Abstract

Rice (*Oryza sativa* L.) is a staple food crop for people worldwide, and a key goal has been to increase its grain yield. An increasing population that relies on a decreasing level of farmland has rendered the traditional method for the isolation and use of genetic loci in rice breeding unsatisfactory. Recently, the rapid development in next-generation sequencing (NGS) has boosted the number of genome sequences for hundreds to thousands of rice varieties. A MutMap strategy and bulk segregation analysis (BSA) has been developed to directly identify candidate genes based on NGS. The genome-wide association analysis (GWAS) has become a commonly used approach towards identifying the genetic loci and candidate genes for several traits that are closely associated with grain yield. The Multi-parent Advanced Generation Inter-Cross population (MAGIC) is introduced here to discuss potential applications for mapping QTLs for rice varietal development. These strategies broaden the capacity of functional gene identification and its application as a complementary method to insert mutants that comprise T-DNA and transposons. High-throughput SNP analysis platforms, such as the SNP array, provide novel strategies for genomic-assisted selections (GAS) for rice genetic improvements. Moreover, accurate genome sequence information enables genome editing for the utilization of key recessive genes that control important agronomic traits. This review summarizes how NGS accelerates rice genetic improvements through the identification and utilization of key functional genes that regulate agronomic traits.

Introduction

The world population will reach 9 billion in 2050, and a 70–100% increase in food production will be required to meet the demand if the population continues to increase at the current rate (Godfray et al., 2010). Rice is a vitally important food crop worldwide (Xing and Zhang, 2010). The limited area of farmland, the occurrence of extreme climates, and emerging biotic and abiotic stresses have driven the need for rice genetic improvements. Additional knowledge of functional genes for agronomic traits is expected to inform the breeding design for rice. Next-generation sequencing (NGS) provides numerous genome resources by re-sequencing diverse accessions, including landraces and varieties. Millions of single nucleotide polymorphisms (SNPs) are available, and they facilitate the genome-wide association study (GWAS). Recently, NGS has enabled rice genetic improvements by identifying hundreds of key functional genes or by establishing associations between SNPs and agronomic traits.
The available methods for identification of functional genes

There are over 40,000 putative genes in rice (Feng et al., 2002; Goff et al., 2002; Sasaki et al., 2002; Yu et al., 2002). Most genes are quantitative trait loci (QTL), each with a minor effect. Currently, a small group of genes that control agronomic traits, such as the yield traits and heading date (which are all controlled by multiple QTLs), have been identified by a forward genetics strategy. In the past two decades, thousands of QTLs have been located (www. gramene.org/qtl), but only approximately 100 of the QTLs were cloned by map-based cloning. Map-based cloning is time-consuming and laborious. Therefore, several alternative approaches have been developed to efficiently discover functional genes, including gene knockdown and gene knockout.

Antisense and RNA interference (RNAi) technologies (Chuang and Meyerowitz, 2000) decrease the expression of target genes, which can lead to phenotypic changes. This method enables a direct mutation of a target gene. Direct gene knockdown is accomplished using engineering techniques, such as Zinc finger nuclease (ZFN), transcription activator-like effector nuclease (TALEN), and clustered regularly interspaced short palindromic repeats-associated protein 9 (CRISPR-Cas9). Random gene knockdown can cause functional mutations, which are randomly produced by a transposon insertion, a T-DNA insertion or treatments with physical factors and chemical mutagens.

Mutant library for functional genomics in rice

T-DNA insertion library

The isolation of genes using T-DNA mutants based on reverse genetics is common, and a series of T-DNA tagging populations have been developed. For example, approximately 100,000 fertile rice T-DNA lines have been tagged in Korea (Jeon et al., 2000; Jeong et al., 2002), and over 30,000 T-DNA insertion lines have been resulted (Wu et al., 2003). Several other groups have reported T-DNA insertion lines in the genetic background of japonica rice (Chen et al., 2003; Sallaud et al., 2003; Sha et al., 2004; Yin and Wang, 2000). No matter how large the T-DNA library is, it does not ensure that all genes will be inserted because the T-DNA insertion preferentially occurs in gene-rich regions (An et al., 2005; An et al., 2003; Barakat et al., 2000; Chen et al., 2003; Wu et al., 2003). Moreover, T-DNA insertions are non-uniformly distributed on each chromosome; they are higher at the distal ends and lower near the centromeres. Additionally, several regions show extreme peaks and valleys of insertion frequency, which suggests hot and cold spots for T-DNA integration (Jeong et al., 2006). Applying this strategy to indica rice has been particularly difficult because the transformation efficiency is not high enough to produce the thousands of requisite transgenic plants. Therefore, other mutant library types are needed to complement the T-DNA library vacancy.

Transposon-derived library

A two-element Activator/Dissociation (Ac/Ds) transposon system has been extended to rice from maize (Enoki et al., 1999; Greco et al., 2001, 2003; Kolesnik et al., 2004). The generation of a Ds transposon population has been achieved using a regeneration procedure that involves culturing seed-derived calli that carry Ac and non-autonomous Ds elements. Ds insertions are randomly distributed throughout the genome, but a hot spot for Ds insertions has been identified on chromosome 7 within a 40-kb region (Kolesnik et al., 2004). However, its insertion preference and unstable insertion across different generations have also been reported (An et al., 2005; Greco et al., 2003).

