Theor Appl Genet (2012) 125:1303–1312
DOI 10.1007/s00122-012-1914-7

ORIGINAL PAPER

Allelic variation for a candidate gene for *GS7*, responsible for grain shape in rice

Gaoneng Shao · Xiangjin Wei · Mingliang Chen · Shaoqing Tang · Ju Luo · Guiai Jiao · Lihong Xie · Peisong Hu

Received: 11 December 2011 / Accepted: 5 June 2012 / Published online: 7 July 2012

Abstract Grain shape is an important component of end-use quality in rice. The genomic location of the grain shape QTL *GS7* was narrowed to lie within a 4.8-kb segment on chromosome 7. The homologous region in cv. Nipponbare contains no annotated genes, while two open reading frames were predicted, one of which (*ORF2*) represented a likely candidate for *GS7* gene on the basis of correlation between sequence variation and phenotype. Semi-quantitative and quantitative RT-PCR analysis of *ORF2* transcription showed that the gene was active in both the leaf and panicle when the cv. D50 allele was present, but not in the presence of the cv. HB277 allele. A microsatellite-based phylogeny and a re-sequencing analysis of *ORF2* among a set of 52 diverse rice accessions suggested that the cv. D50 *GS7* allele may have originated from the *tropical japonica* genepool. The effect on grain length of the alternative alleles at *GS7*and *GS3* showed that combination type 3/A was associated with longer grains than type 1/A. An Indel marker developed within the *ORF2* sequence was informative for predicting grain length.

Communicated by Q. Zhang.

G. Shao, S. Tang and X. Wei contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1914-7) contains supplementary material, which is available to authorized users.

G. Shao · X. Wei · M. Chen · S. Tang · J. Luo · G. Jiao · L. Xie · P. Hu (✉)
State Key Laboratory of Rice Biology,
China National Rice Research Institute,
Hangzhou 310006, China
e-mail: hupeisong@yahoo.com.cn

Introduction

Grain shape, as defined by its length, width and their ratio, is an important measure of end-use quality in rice. Both grain length and width are polygenically inherited (Tan et al. 2000) and controlling loci have been identified on each of the 12 rice chromosomes (http://www.gramene.org). Some of them (*qGL7*, *qGL7-2*, *GW2*, *qSW5*, *GW5*, *GS3* and *GS5*) have been fine mapped and even isolated in recent years. Both *qGL7* and *qGL7-2* underlie the determination of grain length, separated from one another by 13.2 cM on the long arm of chromosome 7, and their locations have each been narrowed to within a ~300-kb segment (Bai et al. 2010; Shao et al. 2010). *GW2*, which affects grain width, lies on chromosome 2 and is known to encode a RING-type E3 ubiquitin ligase. The loss of *GW2* function induces an increase in cell number, resulting in the formation of a larger spikelet hull (Song et al. 2007). The *qSW5* locus on chromosome 5 increases spikelet volume and cell number in the outer glume (Shomura et al. 2008). *GW5* is thought to act within the ubiquitin–proteasome pathway to regulate cell division during grain development, and encodes a nuclear protein identical to the *qSW5* product (Weng et al. 2008). The *GS3* product is a transmembrane protein, and regulates grain size through the action of four functional domains (Fan et al. 2006; Mao et al. 2010). Finally, *GS5* which regulates grain width encodes a putative serine carboxypeptidase and functions as a positive regulator of grain size (Li et al. 2011).

In addition, the domestication of crop plants has been an important factor in the development of human civilization. It has relied heavily on the selection of favorable alleles at a rather small number of so-called "domestication syndrome" genes. Some of these genes have been successfully isolated in recent years. The domesticated allele at the

tomato *fw2.2* gene is responsible for a ~30 % increase in fruit weight and its selection was probably a key event in the crop's domestication (Frary et al. 2000). Variation at *tga1* altered the form of the wild teosinte ear into that of the maize plant; this gene has been shown to belong to an SBP-domain family of transcriptional regulators (Wang et al. 2005). Domestication-related genes have been exploited to understand the process of domestication process in wheat and barley (Simons et al. 2006; Komatsuda et al. 2007). In rice, the domestication-related genes isolated to date include *rc*, *wx* and *badh2.1*, and the acquisition of these sequences has facilitated the elucidation of both the domestication process and the origin of the two major rice genepools *indica* and *japonica* (Konishi et al. 2008). Genes underlying grain shape (including *GW5*, *GS3* and *GS5*) are all thought to have been actively selected during the domestication of rice (Weng et al. 2008; Fan et al. 2009; Li et al. 2011).

