A Pyramid Breeding of Eight Grain-yield Related Quantitative Trait Loci Based on Marker-assistant and Phenotype Selection in Rice (Oryza sativa L.)

Guo Zonga, Ahong Wang a, Lu Wang a, Guohua Liang b, Minghong Gu b, Tao Sang a,c, Bin Han a,*

a National Center for Gene Research and Institute of Plant Physiology and Ecology, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai 200233, China
b Jiangsu Key Laboratory of Crop Genetics and Physiology/Key Laboratory of the Ministry of Education for Plant Functional Genomics, Yangzhou University, 88 Daxue Road, Yangzhou 225009, China
c State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

Received 18 April 2012; revised 17 May 2012; accepted 18 June 2012
Available online 29 June 2012

ABSTRACT

1000-Grain weight and spikelet number per panicle are two important components for rice grain yield. In our previous study, eight quantitative trait loci (QTLs) conferring spikelet number per panicle and 1000-grain weight were mapped through sequencing-based genotyping of 150 rice recombinant inbred lines (RILs). In this study, we validated the effects of four QTLs from Nipponbare using chromosome segment substitution lines (CSSLs), and pyramided eight grain yield related QTLs. The new lines containing the eight QTLs with positive effects showed increased panicle and spikelet size as compared with the parent variety 93-11. We further proposed a novel pyramid breeding scheme based on marker-assistant and phenotype selection (MAPS). This scheme allowed pyramiding of as many as 24 QTLs at a single hybridization without massive cross work. This study provided insights into the molecular basis of rice grain yield for direct wealth for high-yielding rice breeding.

KEYWORDS: Rice; Spikelet number; 1000-Grain weight; Quantitative trait loci; Pyramid breeding; Marker assisted and phenotype selection

1. INTRODUCTION

The future of global food production faces the challenges on less land, less water and fewer inputs. Rice (Oryza sativa L.) is one of the most important crops in the world. An effective rice breeding approach is required to respond to these challenges. 1000-grain weight and spikelet number per panicle are two important components of rice grain yield (Song et al., 2007). Although many quantitative trait loci (QTLs) related to rice 1000-grain weight and spikelet number per panicle have been mapped in different mapping populations (Lin et al., 1996; Lu et al., 1996; Redona and Mackill, 1998; Lai et al., 2003; Li et al., 2004; Fan et al., 2006; Wang et al., 2010; Ding et al., 2011), only a few of them have been cloned and functionally characterized, such as GW2 (Song et al., 2007), GS3 (Mao et al., 2010), GS5 (Li et al., 2011), qSW5 (Shomura et al., 2008), Gn1a (Ashikari et al., 2005), DEP1 (Huang et al., 2009b), and LOG (Kurakawa et al., 2007).

Abbreviations: QTL, quantitative trait loci; RIL, recombinant inbred line; MAS, marker assistant selection; MAPS, marker assisted and phenotype selection; CSSL, chromosome segment substitution line.
* Corresponding author. Tel: +86 21 6484 5260.
E-mail address: bhan@ncgr.ac.cn (B. Han).
The molecular marker-assisted selection (MAS) technology is an effective approach for crop improvement. The two MAS schemes of molecular marker-assisted backcross (Lin et al., 2000; Zhou et al., 2003) and molecular marker-assisted QTL pyramiding (Huang et al., 1997; Ashikari and Matsuoka, 2006) have enabled quick and accurate selection of the lines with the target traits. The MAS-QTL pyramiding approach is based on a strategy to efficiently accumulate beneficial QTLs in a single line. However, conventional QTL pyramiding methods are costly and labor-intensive due to massive backcrossing work.

In our previous study, eight QTLs (Table S1) related to spikelet number and 1000-grain weight were identified through sequencing-based genotyping of 150 rice RILs derived from a cross between two cultivars, *O. sativa* ssp. *indica* cv. 93-11 and *O. sativa* ssp. *japonica* cv. Nipponbare (Wang et al., 2010). For spikelet number per panicle, four QTLs of qSN-1a, qSN-1b, qSN-6, qSN-8 that almost had the same effect (10%) on explaining the phenotypic variation were identified. Another four QTLs responsible for 1000-grain weight, qTGW-5a, qTGW-5b, qTGW-8, qTGW-10, also had nearly same effect (8%) on explaining the phenotypic variation. Table S1 showed that four (qTGW-8, qSN-1a, qSN-6, qSN-8) of eight QTLs had additive effect by the alleles of *indica* 93-11, and another four QTLs (qTGW-5a, qTGW-5b, qTGW-10, qSN-1b) had additive effect by the alleles of *japonica* Nipponbare.

In this research, we validated the effectiveness of qTGW5a, qTGW5b, qTGW10, and qSN1b by using the chromosome segment substitution lines (CSSLs). The results suggested that the qTGW5a and qSN1b might be corresponding to GS5 (Li et al., 2011) and SD1 (Sasaki et al., 2002), respectively. We also used MAS technology to pyramid all the eight QTLs. The pyramided lines increased panicle and spikelet size. We also proposed a novel QTL pyramid breeding scheme with marker-assisted and phenotype selections (MAPS), which improved the previously established SLS-MAS QTL pyramiding scheme (Ribaut and Betran, 1999). Our MAPS scheme could significantly reduce the workload and improve the efficiency of conventional phenotype selection. Using this method, we could theoretically pyramid as many as 24 QTLs with a single cross, which was a significant improvement compared to current QTL pyramiding methods.

2. MATERIALS AND METHODS

2.1. Materials

The RILs and CSSLs were derived from a cross between two cultivars, *O. sativa* ssp. *indica* cv. 93-11 and *O. sativa* ssp. *japonica* cv. Nipponbare (Wang et al., 2010; Xu et al., 2010). The eight QTLs pyramided lines were derived from two different crosses: one was derived from a cross between the RILs AW35 and AW208, and another was derived from a cross between the RILs AW51 and AW208. AW35, AW51 and AW208 were three different lines selected from the RIL population.

