Use of Major Quantitative Trait Loci to Improve Grain Yield of Rice

GUO Long-biao¹, YE Guo-you²
¹(State Key Laboratory of Rice Biology / Chinese National Center for Rice Improvement, China National Rice Research Institute, Hangzhou 310006, China; ²Plant Breeding, Genetics and Biotechnology Division, International Rice Research Institute, College, Los Banos 4031, Laguna, the Philippines)

Abstract: Further improvement of rice productivity remains a challenge. Breeding is perceived as an important option to increase rice yield. However, the genetic progress of grain yield in most rice breeding programs was slow in the last decades. Although great progress in rice genomics and molecular biology has been achieved, the effect of such technological innovations on rice breeding is far small. Marker-assisted selection (MAS) for a few target quantitative trait loci (QTLs) has significant effects in improving qualitative traits, such as disease resistance. The success of MAS has therefore motivated breeders to identify and use major QTLs for yield and yield component traits. In this review, we summarized the recent methods in QTL identification, including novel statistical methods for linkage and association mapping, special population types, and whole-genome sequencing. We reviewed the successful application of marker-assisted gene introgression and gene pyramiding to improve grain yield and discussed the design of efficient MAS schemes to further increase the success rate of breeding programs. The use of well-characterized major QTLs through introgression and gene pyramiding is proven effective in improving grain yield, particularly yield under abiotic stress. Major QTLs that are stable across genetic background and growing environments are often found in less adapted germplasms, such as landraces and wild relatives. Advanced backcross QTL analysis and introgression lines, which integrate QTL discovery and utilization, are important methods for exploiting major QTLs contained in such germplasms. Next-generation sequencing substantially increases mapping resolution and accelerates the identification of casual genes underlying major QTLs. Practical guidelines derived from theoretical and empirical studies are given to guide the design of efficient marker-assisted gene introgression and pyramiding schemes.

Key words: gene pyramiding; marker-assisted backcross; marker-assisted selection; rice; yield

Rice is a staple food for more than half of the world’s population (Delseny et al, 2001; Feng et al, 2013). The estimate of world paddy production in 2011 is 7.2 × 10⁸ t (4.8 × 10⁸ t, milled basis), and the global rice yield must reach 8 × 10⁸ t in 2025 to meet the demand for rice consumption (FAOSTAT, 2010). This additional amount of rice has to be produced by using less land, less water, less labor, and fewer chemical inputs. Therefore, an increase in rice production remains a challenge today. The Green Revolution in the 1960s had greatly increased rice production around the world. At least 50% of the increase in rice production in the past has been achieved because of the adoption of new cultivars. Breeding is expected to be one of the key measures to further increase productivity by improving yield potential and stability. However, the production potential of modern rice cultivars has remained stagnant for the past several decades (Van Nguyen and Ferrero, 2006). Narrow genetic diversity among present breeding lines is a major constraint to the further increase of rice productivity. Domestication and modern breeding have significantly diminished rice genetic diversity. Modern cultivated rice is estimated to retain only 10% to 20% of the genetic diversity compared with its wild relatives.

The potential benefits of using molecular markers linked to the genes of interest in breeding programs, which changed from phenotype-based toward a combination of phenotype- and genotype-based selection, have attracted much attention for more than two decades (Bernardo, 2008; Tester and Langridge, 2010). Marker-assisted selection (MAS) allows for the selection of genes that control traits of interest. This technique is particularly useful in phenotype screening, which is
expensive, technically difficult, or even impossible when using conventional methods. The selection process can be independent from phenotype, which allows selection in off-season nurseries, making the technique more cost effective to grow for more generations per year. Another benefit of MAS is the sharp reduction of required population size because many lines can be discarded in earlier breeding generations after MAS. The efficiency and usefulness of MAS for traits of simple inheritance (i.e., qualitative traits controlled by one or a few genes) have been well proven in many crops, including rice (Collard and Mackill, 2008). The following basic MAS strategies are usually applied: (1) backcrossing favorable alleles into elite germplasm, i.e., marker-assisted-backcrossing (MABC); and (2) stacking multiple genes of different sources, i.e., marker-assisted gene pyramiding (MAGP). The success of MAS has motivated rice breeders to search for QTLs for complex traits, which account for a large proportion of phenotypic variation (major QTLs). However, complex traits are likely to be controlled by minor genes. Interestingly, major QTLs have been often reported for many yield-related traits (Miura et al, 2011). Early studies using MABC and MAGP for improving quantitative traits are less successful (Hospital, 2005; Ye et al, 2009; Ye and Smith, 2010). Some possible reasons are as follows: (1) the putative QTL is a false positive; (2) QTL expression is specific to the testing environment (QTL-by-environment interaction (QEI)); (3) QTLs interact among each other or with genetic background effect (epistasis); and (4) the chromosomal segments detected as QTL hold not only one but several genes. QTLs with small effects are more likely to be false positive. Successful applications of MABC and MAGP for improving yield or yield component traits by using well-characterized major QTLs are recently being reported (Steele et al, 2007; Swamy and Kumar, 2011; Gregorio et al, 2013; Uga et al, 2013). In this paper, we reviewed the methods used for the identification of major QTLs, summarized the successful application of marker-assisted gene introgression and gene pyramiding for improving grain yield, and discussed the design of efficient MAS schemes.

Identification of major QTLs

The detection of the associations between traits of interest and molecular markers is the prerequisite requirement for MAS. Linkage mapping and association mapping (AM) are the two most commonly used approaches to detect these associations. Methods for linkage mapping using biparental populations derived from two inbred lines are well established. The composite interval mapping (CIM) method (Zeng, 1994) is recently the most commonly used method. Many software tools implementing interval mapping (IM) and CIM are freely available, including R/qtl (Broman et al, 2003) and QTL Cartographer (Zeng, 1994). Recently, mixed model variants of CIM have also been developed and used mainly to map QEI (Cooper et al, 2009). AM identifies marker-trait associations caused by linkage disequilibrium (LD), which is the non-random association of alleles at separate loci. AM utilizes ancestral recombination events to identify marker-phenotype associations and provides opportunities for fine mapping that is difficult to achieve through family-based linkage analysis (Mackay and Powell, 2007). This method does not require pedigrees or crosses; thus, data collection is easier. Both linkage mapping and AM have been used to verify QTLs controlling yield and yield-related traits in rice. The following sections briefly review the recent developments in genetic mapping, which can further improve the efficiency of QTL identification and utilization.

