Theor Appl Genet (2012) 125:1717–1726
DOI 10.1007/s00122-012-1948-x

ORIGINAL PAPER

Mapping and characterization of the major quantitative trait locus *qSS7* associated with increased length and decreased width of rice seeds

Xianjin Qiu · Rong Gong · Youbin Tan · Sibin Yu

Received: 18 March 2012 / Accepted: 15 July 2012 / Published online: 5 August 2012

Abstract Seed shape in rice (*Oryza sativa*) is an important factor that determines grain appearance, cooking quality and grain yield. Here, we report a major quantitative trait locus *qSS7* on the long arm of chromosome 7 for seed length, seed width and the ratio of seed length to width, identified using a segregating population derived from a cross between an *indica* variety Zhenshan97 and a chromosomal segment substitution line of a *japonica* variety Cypress within the genetic background of Zhenshan97. The Cypress allele at *qSS7* contributes to an increase in seed length and the ratio of length to width, but a decrease in seed width, without significantly changing seed weight, plant height, heading date or number of spikelets per panicle. Using a large F_2 population generated from a substitution line that carries only a heterozygous single segment surrounding *qSS7*, we delimited the QTL to a 23-kb region containing two annotated genes. Progeny testing of the informative recombinants suggested that this *qSS7* region is a composite QTL in which at least two genes contribute to seed length and width. Sequence comparison and expression analysis of two probable candidate genes revealed differences between the parental lines. These results will facilitate cloning of the gene(s) underlying *qSS7* as well as marker-assisted transfer of desirable genes for seed shape in rice improvement.

Communicated by T. Tai.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1948-x) contains supplementary material, which is available to authorized users.

X. Qiu · R. Gong · Y. Tan · S. Yu (✉)
National Key Laboratory of Crop Genetic Improvement, and College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China
e-mail: ysb@mail.hzau.edu.cn

Introduction

Rice (*Oryza sativa*) is one of the most important crops, providing a staple food for half of the world's population. Seed shape and size have been important breeding traits because they play a significant role in grain appearance, grain milling, cooking quality and grain yield. The various shapes of grains (de-hulled seeds) have different consumer values in diverse areas (Luo et al. 2004). For example, long and slender grain varieties are generally preferred in the international market and most Asian countries, including China and South Asia, while short and round grains are preferred in Japan and Korea (Juliano and Villareal 1993).

Seed shape is determined mainly by its dimensions: seed length (SL), seed width (SW) and the ratio of length to width (RLW). The seed or grain shape is generally recognized as a quantitative trait that is controlled by multiple genes and affected by environmental factors. Numerous quantitative trait loci (QTLs) for SL, SW and RLW have been reported in various mapping populations in rice (Tan et al. 2000; Thomson et al. 2003; Aluko et al. 2004; Li et al. 2004; Rabiei et al. 2004; Govindaraj et al. 2005; Wan et al. 2006; http://www.gramene.org). In these studies, several major QTLs such as those on chromosomes 3 and 5, in addition a few minor QTLs elsewhere, frequently accounted for large proportions of the genetic variation in seed dimensions. So far, most of the loci and their functions remain to be examined, except a small number of genes (e.g., *GS3*, *GW2*, *qSW5*/*GW5*, *GS5*, *SRS3*) associated with grain size or shape that have been identified and

cloned in rice. *GS3*, the first major QTL cloned on chromosome 3 for grain length, encodes a putative transmembrane protein, in which a C-to-A mutation caused a premature stop code resulting in a truncated protein with loss of function (Fan et al. 2006; Takano-Kai et al. 2009). Two other major QTLs controlling seed width and weight in rice, *GW2* on chromosome 2 and *qSW5/GW5* on chromosome 5, encode a RING-type E3 ubiquitin ligase (Song et al. 2007) and a nuclear polyubiquitin-binding protein (Weng et al. 2008; Shomura et al. 2008), respectively. Loss of function in both *GW2* and *qSW5/GW5* led to enhanced seed width and weight but reduced seed length, while loss of function in *GS3* increased seed length and weight, but had little or no effect on seed width (Fan et al. 2006; Takano-Kai et al. 2009; Mao et al. 2010). It seems that *GS3* and *GW2* regulate seed length and width independently during seed development. In contrast to these genes that function as negative regulators of grain shape or size, *GS5*, recently identified on chromosome 5, encodes a putative serine carboxypeptidase regulating cell numbers in mitotic division, and functions as a positive grain size regulator, such that higher expression of *GS5* is associated with larger grains (Li et al. 2011). Similarly, *SRS3* on chromosome 5 encodes a novel kinesin 13 protein (Kitagawa et al. 2010), and might function as a positive regulator of seed shape by regulating cell length in seed formation. Loss of the gene *SRS3* reduced seed length and increased seed width, leading to a small, round seed. Several genes for seed shape or size have also been identified in other plant species. For instance, *AUXIN RESPONSE FACTOR 2* (*ARF2*) in *Arabidopsis* is a transcription factor that mediates gene expression in response to auxin; loss of function in *ARF2* enlarged seeds through extra cell divisions in the integuments (Schruff et al. 2006). *LONGIFOLIA1* (*LNG1*) in *Arabidopsis* encodes novel plant-specific protein enriched with serine residues that enlarged seeds and siliques by regulating longitudinal cell elongation (Lee et al. 2006). Conversely, mutations in either *MINISEED3* or *HAIKU2* in the same pathway led to reduced seed size, which encode a transcription factor and a leucine-rich repeat receptor kinase, respectively (Luo et al. 2005). Characterization of these genes suggests that seed shape or size is governed by cell proliferation and expansion that have complex regulation through diverse mechanisms.

In this study, we found a major QTL (*qSS7*) for seed shape on the long arm of chromosome 7 using segregating populations derived from a chromosomal segment substitution line within the genetic background of an *indica* variety, Zhenshan97 (ZS97). Our main objectives were to map the QTL to a narrow region to enable cloning and to determine whether the QTL independently affects send length and seed width.