2.2. Phenotyping for the CSSLs

We validated the effects of QTLs by using the constructed CSSLs. Each CSSL contained one target QTL. Six individuals of each CSSL and parent were grown in the field. For spikelet number per panicle, the total number of spikelet from the main tiller of each of the six plants was counted. The data of 1000-grain weight was obtained by calculation of all the grains from the six individual plants. Grains were dried in oven for 7 days at 42°C. We detected 200-grain weight, which was then converted to 1000-grain weight. For each CSSL and parent, the 1000-grain weight was calculated for three times. Plants were grown in the field in Shanghai (E 121.4°, N 31.2°) from May to October, 2010, and phenotyping was conducted in laboratory following harvest.

2.3. Phenotyping for the pyramided lines

As the genotype of each QTLs-pyramided individual was different, we investigated spikelet number per panicle and 1000-grain weight in each individual line from about 60 plants. For spikelet number per panicle, the total number of spikelet from the main tiller of each individual line was counted. We used the 200-grain weight of spikelet number per panicle multiplied by 1000-grain weight of each individual to represent its theoretical crop yield. For each RIL and the corresponding parent, about 30 individuals were used for phenotyping. Plants were grown in Sanya (E 109.5°, N 18.2°) field from December 2010 to April 2011 (F3 hybrid of AW35 × AW208, F3 hybrid of AW51 × AW208), and in Shanghai (E 121.4°, N 31.2°) field from May to October, 2011 (F4 hybrid of AW35 × AW208, F4 hybrid of AW51 × AW208).

2.4. Genotyping, linkage map, and QTL analysis

A high-throughput genotyping method was previously developed and used in the 150-rice RILs (Huang et al., 2009a). The RILs were genotyped based on SNPs generated from the whole-genome re-sequencing. A recombination bin map was constructed for each RIL. The recombination maps were aligned to determine recombination bins across the entire population with the minimal bin length of 100 kb adopted. Resulting bins were then treated as a genetic marker for linkage map construction using MAPMAKER/EXP version 3.0b (Lander et al., 2009).

Using this linkage map and phenotypic values, QTL analysis was conducted with the CIM (composite interval mapping) implemented in software Windows QTL Cartographer V2.5 (Wang et al., 2007) (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm). The CIM analysis was run using Model 6 with forward and backward stepwise regression, a window size of 10 cm, and a step size of 2 cm. Experiment-wide significance (*P* < 0.05) thresholds for QTL detection were determined with 1000 permutations. The location of a QTL was described according to its LOD peak location and the surrounding region with 95% confidence interval calculated using WinQTLCart. The epistasis between QTL was estimated.
using R/qtl in the R package (http://www.rqtl.org) (Broman et al., 2003).

2.5. Molecular marker development

SSR and Indel markers were designed according to McCouch et al. (2002) and Hayashi et al. (2006).

2.6. DNA extraction and marker assay

Genomic DNA for each plant was isolated from fresh-frozen leaves using CTAB method (Rogers and Bandich, 1988). The PCR reaction mixture (total volume 20 µL), contained 20 ng of template DNA, 10 mmol/L Tris–HCL, 50 mmol/L KCl, 2.5 mmol/L MgCl$_2$, 5 mmol/L dNTP, 50 pmol primers and 1.0 U Taq polymerase. The reaction cycles are as following: one cycle (95°C, 5 min), 35 cycles (95°C, 45 s; 56°C, 45 s; 72°C, 2 min) and a final extension step (72°C, 10 min). PCR products were subjected to electrophoresis on 3% agarose and stained with ethidium bromide for visualization.

2.7. Development of CSSL GW10 for qTGW10 verification

To develop CSSL for analysis of qTGW10, F$_1$ lines of the RIL-AW77/93-11 were backcrossed with 93-11 for recombinant screening using the 18 DNA markers (Table S2). The AW77 used as the donor parent because it had the Nipponbare allele of qTGW10 with few Nipponbare alleles in the rest of the genome (Fig. S1A). At BC$_2$F$_1$ generation, we carried out genotyping of the lines using the 18 pairs of markers around all the chromosomes. We selected one of the BC$_2$F$_1$ lines which contained the heterozygous allele at qTGW10 against with the purest 93-11 genome background and named it as AW77-6 (Fig. S1B). After one generation of selfing, we selected out one line which contained the Nipponbare homozygous allele at qTGW10 against with the purest 93-11 genome background by using the 4 pairs of markers and named it as GW10 (Fig. S1C). The BC$_2$F$_3$ individuals were used for phenotyping.

3. RESULTS

3.1. QTL verification by using the CSSLs

We validated the effects of three QTLs (qTGW5a, qTGW5b, qSN1b) by using the constructed CSSLs (C48, C37, C42) (Xu et al., 2010). Another QTL (qTGW10) effect was validated by constructing the new line GW10 in this study. The high-resolution genotypes of these four CSSLs were shown in Fig. 1A. The results showed that C48 and C42 had very pure 93-11 genetic background, and the C37 and GW10 had a slightly higher Nipponbare genetic background as compared with C48 and C42. The 1000-grain weight of C48, C37 and GW10 were detected as mean values 30.40 g, 28.52 g and 29.61 g, respectively, which were significantly higher than that of 93-11 (mean value 28.07 g) (Fig. 1B). So the QTLs of qTGW5a, qTGW5b and qTGW10 might have the strong potential to improve 1000-grain weight of 93-11. The spikelet number per panicle of C42 was detected as a mean value 216, which was also significantly higher than that of 93-11 (mean value 165) (Fig. 1C). This result indicated that qSN-1b might have the strong potential to improve spikelet number. The results of 1000-grain weight of C42 and spikelet numbers of C48, C37 and GW10 were shown in Fig. S2.