Linkage mapping by using multiple biparental populations

Only a limited number of QTLs can be identified when using a single biparental population for mapping. In this method, only QTLs for which two parents differ can be detected. Given the limited scope of each study, mapping the same trait in different biparental populations may yield to different QTLs. More than two alleles are likely to segregate per locus, and the direction of QTL effects may vary depending on the genetic background because of epistasis, pleiotropy, and QEI. The probabilities that a QTL will be polymorphic in at least one population increases when several populations derived from more parental materials are considered. Statistical models for QTL mapping of multiple crosses are available. Xu (1998) developed fixed and random models for mapping QTLs in independent F2 populations. Liu and Zeng (2000) extended CIM to combine crosses from multiple inbred lines. The model can accommodate complex cross designs with overlapping or non-overlapping parental lines. Feenstra et al (2006) developed a mixture model to map QTLs in crosses from multiple inbred lines that is more
robust towards non-normal residual variation. The models of Liu and Zeng (2000) and Feenstra et al (2006) have been implemented within the framework of R/qtl (Broman et al, 2003). The method of Xu (1998) can be easily carried out by using general statistical software. Xu (1998) and Xie et al (1998) presented procedures that assume the allelic effect to be random and argued that random allele effect models deal with multiple family QTL mapping naturally. In random models, only one parameter per QTL is estimated, i.e., the variance explained by the QTL. However, the specific parametric distribution of the random effects, usually normal with zero mean and an estimated variance, is difficult to accept when only a small number of alleles or parents that carry these alleles are involved. Random model methods are more appropriate when the number of segregating alleles is large, whereas a fixed-QTL mode approach is preferred to deal with a single or a small number of mapping population. Wu and Jannink (2004) proposed a Bayesian QTL mapping method for mapping additive QTL by using connected population and developed the InterQTL software. Blanc et al (2006) suggested two biometrical models to analyze multiple-line cross QTL mapping experiments. The disconnected model assumes QTL substitution effects as being specific for every single segregating population and therefore fits QTL effects as nested effects within populations. The genetic assumption underlying the connected model is that allele substitution effects are specific for every parental line.

**Association mapping by using breeder’s population**

Massive phenotypic and genotypic information, which is generated through heavy investments in cultivar development, has been available for most of the breeding programs for many years. These programs can be exploited for the identification of marker-trait associations that are relevant to breeding application. Mapping using populations routinely created and tested by breeding programs has several advantages. First, the lines represent more diverse genetic backgrounds that have been found useful. Thus, inferences would apply to a wider germplasm base. Second, inbreds are evaluated in diverse environments as a component of the breeding programs. The use of different environments permits the sampling of a sufficient set of QEIs. Therefore, the results would be applicable across a wider range of future environments. However, plant breeding data are highly unbalanced because inbreds are evaluated in different sets of environments. Moreover, inbreds, which are developed at different stages of a plant breeding program, cannot be assumed to be random members of a homogeneous population. Instead, these inbreds are considered to be a mixture of breeding populations composed of related individuals. Therefore, statistical methods for mapping by using breeder’s populations require accounting for unbalanced data and for pedigree relationships among inbreds.

Parisseaux and Bernardo (2004) presented a mixed-model approach and found that the method is able to detect QTLs highly repeatable across different populations. They assumed that the marker effects are fixed, and the additive polygenic effect unaccounted by the markers is random. The genetic relationships are considered by using the additive relationship matrix composed of twice the coefficient of co-ancestry among inbreds. The additive relationship matrix depends on the pedigree information. Crepieux et al (2004) proposed a method assuming random QTL effect. Along with related random allelic model approaches, the method estimates a variance component associated with the QTL and identifies the marker interval that is likely to contain the QTL. They required a measure of whether QTL alleles in two different inbreds are copies of the same ancestral QTL allele, i.e., identical by descent (IBD). The probability that QTLs are IBD needs to be estimated from information on linked markers and from pedigree records because these alleles are not observable. Different methods have been developed for computing the IBD matrix (Crepieux et al, 2004; Bauman et al, 2008). Random model approaches allow a better evaluation of the overall breeding value of an inbred, as well as the identification of genomic regions associated with the trait. However, these models do not lead to estimates of the mean effect associated with a specific marker allele linked to a QTL. Therefore, the random model approach allows no identification of the favorable QTL alleles for selection.

The breeding and natural populations of inbred crops are not random mating populations. The complex breeding history of breeding populations and the limited gene flow in most wild populations have created complex stratification within the germplasm, wherein associations are common between unlinked loci. Pritchard et al (2000) introduced the so-called structured association to reduce confounding due to population structure. The approach is based on assigning individuals to subpopulations by using a model-based Bayesian
clustering algorithm, STRUCTURE, and carrying out all analyses as conditional on the inferred assignments. The STRUCTURE algorithm is computationally intensive and may be impractical on large datasets. Price et al (2006) suggested that principal component analysis can be used to summarize genome-wide patterns of relatedness. However, as the population is divided in more subgroups, the probability of false positives is reduced at the cost of a reduction in statistical power. Moreover, any method that effectively removes confounding also removes the true positives that are strongly correlated with population structure (Zhao et al, 2007). For instance, if the causal polymorphisms are perfectly correlated with the underlying population structure, distinguishing between true and false positives statistically is impossible, and any attempt to remove the latter will remove the former.

The structured association is effective but may not be sufficient to control the confounding effects (Zhao et al, 2007). Yu et al (2005) introduced a mixed-model approach to control the population structure and the genetic relatedness among inbreds. Similar to other mixed-model-based methods, a random effect to estimate the fraction of the phenotypic variation, which can be explained by genome-wide correlations, is included by assuming that the phenotypic covariance between individuals is proportional to their relative relatedness or kinship. Relative relatedness is estimated by using genome-wide marker data (the K matrix of pairwise kinship coefficients). In addition to this random effect, a fixed effect by using the population assignments produced by the STRUCTURE algorithm (the Q matrix), is included in the model. The Q and K seem to capture different features of the confounding population structure. However, Zhao et al (2007) found that Q is not required in most cases if K is computed by using a method different from the one used by Yu et al (2005).

AM has been proven to be an efficient method in rice using low- and high-density markers. The best examples were presented by Huang et al (2010, 2012) and Zhao et al (2011). Huang et al (2010, 2012) used whole-genome sequencing to identify single nucleotide polymorphisms (SNPs) for association analysis, whereas Zhao et al (2011) used Affymatrix chips with 44 100 SNPs. Huang et al (2010) utilized 517 landraces and about 3.6 million SNPs to analyze marker-trait association for 14 agronomic traits. They identified a total of 37 significant association signals. Association signals for six traits are located close to previously known genes, which were identified by using mutants or recombinant populations. They later reported on genome-wide association studies of flowering time and grain yield traits by using 950 worldwide varieties and detected and identified 32 new loci responsible for flowering time and 10 grain-related traits (Huang et al, 2012). Zhao et al (2011) applied association analysis to 413 diverse accessions of O. sativa from 82 countries for 34 traits. They found SNPs associated with panicle length at 31.7 Mb to 32.7 Mb on chromosome 1 and for amylose content and flowering time at 4.2 Mb to 4.6 Mb on chromosome 6.