Materials and methods

Plant materials and field experiments

A chromosomal segment substitution line (Q043) with long grains was selected from the advanced backcross population (BC_3F_4) derived from a cross between a short-grain *indica* variety (ZS97) as the recurrent parent, and a long-grain tropical *japonica* variety (Cypress) as the donor parent (Yu et al. 2005). Q043 was then crossed with ZS97 to generate a segregating F_2 population for primary mapping. This F_2 population, comprising 166 individuals, and the two parental lines (Q043 and ZS97) were grown in the summer of 2008 at the Ezhou experimental site of Huazhong Agricultural University, China (30.2°N, 114.5°E). The $F_{2:3}$ families for each of the F_2 plants were then grown at the Lingshui experimental site (18.2°N, 108.9°E) in the winter of 2008, planted in rows of ten individuals at a spacing of 16.7 × 26.6 cm.

A single individual containing only one substitution segment heterozygous at the target QTL was chosen from the above F_2 population, and self-pollinated to produce a large segregation population for fine mapping. This population of 1,025 individuals was grown at the Wuhan experimental site (30.4°N, 114.2°E) in the summer of 2009. The recombinant individuals were selected using the markers flanking the QTL, and the phenotypes were evaluated by progeny testing. The progeny for each recombinant was grown in a 3-row plot with 10 plants per row, at the Lingshui site in the winter of 2009. Progeny of some informative plants in which recombination occurred near the QTL was grown in larger quantities, in an 8-row plot with ten plants per row at Wuhan in the summer of 2011, for precise determination of phenotype and genotype. A pair of near-isogenic lines (NIL) having nearly the same genetic background as ZS97, with the exception of the small introduced segment (about 1.5 Mb in length) surrounding the QTL, was also selected and evaluated at Wuhan in the summer of 2011 following the same field management described above for the important recombinants.

Trait measurement

SL, SW, RLW, seed thickness and 1,000-seed weight (TSW) were measured as described in Tan et al. (2000). All measurements were repeated three times for each plant. The average lengths and widths and their ratios were used for phenotype analysis. The other agronomic traits including panicle length, panicle weight, primary branch number per panicle, secondary branch number per panicle, spikelets per panicle, grains per panicle, TSW, yield per plant, plant height and heading date were surveyed for

ZS97 and the NILs following the methods described in Wang et al. (2012), except that spikelet density was measured as the total number of spikelets per panicle divided by its panicle length.

The length and width of inner epidermal cells in the spikelet lemma at the heading stage were analyzed by scanning electron microscopy (SEM). The samples were prepared following the procedure described in Zhu et al. (2010). Samples were fixed in an FAA solution (90 % ethanol, 5 % glacial acetic acid and 5 % formaldehyde). After dehydration in a graded ethanol series and substitution with isoamyl acetate, the samples were critical-point dried by carbon dioxide, coated by sputtering with gold, and observed under a scanning electron microscope (JSM-6390/LV, Japan).

DNA extraction and marker analysis

Young seedling leaves were harvested for DNA extraction using the CTAB method (Murray and Thompson 1980). A bulked segregant analysis (BSA) approach was applied for primary mapping of QTLs in the specific genome region of interest (Michelmore et al. 1991). Two DNA bulks were created by pooling DNA from those individuals with extreme long or short phenotypes from the F_2 population. For genotyping of the parental lines (Q043 and ZS97), 264 simple sequence repeat (SSR) markers evenly distributed over the 12 chromosomes were designed according to the Gramene database (http://www.gramene.org/). For fine mapping, several insertion/deletion (InDel) and single-nucleotide polymorphism (SNP) markers (Table S1) were developed by randomly sequencing about 1-kb fragment every 5 kb along the length of the target region. The genomic region of the candidate genes including the promoter and entire coding region from the parental lines were amplified by PCR using several specific pairs of primers (Table S1). The primers were designed using the software Primer 3 (http://frodo.wi.mit.edu/) based on the reference genome of Nipponbare available at the database (http://rice.plantbiology.msu.edu/). PCR products were digested, purified and sequenced using BigDye Terminator Cycle Sequencing v3.1 (Applied Biosystems, USA) according to manufacturer's specifications. Sequencing reactions were run on an ABI 3730. Sequences were aligned by Sequencer 4.6 (Gene Codes Corporation).

Data analysis

Mapmaker/QTL 1.1b was used to scan QTLs in the F_2 individuals and the corresponding $F_{2:3}$ (Lincoln et al. 1992). Correlation analysis and Duncan's tests for comparison of the NILs and recombinant lines were performed in Statistica (StatSoft, 1999).

Quantitative real-time PCR

The RNA from panicles at various developmental stages was extracted using a TRIzol Reagent Kit (Invitrogen, USA) and treated with DNase I. Quantitative real-time (qRT) PCR was performed using an ABI 7500 Real-Time PCR System according to the manufacturer's instructions. The relative expression of each transcript was obtained by comparison with the expression of the rice *actin* gene and measurements were obtained using the relative quantification method (Livak and Schmittgen 2001). Each measurement was determined in two biological samples with three replicates for each sample. The primers used for qRT-PCR are also listed in Table S1.

Results

Preliminary mapping of qSS7 on chromosome 7

The parental lines (ZS97 and Q043) showed significant differences in the three seed shape parameters (SL, SW, and RLW). Q043 seeds were on average 15.2% longer and 10.2 % narrower than ZS97 (Fig. 1). The segregating population (F_2) from the cross of the two parents showed a binomial or trinomial distribution of seed length (Fig. 1), suggesting a major QTL might be involved in seed length in this F_2 population. Genome-wide genotyping of the parental line (Q043), using 264 SSR markers evenly distributed over the 12 chromosomes, demonstrated that it carries at least eight segments of Cypress located on chromosomes 1, 4, 5, 6, 7 and 12 in the genetic background of ZS9 (Figure S1a). To determine where this major QTL was located, eight individuals with the shortest seeds and eight with the longest seeds were selected from the F_2 population to create two bulked DNA pools. BSA showed that only two markers (RM505 and RM234), closely linked on chromosome 7, detected differences between the two bulked samples (Figure S1b), when using 16 SSR polymorphic markers (RM493, RM562 RM8004, RM273, RM252, RM241, RM470, RM405, RM574, RM586, RM225, RM505, RM234, RM101, RM1047, RM1999) flanking or within the eight introgressions (Figure S1a). These results suggested that the QTL for seed length was in the region RM505–RM234 on chromosome 7.