Tos17 is a retrotransposon in rice. Although its copy number is low in plants (one to five, depending on the cultivar), it can increase up to 30 copies during tissue culture (Hirochika et al., 1996). A previous report describes the generation of 47,196 Tos17-induced insertion mutants in rice (Miyao et al., 2003). Although insertion targets are distributed throughout the chromosomes, they tend to cluster, and 76% of the clusters are located in genic regions. Moreover, Tos populations can only play minor roles in genome-wide mutagenesis because the number of original donor sites is limited (Kim et al., 2004).

EMS-induced mutant library

The establishment of a large cohort of insertion mutants will accelerate investigations of rice gene functions via reverse genetics approaches.
However, the preferred mutagenesis locations of the Ac/Ds-, Tos17- and T-DNA-derived mutant libraries make them incapable of covering all genes. Therefore, physical and chemical mutagenesis strategies are strong candidates to complement their shortcomings. Fast neutrons, gamma rays, and diepoxybutane induce deletions or translocations; ethyl methanesulfonate (EMS) always induces point mutations. IR64, an indica rice, possesses multiple favourable agronomic characteristics (e.g. wide adaptability, high yield potential, tolerance to multiple diseases and pests, and a suitable quality for eating) and is an ideal genotype for identifying mutational phenotype changes (Wu et al., 2005). Approximately 60,000 IR64 mutants have been generated using chemicals [diepoxybutane (DEB) and EMS] and irradiation (fast neutron and gamma ray). Targeting Induced Local Lesions in Genomes (TILLING) is an efficient way to identify point mutations (Comai and Henikoff, 2006). However, TILLING has largely been avoided as a first option towards identifying mutations because the procedure, which includes the specific gene primer design, PCR application, digestion with a crude celery juice extract that contains the CELI nuclease, and electrophoresis with a denaturing polyacrylamide gel, is lengthy.

Next-generation sequencing-based MutMap strategy for gene mutation

MutMap strategies

It is easy to obtain a large number of mutants using the transposon technique and chemical and physical mutagens, but it is difficult to target the mutation that is responsible for the phenotype change. The rapid development of NGS has illuminated and improved the ability to identify nucleotide variations at the genome level. MutMap has been used for gene isolation in rice (Abe et al., 2012). The MutMap strategy frequently involves multiple steps (Fig. 7.1). A rice elite cultivar with

![Figure 7.1](image-url)
a reference genome sequence, such as Hitomebore from Northern Japan, has been mutagenized with the EMS mutagen to yield the M₁ generation, which self-pollinated to yield M₂; the M₁ segregation population yielded the recessive homozygous mutant with an altered agronomic trait, such as heading date, plant height, panicle architecture or grain size. The mutant was crossed with the Hitomebore wild-type to produce the heterozygous F₁, and the F₂ progeny were developed by self-pollination of F₁. The F₂ segregation population comprised over 100 plants. The DNA from 20 mutants was equally bulked and subjected to whole genome sequencing. The depth of the sequence was typically above 10-fold coverage of the rice genome (370 Mb). These mutant sequences were aligned to the reference sequence of the wild-type plants. The SNP-index was calculated, and the candidate SNPs were identified with a SNP-index = 1. The causal SNPs and the candidate genes were determined by further analysis of the Sanger sequence and with bioinformatics. Finally, the candidate gene was verified by overexpression, RNAi or knockout. OsCOA1 and OsRR22, which are two genes that control pale green leaf and salinity tolerance, were identified by the MutMap strategy (Table 7.1) and validated with mutants and transgenic plants (Abe et al., 2012; Takagi et al., 2015).

The mutants with early development lethality or sterility cannot be crossed with the wild-type; this makes them unsuitable for MutMap applications. MutMap+ has been developed by a versatile extension of MutMap to address this problem (Fekih et al., 2013). MutMap+ selects a heterozygous plant from the M₂ generation to give rise to the M₁ segregating population by selfing, which contains the wild-type and homozygous mutants. The DNA bulks (≥ 20 plants) for the wild-type and mutants have been constructed for whole-genome sequencing. The sequences from both bulks were aligned to the reference sequence of the cultivar that was used for mutagenesis. SNP-index plots were calculated for both the mutant and wild-type M₁ bulks. The two SNP-index plots were compared to identify the region with SNP-index = 1 that was specific to the mutant bulk. In other words, the Δ (SNP-index) plot that was obtained by subtracting the wild-type bulk SNP-index value from that of the mutant bulk was a positive value [Δ (SNP-index) > 0]. The candidate SNPs were likely harboured in the genomic region of the peak value of the Δ (SNP-index). The causal SNPs and the candidate genes were then determined by further analysis of the Sanger sequence and with bioinformatics. Finally, the candidate gene was verified by overexpression, RNAi or gene knockout (Fig. 7.1). OsNAP6, which encodes a SufBD protein that regulates dwarfism and pale green leaf, has been identified by the MutMap+ strategy, and its function has been validated by RNAi (Takagi et al., 2015). A gene (Os08g0139100) for albino seeding, which is incapable of growing to maturity, has also been isolated with the MutMap+ method (Takagi et al., 2015). In specific cases where the re-sequenced cultivar/line displays a significant structural variation from the reference genome (Nipponbare), mutations in genome regions that are missing from the reference genome (gaps) cannot be identified by simple alignment. A method called ‘MutMap-Gap’ has been developed (Takagi et al., 2013b), which involves delineating a candidate region that harbours a mutation of interest using the MutMap method, followed by de novo assembly, alignment, and identification of the mutation within the genome gaps. The candidate SNPs are further identified based on SNP-index = 1, and the causal SNPs and candidate gene are determined by the analysis of the Sanger sequence and with bioinformatics. Finally, they are validated with mutants and/or transgenic plants (Fig. 7.1). HIT7, which encodes a NBS-LRR type R protein that is resistant to rice blast fungus, has been isolated using this method (Takagi et al., 2013b).