**Multiparent advanced generation intercross population**

AM using natural or assembled populations usually have high false discovery rate due to the collective effects of population stratification and other factors affecting LD, whereas conventional linkage mapping has a low resolution due to the limited number of meiosis. Special types of populations that render the effect of population structure negligible and have a large number of accumulated recombination events have been devised and applied in QTL mapping. Heterogenous stocks (HS) are developed by repeatedly crossing multiple parental lines over many generations. HS has been successfully used in fine-mapping QTLs by using eight parental strains in mice (Valdar et al, 2006) and Drosophila (Macdonald and Long, 2007). A disadvantage of HS is that the genome of each individual is unique and heterozygous, therefore, the population must be genotyped at high density each time the phenotype is taken. To avoid re-genotype, the recombinant inbred lines (RILs) are generated from multiple parents (Cavanagh et al, 2008). The genomes of the founders are first mixed by several rounds of mating and then inbred to generate a stable panel of inbred lines. The multiparent advanced generation intercross (MAGIC) has been suggested for this type of population (Cavanagh et al, 2008). The large number of parental accessions increases the allelic and phenotypic diversity over traditional RILs, thus potentially increasing the number of QTLs that segregate in the cross. The large number of accumulated recombination events increase the mapping accuracy of the detected QTLs compared with an F2 cross (Valdar et al, 2006). The HS and MAGIC populations have some desirable characteristics different from natural or breeder’s populations for association mapping. For instance, all of the ancestral haplotypes are
available, and the number of generations since the ancestral strain, as well as the mating pattern and population history, are known. Kover et al (2009) described the development of the first set of MAGIC lines in plants (*Arabidopsis thaliana*). RILs descended from a heterogeneous stock of 19 intermated accessions. These lines were derived from an advanced intercross for four generations and inbred for six generations. They showed by simulation that QTLs explaining 10% of the phenotypic variance will be detected in most situations with an average mapping error of about 300 kb. Moreover, if the number of lines were doubled, the mapping error would be under 200 kb. They also demonstrated that several known QTLs for germination data and booting time were mapped with high precision, and a few novel QTLs were identified by using the MAGIC population. Several methods have been developed for analyzing data from HS or MAGIC populations. The key component for all the methods is the probability that the QTL or haplotype in a line was inherited from each of the various discovered lines, which need to be calculated for various points along a chromosome. The HAPPY software is developed for this purpose (Mott et al, 2000). We can test the heterogeneity of founder effects by using this information. In addition, the relationship among individuals or lines can and should be considered through mixed models (Kover et al, 2009). The International Rice Research Institute (IRRI) has developed four MAGIC populations: indica MAGIC (eight indica parents), MAGICplus (eight indica parents with two additional rounds of 8-way F1 inter-crossing), japonica MAGIC (eight japonica parents), and Global MAGIC (16 parents: eight indica and eight japonica). Association mapping using a subset (200 samples) of the indica MAGIC population and genotyping-by-sequencing identified many marker-trait associations including markers in the close regions of several known major genes, as well as QTLs and potentially novel loci associated with essential traits for rice improvement (Bandillo et al, 2013).

**Advanced backcross QTL analysis**

Advanced backcross QTL (AB-QTL) analysis was proposed by Tanksley and Nelson (1996) to simultaneously identify and introgress favorable alleles from unadapted donors into an elite background. The general AB-QTL analysis is composed of the following experimental phases: (1) generating an elite parent through donor hybrid; (2) backcrossing to the elite parent to produce a BC1 population that is subjected to marker or phenotypic selection against undesirable donor alleles; (3) genotyping BC2 or BC3 population with polymorphic molecular markers; (4) evaluating the segregated BC2F2 or BC2F3 population for traits of interest and QTL analysis; (5) selecting target genomic regions bearing useful donor alleles for the production of near isogenic lines (NILs) in the elite genetic background; and (6) evaluating of the agronomic traits of the NILs and elite parent controls in replicated environments.

The effectiveness of AB-QTL in rice was well demonstrated by the parallel AB-QTL studies for yield and yield components conducted by Cornell University. The same wild accession of *Oryza rufipogon* (IRGC 105491) was used as donor parent. The recurrent parents are elite varieties including the high-yielding Chinese hybrid V20/Ce64, the upland *O. sativa* subsp. *japonica* rice variety Caiapo from Brazil, the US long-grain tropical japonica cultivar Jefferson, and the elite tropical cultivar IR64. In these studies, 30% to 50% of QTLs identified in each of the advanced backcross (BC2F2) populations showed improved performance for the target traits. Wild QTLs improve recurrent parent performance by 5% to 20% for most of the characters examined (McCouch et al, 2007). In several cases, specific *O. rufipogon* introgressions were found to be associated with superior performance across several genetic backgrounds and environments. By contrast, some yield and flowering time QTLs are associated with a positive effect in one genetic background and/or environment, but not in others (McCouch et al, 2007). In parallel experiments using recurrent parents from India, Korea, China, and the Philippines, essentially identical results were reported, with similar levels of transgressive variation and comparative percentages of ‘favorable wild QTLs’ coming from other wild or weedy accessions of *O. rufipogon*. Subsequently, this method was used to develop populations with other *O. rufipogon* accessions (Wickneswari et al, 2012). Successful applications of the AB-QTL method have been reported by using another interspecific population, which is derived from crosses between *O. sativa* and the wild species *O. glaberrima* (acc. IRGC#103544 from Mali) (Li et al, 2004), *O. glumaepatula* (Brondani et al, 2002), and *O. nivara* (Eizenga et al, 2013).

**Introgression lines**

Introgression lines (ILs) or chromosomal segment
substitution lines (CSSLs), which are developed by repeated backcrossing, have several distinct advantages over primary mapping populations, such as F$_2$, F$_3$, RILs and double haploids, in detecting QTLs for complex traits (Wang et al, 2013). First, detection capability is increased because of reduced effects of interferences from genetic background. A CSSL carries only one or a few substituted segments from a donor parent. Interactions between donor alleles are limited to those between QTLs on homozygous substituted tracts; thus, these interactions are substantially reduced. Second, NILs can be constructed rapidly through backcrossing and MAS after a target gene is found. Third, high-resolution mapping of putative QTLs as Mendelian factors and further map-based cloning are feasible by using a secondary F$_2$ population derived from a cross between a QTL-containing single segment substitution line (SSSL) and the recurrent parent (Eshed and Zamir, 1994; Nadeau et al, 2000). In addition, segregated populations can be purposely created to detect the effect of interaction between QTLs (epistasis) (Lin et al, 2000). ILs are also useful for breeding purposes because they contain only favorable donor alleles and a low percentage of donor genome. Thus, ILs can be easily and rapidly isolated and transferred into elite varieties. Isolation will be easier if a line has a single segment substitution only. Such ILs are also called SSSLs or 3S lines in literature. For genetic study, isolation will be easier if a collection of SSSLs that cover the entire genome of the donor parent and differ for overlapping regions is developed.

Several sets of ILs were constructed by different research teams. Kubo et al (2002) developed the first set of ILs in rice consisting of 65 ILs using a japonica variety, Asominori as the recurrent parent and an indica rice variety IR24 as the donor. A total of 21 ILs are SSSLs. This population, particularly the SSSLs, has been widely used in mapping a range of traits. Xi et al (2006) reported the construction of a library of 1 123 SSSLs by using Huajingxian 74 (HJX74), an elite indica variety from South China, as a recipient, and 24 accessions, including 14 indica and 10 japonica, collected worldwide as donors. ILs with Lemont and Teqing as recurrent parents were reported by Mei et al (2006). Ebitani et al (2005) constructed 39 ILs by using a japonica cultivar Koshihikari as the recurrent parent and an indica cultivar Kasalath as the donor. Takai et al (2007) developed 44 ILs by using also the japonica cultivar Koshihikari as the recurrent parent and an indica cultivar Nona Bokra as the donor. Two populations of CSSLs were developed by using two sequenced rice cultivars: 93-11, an elite restorer indica cultivar as the recipient; and Nipponbare, a japonica cultivar as the donor. The population developed by Zhu et al (2009) consisted of 103 SSSLs with the total length of the substituted segments in the CSSLs being 2 590.6 cM. The population developed by Xu et al (2010) has 128 CSSLs. Fujita et al (2009) developed 334 ILs derived from crosses between a recurrent parent of an indica rice cultivar IR64 and 10 donor parents, including new plant type lines IR65600-87-2-2-3, IR65598-112-2, IR65564-2-2-3, IR69093-41-2-3-2, IR69125-25-3-1-1, Hoshiaoba, IR66215-44-2-3, IR68522-10-2-2, IR71195-AC1, and IR66750-6-2-1. The number of ILs for the donors varies from 21 to 40. Populations of ILs with wild relatives as donors were also developed. Doi et al (1997) developed 45 ILs of O. glaberrima (Accession IRGC 104038) in the background of a japonica cultivar Taichung 65. Tian et al (2006) developed a set of 159 ILs carrying 67.5% of the genome of O. rufipogon (AA). A population of 89 ILs was developed by Guo et al (2013) by using an elite indica cultivar IR24 as a recipient parent and an accession of the tetraploid species O. minuta (accession No. 101133) with BBCC genome as the donor.