QTL analysis with eight markers surrounding the RM505–RM234 region further demonstrated that the QTL in the 1.5 cM interval of RM21930–RM21936 had simultaneous effect on SL, SW and RLW in both the 166 F_2 individuals and their $F_{2:3}$ families (Fig. 2a; Table 1). This major QTL explained 72.6 % of SL, 41.8 % of SW and 61.6 % of RLW variances in the $F_{2:3}$ families. It was thus tentatively designated *qSS7*. The Cypress allele at *qSS7*

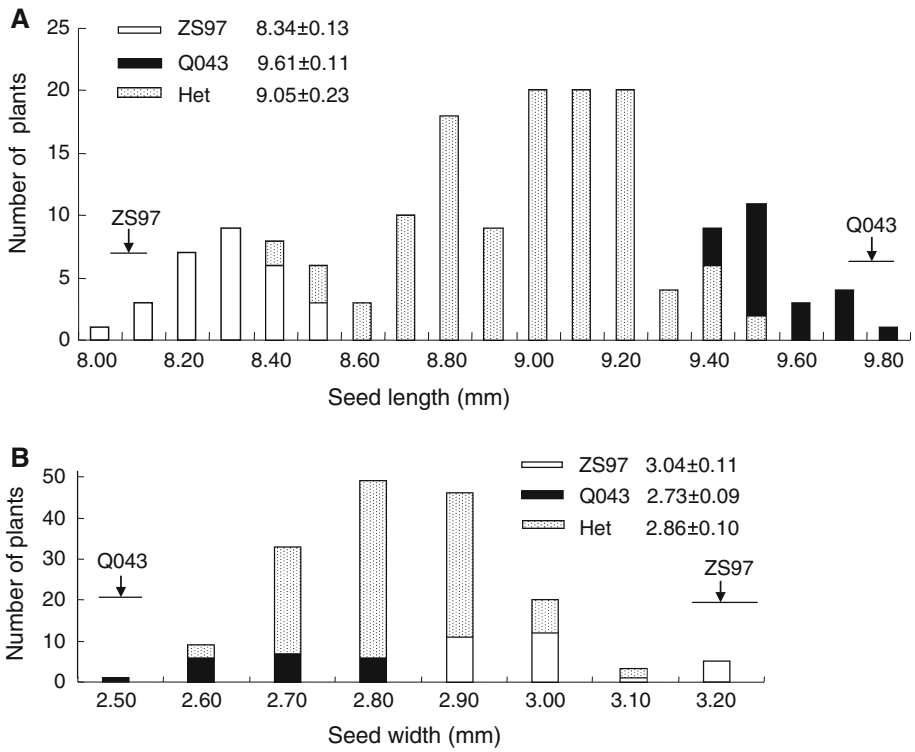

Fig. 1 Frequency distribution of seed length (**a**) and width (**b**) in the $F_{2:3}$ population derived from ZS97 and Q043. *Values* are means ± standard deviations (SD) for three *qSS7* genotypes (ZS97 homozygous, Q043 homozygous, heterozygous) assessed by marker RM21936

contributed to an increase in seed length, but a decrease in seed width. The additive effect of *qSS7* on seed length was approximately 0.63 mm, while the degree of dominance was low (0.12) in the $F_{2:3}$, indicating that the Cypress allele interacts with the ZS97 allele in a largely additive manner (Table 1).

Fine mapping of qSS7

One thousand twenty-five individuals generated from a single F_2 individual heterozygous only at the *qSS7* region were screened with two SSR markers (RM21930 and RM21936) flanking the QTL. Eighteen recombinants in the interval of RM21930–RM21936 were identified, and further genotyped using 14 newly developed SNP markers between ZS97 and Q043 (Fig. 2b; Table S1). Precise phenotyping was performed by progeny testing, which also enabled the determination of genotypes or segregation patterns of *qSS7* in each recombinant. The measurements of seed shape in the progeny of the 18 recombinants and two non-recombinant homozygous for ZS97 alleles and Cyprese alleles designated as controls C1 and C2 respectively are provided in Figure S2. Four recombinants with homozygous ZS97 alleles at *qSS7* showed identical short seeds, eight with homozygous Cypress alleles showed identical long seeds and four with heterozygous *qSS7* alleles displayed segregating seed lengths (Fig. 2c). It is noteworthy that the remaining two recombinants (M3 and M4) showed a segregating seed lengths suggesting they were heterozygous at *qSS7* (Fig. 3c), but with a short-like average. Overall, seed length seemed to be co-segregating with the marker DGS210. In particular, the six most informative recombinants (M3, M4, M8, M14, M15 and M16) allowed us to delimit *qSS7* to a region of approximately 23 kb between the markers GL293 and DGS220 on a BAC OSJNBb0018H10 (Fig. 2). This small region also affected SW and RLW (Fig. 2; Figure S2).

To further delineate the *qSS7* region and to determine whether the recombinant M4 with short-like seed contains a QTL for seed shape, the genotypes of the M4 and M7 progeny were analyzed at the marker GL293. The progeny comprising 71 individuals from the recombinant M4 showed significant differences in seed length and width among the three genotypes at the locus (Fig. 3a, b). Similarly, the progeny comprising 60 individuals for the recombinant M7 showed significant differences in seed length and width among the three genotypes: the heterozygote had an intermediate seed length, longer (and narrower) than the homozygote for ZS97 alleles, but shorter (and wider) than the homozygote for Cypress alleles (Fig. 3c, d). Notably, the distribution of seed shape (i.e., SL) in the progeny of M7 was consistent with that a major QTL for seed shape exhibited by the primary F_2 population of ZS97 and Q043. Interestingly, the progeny test for the recombinant M7 revealed that the average effect of *qSS7* is about threefold greater than in M4 for both seed length and width (Fig. 3). M7 and M4 contain the heterozygous fragments GL293–DGS220 and GL293–DGS210, respectively,