Options for further analysis of MutMap detection

MutMap is an efficient approach for isolating mutated genes, but the prerequisite for this strategy’s use is the easy identification of distinct phenotypic types in a F₂ population that is derived from a cross between a wild-type and mutant. Specifically, a single Mendelian segregation ratio of 3:1 should be observed before MutMap is used to target the mutation. There is always a large candidate region when the MutMap strategy is used for gene identification. For example, the candidate regions of the five genes that were isolated by the MutMap strategy ranged from 2.1 to 5.8 Mb; the number of SNPs in the target region ranged from two to ten (Table 7.1). The SNPs of the coding
<table>
<thead>
<tr>
<th>Mutants</th>
<th>Sequence depth</th>
<th>Candidate region (Mb)</th>
<th>SNPs</th>
<th>Causal SNP</th>
<th>Validation</th>
<th>Gene</th>
<th>Function annotation</th>
<th>Strategy</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pale green leaf</td>
<td>&gt;12x</td>
<td>2.1</td>
<td>7</td>
<td>Leu to Phe</td>
<td>Mutant, RNAi and complementary test</td>
<td>OsCAO1 (Os10g0567400)</td>
<td>Chlorophyll a oxygenase</td>
<td>MutMap</td>
<td>Abe et al., 2012</td>
</tr>
<tr>
<td>Salinity tolerance</td>
<td>20x</td>
<td>5.8</td>
<td>2</td>
<td>Premature stop codon</td>
<td>Tos17 mutant</td>
<td>OsRR22 (Os06g0183100)</td>
<td>B-type response regulator</td>
<td>MutMap</td>
<td>Takagi et al., 2015</td>
</tr>
<tr>
<td>Dwarfism and pale green leaf</td>
<td>Mt:19.4 × Wt:11.7×</td>
<td>5.2</td>
<td>8</td>
<td>Ala to Thr</td>
<td>RNAi</td>
<td>OsNAP6 (Os01g0127300)</td>
<td>SufBD protein</td>
<td>MutMap+</td>
<td>Fekih et al., 2013</td>
</tr>
<tr>
<td>Albino seeding</td>
<td>Mt:12 × Wt:11.4×</td>
<td>4.3</td>
<td>10</td>
<td>Premature stop codon</td>
<td>premature stop codon</td>
<td>Os08g0139100</td>
<td>Homologues of chloroplast precursor differentiation and greening</td>
<td>MutMap+</td>
<td></td>
</tr>
<tr>
<td>Blast fungus resistance</td>
<td>Mt1:6.6x</td>
<td>4.1</td>
<td>4</td>
<td>Premature stop codon</td>
<td>Association analysis</td>
<td>HIT7 (Os09t0327600–01)</td>
<td>NBS-LRR type R protein</td>
<td>MutMap-Gap</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mt2:7.7×</td>
<td>5.8</td>
<td>10</td>
<td>Splicing error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
region induced a premature stop codon, amino acid change and splicing error that frequently caused varied phenotypes and were commonly considered the most likely causal SNPs (Table 7.1). The corresponding gene that carries the causal SNP may be the candidate gene. This has been further validated by another method, which produced mutants or transgenic plants (Fig. 7.1). Notably, the regulatory region is also important for gene function, and mutations in the regulatory region may result in a mutant phenotype. Attention should be directed towards its candidate. Moreover, when the candidate SNPs are located on other genes without the annotation of function (unknown protein), it is difficult to predict the causal SNP and candidate gene. All the genes that harbour the candidate SNPs should be verified with other resource mutants and transgenic plants. Alternatively, a large F2 population should be used to isolate the mutation by map-based cloning.

**Genome-wide quantitative trait loci mapping**

Most agronomic traits are quantitatively inherited in rice, and the natural variations in a segregating population frequently exhibit continuous extremes in high and low phenotypic values. These natural variations provide us with the opportunity to breed super varieties. Generating favourable variations from this diversity is the most important task towards genetic improvements. NGS has more advantages towards QTL mapping than traditional molecular marker techniques. It is ideally suited for a natural collection or for a bi- or multi-parental population.