Whole-genome sequencing-enabled QTL identification

Recent advances in next-generation sequencing (NGS) technologies have revolutionized the ability to obtain massive genomic resources in a rapid and cost-effective manner. A way of applying whole genome sequencing to MAS is by sequencing the genomes of multiple genotypes to identify sequence polymorphisms such as SNPs. After this polymorphism identification phase, large numbers of samples can be genotyped by using more cost-effective methods such as high-density SNP-genotyping arrays and fluorescent allelic discrimination assays (Davey et al, 2011; Chen et al, 2013). NGS can also be used directly in genotyping to avoid marker conversion and to improve the efficiency and cost-effectiveness of genotyping workflow. A few methods developed for gene or QTL identification using NGS are summarized below.

QTL mapping using high density linkage map

Huang et al (2009) developed a high-throughput method for genotyping RILs utilizing whole genome resequencing.
From high-quality sequences obtained for 150 RILs, each equivalent to 0.023 × coverage of the rice genome, a total of 1 493 461 SNPs were detected. This result gives an average density of 25 SNPs per Mb or 1 SNP every 40 kb. A QTL of large effect on plant height is located at a 100 kb region containing the rice ‘green revolution’ gene by using the sequencing-based genetic map. Instead of depending on high-quality sequences of the parents to identify SNPs, Xie et al (2010) developed a parent-independent strategy for genotyping a mapping population on the basis of very low coverage sequencing. An ultrahigh-density linkage map was constructed with 0.05 × depth sequence coverage of an RIL population derived from a cross between two unsequenced rice varieties. With this map, a previously cloned QTL for grain width (GW5) is localized to its presumed genomic region in a bin of 200 kb (Xie et al, 2010). Yu et al (2011) constructed an ultrahigh-density SNP map of the well-studied rice RIL population derived from Zhenshan 97 × Minghui 63 by using the method of Xie et al (2010). The quality of the SNP map was assessed by using several cloned genes including GS3, GW5, and OsC1. QTL analysis of yield and yield-component traits was performed by using the new map in comparison with the results from the traditional RFLP/SSR map. The ultrahigh-density SNP map is shown to be advantageous in QTL detection and resolution. By resequencing 132 RILs of LYP9 and the parental lines, a high-resolution linkage map was constructed (Gao et al, 2013). On the basis of this high-quality map, the genome sequences of the parental lines were significantly improved, and 43 yield-associated QTLs were detected. In particular, DTH8 and LAX1 were identified as candidate genes for two QTLs, namely, qSN8 and qSPB1, respectively. A genetic complementation test demonstrated that DTH8 represents qSN8 (Gao et al, 2013). This study provides a promising strategy to dissect QTLs associated with complex agronomic traits.

**QTL-seq**

A few rapid QTL identification methods are developed by combining the bulk segregant analysis (BSA) (Michelmore et al, 1991) with high-throughput genotyping technologies. Wolyn et al (2004) first proposed an approach named eXtreme Array Mapping, which combines microarray-based genotyping with BSA for QTL mapping. The method is found to be effective for mapping single major QTLs. Brauer et al (2006) and Becker et al (2011) demonstrated the effectiveness of microarray-assisted BSA in mapping major QTLs by conducting studies on yeast and by conducting simulations, respectively. Ehrenreich et al (2010) first applied NGS to BSA for QTL mapping. By utilizing NGS-assisted BSA and microarray-assisted BSA, they mapped a number of QTLs for 17 chemical resistant traits in yeast, and the results showed that the methods are applicable to various quantitative traits with different levels of genetic complexity, ranging from simple traits influenced by a major locus to very complex traits affected by at least 20 loci. Magwene et al (2011) proposed a statistical framework for QTL mapping based on NGS-assisted BSA. A method named QTL-seq was developed by Takagi et al (2013a) to identify QTLs rapidly by using progeny derived from crosses made between genetically different varieties. In the QTL-seq, two bulks of DNA were used. H-bulk which had 20 to 50 progenies that show high phenotype values, and L-bulk which had 20 to 50 progenies that show low phenotype values, were separately applied to whole genome sequencing. The resulting short reads were aligned to the reference sequence of either of the parents, and the SNP-index plots of H-bulk and L-bulk were compared. Genomic regions displaying contrasting patterns of SNP-index plots between the two bulks indicate the positions of QTLs. Takagi et al (2013a) applied QTL-seq to rice RILs and F2 populations and successfully identified QTLs for important agronomic traits, such as partial resistance to the fungal rice blast disease and seedling vigor. A simulation study showed that QTL-seq is able to detect QTLs over wide ranges of experimental variables (Takagi et al, 2013a). Yang et al (2013) identified six QTLs for seedling cold tolerance in rice by using a large F2 population (10 800 individuals).

NGS-assisted BSA does not require a precise trait value of each individual but only the identification of the individuals that exhibit opposite extreme phenotypes. BSA is less sensitive to the occasional phenotyping mistake (Schneeeberger et al, 2009). Therefore, the requirement for phenotyping in NGS-assisted BSA is less stringent than in traditional QTL mapping methods. Thus, QTL mapping by NGS-assisted BSA can be carried out based on a population of individuals rather than lines. Another merit of using NGS-assisted BSA for QTL mapping is that the genomic sequence data obtained by NGS allow identification of the allelic variation (polymorphisms) between parents, which can facilitate subsequent fine mapping and positional
cloning of the QTLs. With the large number of identified polymorphisms, appropriate markers can be easily developed for the marker-assisted breeding of ILs, which are generally required for the fine mapping of QTLs. In addition, the identified sequence variations in genes, combined with the gene annotation and gene expression from other sources, are helpful for the identification of candidate genes of the QTLs. In summary, NGS-BSA can identify large numbers of markers linked to the target genes or QTLs. On the basis of these linked markers, the target genes or QTLs can be directly mapped by referring to reference genome sequences.

