplasm. However, subjective evaluation and empirical selection still play a considerable role in conventional breeding. Scientific breeding needs less experience and more science. MAB has brought great challenges, opportunities and prospects for conventional breeding. As a new member of the whole family of plant breeding, however, MAB, as transgenic breeding or genetic manipulation does, cannot replace conventional

breeding but is and only is a supplementary addition to conventional breeding. High costs and technical or equipment demands of MAB will continue to be a major obstacle for its large-scale use in the near future, especially in the developing countries (Collard and Mackill, 2008; Ribaut et al., 2010). Therefore, integration of MAB into conventional breeding programs will be an optimistic strategy for crop improvement in the future. It can be expected that the drawbacks of MAB will be gradually overcome, as its theory, technology and application are further developed and improved. This should lead to a wide adoption and use of MAB in practical breeding programs for more crop species and in more countries as well.

Author details

Guo-Liang Jiang

Plant Science Department, South Dakota State University, Brookings, USA

References

[1] Anderson, J.A., S. Chao, and S. Liu, 2007: Molecular breeding using a major QTL for Fusarium head blight resistance in wheat. Crop Sci. 47, S-112–119.

[2] Babu, R., S.K. Nair, B.M. Prasanna, and H.S. Gupta. 2004. Integrating marker-assisted selection in crop breeding – Prospects and challenges. Current Science 87: 607-619.

[3] Babu, R., S.K. Nair, A. Kumar, S. Venkatesh, J.C. Sekhar, N.N Singh, G. Srinivasan, and H.S. Gupta. 2005. Two-generation marker-aided backcrossing for rapid conversion of normal maize lines to quality protein maize (QPM). Theor. Appl. Genet. 111: 888-897.

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

[5] Bernardo, R. 2010. Genomewide selection with minimal crossing in self-pollinated crops. Crop Sci. 50: 624-627.

[6] Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47: 1082-1090.

[7] Bilyeu, K., L. Palavalli, D.A. Sleper, and P. Beuselinck. 2006. Molecular genetic resources for development of 1% linolenic acid soybeans. Crop Sci. 46: 1913-1918.

[8] Bonnett, D.G., G.J. Rebetzke, and W. Spielmeyer. 2005. Strategies for efficient implementation of molecular markers in wheat breeding. Mol. Breeding 15: 75-85.

[9]   Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32: 314-331.

[10] Bradbury, L.M.T., T.L. Fitzgerald, R.J. Henry, Q. Jin, and Waters, D.L.E. (2005a). The gene for fragrance in rice. Plant Biotech. J. 3: 363–370.

[11] Bradbury, L.M.T., R.J. Henry, Q. Jin, R.F. Reinke, and Waters, D.L.E. (2005b). A perfect marker for fragrance genotyping in rice. Mol. Breeding 16:279-283.

[12] Buerstmayr, H., T. Ban, and J.A. Anderson. 2009. QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: a review. Plant Breeding 128: 1-26.

[13] Castro, A.J., F. Capettini, A.E. Corey, T. Filichkina, P.M. Hayes, A. Kleinhofs, D. Kudrna, K. Richardson, S. Sandoval-Islas, C. Rossi, and H. Vivar. 2003. Mapping and pyramiding of qualitative and quantitative resistance to stripe rust in barley. Theor. Appl. Genet. 107: 922-930.

[14] Collard, B.C.Y., and D.J. Mackill. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Phil. Trans. R. Soc. B 363: 557-572.

[15] Collard, B.C.Y., M.Z.Z. Jahufer, JB Brouwer, and E.C.K. Pang. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142: 169-196.

[16] Concibido, V.C., B. La Vallee, P. Mclaird, N. Pineda, J. Meyer, L. Hummel, J. Yang, K. Wu, and X. Delannay. 2003. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor. Appl. Genet. 106: 575-582.

[17] Cregan, P.B., J. Mudge, E.W Fickus, D. Danesh, R. Denny, and N.D. Young. 1999. Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. Theor. Appl. Genet. 99: 811-818.