**Genome-wide association study**

The GWAS is a routinely used method for identifying the association between markers and traits in humans and plants. Recently, several NGS projects have been widely applied to rice genomic studies, which have provided millions of SNPs. Because of their abundance and uniform distributions throughout the genome, SNPs are becoming the most desirable molecular markers for QTL scanning. Thus, GWAS has been regularly applied for different traits in rice. The agronomic traits have been investigated by several research groups using GWAS. A mixed linear model (MLM) has been introduced as an improved method to simultaneously account for the population structure and unequal relatedness among individuals, which should decrease false-positive associations (Zhang et al., 2010). A report by Huang et al. (2010) describes the identification of approximately 3.6 million SNPs through the sequencing of 517 rice landraces; it also details the construction of a high-density haplotype map of the rice genome. GWAS was performed for 14 agronomic traits in a population of the *Oryza sativa indica* subspecies. The association population was extended to a larger and more diverse sample of 950 worldwide rice varieties, and GWAS was performed for 11 grain-related traits (Huang et al., 2011). GWAS was performed for agronomic traits and metabolites by other groups (Bai et al., 2016; Chen et al., 2014b; Crowell et al., 2016; Magwa et al., 2016a,b; Zhao et al., 2011). Huang et al. (2015) developed an integrated genomic approach using the SNP data that was obtained by NGS to construct a genome map for 1495 elite hybrid rice varieties and their inbred parental lines. In total, 38 agronomic traits were investigated, and 130 associated loci were identified. Based on the in-depth analyses of the effects of the heterozygous genotypes of these loci, they found that high-yielding hybrid varieties had several superior alleles, but most parental inbred lines had a small number. Therefore, they concluded that the accumulation of numerous rare superior alleles with positive dominance was an important contributor to the heterotic phenomena. Huang et al. (2016) reported the genomic sequencing of 10,074 F2 lines from 17 representative hybrid rice crosses and constructed a dense genotype map with 347,803 recombination events using the NGS-based genotyping method. They classified modern hybrid rice varieties into three groups, which represented different hybrid breeding systems, and obtained support from several loci for a partial dominance of the heterozygous locus for yield-related traits. These results address the genomic architecture of heterosis and rice hybrid breeding. Han et al. (2016) performed an association analysis of 11 known flowering genes at the SNP and haplotype levels. The haplotype-level association analysis detected 7 of the 11 known genes, including the only four detected at the SNP level. Notably, the haplotype-level association strongly increased the power of QTL mapping. Moreover,
the haplotype-level association mapping suggested favourable haplotypes for agronomic traits. However, genome-wide haplotype-level association mapping continues to be a difficult procedure.

The resolution of GWAS is dependent on the SNP local linkage disequilibrium (LD). The average LD decay in cultivated rice has been extended from 100 kb to 200 kb (Huang et al., 2010; Mather et al., 2007; McNally et al., 2009). Although it is difficult to map a QTL to a single gene, this resolution can provide valuable information for marker-selected breeding.

Recently, a Multi-parent Advanced Generation Inter-Cross population (MAGIC) was introduced to discuss potential applications for mapping QTLs and for rice varietal development. MAGIC that is derived from a cross between multiple parents has a genotypic diversity and reduced linkage drag (Leung et al., 2015). There is a clear genealogy among the lines and between the generations, which can effectively avoid the population structure. Several MAGIC populations that were derived from diverse elite parental lines have been used to detect QTLs for different traits in rice. The MAGIC genetic map was constructed using the SNP markers based on NGS and/or SNP array (Bandillo et al., 2013; Meng et al., 2016a,b). A GWAS with these MAGIC populations had a higher detection power than with the assembled populations/germplasm collection and had a higher resolution than the bi-parental populations (Meng et al., 2016a). Moreover, the Hidden Markov Model (HMM) was used for a QTL analysis in MAGIC, which led to a low false positive rate and a low bias of the estimated QTL effect sizes. An implementation of the approach is available as an R package (Verbyla et al., 2014). Currently, GWAS is primarily performed based on the SNPs. Here, we strongly suggest the use of bin-based GWAS for QTL mapping; this may improve the precision of the QTL location and reduce the false positive associations compared with those of the SNP-based GWAS.

**Quick QTL mapping via bulk segregation analysis**

F₂ and/or BC₁F₂ are generally used for QTL mapping. To further isolate the candidate genes that underlie the QTLs that are mapped in the F₂, RIL or BC₁F₂ primary populations, advanced populations, such as near isogenic lines (NILs), are required. However, five to six cropping seasons are necessary to develop the advanced population, which is a lengthy time period. Fortunately, the BSA strategy (Takagi et al., 2013a) has become more vigorous towards identifying QTLs with the NGS application in rice, particularly for major QTLs. Two DNA bulks are constructed by selecting extremely different phenotypic groups from the mapping populations. The QTL region is quickly mapped with an in-depth comparative analysis of the NGS-based sequence. Specifically, if the allele frequency in one region is very different between these two bulks, the region is most likely to contribute to the phenotypic difference. Applications of this strategy will quickly provide valuable information for breeders and will shorten the required times for major QTL isolations in rice.