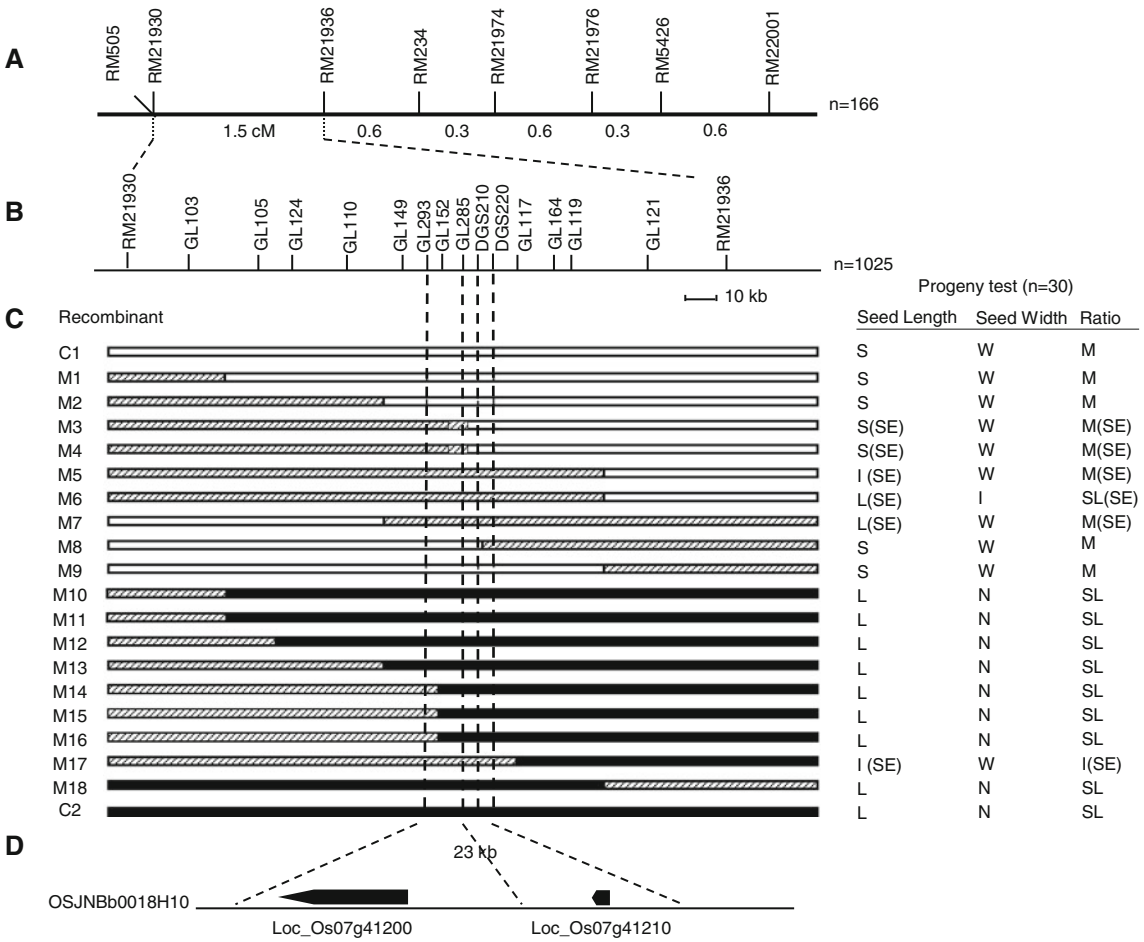

Fig. 2 Fine mapping of *qSS7*. **a** The genetic linkage map of the *qSS7* region based on 166 F_2 plants. *Numbers under the line* indicate genetic distance between adjacent markers. **b** High-resolution map of the *qSS7* region from 1,025 F_2 individuals. **c** Progeny testing of 18 recombinants between RM21930 and RM21936 narrowed *qSS7* to a 23-kb region between GL293 and DGS220. The non-recombinant homozygous for the ZS97 alleles and the Cyprese alleles in the interval were designated as the controls C1 and C2, respectively. Scores for seed length: S (short), SE (segregating) or I (intermediate), and L (long); for seed width: W (wide), I (intermediate), and N (narrow); for ratio of length to width: M (medium), SE (segregating) or I (intermediate), and SL (slender). **d** Two predicted candidate genes in the target region based on the genome annotation databases

Table 1 The parameters of the QTL (*qSS7*) in the interval RM21930–RM21936 for three seed shape traits in F_2 and $F_{2:3}$ populations

Trait	Population	*LOD*	*A*	*D*	*D/A*	R^2 %
SL	F_2	35.8	0.60	0.10	0.16	63.7
	$F_{2:3}$	43.7	0.63	0.08	0.12	72.6
SW	F_2	12.6	−0.12	−0.03	0.23	29.5
	$F_{2:3}$	19.5	−0.15	−0.03	0.20	41.8
RLW	F_2	28.8	0.40	0.06	0.15	55.1
	$F_{2:3}$	32.7	0.38	0.04	0.11	61.6

LOD logarithm of odds, *A* additive effect, *D* dominant effect, *D/A* dominant degree, R^2 % percentage of total phenotypic variance explained by the QTL

within an overlapping region (GL293–GL285) surrounding *qSS7*. The inconsistency of QTL effect in the M7 and M4 progeny indicates a major gene that may contain *qSS7* in the non-overlapping segment (GL285-DGS220) held by M7, and a minor gene in the overlapping segment. The *qSS7* region appears to be a composite QTL containing at least two linked genes with different effects regulating seed length and width.

Candidate genes in the target region

The 23-kb target region contains two common predicted genes (LOC_Os07g41200, LOC_Os07g41210) based on three genome annotation databases (http://rice.plantbiology.msu.edu/; http://ricegaas.dna.affrc.go.jp/; http://linux1.softberry.com/) (Fig. 2d). The former gene encodes a protein with unknown function, and the latter encodes a conserved hypothetical protein. It is noteworthy that two other predicted genes each encoding a hypothetical protein in this region were annotated by RiceGAAS using the monocot