We searched for any known genes that could be related to 1000-grain weight in the substituted fragment of the C48 line. Genotyping of C48 showed that its genomic region from 1.44 Mb to 5.33 Mb on chromosome 5 is substituted by Nipponbare alleles in 93-11 genetic background based on the IRGSP V4.0 release. The GS5 just located at 3.45-Mb region on chromosome 5. Comparative analysis of the protein sequences of GS5 between Nipponbare and 93-11 revealed several amino acid differences (Fig. S3), and these differences are key factors that affect grain size and 1000-grain weight (Li et al., 2011). So, C48, which carried a qTGW5a allele, might contain the Nipponbare allele of GS5. The increased 1000-grain weight in C48 might result from the effect of GS5.

We also found that the substituted fragment in C42 contains the SD1 locus (Sasaki et al., 2002). Ashikari et al. (2005) report that significant difference of spikelet number per panicle between Koshihikari and NIL-sd1 is due to SD1 gene. SD1 affects both plant height and spikelet number per panicle. SD1 is located at 38.38 Mb on chromosome 1 (based on IRGSP V4.0 release). Genotyping of the C42 showed that its genomic region from 34.03 Mb to 43.63 Mb on chromosome 1 was substituted by Nipponbare alleles in 93-11 genetic background. We suggested that SD1 might be the candidate gene of qSN-1b. The phenotype data of the C42 was shown in Fig. S4. The significant difference of spikelet number per panicle between C42 and 93-11 was likely to be resulted from the effect of SD1.

3.2. QTL pyramiding

On the purpose of breeding high-yielding rice varieties, we employed a QTL pyramid breeding strategy. The eight QTLs, which were mapped previously (Wang et al., 2010), were combined through crossing between some RILs lines from the Nipponbare/P3-11 RIL populations.

We constructed two hybrids using selected RILs to pyramid the eight QTLs. One was AW35 × AW208, the other was AW51 × AW208. Crossing between the selected RILs had some advantages over the crossing between the two parents: 1) Avoiding the incomplete fertility of japonica × indica rice hybrids (Chen et al., 2010); 2) Reducing the number of QTLs need to be selected. Crossing between Nipponbare and 93-11 needed to select all the eight QTLs by MAS. However, crossing between some selected RILs would make selection of target QTLs easier, because the less number of QTLs was to be selected. The sequencing-based genotyping results of AW35, AW51 and AW208 were shown in Fig. S5. Based on the genotypes of these RILs, we could could select the lines with eight QTLs pyramid in the offspring of these two hybrids.
We developed 16 pairs of unique INDEL or SSR markers for mapping the eight QTLs (Table 1), and each pair of two markers covered one QTL. We selected five lines out of about 2000 lines in the F2 populations of 

\[
AW35 \times AW208 \text{ and } AW51 \times AW208
\]

by MAS. Two lines of JH1-101 and JH2-85 had the corresponding homozygous alleles in all the eight QTLs and three lines of JH1-70, JH1-103 and JH2-118 had only one heterozygous marker (Fig. 2). We further planted about 120 seeds of each selected line, since the JH1-70, JH1-103 and JH2-118 lines were still needed to be selected by molecular markers. We randomly chose about 60 different individuals to investigate spikelet number per panicle, 1000-grain weight and theoretical crop yield (spikelet number per panicle multiplied by 1000-grain weight) (Fig. 3).

The three traits were widely distributed in F3 population of these two hybrids. The distribution of these three traits was sharpened after pyramiding the eight QTLs responsible for spikelet number per panicle and 1000-grain weight by MAS. The pyramided lines showed an obvious bias to more spikelet number per panicle and more theoretical crop yield, but the line distribution did not show an obvious bias in 1000-grain weight. This might be due to the relationship between spikelet number per panicle and 1000-grain weight. The two traits in the RILs were investigated and they had a certain degree of negative correlation whose coefficient was \(-0.23\) (Fig. S6). This negative correlation might be explained by the source-sink concept (Venkateswarlu and Visperas, 1987), where the source was the potential capacity for photosynthesis and the sink was the potential capacity to utilize the photosynthetic products. If the sink was small, the yield could not be

Fig. 1. The genotype and phenotype of the CSSLs and the elite parent 93-11.
A: the high-throughput genotyping results of the CSSLs (C48, C37, C42 and GW10), blue represented homozygous for Nipponbare alleles; red represented homozygous for 93-11 alleles and light green represented uncertain alleles because of the lack of marker information; four mapped QTLs were also indicated; B: 1000-grain weight of the CSSLs and the elite parent 93-11. Error bars represented standard deviation (SD) and the \(P\) value was obtained by \(t\) test; C: Spikelet number per panicle of the CSSL and the elite parent 93-11. Error bars represented standard deviation (SD) and the \(P\) value was obtained by \(t\) test.

(AW35 × AW208 and AW51 × AW208) by screening only four QTLs.
high; and even if the sink was large, the yield could not be high if the source capacity was limited.