**Genomic selection for breeding**

**Genetic diagnosis**

The selection of parental lines is a key step in the genetic research and breeding processes of rice (Chen et al., 2013). Breeders often choose varieties with diverse genetic backgrounds for their breeding programs. For example, to screen QTLs for different agronomic traits, a population developed from the varieties with dramatically different backgrounds will enhance the mapping power and will map numerous QTLs. Phylogenetic analysis is useful with this breeding protocol. Rice varieties can be clustered into several subgroups based on the SNP analysis. The distance in the neighbouring tree indicates the genetic similarities among these varieties. The cluster information can be used to guide the parental line selections for genetic studies and rice breeding, such as in the construction of MAGIC populations for genetic studies. Parent founders with higher diversity may provide more opportunities for recombination, which enables efficient and precise explorations of multiple desirable traits (Bandillo et al., 2013).

On the other hand, for hybrid breeding, we can perform genetic diagnosis through the SNP analysis to select the parental lines that have superior alleles to control yield-related traits, such as the plant height, heading date, panicle or grain shape. Exploitation of hybrid vigour (heterosis) has been accomplished in rice production. Several
yield-related genes, such as \( sd1 \), \( Ghd7 \), \( Ghd8 \), \( gs3 \), \( dep1 \), \( ipa1 \), and so on (Table 7.2), have been identified as being critical for hybrids and as having partial heterotic effects (Huang et al., 2015, 2016). Huang et al. (2016) found that genes for yield-related traits were enriched for positive dominance effects, which indicated that superior alleles could be optimally utilized by complementing parents. Furthermore, most of these superior alleles were contributed by the male parents. These findings will be helpful towards rational genetic designs for producing rice hybrids. For example, using SNP analysis, we can choose a more suitable male parent with the desired alleles for \( sd1 \), \( gs3 \), \( tac1 \) and \( Ghd8 \) and a more suitable female parent with superior alleles for \( nal1 \), \( ipa1 \), and \( dep1 \). With these favourable alleles, the expected hybrid would exhibit a high-yield potential with an ideal plant architecture.

### Genomic-assisted selection

Several NGS projects have been widely applied to rice genomic studies (Huang et al., 2010, 2011, 2012; McNally et al., 2009; Xu et al., 2011), which have generated millions of SNPs. Benefiting from NGS, high-throughput SNP arrays have been developed to genotype numerous samples in a short timeframe by identifying thousands of SNPs that are distributed in the whole genome. Various SNP array platforms based on different principles are available in rice. An Affymetrix 44K SNP array has been developed and successfully used in GWAS (Famoso et al., 2011; McCouch et al., 2010; McNally et al., 2009; Zhao et al., 2011). However, most SNPs on the array are derived from the OryzaSNP project with 20 varieties (McNally et al., 2009), which may not adequately represent the rice species. Various SNP chips based on the

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosome</th>
<th>Trait</th>
<th>Yield-related phenotype</th>
<th>Heterosis</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( sd1 )</td>
<td>1</td>
<td>PH</td>
<td>High YD, reduced PH</td>
<td>+</td>
<td>Spielmeyer et al., 2002</td>
</tr>
<tr>
<td>( hd3a )</td>
<td>6</td>
<td>HD</td>
<td>High YD, delayed HD</td>
<td>+</td>
<td>Kojima et al., 2002</td>
</tr>
<tr>
<td>( Hd1 )</td>
<td>6</td>
<td>HD</td>
<td>High YD, delayed HD</td>
<td>+</td>
<td>Yano et al., 2000</td>
</tr>
<tr>
<td>( Ghd7 )</td>
<td>7</td>
<td>HD</td>
<td>High YD, delayed HD</td>
<td>+</td>
<td>Xue et al., 2008</td>
</tr>
<tr>
<td>( Ghd8 )</td>
<td>8</td>
<td>HD</td>
<td>High YD, delayed HD</td>
<td>+</td>
<td>Yan et al., 2011</td>
</tr>
<tr>
<td>( Ghd7.1 )</td>
<td>7</td>
<td>HD</td>
<td>High YD, delayed HD</td>
<td>+</td>
<td>Yan et al., 2013</td>
</tr>
<tr>
<td>( gs3 )</td>
<td>3</td>
<td>GL and KGW</td>
<td>Increased GL and KGW</td>
<td>+</td>
<td>Fan et al., 2006</td>
</tr>
<tr>
<td>( GS5 )</td>
<td>5</td>
<td>GW and KGW</td>
<td>Increased GW and KGW</td>
<td>+</td>
<td>Li et al., 2011</td>
</tr>
<tr>
<td>( GW2 )</td>
<td>2</td>
<td>GW and KGW</td>
<td>Increased GW and KGW</td>
<td>+</td>
<td>Song et al., 2007</td>
</tr>
<tr>
<td>( GW5 )</td>
<td>5</td>
<td>GW and KGW</td>
<td>Increased GW and KGW</td>
<td>+</td>
<td>Weng et al., 2008</td>
</tr>
<tr>
<td>( TGW6 )</td>
<td>6</td>
<td>KGW</td>
<td>Increased YD and KGW</td>
<td>+</td>
<td>Ishimaru et al., 2013</td>
</tr>
<tr>
<td>( gn1a )</td>
<td>1</td>
<td>TGN</td>
<td>Increased YD and KGW</td>
<td>+</td>
<td>Ashikari et al., 2005</td>
</tr>
<tr>
<td>( osgi )</td>
<td>1</td>
<td>GN</td>
<td>Increased YD and GN</td>
<td></td>
<td>IZaw et al., 2011</td>
</tr>
<tr>
<td>( dep1 )</td>
<td>9</td>
<td>GN</td>
<td>Increased YD, SPP and GN</td>
<td>+</td>
<td>Huang et al., 2009</td>
</tr>
<tr>
<td>( nal1 )</td>
<td>4</td>
<td>GN</td>
<td>Increased YD, SPP and GN</td>
<td>+</td>
<td>Fujita et al., 2013</td>
</tr>
<tr>
<td>( MOC1 )</td>
<td>6</td>
<td>Tillering</td>
<td>Increased YD and tillers</td>
<td>+</td>
<td>Li et al., 2003</td>
</tr>
<tr>
<td>( LAX1 )</td>
<td>1</td>
<td>Tillering</td>
<td>Increased YD and tillers</td>
<td>+</td>
<td>Komatsu et al., 2001</td>
</tr>
<tr>
<td>( htd1/osc7 )</td>
<td>4</td>
<td>Tillering</td>
<td>Increased YD and tillers</td>
<td>+</td>
<td>Kulkarni et al., 2014</td>
</tr>
<tr>
<td>( ipa1/osspl14 )</td>
<td>8</td>
<td>Ideal architecture</td>
<td>Increased YD, an ideal architecture</td>
<td>+</td>
<td>Miura et al., 2010</td>
</tr>
<tr>
<td>( tac1 )</td>
<td>9</td>
<td>Tiller angle</td>
<td>increased YD, a compact plant status</td>
<td>+</td>
<td>Yu et al., 2007</td>
</tr>
</tbody>
</table>