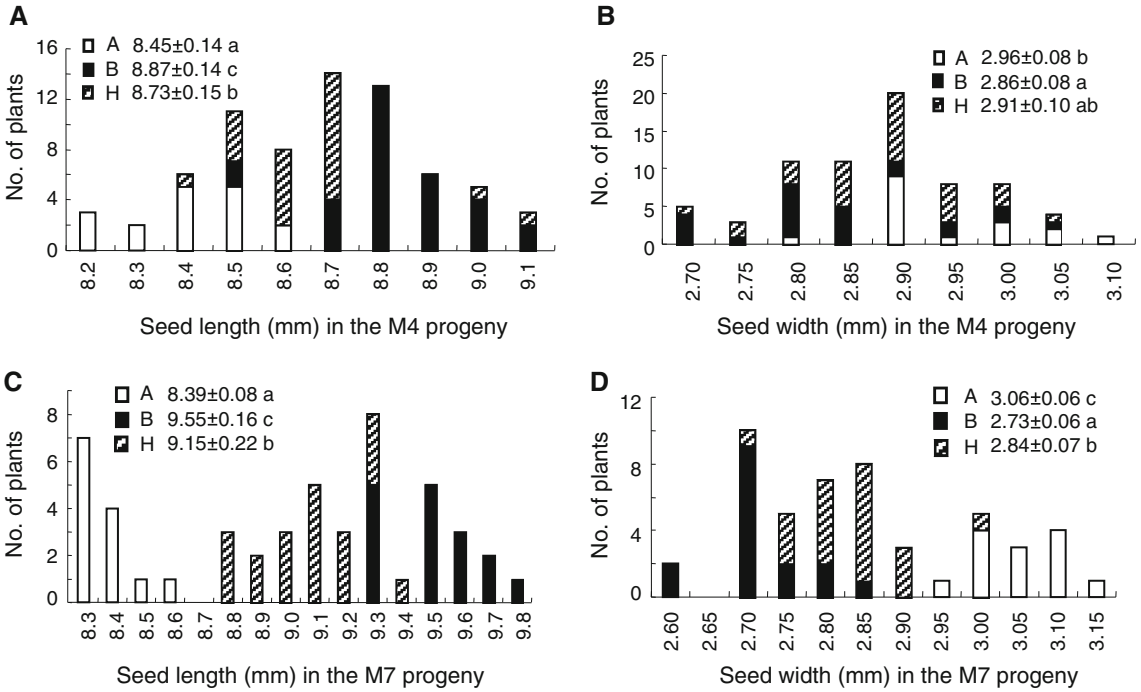

Fig. 3 Progeny testing of two recombinants M4 (**a**, **b**) and M7 (**c**, **d**) showed significant differences in seed length and width among the three genotypes at GL293 in the *qSS7* region. Three genotypes A, B and H in each panel represent the homozygote for ZS97 alleles, Cypress alleles, and the heterozygote at the locus, respectively. Mean ± SD for each genotype is provided, with *different letters* (*a*, *b* and *c*) indicating significant difference among the genotypes at $P < 0.05$

plant model (http://ricegaas.dna.affrc.go.jp/). However, there are no full-length cDNAs or ESTs responding to the predicted genes except LOC_Os07g41200 in the public databases (http://rapdb.dna.affrc.go.jp/; http://rice.plantbiology.msu.edu/). BLAST searching identified that LOC_Os07g41200 has one closely related homolog LONGIFOLIA1, which affect longitudinal expansion of seeds in *Arabidopsis* (http://www.ncbi.nlm.nih.gov/). The genomic fragments of approximately 7.5 kb for LOC_Os07g41200, and 1.8 kb for LOC_Os07g41210, as well as about 1 kb covering each of the other two predicted genes were sequenced. Sequence comparison between ZS97 and Q043 demonstrates a number of nucleotide differences in LOC_Os07g41200 and LOC_Os07g41210 (Figure S3), but no difference in the other predicted genes. There are 26 Indels and SNPs located in the promoter region and 14 polymorphisms in the coding region of LOC_Os07g41200, and 30 in the promoter and nine polymorphic sites in the coding region of LOC_Os07g41210. A few are worth mentioning as possible causal sites for the QTL. Two nucleotide substitutions in the coding region of LOC_Os07g41200 change the predicted amino acids. For LOC_Os07g41210, four nucleotide substitutions in the coding region change the predicted amino acids, one of which, a single nucleotide transition (G-to-A), results in a Tryptophan codon (TG**G**) in Q043 becoming a termination codon (TG**A**) in ZS97 (Figure S3). This premature termination causes a 30-aa truncation at the C-terminus of the predicted protein in ZS97.

Expression of the candidate genes

Expression analysis of *LOC_Os07g41200* and *LOC_Os07g41210* based on the microarray data of the tissues including root, stem, leaf, panicle and endosperm collected throughout the life cycle of the variety ZS97 (Wang et al. 2010), indicated that *LOC_Os07g41200* was expressed in various tissues, especially with the highest level in young panicles; however, *LOC_Os07g41210* was relatively low or hardly detected in the tissues assayed (Figure S4; http://crep.ncpgr.cn/). Quantitative real-time PCR analysis showed that expression patterns of both genes in Q043 are different from those in ZS97. In particular, *LOC_Os07g41200* in Q043 is expressed at higher levels than in ZS97 through almost the entire panicle development stage (panicles ranging 3–23 cm in length), except the early stage (panicles ≤2 cm in length) (Fig. 4a). In contrast, the expression levels of *LOC_Os07g41210* in ZS97 are higher at most of the panicle development stages except the late one (panicles reaching 23 cm in length) than those in Q043 (Fig. 4b). These results suggest that *LOC_Os07g41200* might be involved in panicle development in rice.

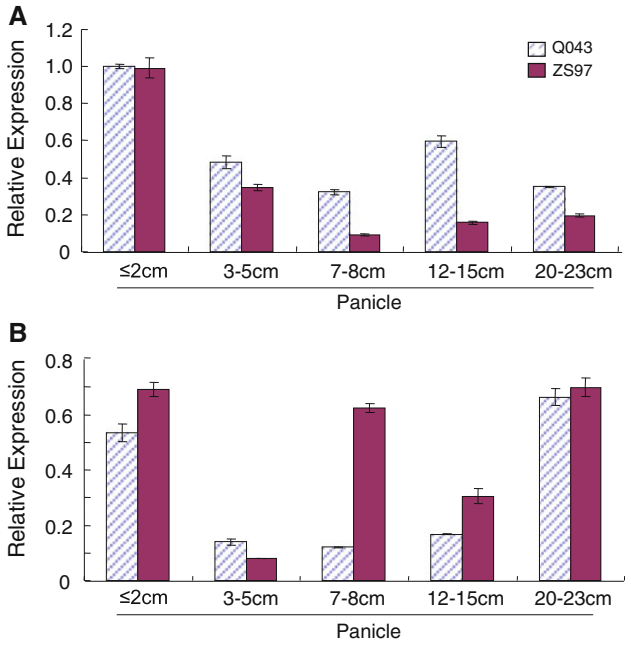

Fig. 4 Relative expression of two candidate genes **a** LOC_Os07g41200 and **b** LOC_Os07g41210 at five panicle development stages by qRT-PCR. The *actin* gene was used as an internal control. Mean values and standard error (*bar*) were determined by two biological replicates measured three times for each sample

Comparison of NILs

The phenotypes of two NILs differing only at the *qSS7* region were evaluated. The NILs carrying the Cypress and ZS97 alleles at the *qSS7* locus were designated NIL (Cyp) and NIL (ZS), respectively. Compared with NIL (ZS), NIL (Cyp) showed significantly increased SL, RLW and panicle length, decreased seed width and spikelet density, and a small but statistically significant difference in heading date ($P = 0.03$) (Table 2). However, no significant differences were observed between the NILs in other traits such as seed thickness, number of spikelets per panicle, TSW and plant height (Table 2). These results validate the QTL analysis showing Cypress *qSS7* allele enhanced SL and RLW but reduced SW.