[18] Crosbie, T.M., S.R. Eathington, G.R. Johnson, M. Edwards, R. Reiter, S. Stark, R.G. Mohanty, M. Oyervides, R.E. Buehler, A.K. Walker, R. Dobert, X. Delannay, J.C. Pershing, M.A. Hall, and K.R. Lamkey. 2006. Plant breeding: past, present and future. In: K.R. Lamkey and M. Lee (eds.), Plant Breeding: the Arnel R. Hallauer International Symposium, Blackwell Publishing, Oxford, UK, pp. 3-50.

[19] Dwivedi, S.L., J.H. Crouch, D.J. Mackill, Y. Xu, M.W. Blair, M. Ragot, H.D. Upadhyaya, and R. Oritiz. 2007. The molecularization of public sector crop breeding: progress, problems and prospects. Advances in Agronomy 95: 163-318.

[20] Eathington, S.R. 2005.Practical applications of molecular technology in the development of commercial maize hybrids. In: Proceedings of the 60th Annual Corn and Sorghum Seed Research Conferences, American Seed Trade Association, Washington, D.C.

[21] Edwards, D., J.W. Forster, D. Chagne, and J. Batley. 2007. What is SNPs? In: N.C. Or-aguzie, E.H.A. Rikkerink, S.E. Gardiner and H.N. de Silva (eds.), Association Mapping in Plants. Springer, Berlin, pp 41-52.

[22] Farooq, S., and F. Azam. 2002a. Molecular markers in plant breeding-I: Concepts and characterization. Pakistan J Biol. Sci. 5: 1135-1140.

[23] Farooq, S., and F. Azam. 2002b. Molecular markers in plant breeding-II: Some pre-requisites for use. Pakistan J Biol. Sci. 5: 1141-1147.

[24] Flint-Garcia, S.A., L.L. Darrah, M.D. McMullen, and B.E. Hibbard. 2003. Phenotypic versus marker-assisted selection for stalk strength and second-generation European corn borer resistance in maize. Theor. Appl. Genet. 107: 1331-1336.

[25] Frisch, M., M. Bohn, and A.E. Melchinger. 1999. Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci. 39: 1295-1301 (See also Errata, Frisch et al., 1999. Crop Sci. 39: 1903).

[26] Fulton, T.M., P. Bucheli, E. Voirol, J. Lopez, V. Peetiard, and S.D. Tanksley. 2002. Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavor, identified in four advanced backcross pop-ulations of tomato. Euphytica 127: 163-177.

[27] Goddard, M.E., and B.J. Hayes. 2007. Genomic selection. J. Anim. Breed. Genet..124: 323-330.

[28] Guo, Z., D.M. Tucker, J. Liu, V. Kishore, and G. Gay. 2011. Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor. Appl. Genet. 124: 261-275.

[29] Gupta, P.K., J.K. Roy, and M. Prasad. 2001. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Current Science 80: 524-535.

[30] Heffner, E.L., M.E. Sorrells, and Jannink, J.-L. 2009. Genomic selection for crop im-provement. Crop Sc. 49: 1-12.

[31] Heffner, E.L., A.J. Lorenz, Jannink, J.-L. and M.E. Sorrells. 2010. Plant breeding with genomic selection: gain per unit time and cost. Crop Sci. 50: 1681-1690.

[32] Heffner, E.L., Jannink, J.-L., H. Iwata, E. Souza, and M.E. Sorrells. 2011. Genomic se-lection accuracy for grain quality traits in biparental wheat populations. Crop Sci. 51: 2597-2606.

[33] Hospital, F. 2003.Marker-assisted breeding. In: H.J. Newbury (ed.), Plant molecular breeding. Blackwell Publishing and CRC Press, Oxford and Boca Raton, pp. 30-59.

[34] Hospital, F., C. Chevalet, and P. Mulsant. 1992. Using markers in gene introgression breeding programs. Genetics 231: 1199-1210.