We investigated the genotypes of AW35/C2 AW208 F3 population with eight QTLs pyramided. As we already had the genotyping results of AW35 and AW208, we just needed to design a few primers at the specific alleles in order to confirm the whole genome genotype results. Without the 16 pairs of primers we used to pyramid the eight QTLs, we chose another 64 pairs of primers among the whole genome to test the genotype of the F3 population (Fig. 4), and the sequence of the primers were shown in Table S3. Through the high-throughput genotyping results of AW35 and AW208, we found that the

<table>
<thead>
<tr>
<th>Name</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Chr.</th>
<th>Physical map position (Mb)</th>
<th>QTL covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN1aU</td>
<td>CAACCGTGCAACCCGTAGTCG</td>
<td>CAACCGTGCAACCGTAGCTC</td>
<td>1</td>
<td>2.90</td>
<td>qSN-1a</td>
</tr>
<tr>
<td>SN1aID</td>
<td>GCGAAACGACATGCAAAAAAA</td>
<td>GCGAAACGACATGCAAAAAAA</td>
<td>1</td>
<td>4.64</td>
<td></td>
</tr>
<tr>
<td>SN1bJU</td>
<td>ATACACACCAAACACGCGA</td>
<td>ATACACACCAAACACGCGA</td>
<td>1</td>
<td>36.61</td>
<td>qSN-1b</td>
</tr>
<tr>
<td>SN1bJD</td>
<td>TCACACACCAAACACGCGA</td>
<td>TCACACACCAAACACGCGA</td>
<td>1</td>
<td>38.08</td>
<td></td>
</tr>
<tr>
<td>TGW5aJU</td>
<td>CGGTAACCCGTTCACAAACAC</td>
<td>CGGTAACCCGTTCACAAACAC</td>
<td>5</td>
<td>1.18</td>
<td>qTGW-5a</td>
</tr>
<tr>
<td>TGW5aJD</td>
<td>TGGAACGTCACACAGTACCACT</td>
<td>TGGAACGTCACACAGTACCACT</td>
<td>5</td>
<td>2.25</td>
<td></td>
</tr>
<tr>
<td>TGW5bJU</td>
<td>CGGTAACCCGTTCACAAACAC</td>
<td>CGGTAACCCGTTCACAAACAC</td>
<td>5</td>
<td>26.24</td>
<td>qTGW-5b</td>
</tr>
<tr>
<td>TGW5bJD</td>
<td>CGGTAACCCGTTCACAAACAC</td>
<td>CGGTAACCCGTTCACAAACAC</td>
<td>5</td>
<td>28.08</td>
<td></td>
</tr>
<tr>
<td>SN6JU</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>6</td>
<td>10.02</td>
<td>qSN-6</td>
</tr>
<tr>
<td>SN6ID</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>6</td>
<td>15.48</td>
<td></td>
</tr>
<tr>
<td>SN8IU</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>8</td>
<td>2.21</td>
<td>qSN-8</td>
</tr>
<tr>
<td>SN8ID</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>8</td>
<td>3.31</td>
<td></td>
</tr>
<tr>
<td>TGW8JU</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>8</td>
<td>25.45</td>
<td>qTGW-8</td>
</tr>
<tr>
<td>TGW8ID</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>8</td>
<td>26.68</td>
<td></td>
</tr>
<tr>
<td>TGW10JU</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>10</td>
<td>11.60</td>
<td>qTGW-10</td>
</tr>
<tr>
<td>TGW10JD</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>AGCTGTACCAATGGAGCTAC</td>
<td>10</td>
<td>12.35</td>
<td></td>
</tr>
</tbody>
</table>

Table S3

Table 1

Markers used to pyramided the eight QTLs

- Information from IRGSP V4.0.

A. PCR results of the two parental lines and five lines selected from the F3 population. The names of the lines were indicated on top and names of the markers were represented below. B: diagrammatic representation of the PCR results of the lines selected from the F3 populations. AW35, AW51 and AW208 were genotyped by high-throughput sequencing. Three genotype classes of each QTL: blue represented homozygous for Nipponbare alleles; red represented homozygous for 93-11 alleles, yellow represented heterozygous.
Fig. 3. Distribution of spikelet number per panicle, 1000-grain weight and theoretical crop yield.
A: distribution of spikelet number per panicle in the eight QTLs pyramiding F$_3$ population (selected by DNA markers) contrasted to the results in the unselected F$_3$ population and the corresponding parents; B: distribution of 1000-grain weight in the eight QTLs pyramiding F$_3$ population (selected by DNA markers) contrasted to the results in the unselected F$_3$ population and the corresponding parents; C: distribution of theoretical crop yield (1000-grain weight multiply spikelet number per panicle) in the eight QTLs pyramiding F$_3$ population (selected by DNA markers) contrasted to the results in the unselected F$_3$ population and the corresponding parents.
heterozygous rate of their F\textsubscript{1} hybrids was 38%, and the heterozygous rate of their F\textsubscript{3} was about 1%—6.5%. About 7%—9% of the genome in F\textsubscript{3} was uncertain alleles. Combining the phenotype (Table S4) and genotype results of the eight QTLs pyramiding F\textsubscript{3} population, we conducted the QTL analysis of 1000-grain weight and spikelet number per panicle by using Windows QTL Cartographer V2.5 (Fig. S7). We did not find out any novel QTL to be associated with the two traits besides four peaks showed in Fig. S7 for the trait of spikelet number per panicle, which were located in the region with uncertain genotype alleles.

We randomly chose about 30 different lines in AW35 × AW208 F\textsubscript{3} population with eight QTLs pyramided, and planted about 24 seeds of each selected line to consist AW35 × AW208 F\textsubscript{4} population with eight QTLs pyramided. We further investigated plant and panicle architecture of AW35 × AW208 F\textsubscript{4} population with eight QTLs pyramided. Fig. 5 showed the phenotype of the parents and five lines.