**MutMap and related methods**

Artificially induced mutants have been a valuable resource for the identification of genes underlying agronomic traits in rice. Large collections of mutants are produced by different research teams. Methods combining NGS and BSA for the rapid identification of the causal gene mutation for a mutant phenotype have been proposed to complement the commonly used, but time-consuming positional cloning. Schneeberger et al (2009) proposed the SHOREmap method, which used large (500) mutant F2 progenies from a cross between a mutant in one accession (i.e., the Columbia-0 (Col-0) accession of Arabidopsis) with a wild-type plant in another accession (Ler) to identify the causal mutation in one NGS run of 20 × coverage of the genome. Subsequently, pools containing considerably fewer segregants were also successfully used to identify genes involved in microRNA precursor processing, cell wall, and growth regulation (Cuperus et al, 2010; Austin et al, 2011; Uchida et al, 2011). The MutMap method, which was proposed by Abe et al (2012), is based on the crossing of a mutant of interest to its wild type (the parental line used for the mutagenesis), followed by selfing of F1 individuals to generate F2 progeny. DNA from about 20 F2 individuals, which show the mutant phenotype, is pooled in an equal ratio and subjected to whole genome sequencing with depth of more than 10 × coverage. The short reads are then aligned to the reference genome sequence, which is constructed for the variety used for mutagenesis. Given that the causal SNP is shared by the F2 mutant progeny, all re-sequenced short reads covering such SNP are expected to have a nucleotide that is different from the reference sequence. By contrast, SNPs that are irrelevant to the phenotype under consideration should segregate in a 1:1 ratio among the F2 progeny. Consequently, about half of the short reads covering such positions contain a nucleotide that is different from the reference genome. An SNP-index is computed to quantify the proportion of short reads with nucleotides different from the reference sequence (SNPs). If the entire short reads covering a particular genomic position share a SNP that deviates from the reference, then the SNP index is defined as 1. By contrast, if only half of the short reads share such a SNP, then the SNP index is 0.5. The SNP index is calculated for all the SNPs incorporated by mutagenesis, and the relationship between SNP index and genomic position is graphically plotted. The genomic region showing a unique SNP-index peak (SNP-index = 1) corresponds to the position of the causal mutation responsible for the phenotype of the candidate mutant. A highly similar method was proposed independently by Zhu et al (2012). Fekih et al (2013) described MutMap+ as an extension of MutMap. MutMap+ identifies causal mutations by comparing SNP frequencies of bulked DNA of mutant and wild-type progeny of M3 generation derived from the selfing of an M2 heterozygous individual. MutMap+ does not necessitate artificial crossing between mutants and the wild-type parental line. This method is therefore suitable for identifying mutations that cause early developmental lethality, sterility, or generally hamper crossing. MutMap+ offers the same advantage as the classical MutMap protocol. Given that MutMap+ is based on selfing, it enables precise and robust phenotyping of minor effect traits. When the resequenced variety/line displays significant structural variation from the reference genome, mutations in the genome regions missing from the reference (gaps) cannot be identified by MutMap or MutMap+, which relies on simple alignment. Takagi et al (2013b) described a method called ‘MutMap-Gap’, which involves delineating a candidate region that harbors a mutation of interest by using MutMap, followed by de novo assembly, alignment, and identification of the mutation within genome gaps. They then applied MutMap-Gap to isolate the blast-resistance gene Pii from the rice variety Hitomebore by using mutant lines that have lost the Pii function.

**Marker-assisted backcross (MABC)**

**Basic principle**

In MABC, markers are used during repeated backcrossing to select the presence of the target gene (foreground selection), to select against donor genome contribution (background selection), and to reduce the
introgressed segment size and consequently linkage drag. Fig. 1 illustrates an MABC scheme. The frequency of the desired allele is 0.5 in BC1 population of two homozygous genotypes. The frequency of the desired allele can be increased by effective selection. When linked markers are used for selection, the frequency in the selected population can be substantially higher than 0.5 and approach 1 (unity), which happens when the recombination rate is 0 (perfect or diagnostic marker). Tanksley et al (1989) stated that a sufficiently high proportion of the recurrent genome is recovered after three generations of marker-assisted background selection (MABS). Hospital et al (1992) expected the preservation of two backcross generations because of MABS. Frisch et al (1999) demonstrated that the number of backcross generations required for the introgression of one target gene is reduced by two to four backcross generations, depending on the genetic difference between the recurrent and donor parents. Frisch and Melchinger (2001) showed that the preservation of three backcross generations is also a realistic goal for simultaneous introgression of two genes.

**Successful application in improving grain yield**

MABC has been used extensively for introgressing resistance to biotic stress (Ye et al, 2009). However, reports on the use of MABC to develop the superior lines/varieties for yield related traits are limited. Scientists at the IRRI have identified a single major QTL (Saltol) on the short arm of chromosome 1, which explains much of the salt tolerance variation in a segregating rice population (Bonilla et al, 2002). In subsequent studies, various markers closely linked to Saltol were identified and used to transfer the QTL to commercial varieties in Bangladesh, India, Vietnam, and the Philippines. A minimum of four popular varieties presently carry the Saltol QTL. Introgression of Saltol into BR11, BRRI dhan 28, BRRI dhan 29, and IR64 were completed by collaboration between IRRI and Bangladesh Rice Research Institute. The introgressed lines of BR11-Saltol and BRRI dhan28-Saltol were tested in salt-affected coastal regions of the Philippines, Bangladesh, and India during the last two seasons of 2011 to 2012 (Gregorio et al, 2013). The drought breeding program at IRRI has identified 11 major QTLs for yield under drought by using different mapping populations. These QTLs are validated across different backgrounds and environments, and all of them show a significant effect across different environments. The first major-effect QTL identified at IRRI, qDTY12.1, has been evaluated in 21 field trials in the Philippines and eastern India. The effect of this QTL is relatively stable across the environments (Bernier et al, 2009). Interestingly, the effect of qDTY12.1 increases with an increase in severity of stress (Bernier et al, 2009). The major effect QTLs qDTY3.1, qDTY2.2, qDTY4.1, qDTY9.1, and qDTY10.1 were evaluated at IRRI during 2010 and 2011 dry seasons, and their effects are consistent. The markers linked to the major-effect QTLs are validated on a panel of drought-tolerant lines to confirm their presence in a larger set of lines. Notably, the major-effect QTL qDTY12.1 is present in 85% of the lines. Meanwhile, qDTY3.2, qDTY2.1, qDTY3.1, qDTY1.1, qDTY8.1, and qDTY1.2 are present in more than 50% of the lines. The study also indicates the presence of at least one major-effect grain-yield QTL in every drought panel line. qDTY12.1, qDTY3.1, qDTY2.2, qDTY9.1, qDTY10.1, qDTY4.1, and qDTY1.1 increase grain yield under stress conditions and have no any adverse effect on grain yield under non-stress conditions. qDTY12.1 has been successfully introgressed in the background of Vandana. Vandana-introgressed lines with qDTY12.1 show a yield advantage of 0.5 t/hm² over drought-tolerant cultivar Vandana under drought conditions and have a yield similar to that of Vandana under normal irrigation situations (Swamy and Kumar, 2011). qDTY2.2, qDTY4.1, qDTY9.1, and qDTY10.1.

![Fig. 1. An example of marker-assisted backcross (MABC) scheme with P1 and P2 as the recurrent and donor parents, respectively (Ye et al, 2009).](image-url)
have been introgressed in an IR64 background. Meanwhile, qDTY1.1, qDTY2.1, qDTY2.2, qDTY3.1, qDTY4.1, qDTY9.1, qDTY10.1, and qDTY12.1 are being introgressed in several popular rice mega-varieties. In general, the major-effect QTLs identified for grain yield under drought have a genetic gain of 10% to 30%, with a yield advantage of 150 to 500 kg/hm² over the recipient parents (Swamy and Kumar, 2011). Similarly, Uga et al (2011) identified a QTL for deep rooting on chromosome 9 (denoted as DRO1) by using a mapping population that was derived from the cross between a shallow-rooting cultivar IR64 and a deep-rooting cultivar Kinandang Patong from the Philippines. Cloning and characterization of DRO1 show that the QTL functions the downstream of the auxin-signaling pathway and controls the gravitropic curvature in rice roots (Uga et al, 2013). Under upland conditions with drought stress, a NIL with a functional allele of DRO1 introduced from the deep-rooting cultivar Kinandang Patong has deeper roots and a significantly higher grain yield than the shallow-rooting parental variety IR64, which contains a non-functional allele of DRO1. When tested under simulated conditions of moderate drought, IR64 yields decline by almost 60%, whereas the NIL with a functional allele of DRO1 suffers only a 10% yield loss. Under extreme drought, IR64 completely fails, but the NIL continues to produce grains at about 30% of the yield of unstressed rice plants growing in normal conditions. QTLs associated with root traits from a DH line of CT9993 × IR62266 are being introgressed into RD6, a popular rice variety in Thailand. QTLs linked to root traits from chromosomes 4, 8, and 9 are being introgressed into IR20 and IR64 by using MAS in Tanu, Coimbatore. NILs of IR62266 introgressed with QTL for root penetration ability and basal root thickness from chromosome 4 of CT9993 were recently field-tested, and results show improved performance of the NILs compared with the recurrent parent (Babu 2012).