To investigate whether cell expansion was affected in the NILs, the length and width of the inner epidermal cells of the outer glume (lemma) enclosed within a square that had the same number of cells ($n = 30$) was measured by SEM (Fig. 5b, c). The SEM analysis showed that NIL (Cyp) had an average cell length of 114.7 ± 27.1 μm, significantly longer than those of ZS97 (averaging 100.9 ± 23.6 μm). Conversely, the average cell width in NIL (Cyp) was 56.9 ± 14.2 μm, significantly narrower than in ZS97, which was 64.5 ± 8.7 μm (Fig. 5d). The NIL (Cyp) epidermal cells were 13.7 % longer and 11.8 % narrower than their counterparts, suggesting that the longer, narrower cells of the NIL largely accounted for the differences in seed shape (Fig. 5a). These results indicate that *qSS7* may have function in regulating both cell length and width elongation in the lemma.

Table 2 Comparison of seed shape and other traits in the NILs that contain different alleles at *qSS7*

Trait	NIL(ZS)	NIL (Cyp)	*P* value
Seed length (mm)	8.47 ± 0.07	9.52 ± 0.06	0.000
Seed width (mm)	2.96 ± 0.12	2.79 ± 0.04	0.001
Ratio of length to width	2.86 ± 0.12	3.42 ± 0.06	0.000
Seed thickness (mm)	2.01 ± 0.06	2.02 ± 0.04	0.666
Panicle length (cm)	23.1 ± 1.06	25.1 ± 1.21	0.000
Panicle weight (g)	2.25 ± 0.41	2.15 ± 0.40	0.321
Primary branch number	9.5 ± 1.18	9.5 ± 1.04	0.958
Secondary branch number	20.3 ± 4.23	18.5 ± 3.71	0.095
Grains per panicle	75.2 ± 18.58	68.3 ± 16.75	0.131
Spikelets per panicle	108.2 ± 16.35	101.0 ± 14.56	0.073
Spikelet density	4.7 ± 0.61	4.0 ± 0.54	0.000
1,000-seed weight (g)	26.4 ± 0.77	26.9 ± 0.34	0.052
Yield per plant (g)	15.3 ± 6.44	16.3 ± 5.61	0.502
Productive tiller	9.9 ± 2.80	10.5 ± 2.91	0.393
Heading date (d)	64.9 ± 2.41	63.5 ± 2.52	0.030
Plant height (cm)	84.6 ± 3.17	85.1 ± 4.07	0.580

Phenotypes are given as mean ± standard deviation of 30 individuals for each line. A *t* test was applied to generate *P* values

Discussion

qSS7 for seed shape has little effect on seed weight

Seed shape is one of the most important attributes of grain weight and appearance that determines the market value of rice. The *qSS7* on the long arm of chromosome 7 described here acts as a major QTL affecting seed shape parameters. A consistent QTL for grain length has been mapped to an approximately 300-kb or larger interval (Amarawathi et al. 2008; Shao et al. 2010) or adjacent region but with only a small amount of the SL variation explained in the populations (Bai et al. 2010). *qSS7* is simultaneously associated with SL, SW and RLW, but has little or no effect on TSW (Table 2). The effect of *qSS7* is different from previously reported QTLs and/or cloned genes for seed shape which affect TSW to some extent. For example, *GS3* and *SRS3* had major effects on both grain length and weight (Fan et al. 2006; Kitagawa et al. 2010), and *GW2* and *qSW5* act as major QTLs for both grain width and weight (Song et al. 2007; Shomura et al. 2008). Notably, *qSS7* has an opposite effect on SL and SW: the Cypress allele contributes to an increase in SL and decrease in SW but does not alter seed