Fig. 4. Markers used to test the genotype of AW35 × AW208 F\textsubscript{3} population and the genotyping results. Blue represented homozygous for Nipponbare alleles; red represented homozygous for 93-11 alleles, yellow represented heterozygous and light green represents uncertain alleles because of the lack of marker information. A: markers used to detect the genotypes of AW35 × AW208 F\textsubscript{3} population and the genotypes of the two parental lines were indicated on 12 chromosomes; B: genotyping results of 37 lines from the AW35 × AW208 F\textsubscript{3} population and the corresponding parents. The sites of eight QTLs were indicated on the chromosomes and blue represented the QTL effect came from Nipponbare alleles, red represented the QTL effect came from 93-11 alleles.
randomly chosen from AW35 × AW208 F₄ population with eight QTLs pyramided. We found out that after pyramiding eight QTLs, the panicle and spikelet architecture of the five F₄ lines were almost the same, and all of the five lines had bigger panicles and spikelet than the elite parent 93-11. At the meantime of observing bigger panicles and spikelet, which represented better yield to some extent, we could still find out some differences between the lines with eight QTL pyramided in other phenotypes, such as plant height, plant architecture, etc. So we observed lines with better plant architecture on the basis of better panicle architecture.

All the results above fully demonstrated the significant effects of the eight QTLs that had mapped and directly proved effectiveness of our QTL pyramiding strategy in agricultural production.

3.3. MAPS breeding scheme

We further optimized the QTL pyramiding process and proposed a novel QTL pyramid breeding scheme with marker assisted and phenotype selections (MAPS). The MAPS crop breeding scheme was illustrated in Fig. 6. We would use conventional breeding procedures at first to identify elite parental lines and develop the suitable segregating populations for MAS (Fig. 6, Phase I). One of the most important parameters in parental lines selection would be that there must be enough polymorphisms between the parental lines (Ribaut and Betran, 1999). Using high-throughput sequencing technology, polymorphisms between the parental lines could be rapidly detected, which would be very helpful for plant breeders when choosing breeding parents. Following parental line selection, we would cross them and use hundreds of F₂ plants to establish linkage map. Based on trait investigation of F₂:₃ families and the genetic linkage map, we would identify several QTLs conferring these traits (Fig. 6, Phase II) (Li et al., 2000; Zhang and Xu, 2004). Since the genomic sequences of the breeding parents were available, after QTL identification, we could design PCR-based markers, which would represent polymorphisms closely linked to the desired QTLs; subsequently, we would use the marker genotypes to represent the corresponding QTL genotypes. Instead of having to go through the extremely time consuming process of planting thousands of F₂ seeds in the field, these F₂ seeds would be germinated under laboratory conditions and screened using above mentioned PCR-based markers to identify the genotypes of all the QTLs in each F₂ seedling. After genotype selection by PCR-based markers, we would narrow down several hundred F₂ lines containing positive-effect QTLs, which could be homozygous or heterozygous, from several thousand F₂ lines. Instead of planting thousands of F₂ lines for phenotype selection, only the several hundred F₂ lines that contained positive-effect QTLs would be planted. As a result,
the workload of phenotype selection would be significantly reduced. We would then use the same phenotype selection procedure as the conventional hybridization and selection scheme. There would usually be less than one hundred lines that passed the phenotypic selection. We would collect seeds of each line separately, and all of the seeds would constitute the F3 population for further genotype and phenotype selection. The same processes of genotype and phenotype selection described above would be repeated for subsequent generations until genotypes of all the selected positive-effect QTLs became homozygous. Then genotype selection would be stopped and the phenotype selection should be continued until the new elite lines with better traits no longer showed segregation.

Our MAPS scheme, albeit having the same basic selection principles as SLS-MAS (Ribaut and Betran, 1999), represented a significant advancement from the previously described method. In the previously described SLS-MAS QTL pyramiding scheme, as strong selective pressure was only present for one generation, it was unlikely that large numbers of QTLs would become homozygous after one round of self-pollination. The probability of such event happening can be modeled as \( (1/4)^{N_q} \) (described in detail later) with \( N_q \) representing the number of desirable QTLs. Although we could theoretically select out the lines with all the QTLs as early as in the F2 generation if the F2 population was large enough, we had to take time, cost and labor intensity into account in

---

**Fig. 6. General MAPS scheme.**

Phase I represented selection of parental lines from elite varieties with desirable trait(s) using conventional phenotypic selection. Phase II depicted the process of QTL identification for target traits (in this example, target traits were yield and resistance). The genetic linkage map would be constructed using several hundred F2 plants. The phenotypic evaluations under target environment would be observed in F2 families. After QTL identification, PCR-based markers, which would represent polymorphism closely linked to the selected QTL peaks, would be prepared for the following MAS procedure. Phase III represented the process of QTL pyramiding. Several thousand F2 seedlings would be planted in the lab for genotype selection using PCR-based markers mentioned above. After genotype selection, several hundred positive-effect-QTL-containing lines, which could be heterozygous or homozygous, would be selected out from several thousand of F2 lines. Selected lines would be then transferred into the field for phenotype selection based on local needs. After phenotype selection, there might be only less than 100 lines matching breeder’s requirements. Following self-pollination, seeds of each selected line would be collected separately, and all of the seeds would constitute the F3 population for further genotype and phenotype selection. The same processes of genotype and phenotype selection described above would be repeated for subsequent generations until genotypes of all the selected positive-effect QTLs became homozygous. Then genotype selection would be stopped and the phenotype selection should be continued until the new elite lines with better traits no longer showed segregation.
practical crop breeding procedure. Under our breeding scheme, the probability of retaining positive-effect QTLs, which could be in heterozygous or homozygous subjects, was significantly increased and could be modeled as \( (3/4)^{N_Q} \) (described in detail later) with \( N_Q \) representing the number of desirable QTLs. This dramatic increase in our ability to retain positive-effect QTLs would allow us to pyramid much larger numbers of QTLs at the same time. In SLS-MAS scheme, strong selective pressure was only present for one generation, which limited the number of QTLs that could be pyramided; as a result, no more than six QTLs could be pyramided with a reasonable cost. In order to obtain desirable, complex agronomic traits, scientists routinely had to pyramid large numbers of QTLs. We extended the MAS procedure into multiple generations and created this novel MAPS QTL pyramiding scheme. Lines with undesirable traits could be eliminated at seedling stage by MAS. As a result, the phenotyping workload was significantly reduced. More importantly, our MAPS scheme could allow us to pyramid as many as 24 QTLs at a single cross work (described in detail later).