In the present report, we describe the fine mapping of *GS7* and the identification of a likely candidate gene. We also report a diversity analysis of the candidate gene among a collection of rice varieties.

Materials and methods

Plant materials

A recombinant inbred line (RIL) population of 190 lines was bred from the cross cv. D50 × cv. HB277 (Shao et al. 2009). Using a set of 102 informative microsatellites distributing over all 12 chromosomes, two genetic stocks were developed from these RILs. The first, a residual heterozygous line (RHL), contained a heterozygous segment flanked by the microsatellite loci RM11 and RM134 on chromosome 7, but was homozygous throughout most of the rest of the genome (Fig. 1a) (Shao et al. 2010). The second was bred from an RIL in which 60 % of the genome had been inherited from cv. HB277, while the segment between RM11 and RM134 had been inherited from cv. D50. After successive crosses with cv. HB277, marker-assisted selection was applied to derive a pair of near isogenic lines (NILs) which differed mainly for the RM11–RM134 segment (Fig. 1b). The presence of the cv. D50 allele was associated with the formation of more slender, longer grains (Fig. 2). The fine mapping of *GS7* was based on 1,000 selfed progenies of the RHL that were screened genotypically with the markers Indel1 and RM21945, which defined the segment known to contain *GS7*.

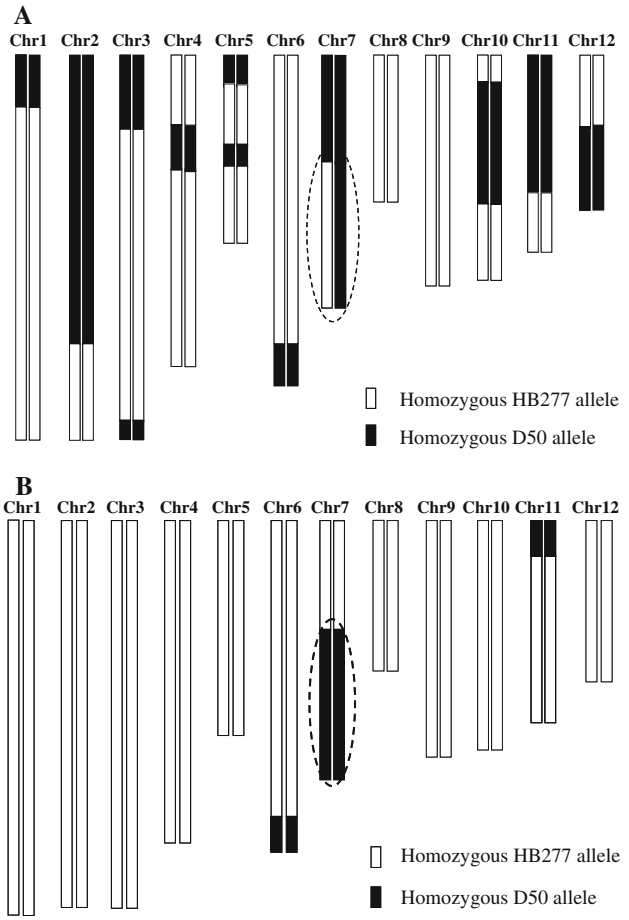

Fig. 1 Graphical genotypes of **a** RHL and **b** the NIL pair

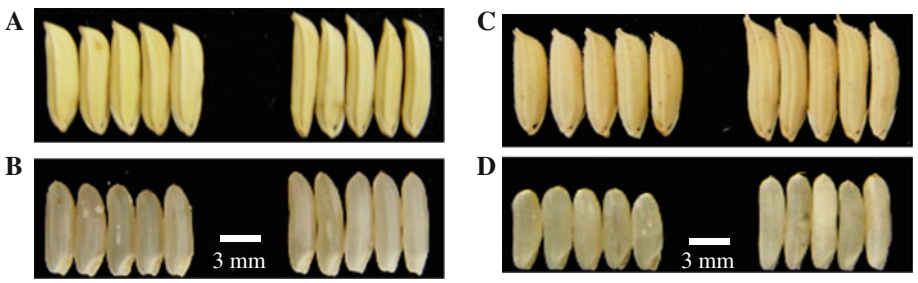

Fig. 2 Grain shape as influenced by the identity of the *GS7* allele present. **a**, **b** Grains formed by RHL progeny carrying the cv. HB277 allele on the *left*, and by those carrying the cv. D50 allele on the *right*. **c**, **d** Grains formed by the NIL carrying the cv. HB277 allele on the *left*, and by the NIL carrying the cv. D50 allele on the *right*. **a**, **c** Grains prior to dehulling; **b**, **d** after dehulling