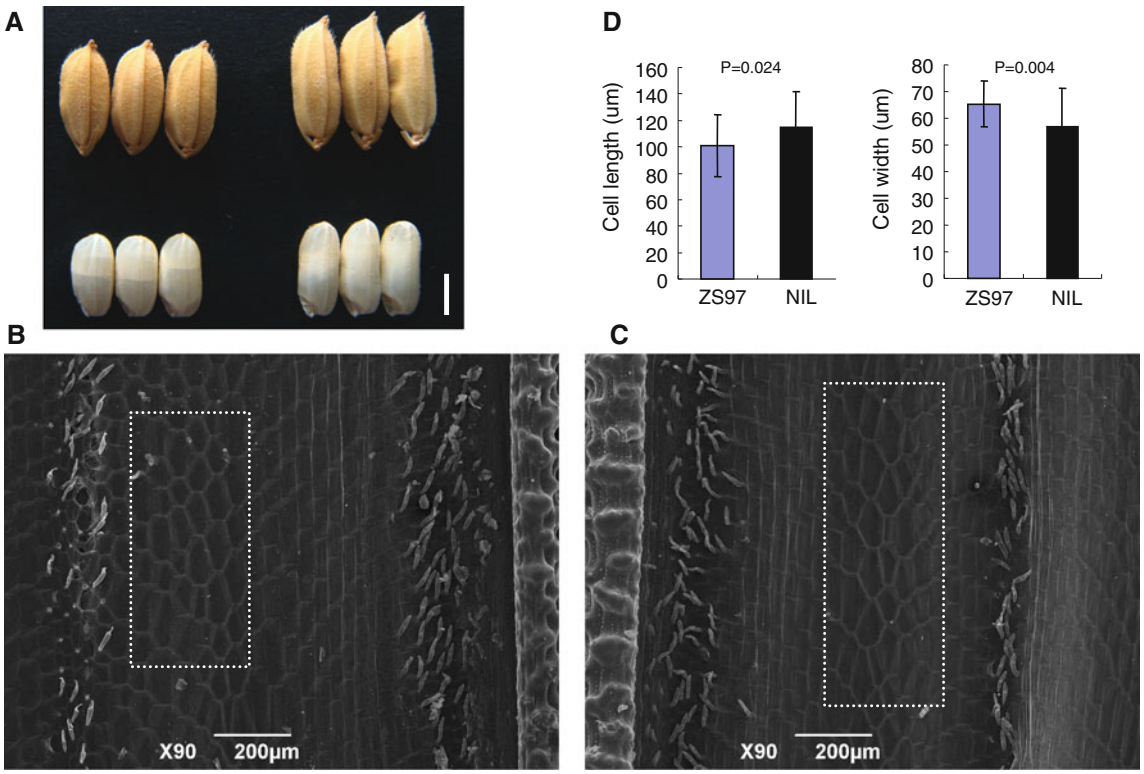

Fig. 5 Seed differences between ZS97 and NIL. **a** Mature seeds and grains of ZS97 (*left*) and the NIL carrying Cypress alleles (*right*), *scale bar* 3 mm. Scanning electron microscopy shows the differences in inner epidermal cells of the lemma between **b** ZS97 and **c** the NIL. **d** Mean ± SD ($n = 30$) cell length and width in the *boxed regions* of **b** ZS97 and **c** the NIL. *P* values were generated by a *t* test

thickness, and consequently does not change the seed volume or weight. This finding suggests that *qSS7* might regulate seed length and width independent of seed weight. This notion is supported by evidence from correlation analysis in the $F_{2:3}$ population, in which SL was highly positively correlated with RLW ($r = 0.87$, $P < 0.01$) and negatively correlated with SW ($r = -0.56$, $P < 0.01$), but not associated with TSW ($r = 0.10$, $P > 0.5$). This makes *qSS7* of particular interest because the way it acts on SL and SW in opposing directions differs from the many previously reported seed shape-related genes. *GS3* for instance, has a major effect on seed length and a minor effect in the same direction on seed width, and *qSW5/GW5* plays a significant role in seed width but has a less significant effect on seed length (Weng et al. 2008; Shomura et al. 2008). The gene(s) underlying *qSS7* appear to have differential modulation in growth (or growth arrest) along the longitudinal and transverse axes of the seed. Interestingly, the panicles of NIL (ZS) and NIL (Cyp) were significantly different in length, suggesting that *qSS7* may regulate cellular elongation within the panicle as well. In fact, several genes (i.e., *SG1*, *EP2*, *EP2* and *qPE9-1*) recently isolated for panicle traits (i.e., panicle length, panicle erectness) have also been reported to affect grain shape (Piao et al. 2009; Zhou et al. 2009; Zhu et al. 2010; Nakagawa et al. 2012). For example, the *ep3* mutant of rice showed erect and short panicles and wide grains, but no differences in grain length and weight as compared to the wild type (Piao et al. 2009). The mutation alleles of *EP2* and *qPE9-1* caused an erect- and short-panicle phenotypes, as well as a short-grain phenotype with a decrease in grain length and weight, but an increase in grain width and grain thickness in the *ep2* mutant (Zhu et al. 2010), while no obvious changes in grain width and grain thickness in the *qpe9-1* mutant (Zhou et al. 2009). These results together with our study suggest that some common mechanisms are likely to control both panicle length and grain shape in rice.

Implication of qSS7

The development and utilization of hybrid rice has made a great contribution to food production in China and other Asian countries. However, the appearance quality (i.e., grain shape) of most *indica* hybrids has failed to keep pace with current demands due to the short-grain female parents (Fan et al. 2006). In terms of breeding perspectives to make female parents have the long-grain phenotype, *GS3* can be applied to improve grain yield and appearance quality, but has limitations for use in hybrids because the heterozygote always expresses the short-grain phenotype due to its

recessive long-grain allele. Comparatively, *qSS7* could modify seed dimensions (i.e., SL and SW) antagonistically leading to a much longer and narrower seed, which is more useful for breeding hybrids with slender seed due to the partial dominant allele for the long grain phenotype. Thus, the cloning and characterization of *qSS7* currently underway will not only be helpful in understanding mechanisms regulating seed shape, but may also facilitate combining *qSS7* with other desirable genes to improve both grain yield and appearance quality in rice hybrids.

Fine mapping clearly shows that *qSS7* is located in a small region containing at least two possible candidate genes. Interestingly, LOC_Os07g41200 has one closely related *Arabidopsis* homolog *LNG1* that has been shown to regulate longitudinal cell elongation (Lee et al. 2006). This is consistent with our hypothesis that *qSS7* may have a function in regulating cell length elongation in the lemma (Fig. 5). In addition, *LOC_Os07g41200* is specifically highly expressed in the seeds or panicles of rice (Wang et al. 2010). Therefore, the role of this gene in seed shape needs to be explored in rice. It is also noteworthy that the sub-fragment GL285–DGS220 of the *qSS7* region that harbors *LOC_Os07g41210* seemed to have a relatively large effect on seed shape (Figs. 2, 3). Moreover, this predicted gene has a premature termination causing a truncated protein in ZS97. Therefore, *LOC_Os07g41210* cannot be ruled out as a candidate underlying the QTL. Further, transgenic complementation experiments and RNA interference of the potential candidate genes and association analysis of the sequence variations of these genes in a large panel of rice germplasm are necessary to test the definitive role either or both of these genes play in determining seed shape.

Acknowledgments This work was supported by grants from National Program on Key Basic Research Project, National Special Program for Research of Transgenic Plant of China, and the National Natural Science Foundation of China.

References

Aluko G, Martinez C, Tohme J, Castano C, Bergman C, Oard JH (2004) QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima*. Theor Appl Genet 109:630–639

Amarawathi Y, Singh R, Singh AK, Singh VP, Mohapatra T, Sharma TR, Singh NK (2008) Mapping of quantitative trait loci for basmati quality traits in rice (*Oryza sativa* L.). Mol Breeding 21:49–65

Bai XF, Luo LJ, Yan WH, Kovi MR, Zhan W, Xing YZ (2010) Genetic dissection of rice grain shape using a recombinant inbred line population derived from two contrasting parents and fine mapping a pleiotropic quantitative trait locus qGL7. BMC Genet 11:16

Fan CC, Xing YZ, Mao HL, Lu TT, Han B, Xu CG, Li XH, Zhang Q (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112:1164–1171

Govindaraj P, Arumugachamy S, Aheswaram M (2005) Bulked segregant analysis to detect main QTL associated with grain quality parameters in Basmati370/ASD 16 cross in rice (*Oryza sativa* L.). Euphytica 144:61–68