All the theoretical calculations provided below were based on Mendelian ratios and it was assumed that one PCR marker was enough to represent one target QTL. When pyramiding \( N_Q \) number of desirable QTLs, \( F_1 \) hybrids were heterozygous at all the QTLs, which would require marker selection assay on all the \( N_Q \) QTLs in \( F_2 \) population. After one self-pollination, some of the QTLs would become positive-effect homozygous, meaning the number of QTLs to be genotypically selected in \( F_3 \) population would be less than or equal to \( N_Q \). Assuming \( H_{F_3} \) was the number of heterozygous QTLs at \( F_{n-1} \) generation, and these \( H_{F_3} \) QTLs would still require selection in \( F_2 \) population, then \( H_{F_3} \leq H_{F_{n-1}} \leq N_Q \) and \( H_{F_3} = N_Q \). So if we select out lines with \( N_Q \) QTLs pyramided in \( F_2 \) population with high enough probability, we could also observe lines with \( N_Q \) QTLs pyramided in the subsequent pedigree selection. The probability of obtaining \( l \) number of lines that were pyramided with \( N_Q \) positive-effect QTLs including either homozygous or heterozygous subjects from \( F_n \) lines in \( F_3 \) population could be modeled as \( p_{F_3} = C_{x_n}^l p_{F_n}^l (1- p_{F_n})^{N_x-l} \), \( p_{F_3} \) from formula:

\[
\begin{align*}
\sum_{l=0}^{H_{F_3}} C_{H_{F_{n-1}}}^l (p_{peho})^l (p_{he})^{H_{F_{n-1}}-l} &= (p_{peho} + p_{he})^{H_{F_{n-1}}}, \quad \text{represented the probability of } H_{F_{n-1}} \text{ heterozygous QTLs at } F_{n-1} \text{ generation turning into positive-effect homozygous or remained heterozygous after self-pollination}. \\
p_{peho} \text{ represented the probability of one heterozygous QTL turning into positive-effect homozygous after self-pollination}. \\
p_{he} \text{ represented the probability of one heterozygous QTL remaining heterozygous after self-pollination.}
\end{align*}
\]

\( p_{F_3} = C_{x_n}^l p_{F_n}^l (1- p_{F_n})^{N_x-l} \) was a typical binomial distribution, \( \mu = X_{F_n} p_{F_n}, \sigma^2 = X_{F_n} p_{F_n} (1 - p_{F_n}) \), when \( X_{F_n} p_{F_n} > 5, X_{F_n} (1 - p_{F_n}) > 5 \), binomial distribution could be considered as normal distribution. Assuming we needed to select out at least \( B \) different lines that would contain \( N_Q \) positive-effect QTLs including either homozygous or heterozygous subjects in each generation for further phenotype selection at 95% as the confidence level, \( \sum_{l=0}^{B-1} p_{F_3} \leq 5\%, \) then \( \mu - u \sigma \geq B - 1 \), when doing one-tailed test, \( u = 1.64, \) so, \( X_{F_n \text{min}} = (1.64 \sqrt{1 - p_{F_n}} + \sqrt{1.64^2(1 - p_{F_n}) + 4(B-1)^2}/4p_{F_n} \). In formula \( E_{F_n} = B/X_{F_n \text{min}}, \) \( E_{F_n} \) represented the efficiency of selecting out \( B \) different lines with \( N_Q \) QTLs, either positive-effect homozygous or heterozygous, from at least \( X_{F_n} \) lines at 95% confidence level. The most important difference between our scheme and SLS-MAS scheme originated from the difference of \( p_{F_2} \) value. Since only subjects with homozygous positive-effect QTLs were selected, the \( p_{F_2} \) value in SLS-MAS scheme should be calculated in the following formula: \( p_{F_2} = (p_{peho} + p_{he})^{N_Q} \). On the other hand, with our selection scheme, \( p_{F_2} = (p_{peho} + p_{he})^{N_Q} \).

Under Mendelian inheritance, \( p_{peho} = 1/4, p_{he} = 1/2, \) thus in our MAPS \( p_{F_2} = (3/4)^{N_Q} \), in SLS-MAS \( p_{F_2} = (1/4)^{N_Q} \). We took \( B = 100 \) for example, which meant we needed at least 100 different lines for phenotype selection assay. In MAPS, the \( E_{F_2} \) curve in MAPS was illustrated in Fig. 7A. In the case of pyramiding 24 QTLs in one recombinant population, the efficiency for selecting at least 100 different lines having all the 24 positive-effect QTLs (either positive-effect homozygous or heterozygous) would be 0.86\% \( p_{F_2} \), which would require the \( F_2 \) population to include about 120,000 plants. Therefore, although we could theoretically pyramid as many QTLs as needed, the maximum number of QTLs should be less than 24 because of the low probability. With the previously established SLS-MAS scheme, if 24 QTLs were to be pyramided in one recombinant population, the efficiency of selecting at least 100 different lines with all the 24 positive-effect homozygous QTLs would be 3.04E−15. This would require the \( F_2 \) population to include approximately 3.29E+16 plants for MAS in order to ensure that all 24 positive-effect QTLs were still present in the population. As \( H_{F_2} \leq H_{F_{n-1}} \), in formula \( m_{H_{F_{n-1}}} = H_{F_{n-1}} - H_{F_3}, \) \( M_{F_2} \) represented the number of homozygous QTLs, either positive-effect in \( F_n \) generation selected from \( H_{F_{n-1}} \), number of heterozygous QTLs \( (3/4)^{N_Q} \) in \( F_{n-1} \) generation after self-pollination. The probability distribution of obtaining \( M_{F_2} \) under different \( H_{F_{n-1}} \) was modeled (Fig. 7B, Table S5). The distribution was also a binomial distribution, and we could transfer it into a typical binomial distribution: \( P(M_{F_2}, H_{F_{n-1}}) = C_{H_{F_{n-1}}}^m (p_{peho})^m (p_{he})^{H_{F_{n-1}} - m} = (3/4)^{H_{F_{n-1}}} C_{M_{F_2}}^{H_{F_{n-1}}} (1/3)^{M_{F_2}} (2/3)^{H_{F_{n-1}}} \). \( \mu = (3/4)^{H_{F_{n-1}}} \times H_{F_{n-1}} \times 1/3, \sigma^2 = (3/4)^{H_{F_{n-1}}} \times H_{F_{n-1}} \times 1/3 \times 2/3, \) so when \( M_{F_2} = [H_{F_{n-1}}/3], \) \( P(M_{F_2}, H_{F_{n-1}}) \) was maximal, which meant in the following generation of \( H_{F_{n-1}} \) QTLs heterozygous lines, \( [H_{F_{n-1}}/3] QTLs \) would become positive homozygous with the highest probability.