*Gene in lower case means the negative regulation of rice yield; **Gene in capital format means the positive controlling. +, means that partial dominance or overdominance effect has been reported within the gene (Huang et al., 2015, 2016). YD, yield; PH, plant height; HD, heading date; GL, grain length; GW, grain width; KGW, kilo-grain weight; TGN, total grain number; GN, grain number; SPP, spikelets per panicle.
Illumina GoldenGate SNP Chip have also been developed for genetic analysis (Chen et al., 2011; Yamamoto et al., 2010; Zhao et al., 2010). SNP arrays that are specific for rice genotyping have recently been designed based on the Illumina Infinium technology; they include RICE6K (Yu et al., 2014) and RiceSNP50 (Chen et al., 2014). RICE6K is based on the low-coverage genome sequences (~1x) of the 520 rice germplasm collections (Huang et al., 2010). It contains 5556 SNP sites and accommodates 80 functional markers that cover 40 rice genes controlling traits such as grain yield, grain quality, heading date, hybrid fertility and stress resistance (Yu et al., 2014). RiceSNP50, a new high-density array, was developed after the S20 accessions (~1x) (Huang et al., 2010) and 281 accessions were sequenced (2.5–15x) (Chen et al., 2014). RiceSNP50 has been synthesized with 51,478 SNPs that are evenly distributed along the whole rice genome. These SNP arrays have become useful for rice genetic research and breeding (Chen et al., 2014; Hu et al., 2013; Tan et al., 2013; Yu et al., 2014).

Marker-assisted selection (MAS) is a valuable tool for improving target traits with DNA markers that are linked to the genes/QTLs underlying the traits (Xu and Crouch, 2008). A promising MAS approach, known as genomic-assisted selection (GAS), has been widely used for genetic improvements in rice (Heffner et al., 2009; Nakaya and Isobe, 2012). GAS is a high-throughput and efficient approach that uses genome-wide SNP markers to improve the agronomic traits; GAS completes the foreground and background examination in a single SNP array assay. During the breeding process with backcross (BC), breeders aim to introduce favourable alleles from a donor parent into the genome of a desired variety without changing the genetic background. The progeny that carry the target genes would then be selected during the early generations of the BC populations (positive selection), while the genetic background of the progeny that carry the donor alleles of interest would be screened to minimize the donor background in advanced BC populations, such as BC_4^1. Rice SNP arrays, such as RICE6K and RiceSNP50, are able to achieve the goals of foreground and background selection. Closely linked low-density SSR markers or functional markers are typically used for positive selection. For GS3 (Fan et al., 2006, 2009; Mao et al., 2010), a major gene that controls grain length, a SNP between C and A in the second exon induces the differentiation of the rice grain length. A total of 180 varieties with genetic diversity have been used for association analysis with the C/A mutation and grain length; they have confirmed the causal mutation to be a SNP (Fan et al., 2009). Based on the causal SNP, a CAPS marker, SF28, has been developed, and it works well in the MAS of the rice grain length (Fan et al., 2009; Lu et al., 2013; Wang et al., 2011). We can also use the RICE6K SNP array for this selection because it accommodates 80 functional probes for 40 genes, including GS3. For an advanced BC population, the individuals that possess the target alleles are required for the genetic background selection. For example, 29 rice individuals with introgressions of Pi1 (Hua et al., 2012) and Pi2 (Zhou et al., 2006) into Kongyu 131, a japonica cultivar that is popular in north-east China, were examined by RICE6K (Yu et al., 2014). The results showed that several improved lines with minimal donor parental backgrounds could be selected, but the target genomic regions also contained large, dragged fragments from the donor parent. These results suggest that closely linked markers or gene functional markers should be used for selection and should be performed on early generations. As another example, the RiceSNP50 array was used in rice BC breeding to assist a selection of the fragment with Pigm(t) (Deng et al., 2006) from Gumei 4 (a rice variety that possesses broad-spectrum resistance) into an elite parental line, R608. The results showed that two out of four BC4_4^1 individuals were identified as possessing the Pigm(t) allele, and one improved line that contained one additional fragment from the donor parent was selected for further breeding (Chen et al., 2014). These application examples indicate that the high efficiencies of Rice SNP arrays make them applicable to rice GAS breeding programmes.