OsPPKLI1 gene, which is associated with grain length, was introgressed into 93-11. Under field conditions (13.4 m² plot of 320 plants), a NIL containing the desirable allele qgl3 shows an increase of 16.20% in yield and resulting in an increase of 19.68% in grain length, 1.15% in grain width, 8.25% in grain thickness, 37.03% in filled-grain weight, and 11.76% in panicle length (Zhang et al, 2012). Three hybrids between the NIL and three commercial photothermo-sensitive male sterile lines show increases of 10.12% to 13.48% in grain yield, compared with the corresponding three hybrids developed by using 93-11 as the male parent (Zhang et al, 2012).

Designing efficient schemes

Foreground selection for the target genes and background selection for the reduction of the contribution of donor genome and linkage drag must be combined to obtain an acceptable cultivar by using MABC. Ye et al (2009) provided a summary of the key theoretical and empirical results of MABC studies. These results, which can be used in guiding the design of efficient schemes, are stated below (heavily borrowed from Ye et al (2009)). For foreground selection, perfect markers or the closest flanking markers were highly useful. For markers identified using coarse mapping, various markers, instead of only two of the closest flanking markers, are used to reduce the chance of losing the target QTL because of the uncertain location of the target QTL. The following are several useful points for background selection for practical breeders: (1) four independent markers per chromosome, which do not carry the target gene are sufficient; (2) the use of equally spaced markers reduces the required population size; (3) background selection is more efficient in an advanced generation; and (4) multiple-stage selection is required at each generation by exploring different types of marker genotypes, thus reducing the number of marker genotyping. The following conclusions on practical breeding to reduce linkage drag are obtained: (1) A small flanking marker distance is advantageous. Selecting distant markers over several successive backcross generations cannot provide a better reduction of linkage drag than using close markers. (2) When flanking markers are used, symmetric marker brackets (i.e., flanking markers that are equally distant from the target gene) are preferable. (3) When the distance between flanking markers is short (< 20 cM), the number of backcross generations performed has little effect on the reduction of donor segment length. (4) The probability of having a smaller intact segment is greater with selection in an early generation than with selection in an advanced generation because crossover events in subsequent generations after selection may result in the reduction of the intact chromosome segment. (5) More successive backcrosses is beneficial in reducing the required population size, that is, a first-generation single recombinant should be selected on one side of the target gene and a second-generation single recombinant should be selected on the other side. Allowing more than two generations
permits an even further reduction of the total number of individuals needed.

The logical steps of MABC are as follows: (1) Select individuals carrying the target allele. (2) Select individuals homozygous with recurrent parent genotype at loci close to the target gene or markers linked to it. (3) Select individuals homozygous for recurrent parent genotype at few (i.e., 2) marker loci on the chromosome carrying the target allele. (4) Finally, select individuals that are homozygous for recurrent parent genotype at other marker loci of the other chromosomes (3 to 4 independent markers per chromosome). Ye et al. (2009) suggested that the background control and the reduction of segment size should be separated to reduce the cost of MABC. This suggestion implies that lines with the target gene are developed first by foreground selection. Subsequently, chromosomes without the target gene are developed by background selection and then are phenotypically tested for key agronomic traits to confirm the presence of significant undesirable linkage drag. The best line can be used as donor to reduce the segment size around the target gene by MABC if the undesirable linkage drag is proven significant. This strategy can prevent unnecessary time and resource expenditure in reducing/eliminating non-existing or insignificant linkage drag. Furthermore, this strategy can potentially utilize the possible beneficial alleles linked to the target allele.

**Marker-assisted gene pyramiding (MAGP)**

**Basic principle**

For quantitative traits such as yield, a single QTL rarely accounts for a significantly large portion of the trait variation. Several QTLs with additive effects need to be stacked to achieve significant improvement. Gene pyramiding is a method aimed at assembling multiple desirable genes from multiple parents into a single genotype. The end product of a gene pyramiding program is a genotype with all of the target genes. MAGP has gained popularity in the last decades (Ye and Smith, 2010). As foreground selection in the context of MABC, markers are used to ensure the presence of the target QTLs. With the recombination frequencies among genes and markers, the frequencies of various satisfactory and desirable genotypes in a population can be predicted according to simple genetic principles. The minimum population size to ensure the presence of the desired genotypes can be calculated using a simple statistical computation. In principle, pyramiding multiple genes is achieved by crossing parental lines with complementary desirable genes and then selecting the desired recombinants among the progeny populations (Allard, 1999). Given that breeding is time-consuming, breeders aim to combine as many desirable alleles as possible in a single breeding cycle (from crossing to the generation of near-homozygous breeding lines). Once the number of genes to be assembled is known, gene pyramiding then aims to obtain near-homozygous breeding lines that are fully homozygous for the desirable alleles of the target genes by using the minimum number of generations, as well as the lowest genotyping and phenotyping costs. Fig. 2 provides an example of a MAGP scheme.

**Successful application in improving yield and yield related traits**

Numerous successful gene pyramiding cases for improving disease resistance in rice and other crops have been reported (Ye and Smith, 2010). However, limited applications of MAGP in improving quantitative traits are successful. As such, this area holds much potential, as indicated by the examples in the following sections. Considering the time requirement of MAGP, more successful applications are expected to be reported in the following years.

Steele et al. (2006) initiated a marker-assisted introgression breeding program in India to improve rice drought tolerance of the Indian upland rice variety Kalinga III, by introgressing desirable QTLs for root traits identified in Azucena, an upland japonica variety from the Philippines. Five segments on different chromosomes were targeted for introgression; four
segments carried QTLs for improved root morphological traits (root length and thickness), and the fifth carried a recessive QTL for aroma. BC3 lines with complementary target QTLs were crossed to pyramid the five QTLs. A set of pyramided lines with different numbers of target QTLs was selected and tested at 60 field trials in eastern India over six years. QTL 9, which has a demonstrated effect of increasing root length, has a significant effect of 0.2 t/hm² on grain yield (Steele et al, 2007). Lines with two or more Azucena alleles at root QTLs show a mean increase of 0.4 t/hm². QTL 7 is associated with a mean increase of 0.9 t/hm² in combination with the other QTLs (Steele et al, 2007). The combination of all four QTLs in PY84 lead to an increased grain yield of 1.0 t/hm² (Steele et al, 2007) and was released in Jharkhand state as Birsa Vikas Dhan 111, which is the first-released drought-tolerant rice cultivar bred through MAS for improved roots (Steele et al, 2013).