Time required in MAPS is comparable to that in conventional breeding for self-pollinated species or in SLS-MAS. In conventional breeding for self-pollinated species, new lines exhibiting stable inheritance could be created in about ten generations (George, 2006). Thus under our breeding scheme, new lines with stable inheritance could be generated in about ten generation. The number of PCRs to be performed could represent the cost of PCR-based marker selection in our crop breeding program. We could calculate the cost at \( F_n \) generation in the following formula: \( \text{Cost}_{F_n} = X_{F_n} \sum_{l=0}^{N_Q} (3/4)^l \).
the probability we selected out at least 100 different lines with the corresponding homozygous genotype before new generations, which needed to be further selected in $F_{n+1}$ generation after self-pollination. For the formula $P_{(M_{Fn},H_{Fn})} = C_{M_{Fn}}^{M_{Fn}} (1/4)^{M_{Fn}} (1/2)^{H_{Fn}}, \ P_{(M_{Fn},H_{Fn})}$ represented the probability distribution of $M_{Fn}$ under different $H_{Fn}$, $H_{Fn}$ represented the number of heterozygous QTLs in $F_{n+1}$ generation after self-pollination. For example, $P_{(12,18)}$ would represent the probability of obtaining 12 positive-effect homozygous QTLs with self-pollination, when the previous generation contained 18 heterozygous QTLs. The inset represented a zoom-in version of the boxed area.

\[
\lim_{N_0 \to \infty} \text{Cost}_{F_n} = 4\text{Cost}_n \text{ represented the maximum of cost, so Cost}_{F_n} \approx 4\text{Cost}_n \text{ could be used to represent the real cost at } F_n \text{ generation to some extent.}
\]

Based on the analysis described above, we took 24 and 16 QTLs pyramiding as examples to demonstrate the process of our MAPS QTL pyramiding crop breeding scheme (Table 2). All of the QTLs would be pyramided in less than 10 generations, which would be before the lines exhibited stable inheritance. In other words, the desirable QTLs would be pyramided with the corresponding homozygous genotype before new lines with no phenotypic segregation could be generated. Moreover, as shown in Table 2, 16 QTLs pyramiding was much more cost-effective than 24 QTL pyramiding, thus breeders should choose the number of QTLs to be pyramided in one recombination based on resources available to them.

4. DISCUSSION

In this research, effects of $qTGW5a$, $qTGW5b$, $qTGW10$, and $qSN1b$ were validated by the constructed CSSLs. So the further gene investigation of these four QTLs could be done in the future.

The QTL validation results suggest the candidate gene of $qTGW5a$ and $qSN1b$ are $GS5$ and $SD1$, respectively. For $qSN1b$, we have known that $SD1$ alleles from Nipponbare and Kasalath are two important alleles, and the $SD1$ allele from Kasalath can improve the plant height of Nipponbare (Asano et al., 2011). In our results, the $SD1$ allele from Nipponbare might improve the plant height of 93-11, which means $SD1$ allele from 93-11 is different from that from Nipponbare or Kasalath. Comparison of the DNA sequences of $SD1$ between Nipponbare and 93-11 reveals several nucleotide changes and one of these changes lead to a pre-stop codon in 93-11 (Wang et al., 2010). So there must be at least three different $SD1$ alleles in natural populations. We will construct a CSSL in the future, which contains $SD1$ allele from Kasalath in the genetic background of 93-11, and would test whether this NIL have more spikelet number per panicle and higher plant height than C42.

Our QTL pyramiding results prove the effectiveness of our QTL pyramiding process and suggest the eight QTLs could be direct wealth for high-yielding rice breeding. The MAPS breeding scheme proposed based on our experimental results takes the advantage of conventional hybridization and selection breeding and MAS technology. The reason we emphasize the importance of conventional phenotype selection in our MAPS breeding scheme can be summarized as the following. 1) For important traits with known and selected QTLs, we use conventional phenotype selection to further accumulate undetected minor effect QTLs so that we can generate better lines. 2) For important traits with unknown or unselected QTLs, the corresponding QTL information is either unavailable or not included in MAS; as a result, it is important to continue to use the conventional phenotype selection method to identify lines with superior traits. Hence, it is very difficult to create a new rice variety by only genotype selection.