Genome editing
Genome editing is a precise approach for targeting genetic manipulation and is becoming a popular tool for rice genetic improvements. Genome editing relies on accurate genome sequence information to target DNA fragments for mutations, which can result in a DNA insertion, deletion, replacement or reverse replication. ZFN,
TALEN and CRISPR-Cas9 are the three preferred genome-editing tools (Barrangou, 2012; Bibikova et al., 2003; Bogdanove and Voytas, 2011; Pennisi, 2013).

In rice, several important genes that negatively control agronomic traits have been targeted through genome editing (Table 7.3). In 2012, TALEN was used to improve bacterial blight resistance by mutating the OsSWEET14 promoter region, which harbours the cis-elements that interact with pathogens (Li et al., 2012). This mutation decreased the binding effect between the promoter and the effectors that were secreted by the bacterial blight pathogen. The following year, the genes that showed negative effects on the rice architecture, yield, and quality were targeted by TALEN and CRISPR (Shan et al., 2013a,b), with the aim of improving the rice yield and quality. The examples of genome editing applications to rice indicate that the induction-specific mutations in the negative regulatory genes can rapidly produce desirable agronomic traits. There are numerous, well-performing lines that contain one or more undesirable traits for the rice breeding process. There are also dozens of accessible genes that can be edited to improve these traits. Heading date is a vital trait for rice diversification and domestication (Izawa, 2007; Meyer and Purugganan, 2013). The molecular regulatory network for flowering is becoming clearer and more complete. Several important flowering genes have been extensively studied (Xue et al., 2008; Yan et al., 2011; Yano et al., 2000); they provide flexibility towards manipulating the heading date by providing different gene combinations for selection. For example, we previously observed that different allele-combinations of the major heading date genes [Ghd7 (Xue et al., 2008), Ghd8 (Yan et al., 2011) and Hd1 (Yano et al., 2000)] largely defined eco-geographical adaptation in rice cultivars (Zhang et al., 2015). Therefore, we suppose that by targeting Ghd7, Ghd8, or Hd1, we can rapidly design the flowering time of the desired varieties to adapt to different ecological regions based on the combinations of these genes. On the other hand, modifications of the R genes, such as Xa5 (Iyer-Pascuzzi et al., 2008; Iyer and McCouch, 2004) and Xa13 (Antony et al., 2010; Chu et al., 2006; Yuan et al., 2009), would cause the xa5 and xa13 recessive genes to function particularly well.

In the rice bacterial blight system, the dominant allele, Xa5, is necessary and sufficient for pathogen susceptibility, but xa5 is a completely recessive gene that contributes to bacterial blight resistance (Iyer-Pascuzzi et al., 2008). Hence, we believe that through genome editing, we can efficiently obtain improved lines with the xa5xa5 or xa13xa13 homozygous alleles, which would enhance bacterial blight resistance.

Gene replacements and insertions have recently been achieved in rice (Li et al., 2016). Using a pair of single guide RNAs (sgRNAs) that targeted adjacent introns and a donor DNA