Homozygous derivatives of each recombinant involving this segment were grown in a randomized block design in the field as six rows of ten plants each. To further narrow the size of the segment containing *GS7*, a second set of 10,000 F_2 progenies of RHL was produced. A set of 60 $F_{2:3}$ individuals bred from each F_2 recombinant was grown in the field and genotyped. Finally, the homozygous derivatives of each recombinant were used to assess the correlation between genotype and grain shape. For these latter experiments, eight plants per recombinant line (RHL-F_4) were grown in the field.

Trait assessment

Plant height (PH), heading date (HD), panicle length (PL), panicle number per plant (PN), the number of filled grains per panicle (NFGP), the total number of spikelets per panicle (TNSP), spikelet fertility (SF) and 1,000-grain weight were determined from eight field-grown plants of RHL and the NIL pair. PL was defined as the separation between the panicle neck and its tip (excluding the awn), while SF reflected the performance of all the panicles on each plant per line. The grains were allowed to dry naturally following harvest, after which they were dehulled to determine grain shape. The rapid analysis system SC-E was applied to a sample of 20 grains per line to obtain mean values for grain of: length, width, the ratio between length and width, circumference, surface area, roundness and equivalent diameter.

Genotypic analysis

DNA was extracted from seedling leaves, following the protocol described by Lu and Zheng (1992). Each 10 μL PCR contained 1 μL 10× PCR buffer (25 mM $MgCl_2$), 0.8 μL 2 mM dNTP, 1 μL of each primer (5 μM), 0.25 μL 2 U/μL *Taq* DNA polymerase and 1 μL template DNA. The cycling regime consisted of an initial denaturation of 94 °C/2 min, followed by 30 cycles of 94 °C/45 s, 55 °C/45 s and 72 °C/60 s, with a final extension of 72 °C/8 min. The PCR products were electrophoretically separated through non-denaturing 6 % polyacrylamide gels (Shi et al. 2005). Microsatellite primer sequences were obtained from the Gramene database (http://www.gramene.org).

Transcription analysis

RNA was isolated from three plants of each line and the samples were pooled. RNA was extracted from the flag leaf and young panicle using the RNAiso Plus reagent (Takara) and treated with RNase-free DNaseI (Takara) to remove any contaminating genomic DNA. About 1 μg of total RNA was converted into cDNA using an M-MLV RTase cDNA Synthesis kit (Takara), according to the manufacturer's instructions. A portion of the *OsACT1* sequence provided the reference for semi-quantitative RT-PCR experiments, in which the PCR regime comprised an initial denaturation step (95 °C/4 min), followed by 25 cycles of 95 °C/30 s, 55 °C/30 s and 72 °C/30 s, ending with an extension step of 72 °C/10 min (38 cycles for the target gene). For quantitative RT-PCR, the same cDNA template was amplified using 2× SYBR Green PCR Master Mix (Takara) on a Roche Lightcycler 480II Real-Time PCR System. The relative expression level of each transcript was obtained by normalization against the *OsACT1* signal, based on the $2^{-\triangle\triangle CT}$ method. Here, the PCR regime comprised an initial denaturation step (95 °C/4 min), followed by 40 cycles of 95 °C/15 s, 55 °C/30 s. Three independent RNA extractions were produced from each pooled sample of three plants per line, and each derived cDNA sample was then subjected to three technical replications of quantitative RT-PCR.

Sequencing and haplotype analysis

The full-length genomic DNA sequence of the candidate gene was determined by dividing it into several overlapping segments. Sequencing primers were designed according to the sequence of cv. Nipponbare in the target region. The resulting amplicons were separated through a 1.2 % agarose gel and recovered using a TIAN gel Midi Purification kit (TIANGEN). The recovered DNA was cloned into pGEM-T Easy Vector (Promega) and transformed into *E. coli* competent DH5α cells. The aligned sequences across the *ORF2* were imported into the TASSEL program to extract all polymorphisms for constructing gene haplotypes. The candidate gene region was scanned for its gene content using RiceGAAS (http://www.ricegaas.rgp.dna.affrc.go.jp).