IRRI’s drought breeding team has successfully developed pyramided lines containing different combinations of the four major drought QTLs (qDTY2.2, qDTY4.1, qDTY9.1, and qDTY10.1) (Swamy and Kumar, 2011). Introgressed lines with three and two QTLs in an IR64 background show yield advantages of 1.2 to 2.0 t/hm² under drought, yield similar to IR64 under normal irrigated situations, and possess quality traits similar to those of IR64 (Swamy and Kumar, 2011).

Ando et al (2008) developed a rice line with two major QTLs, qSBN1 (for the secondary branch number on chromosome 1) and qPBN6 (for the primary branch number on chromosome 6). The pyramided line produces more spikelets than those with only one of the QTLs. Wang et al (2012) developed NILs containing one or more target genes via transfer of 93-11 alleles at qHD8, qHD7, and qHD6.1, and the GS3 gene for grain size, into Zhenshan 97. The pyramid line, NIL (qHD8 + GS3), shows higher yield potential, longer grains, and a more suitable heading date than Zhenshan 97. NIL (qHD7 + GS3 + qHD6.1) can also increase yield and leaf size, but considerably decrease days to heading. Zong et al (2012) reported that they developed pyramided rice lines with eight yield-related QTLs in the background of 93-11. The pyramided lines show increased panicle and spikelet size.

A large-scale marker-assisted gene pyramiding program in rice is underway at IRRI and the Chinese Academy of Agricultural Sciences (CAAS), Beijing, China to improve drought tolerance. Introggression lines containing un-related QTLs for drought tolerance were crossed to produce segregating F2 populations, which were then screened under both stress and non-stress conditions to identify superior individuals with desirable QTL or QTL combinations and good yield potential. At IRRI, 10 F2 populations developed by this method were screened under very severe lowland drought, resulting in 560 drought-tolerant pyramiding lines. At CAAS, a total of 56 F2 populations were developed, from which 1 207 drought-tolerant pyramiding lines were selected (Li et al, 2005).

Yield components and morphological traits related to yield have been targeted because they are expected to be genetically simpler than yield and the QTLs identified for them may be used in improving yield via MAS. Ashikari et al (2005) identified a major QTL for grain number (Gn1) and a QTL for plant height (Ph1) by using the progeny from the cross between a japonica rice, Koshihikari, and an indica rice, Habataki. NILs under the Koshihikari genetic background were developed and used to combine both beneficial traits. The two lines were crossed, and a pyramiding line carrying Gn1 and Ph1 (sd1) was selected from the progenies by using MAS. The pyramided line showed increased grain production (23%) and reduced plant height (20%) compared with Koshihikai. Qian et al (2007) summarized the development of a collection of pyramided lines containing different numbers of QTLs (2 to 9) for tiller number at the China National Rice Research Institute, Hangzhou, China. Pyramided lines with desired tilling were obtained for super-rice breeding. Phenotypic and genotypic information were used to develop a genotype-phenotype database, which allows for predicting the outcomes of pyramiding. Ohsumi et al (2011) reported that NIL (SBN1), NIL (PBN6), and NIL (SBN1 + PBN6) produce 28% to 37%, 9% to 16%, and 62% to 65% more spikelets per panicle than the recurrent parent, respectively. However, the increased spikelet number per panicle does not lead to the significant, same-level increase in grain yield, probably because of compensation among different yield components. The pyramided lines had only 4% to 12% higher yield. Wang et al (2012) reported on the development of NILs with desirable alleles from Zhenshan 97 and 93-11 under the Zhenshan 97 background by using four backcross inbred lines derived from the two cultivars. Although the heading date is slightly delayed, NIL with desirable allele at qHD8 increases the yield per plant by over 50% and enlarges leaf size by nearly 30%. The two-gene
pyramided line with favorable alleles at both \( qHD8 \) and \( GS3 \) can increase yield-related traits and also enhance the grain size to make it qualify as a long-grain type, which is preferred by the majority of consumers in most Asian countries. The three-gene pyramided line \( (qHD7 + GS3 + qHD6.1) \) also increases yield and leaf size, but decreases days to heading considerably compared with the two-gene line with \( qHD7 \) and \( GS3 \). Zong et al (2012) pyramided eight QTLs for spikelet number per panicle and 1000-grain weight into a single line by using four RILs. Compared with the parent variety 93-11, the population of pyramided lines has more spikelets per panicle, but shows similar 1000-grain weight. Phenotypic selection allows them to identify a few lines with increased panicle and spikelet size.

**Designing efficient schemes**

Total cost and duration are two principal criteria in the design and comparison of pyramiding strategies. When the number of parental lines containing the desirable genes (founding parents) exceeded three, several crossing schemes can result in the generation of the target genotype, and the best genotype must be used. Therefore, the gene pyramiding scheme can be divided into two parts. The objective of the first part is to cumulate one copy of all target genes in a single genotype, which is called the root genotype. The second part aims to fix the target genes into a homozygous state, i.e., to derive the target genotype from the root genotype (called the fixation scheme). Servin et al (2004) labeled these two parts as pedigree and fixation, respectively. Numerous factors must be taken into consideration in designing a gene-pyramiding program. Several of the important factors, such as the characterization of the target traits/genes, reproductive characteristics of the crop and parental lines, linkage between markers and genes, and the operation capital, were discussed in detail by Ye and Smith (2010). With the assumption that every founding parent is involved in only one cross in the gene-pyramiding scheme, Servin et al (2004) described an algorithm for building every possible succession of paired crosses leading to the target genotype. They developed a computer program to generate all the possible schemes and associated minimal and largest population sizes that can be handled at any segregating generation or step during the pyramiding process. The number of possible satisfactory schemes rapidly increases with the number of genes. Even with the computer program, evaluating all the satisfactory schemes when the number of loci is more than 12 is virtually impossible. Ye and Smith (2010) summarized the practical guidelines given by different authors for designing the crossing scheme as follows: (1) parental with fewer target genes enters the crossing schedule earlier; (2) crossing causing strong repulsion linkage is conducted first; (3) the number of crosses made should be increased in each generation; and (4) backcross should be used before assembling genes to reduce population size. Ye and Smith (2010) also summarized the methods for enhancing the efficiency of the fixation step: (1) crossing between selected individuals; (2) crossing the root genotype to a genotype with desirable genes; and (3) advancing all satisfactory genotypes at each generation. The rationales behind these recommendations can be found in Ye and Smith (2010).

**Prospective**

The rapid development in gene sequencing, functional genomics, and comparative genetics have led to the development of remarkable numbers of molecular markers. The development of highly automated genotyping systems not only significantly accelerates genotyping process and improves genotyping quality, but also, more importantly, dramatically reduces the cost of genotyping. QTL mapping has led to the identification of numerous QTLs for yield-related traits (http://www.gramene.org/). The development of techniques for QTL validation and analysis has facilitated the deciphering of the genetic basis of yield traits. Coordinated efforts in rice functional genomics have led to the identification of 23 genes corresponding to such QTLs (Miura et al, 2011).