[35] Hospital, F., and G. Decoux. 2002. Popmin: a program for the numerical optimization of population sizes in marker-assisted backcross breeding programs. Journal of Heredity 93: 383-384.

[36] Huang, N., E.R. Angeles, J. Domingo, G. Magpantay, S. Singh, G. Zhang, N. Kumaravadivel, J. Bennet, and G.S. Khush. 1997. Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor. Appl. Genet. 95: 313-320.

[37] Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics 9: 166-177.

[38] Jefferies, S.P., B.J. King, A.R. Barr, P. Warner, S.J. Logue, and P. Langridge. 2003. Marker-assisted backcross introgression of the *Yd2* gene conferring resistance to barley yellow dwarf virus in barley. Plant Breeding 122: 52-56.

[39] Jiang, G.-L., Shi, J., & Ward, R. W. (2007a). QTL analysis of resistance to Fusarium head blight in the novel wheat germplasm CJ 9306. I. Resistance to fungal spread. Theor. Appl. Genet. 116: 3-13.

[40] Jiang, G.-L., Dong, Y., Shi, J., & Ward, R. W. (2007b). QTL analysis of resistance to *Fusarium* head blight in the novel wheat germplasm CJ 9306. II. Resistance to deoxynivalenol accumulation and grain yield loss. Theor. Appl. Genet. 115: 1043-1052.

[41] Johnson, R. 2004. Marker assisted selection. In: J. Jannick (Ed.), Plant Breed. Rev. Vol. 24, Part 1, 293-309.

[42] Joshi, R.K., and S. Nayak. 2010. Gene pyramiding – A broad spectrum technique for developing durable stress resistance in crops. Biotechol. Mol. Biol. Rev. 5: 51-60.

[43] Konieczny, A., and F. Ausubel. 1993. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR based markers. The Plant Journal 4: 403-410.

[44] Kumpatla, S.P., R. Buyyarapu, I.Y. Abdurakhmonov, and J.A. Mammadov. 2012. Genomics-assisted plant breeding in the 21st century: technological advances and progress. In: I.Y. Abdurakhmonov (ed.), Plant Breeding, InTech, pp 131-184.

[45] Lecomte, L., P. Duffé, M. Buret, B. Servin, F. Hospital, and M. Causse. 2004. Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. Theor. Appl. Genet. 109: 568-668.

[46] Li, X., Y. Han, W. Teng, S. Zhang, K. Yu, V. Poysa, T. Anderson, J. Ding, and W. Li. 2010. Pyramided QTL underlying tolerance to Phytophthora root rot in mega-environment from soybean cultivar 'Conrad' and 'Hefeng 25'.Theor. Appl. Genet. 121: 651-658.

[47] Liu, B.-H. 1991. Development and prospects of dwarf male-sterile wheat. Chinese Science Bulletin, 36(4), 306.

[48] Lorenzana, R.F., and R. Bernardo. 2009. Accuracy of genotypic predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120: 151-161.

[49] Luo, Y., J.S. Sangha, S. Wang, Z. Li, J. Yang, and Z. Yin. 2012. Marker-assisted breeding of *Xa4*, *Xa21* and *Xa27* in the restorer lines of hybrid rice for broad-spectrum and enhanced disease resistance to bacterial blight. Mol. Breeding DOI 10.1007/s11032-012-9742-7.

[50] Meuwissen, T. 2007. Genomic selection: marker assisted selection on a genome wide scale. J. Anim. Breed. Genet. 124: 321-322.

[51] Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome wide dense marker maps. Genetics 157: 1819-1829.

[52] Miedaner, T., F. Wilde, B. Steiner, H. Buerstmayr, V. Korzun, and E. Ebmeyer, 2006. Stacking quantitative trait loci (QTL) for Fusarium head blight resistance from non-adapted sources in an European elite spring wheat background and assessing their effects on deoxynivalenol (DON) content and disease severity. Theor. Appl. Genet. 112: 562-569.