So far, four schemes could be used in QTL pyramiding practice. We take 24 QTLs pyramiding as an example to demonstrate the differences between the four methods.
of the 24 QTLs (Q1 through Q12) have the positive-effect alleles from one parent P1, the other twelve (Q13 through Q24) have the positive-effect alleles from another parent P2. The first scheme by crossing the near-isogenic lines (NILs) was shown in Fig. 8A. We could first construct NILs of Q1 through Q12 by choosing P2 as the recurrent parent, then cross the NILs and use MAS to get the lines with 24 QTLs pyramided. This scheme would be very costly and laborious because of the massive backcrossing work. It would take approximately 16 generations to accomplish the final goal if there were enough human and financial resources. The second scheme, direct backcrossing, was shown in Fig. 8B. We chose P2 as the recurrent parent to cross with P1, and use MAS approach to confirm that the genotype of Q1 through Q12 would always be heterozygous, and the genotype of Q13 through Q24 would come to P2 homozygous. Due to the extremely low probability of observing heterozygosities in all the 12 QTLs (Q1 through Q12) after backcrossing, it was unlikely that 24 QTLs could be pyramided using this scheme. The third scheme, SLS-MAS, was shown in Fig. 8C. Due to the extremely low probability of observing all the 24 homozygous QTLs with positive-effect at F2 generation, it was unlikely that 24 QTLs could be pyramided using this scheme. Our breeding scheme, MAPS, was a more efficient way to pyramid 24 QTLs (Fig. 8D). Plant breeders could extend the MAS procedure into several generations, and this could not only increase the number of QTLs pyramided, but also greatly reduce the marker selection workload. In MAPS scheme, plant breeders use the single-seed-descendent method to obtain the final breeding lines. Based on the theory provided above, the permanent breeding lines could be generated in about 10 generations. By using our MAPS QTL pyramiding scheme, we cannot only overcome the shortcomings of crossing NILs, but also efficiently pyramid the QTLs we need without any backcross works. The advantage of our breeding scheme is especially obvious when pyramiding a large number of QTLs. As
Fig. 8. Comparison of four methods to pyramid 24 QTLs.

A: QTL pyramiding by crossing the NILs; B: QTL pyramiding by direct backcrossing; C: QTL pyramiding by SLS-MAS; D: QTL pyramiding by MAPS. The positive-effect QTLs, Q1 through Q12 came from parent P1, and positive-effect QTLs Q13 through Q24 came from the other parent P2. Four genotype classes of each QTL: blue represented homozygous for P1 alleles, red represented homozygous for P2 alleles, yellow represented heterozygous alleles, and green represented positive-effect homozygous alleles or heterozygous alleles. p value under the genotype represented the probability that the genotype could be selected from the previous generation. p values in A–C were calculated by the formula in the figure, and p values in D were calculated by the formula in Table 2. F represented the generation after hybridization. indicated self-pollination step.
demonstrated in our modeling, the workload of pyramiding 24 QTLs with the traditional backcrossing strategies is unmanageable. Therefore, our proposed breeding scheme represents a significant improvement from previous schemes in rapidly generating new rice varieties with complex favorable traits.

With the rapid development of high-throughput genotyping, genome-wide association studies (GWAS) are playing an increasingly important role in QTL mapping (Nordborg and Weigel, 2008; Gore et al., 2009; Atwell et al., 2010; Huang et al., 2010). When combine the GWAS results with our MAPS QTL pyramiding scheme, we can breed new crop lines in a novel way (Fig. 9). We take rice GWAS results as an example. After obtaining the high-throughput genotyping and phenotyping results of a number of landraces (Fig. 9, Phase I), we can map a number of QTLs for various traits by GWAS (Fig. 9, Phase II). Combining the QTL mapping results and high-throughput genotyping results, the identification of breeding parents will be much easier (Fig. 9, Phase III). We could cross any two lines with desirable QTLs to develop suitable segregating populations for MAS, and then use our MAPS QTL pyramiding scheme to create lines with superior traits (Fig. 9, Phase IV). In the event of more than two parents were needed to cover all the desirable QTLs, it is possible to sequentially pyramid these QTLs using our breeding scheme.

The QTLs detected in this study provide a source of information about the natural genetic variation underlying yield-related traits of rice. The MAPS QTL pyramid breeding scheme proposed in this research combining MAS with phenotype selection would be widely used in crop breeding in the near future.
ACKNOWLEDGEMENTS

We thank Professor Qiaoquan Liu for providing the CSSLs. We thank Benguo Gu, Junping Shi, Yan Zhou, Canyang Li for assistance with field management. This work was supported by the Ministry of Science and Technology (Grant No. 2011CB100205), and the Ministry of Agriculture of China (Grant Nos. 2011ZX08001-004 and 2011ZX08009-002), and the National Natural Science Foundation of China (Grant No. 31121063).

SUPPLEMENTARY DATA

Fig. S1. Genotyping results of AW77, AW77-6 and GW10. Fig. S2. Phenotype of the CSSLs and the elite parent 93-11. Fig. S3. The predicted protein sequences of the GS5 alleles in Nipponbare and 93-11. Fig. S4. Plant height of C42 and corresponding parent 93-11. Fig. S5. High-throughput genotyping results of AW35, AW51 and AW208. Fig. S6. The relationship between spikelet number per panicle and 1000-grain weight.

Table S1. Eight QTLs identified from the analysis of the rice recombinant inbred lines associated with 1000-grain weight and spikelet number per panicle. Table S2. Markers used to construct GW10. Table S3. Markers used to test the genotype of the eight QTLs pyramiding AW35 × AW208 F3 population. Table S4. Phenotyping results of AW35 × AW208 F3 population and the corresponding parents. Table S5. The probability distribution of obtaining \( M_{F_n} \) under different \( H_{F_n}+1 \).

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jgg.2012.06.004.

REFERENCES