Juliano BO, Villareal CP (1993) Grain quality evaluation of world rice. International Rice Research Institute, Manila

Kitagawa K, Kurinami S, Oki K, Abe Y, Ando T, Kono I, Yano M, Kitano H, Iwasaki Y (2010) A novel kinesin 13 protein regulating rice seed length. Plant Cell Physiol 51:1315–1329

Lee YK, Kim GT, Kim IJ, Park J, Kwak SS, Choi G, Chung WI (2006) LONGIFOLIA1 and LONGIFOLIA2, two homologous genes, regulate longitudinal cell elongation in Arabidopsis. Development 133:4305–4314

Li J, Xiao J, Grandillo S, Jiang L, Wan Y, Deng Q, Yuan L, McCouch SR (2004) QTL detection for rice grain quality traits using an interspecific back-cross population derived from cultivated Asian (*O. sativa* L.) and African (*O.glaberrima* S.) rice. Genome 47:697–704

Li YB, Fan CC, Xing YZ, Jiang YH, Luo LJ, Sun L, Shao D, Xu CJ, Li XH, Xiao JH, He YQ, Zhang QF (2011) Natural variation in GS5 plays an important role in regulating grain size and yield in rice. Nat Genet 43:1266–1269

Lincoln S, Daly M, Lander ES (1992) Mapping genes controlling quantitative traits with MAPMAKER/QTL 1.1. Whitehead Institute Technical Report, 2nd edn. Whitehead Institute, Cambridge, Mass

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25:402–408

Luo YK, Zhu ZW, Chen NG, Dunan BW, Zhang LP (2004) Grain types and related quality characteristics of rice in China. Chinese J Rice Sci 18:135–139

Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A (2005) MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. Proc Natl Acad Sci USA 102:17531–17536

Mao HL, Sun SY, Yao JL, Wang CR, Yu SB, Xu CG, Li XH, Zhang QF (2010) Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. Proc Natl Acad Sci USA 107:19579–19584

Michelmore RW, Oaran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked analysis: a rapid method to detect markers in specific genomic region by using segregating population. Proc Natl Acad Sci USA 88: 9828–9832

Murray MG, Thompson WF (1980) Rapid isolation of high molecular-weight plant DNA. Nucl Acid Res 8:4321

Nakagawa H, Tanaka A, Tanabata T, Ohtake M, Fujioka S, Nakamura H, Ichikawa H, Mori M (2012) SHORT GRAIN1 decreases organ elongation and brassinosteroid response in rice. Plant Physiol 158:1208–1219

Piao R, Jiang W, Ham TH, Choi MS, Qiao Y, Chu SH, Park JH, Woo MO, Jin Z, An G, Lee J, Koh HJ (2009) Map-based cloning of the ERECT PANICLE 3 gene in rice. Theor Appl Genet 119: 1497–1506

Rabiei B, Valizadeh M, Ghareyazie B, Moghaddam M, Ali AJ (2004) Identification of QTLs for rice grain size and shape of Iranian cultivars using SSR marker. Euphytica 137:325–332

Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scorr RJ (2006) The AUXIN RESPONSE FACTOR2 gene for Arabidopsis links auxin signaling, cell division, and the size of seeds and other organs. Development 133:251–261

Shao GN, Tang SQ, Luo J, Wei XJ, Tang A, Wu JL, Zhuang JY, Hu PS (2010) Mapping of qGL7-2, a grain length QTL on chromosome 7 of rice. J Genet Genomics 37:523–531

Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M (2008) Deletion in a gene associated with grain size increased yield during rice domestication. Nat Genet 40: 1023–1028

Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet 39:623–630

StatSoft (1999) Statistica. StatSoft Incorporated, Tusla

Takano-Kai N, Jiang H, Kubo T, Sweeney M, Matsumoto T, Kanamori H, Padhukasahasram B, Bustamante C, Yoshimura A, Doi K, McCouch S (2009) Evolutionary history of GS3, a gene conferring grain length in rice. Genetics 182:1323–1334

Tan YF, Xing YZ, Li JX, Yu SB, Xu CG, Zhang QF (2000) Genetic bases of appearance quality of rice grain in Shanyou 63, an elite rice hybrid. Theor Appl Gent 101:823–829

Thomson MJ, Tai TH, McClung AM, Lai XH, Hinga ME, Lobos KB, Xu Y, Martinez CP, McCouch SR (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theor Appl Genet 107:479–493

Wan XY, Wan JM, Jiang L, Wang JK, Zhai HQ, Weng JF, Wang HL, Lei CL, Wang JL, Zhang X, Cheng ZJ, Guo XP (2006) QTL analysis for rice grain length and fine mapping of an identified QTL with stable and major effects. Theor Appl Genet 112: 1258–1270

Wang L, Xie W, Chen Y, Tang W, Yang J, Ye R, Liu L, Lin Y, Xu C, Xiao J, Zhang Q (2010) A dynamic gene expression atlas covering the entire life cycle of rice. Plant J 61:752–766

Wang P, Zhou GL, Cui KH, Li ZK, Yu SB (2012) Clustered QTL for source leaf size and yield traits in rice (*Oryza sativa* L.). Mol Breeding 29:99–113

Weng JF, Gu SH, Wan XY, Gao H, Guo T, Su N, Lei CL, Zhang X, Cheng ZJ, Guo XP, Wang JL, Jiang L, Zhai HQ, Wan JM (2008) Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. Cell Res 18: 1199–1209

Yu SB, Mu JX, Zhao SJ, Zhou HJ, Tan YB, Xu CG, Luo L, Zhang Q (2005) Development and selection of introgression lines with the identical genetic background of varieties Zhenshan 97 and 9311. Mol Plant Breeding 3:629–636 (in Chinese with English abstract)

Zhou Y, Zhu J, Li Z, Yi C, Liu J, Zhang H, Tang S, Gu M, Liang G (2009) Deletion in a quantitative trait gene qPE9-1 associated with panicle erectness improves plant architecture during rice domestication. Genetics 183:315–324

Zhu KM, Tang D, Yan CJ, Chi ZC, Yu HX, Chen JM, Liang JS, Gu MH, Cheng ZK (2010) EREECT PANICLE2 encodes a novel protein that regulates panicle erectness in *Indica* rice. Genetics 184:343–350