Earlier works in introgression and pyramiding of QTLs are less successful in terms of achieving expected improvement. One of the important causes is the persistence of the linkage phase between the target QTL and its linked markers across multiple populations. Markers linked to the QTL identified through linkage mapping using one or a few populations may not be useful in gene pyramiding because different subsets of QTL will be polymorphic in each population, and the linkage phases between a marker and QTL alleles can differ even between closely related genotypes. Another reason is that the targeted QTLs are mapped only coarsely by using small conventional bi-parental mapping populations and a limited number of markers. Given that the QTLs were not precisely mapped, large
regions of donor chromosomes were transferred. The transferred chromosomal segments may hold not only one, but several genes. The recombination between those genes will then modify the effect of the targeted QTLs. Moreover, unfavorable linkage drag may be caused by the unintentional introduction of undesirable alleles. Recent advances in genomics will help address these two issues. Low-cost, high-throughput genotyping, as expected, will allow the use of abundant markers and large populations to fine-map the QTLs of interest and develop functional markers. This approach will greatly facilitate the applications of MABC and MAGP.

The recent success of MABC and MAGP in improving complex traits as exemplified in the drought and salt tolerance cases could be contributed to the use of well-characterized major QTLs, which tend to be less sensitive to environment and genetic background. No commonly accepted criterion is established to group QTLs/genes into major or minor classes. A major QTL needs to explain at least 20% of the total phenotypic variance to have a reasonably high chance of success when used in MABC and/or MAGP. Evidently, MABC and MAGP should target major QTLs. Although QTLs without additive effect can possibly carry sustainable favorable epistatic effects, identifying them is difficult. Only QTLs with confirmed large effects should be targeted for pyramiding. The use of major QTLs makes QEI less an issue. Major QTLs have desirable effects across a wide range of environments, despite different effects in different environments. However, identifying major yield-enhancing QTLs in elite gene pools is unlikely because of the fixation of genes with large effects by numerous round artificial selections. Nevertheless, major QTLs for grain yield in rice are more often detected under stressed environments probably because modern breeding has not been intensively conducted for yield under unfavorable conditions. Therefore, improving yield under stress by using MAS for major QTLs is beneficial.

Major yield-enhancing QTLs have been often reported by researchers on exotic germplasms ( Tanksley and McCouch, 1997). Interestingly, the linkage segment between markers and QTLs tends to be more persistent if a QTL is derived from a gene pool distinct from that used by the breeders. Thus, markers linked to novel alleles from exotic germplasm or wild relatives are more likely to be successfully implemented. However, the chance of desirable QTLs linking to undesirable genes is high, because exotic germplasms usually have more undesirable characteristics compared with elite germplasms. We recommend developing and using ILs for the identification and application of favorable QTLs contained in exotic germplasms. Constructing NILs for each of the target QTLs before taking MABC and/MAGP for developing new varieties may also be necessary. By transferring QTL into a common background, the effect of each of the QTL can be estimated more precisely.

We expect that the use of ILs for the identification and application of favorable QTLs contained in exotic germplasms, including wild relatives, will be proven an efficient method for improving grain yield in rice.

To identify genes/QTLs contributing to grain yield without the genetic and phenotyping complexity associated with yield, numerous researchers have chosen to work on yield-related traits, particularly yield components and morphological traits that determine plant type and architecture. The success of this strategy is demonstrated by the cloning of 23 genes for different yield-related traits (Zou and Li, 2013). The use of these well-characterized and cloned genes in improving yield has started. However, significant improvement of grain yield in farm environments has yet to be reported. A nonlinear relationship exists between yield-related traits and the yield, implying that high-level gene-gene interaction will be seen at yield level, despite the yield-related traits being under additive genetic control. Yield is the product, instead of sum, of the yield component traits, including number of panicles per plant, number of spikelets per panicle, and 1000-grain weight. Yield cannot be significantly improved by manipulating any single-related trait, which suggests that pyramiding QTLs for two or more yield-related traits are necessary. Considering the importance of gene-gene interaction, step-wise pyramiding may present a good strategy. This approach allows the utilization of favorable interactions between QTLs, which are hard to predict. Moreover, many of the cloned genes have been shown to affect more than one trait (pleiotrophy). For instance, MOC1, IPA1, DEP1, and Ghd7 perform important functions in regulating tillering, plant height, and panicle development (Zou and Li, 2013). A gene may even be involved in the regulation of seemingly unrelated or separated biological processes (Zou and Li, 2013). This finding further demonstrates the advantage of step-wise pyramiding strategy.

Although rice appears to have more major trait-enhancing QTLs than other crops, more QTLs/genes with small effects also need to be utilized to hasten
yield improvement. A very promising MAS method, known as genomic selection (GS) or genome-wide selection, has been recently introduced for using all trait-affecting genes to improve quantitative traits (Meuwissen et al., 2001; Hayes et al., 2012). Simulation and empirical studies in self-pollinated crops, including wheat, barley and oat, have demonstrated the great potential of GS in improving quantitative traits (Jannink et al., 2010). GS uses genome-wide markers to predict the breeding (genotypic) values of the selection candidates. Once an accurate prediction model is developed using a reference population with genotyping and phenotypic observations, the model is then to be used in genotypes belonging to the selection population with genotyping but no phenotypic records. We are currently developing a discussion paper on improving grain yield using GS.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31221004 and 31271700), National Basic Research Program of China (Grant No. 2013CBA01405) and the Chinese 863 Program (Grant No. 2012AA10A302).

REFERENCES


Liu Y, Zeng Z B. 2000. A general mixture model approach for
mapping quantitative trait loci from diverse cross designs
taxit locus effect and frequency using synthetic recombinant
Mackay I, Powell W. 2007. Methods for linkage disequilibrium
segregant analysis using next generation sequencing. PLoS
E, Edwards J, Moncada P, Xiao J H, Garris A, Tai T, Martinez C,
Through the genetic bottleneck: O. rufipogon as a source of trait-
Mei H W, Xu J L, Li Z K, Guo L B, Wang Y P, Ying C S,
Luo L J. 2006. QTLs influencing panicle size detected in two
reciprocal introgressive line (IL) populations in rice (Oryza
total genetic value using genome-wide dense marker maps.
linked to disease-resistance genes by bulked segregant
analysis: A rapid method to detect markers in specific genomic
regions by using segregating populations. Proc Natl Acad Sci
USA, 88(21): 9828–9832.
Miura K, Ashikari M, Matsuoka M. 2011. The role of QTLs in the
method for fine mapping quantitative trait loci in outbred animal
complex genetic traits with chromosome substitution strains. Nat
Ohsumi A, Takai T, Ida M, Yamamoto T, Arai-Sanoh Y, Yano M,
near-isogenic lines with increased spikelet number. Field
Parisseaux B, Bernardo R. 2004. In silico mapping of quantitative
Price A L, Patterson N J, Plenge R M, Weinblatt M E, Shadick N
A, Reich D. 2006. Principal components analysis corrects for
stratification in genome-wide association studies. Nat Genet,
Priched J, Stephens M, Donnelly P. 2000. Inference of population
structure using multilocus genotype data. Genetics, 15: 945–959.
Qian G, Guo L B, Yang C D. 2007. Gene-Based Designed
Schneeberger K, Ossowski S, Lanz C, Jaul T, Petersen A H,
SHOREmap: Simultaneous mapping and mutation identification
Xie W B, Feng Q, Yu H H, Huang X H, Zhao Q, Xing Y Z, Yu S
Xu J J, Zhao Q, Du P N, Xu C W, Wang B H, Feng Q, Liu Q Q,
Valdar W, Flint J, Mott R. 2006. Simulating the collaborative cross:
Wickneswari R, Bhuiyan M A R, Kalluvettankuzhy K S, Lim L S,
Wu X L, Jannink J L. 2004. Optimal sampling of a population to