[53] Moreau, L., A. Charcosset, F. Hospital, and A. Gallais. 1998. Marker-assisted selection efficiency in populations of finite size. Genetics 148: 1353-1365.

[54] Mullis, K. 1990. The unusual origin of the polymerase chain reaction. Scientific American 262 (4): 56–61, 64–65.

[55] Nakaya, A., and S.N. Isobe. 2012. Will genomic selection be a practical method for plant breeding? Annals of Botany, DOI:10.1093/aob/mcs109.

[56] Newbury, H.J. 2003. Plant Molecular Breeding. Blackwell Publishing, CRC Press, Birmingham, UK.

[57] Payne, P.I., M.A. Nigtingale, A.F. Krattiger, and L.M. Holt. 1987. The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. Journal of the Science of Food and Agriculture 40: 51-65.

[58] Pham, A.-T., J.G. Shannon, and K.D. Bilyeu. 2012. Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. Theor. Appl. Genet. 125: 503–515.

[59] Pumphrey, M. O., R. Bernardo, and J. A. Anderson, 2007: Validating the *Fhb1* QTL for Fusarium head blight resistance in near-isogenic wheat lines developed from breeding populations. Crop Sci. 47: 200-206.

[60] Ragot, M., M. Biasiolli, M.F. Dekbut, A. Dell'orco, L. Margarini, P. Thevenin, J. Vernoy, J. Vivant, R. Zimmermann, and G. Gay. 1995. Marker-assisted backcrossing: a practical example. In: A. Berville and M. Tersac (eds.), Les Colloques, No. 72, INRA, Paris, 45-46.

[61] Ragot, M., G. Gay, J.P. Muller, and J. Durovray. 2000. Efficient selection for the adaptation to the environment through QTL mapping and manipulation in maize. In: J.M.

Ribaut and D. Poland (eds.), Molecular Approaches for the Genetic Improvement of Cereals for Stable Production in Water-limited Environments. CIMMYT, Mexico, pp 128-130.

[62] Ribaut, J.M., and J. Betran. 1999. Single large-scale marker-assisted selection (SLS-MAS). Mol. Breeding 5: 531-541.

[63] Ribaut, J.M., G. Edmeades, E. Perotti, and D. Hoisington. 2000. QTL analysis, MAS results and perspectives for drought-tolerance improvement in tropical maize. In: J.M. Ribaut and D. Poland (eds.), Molecular Approaches for the Genetic Improvement of Cereals for Stable Production in Water-limited Environments. CIMMYT, Mexico, pp 131-136.

[64] Ribaut, J.M., C. Jiang, and D. Hoisington. 2002. Simulation experiments on efficiencies of gene introgression by backcrossing. Crop Sci. 42: 557-565.

[65] Ribaut, J.M., M.C. de Vicente, and X. Delannay. 2010. Molecular breeding in developing countries: challenges and perspectives. Current Opinion in Plant Biology 13: 1-6.

[66] Richardson, K.L., M.I. Vales, J.G. Kling, C.C. Mundt, and P.M. Hayes. 2006. Pyramiding and dissecting disease resistance QTL to barley stripe rust. Theor. Appl. Genet. 113: 485-495.

[67] Rogers, W.J., P.I. Payne, and Harinder K. (1989). The HMW glutenin subunit and gliadin compositions of German-Grown wheat varieties and their relationship with bread-making quality. Plant Breeding 103: 89-100.

[68] Sebolt, A.M., R.C. Shoemaker, and B.W. Diers. 2000. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci. 40: 1438-1444.

[69] Semagn, K., A. Bjornstad, and M.N. Ndjiondjop. (2006a). An overview of molecular marker methods for plants. Afr. J. Biotechnol. 5: 2540-2568.

[70] Semagn, K., A. Bjornstad, and M.N. Ndjiondjop. (2006b). Progress and prospects of marker assisted backcrossing as a tool in crop breeding programs. Afr. J. Biotechnol. 5: 2588-2603.