<table>
<thead>
<tr>
<th>Mutation types</th>
<th>Genes</th>
<th>Aims</th>
<th>Tool</th>
<th>Repair</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene disruption</td>
<td>OsSWEET14</td>
<td>Bacterial blight resistance</td>
<td>TALEN/CRISPR</td>
<td>NHEJ</td>
<td>Li et al., 2012</td>
</tr>
<tr>
<td>OsBADH2</td>
<td>Rice with fragrance</td>
<td>TALEN/CRISPR</td>
<td>NHEJ</td>
<td>Shan et al., 2013a,b, 2015</td>
<td></td>
</tr>
<tr>
<td>Waxy</td>
<td>Lower levels of amylose</td>
<td>CRISPR</td>
<td>NHEJ</td>
<td>Ma et al., 2015</td>
<td></td>
</tr>
<tr>
<td>ROC5</td>
<td>Rolling leaves rice</td>
<td>CRISPR</td>
<td>NHEJ</td>
<td>Feng et al., 2013</td>
<td></td>
</tr>
<tr>
<td>SD1</td>
<td>Semi-dwarfing breeding</td>
<td>TALEN</td>
<td>NHEJ</td>
<td>Shan et al., 2013a</td>
<td></td>
</tr>
<tr>
<td>DEP1</td>
<td>Erect panicle breeding</td>
<td>TALEN</td>
<td>NHEJ</td>
<td>Li et al., 2016; Shan et al., 2013a</td>
<td></td>
</tr>
<tr>
<td>Gn1a</td>
<td>High yield</td>
<td>TALEN</td>
<td>NHEJ</td>
<td>Li et al., 2016; Shan et al., 2013a</td>
<td></td>
</tr>
<tr>
<td>GS3</td>
<td>Bacterial blight resistance</td>
<td>CRISPR</td>
<td>NHEJ</td>
<td>Li et al., 2016</td>
<td></td>
</tr>
<tr>
<td>IPA1</td>
<td>Bacterial blight resistance</td>
<td>CRISPR</td>
<td>NHEJ</td>
<td>Li et al., 2016</td>
<td></td>
</tr>
<tr>
<td>EPSPS</td>
<td>Glyphosate resistance</td>
<td>CRISPR</td>
<td>NHEJ</td>
<td>Li et al., 2016</td>
<td></td>
</tr>
<tr>
<td>Gene replacement</td>
<td>EPSPS</td>
<td>Glyphosate resistance</td>
<td>CRISPR</td>
<td>NHEJ</td>
<td>Li et al., 2016</td>
</tr>
</tbody>
</table>
template with the same sgRNAs sites, gene replacements were achieved for an endogenous gene, 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), in rice at a frequency of 2.0%. Moreover, the gene insertion frequency was 2.2% with this intron targeting method. Genetically modified plants with the OsEPSPS gene substitution were glyphosate resistant. These achievements for target insertions in rice genes may assist breeders with the introduction useful gene clusters as breeding modules, which could accelerate genetic improvements for the utilization of multiple genes.

The promise of genome editing in rice is its fast and accurate modifications/introductions of ideal agronomic traits for genetic improvements. However, there are always critical limitations, such as off-target effects. Procedures to minimize the off-target effects have been developed and include the double nickase system (Mali et al., 2013) and an inactivated Cas9 (dCas9) that is fused to FokI (Guilinger et al., 2014). Nevertheless, off-target editing can still occur. Another limitation is food safety. Whether the improved lines with the targeted DNA modifications/introductions can be described as genetically modified organisms (GMOs) remains a critical question (Hartung and Schiemann, 2014; Lusser and Davies, 2013; Pauwels et al., 2014). Araki and Ishii (2015) suggest that the genome editing events that result in small insertions/deletions through the non-homologous end joining (NHEJ) repair pathway should not be regulated as GMOs, while genome editing that induces a foreign template by homologous recombination should be considered as a GMO. This issue remains controversial (Schaart et al., 2016; Wolt et al., 2016).

**Omics integration will push rice genetic improvement forward**

Several NGS-based omics systems have been established, and high-throughput omics measurements are routinely explored in rice, including genomics, transcriptomics, and epigenomics. Additionally, proteomics (Chen et al., 2016; Das et al., 2016), metabolomics (Chen et al., 2014) and phenomics (McCouch et al., 2016; Yang et al., 2014, 2015) have been established for high-throughput measurements of phenotypic values, including biochemical, physiological and agronomic traits. Determining the association between the genome variation [SNP, Insertion/deletion (InDel) and copy number variation (CNV)] and trait performance is the crucial step towards estimating genetic effects for rice breeding. GWAS for metabolites has detected more loci with higher resolution than agronomical traits, and 36 candidate genes have been identified, five of which are functionally validated in rice (Chen et al., 2014). However, the relationship between the metabolites and agronomical traits is unclear for most cases. Comparison of the GWAS for agronomics with that for the metabolites, such as assimilates and phytohormones, would address uncertainties by providing compensatory or supporting information for one dataset. False positives are less likely when multiple sources of evidence point to the same gene or pathway. A complete biological model is likely to be discovered only when omics at different levels are jointly considered in an analysis. Integration of these omics datasets could categorize them into two clades – a meta-dimensional analysis and a multistaged analysis (Ritchie et al., 2015). Meta-dimensional analysis is a method where different scales of data are simultaneously combined to identify a complex. Meta-dimensional models involve multiple variables from different data types, but the models for a multistaged analysis are constructed using two different data scales at a time, in a stepwise or linear manner. The main objective of this method is to divide the analysis into multiple steps, where the associations are initially determined between the different data types and subsequently between the data type and trait phenotype of interest (Ritchie et al., 2015). Developing powerful statistical methods will bridge the gap between our ability to generate vast amounts of data and our biological understanding to uncover the genetic basis that underlies the large datasets. A systems genomics approach can provide a more thorough and informative interrogation of genotype-to-phenotype associations.

**Acknowledgements**

This work was partially supported by natural science foundation of China (31571751) and the Natural Science Foundation of Hubei province, China (2015CFA006).
References