[71] Servin, B., and F. Hospital. 2002. Optimal positioning of markers to control genetic background in marker-assisted backcrossing. The Journal of Heredity 93: 214-217.

[72] Servin, B., O.C. Martin, M. Mezard, and F. Hospital. 2004. Toward a theory of marker-assisted gene pyramiding. Genetics 168: 513-523.

[73] Shi, A., P. Chen, D. Li, C. Zheng, B. Zhang, and A. Hou. 2009. Pyramiding multiple genes for resistance to soybean mosaic virus in soybean using molecular markers. Mol. Breeding 23: 113-124.

[74] Singh, S., J.S. Sidhu, N. Huang, Y. Vikal, Z. Li, D.S. Brar, H.S. Dhaliwal, and G.S Khush. 2001. Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*)

using marker-assisted selection into indica rice cultivar PR106. Theor. Appl. Genet. 102: 1011-1015.

[75]   Sobrino, B., M. Briona, and A. Carracedoa. 2005. SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Science International 154: 181-194.

[76]   Somers, D.J., J. Thomas, R. DePauw, S. Fox, G. Humphreys, and G. Fedak. 2005. Assembling complex genotypes to resist *Fusarium* in wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 111: 1623-1631.

[77]   Song, Q., G. Jia, Y. Zhu, D. Grant, R.T. Nelson, , Hwang, E.-Y., D.L. Hyten, and P.B. Cregan. 2010. Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Sci. 50: 1950-1960.

[78]   Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98: 503-517.

[79]   Stuber, C.W., M. Polacco, and M.L. Senior. 1999. Synergy of empirical breeding, marker-assisted selection and genomics to increase crop yield potential. Crop Sci. 39: 1571-1583.

[80]   Tanksley, S.D., and J.C. Nelson. 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.

[81]   Tanksley, S.D., N.D. Young, A.H. Paterson, and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: new tools for an old science. Bio. Technology 7: 257-263.

[82]   Visscher, P.M., R. Thompson, and C.S. Haley. 1996. Confidence intervals in QTL mapping by booststrapping. Genetics 143: 1013-1020.

[83]   Wang, B., and P.W. Chee. 2010. Application of advanced backcross quantitative trait locus (QTL) analysis in crop improvement. Journal of Plant Breeding and Crop Science 2: 221-232.

[84]   Wang, X., Jiang, G.-L., M. Green, R.A. Scott, D.L. Hyten, and P.B. Cregan. 2012. Quantitative trait locus analysis of saturated fatty acids in a population of recombinant inbred lines of soybean. Mol. Breeding, DOI 10.1007/s11032-012-9704-0.

[85]   Wong, C.K., and R. Bernardo. 2008. Genome wide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor. Appl. Genet. 116: 815-824.

[86]   Xu, Y. 2010. Molecular plant breeding. CAB International.

[87]   Ye, G., and K.F. Smith. 2008. Marker-assisted gene pyramiding for inbred line development: Basic principles and practical guidelines. Intern. J. Plant Breed. 2: 1-10.

[88]   Yi, C., W. Guo, X. Zhu, L. Min, and T. Zhang. 2004. Pyramiding breeding by marker-assisted recurrent selection in upland cotton II. Selection effects on resistance to *Helicoverpa armigera*. Scientia Agricultura Sinica 37: 801-807.

[89] Young, N.D. 1999. A cautiously optimistic vision for marker-assisted breeding. Mol. Breeding 5: 505-510.

[90] Young, N.D., and S.D. Tanksley. 1989. RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. Theor. Appl. Genet. 77: 353-359.

[91] Zhao, X., G. Tan, Y. Xing, L. Wei, Q. Chao, W. Zuo, T. Lübberstedt, and M. Xu. 2012. Marker-assisted introgression of *qHSR1* to improve maize resistance to head smut. Mol. Breeding, DOI 10.1007/s11032-011-9694-3

[92] Zhong, S., J.C.M. Dekkers, R.L. Fernando, and Jannink, J.-L. (2009). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182: 355-364.