Germplasm panel, GS3 and GS7 genotyping and microsatellite-based diversity analysis

A set of 52 rice cultivars (26 *indica* and 26 *japonica*) of diverse geographical origin was assembled to assess allelic variation within the *GS7* candidate gene and at *GS3*. Two markers were developed to assay variation at each of *GS7* and *GS3*. The former (FMGS7) targeted an indel lying in the upstream sequence. The latter (FMGS3) exploited the C/A polymorphism described by Fan et al. (2009) in the form of a *PstI*-based CAPS marker. A set of 24 microsatellite markers (http://www.gramene.org) was chosen to genotype the 52 rice entries. MEGA v4.1 software was combined with PowerMarkerVer 3.25 to generate an UPGMA-based phylogeny based on genetic distances (Liu and Muse 2004).

Statistical analysis of data

Mean phenotypic values were compared using the Student's *t* test. The correlation between genotypes and grain shape was carried out using a generalized linear model (GLM) implemented within the SAS statistical software package. A recurrent substitution mapping strategy as described by Paterson et al. (1990) was used for the fine mapping of *GS7*.

Results

Trait evaluation and the validation of GS7

Analysis of the cv. D50/cv. HB277 RIL population and its derived RHL population revealed that a grain shape quantitative trait locus (QTL) lies within a 278-kb segment of chromosome 7 flanked by Indel1 and RM21945, which harbors the grain length QTL *qGL7-2* (Shao et al. 2010). Since the QTL also affected grain width and the ratio of length to width (Fig. 2; Tables 1 and S1), it was named *GS7*. In addition, variations with respect to grain circumference, area, roundness and equivalent diameter were also associated with this QTL (Table 1). The robustness of *GS7* was tested both among RHL derivatives and by comparing the NIL pair. With respect to the latter, genotyping showed them to be >90 % identical to one another and that their 1,000-grain weight was indistinguishable. Similarly the RHL material did not vary with respect to 1,000-grain weight. Among the yield-related traits investigated, only PN was significantly correlated with *GS7* ($P < 0.05$), with the better performance associated with the presence of the cv. D50 allele.

Fine mapping of GS7

The initial round of fine mapping of *GS7* based on 1,000 progenies of RHL produced a set of 13 recombinants between Indel1 and RM21945 (Fig. 3b). Three additional informative markers (Indel2, RM21936 and RM21943) were developed to allow for a more precise genotypic description of these recombinants (Table S2; Fig. 3b). Four recombinants were identified in the key region between Indel2 and RM21936, and these were subsequently phenotyped using RHL-F_4 lines. The contrast between groups C1 and C2 allowed the size of the segment harboring *GS7* to be narrowed to a 93-kb region (Fig. 3b). The screening of 10,000 further RHL progenies (heterozygous for the Indel2–RM21936 segment) produced 26 further recombinants (Fig. 3c), which were genotyped with respect to nine additional indel markers (Table S2). The recombinants that

Table 1 Variation with respect to grain shape and other yield traits, as affected by the identity of the *GS7* allele present

Trait	RHL		NIL	
	RHL-D	RHL-H	NIL-D	NIL-H
GL	7.84 ± 0.05**	7.32 ± 0.07	7.53 ± 0.12**	6.78 ± 0.15
GW	2.21 ± 0.03**	2.34 ± 0.05	2.02 ± 0.01**	2.14 ± 0.03
L/W	3.57 ± 0.06**	3.13 ± 0.08	3.75 ± 0.07**	3.18 ± 0.06
Circumference	18.32 ± 0.13**	17.19 ± 0.19	17.65 ± 0.24**	16.05 ± 0.26
Area	14.02 ± 0.18*	13.51 ± 0.41	12.63 ± 0.15*	11.91 ± 0.33
Roundness	1.74 ± 0.03**	1.59 ± 0.03	1.81 ± 0.01**	1.57 ± 0.16
ED	4.22 ± 0.03*	4.14 ± 0.06	4.01 ± 0.01*	3.89 ± 0.05
TGW	29.6 ± 0.03	29.5 ± 0.24	24.4 ± 0.12	24.3 ± 0.08
PH	111 ± 3.41	114 ± 1.63	109 ± 3.58	106 ± 2.92
HD	106 ± 1.58	106 ± 0.98	103 ± 1.81	104 ± 2.45
PL	20.7 ± 1.2	20.8 ± 1.2	27.0 ± 1.4	27.1 ± 2.1
PN	9.3 ± 1.1*	8.4 ± 1.2	7.9 ± 1.2*	7.0 ± 1.0
NFGP	73 ± 15	72 ± 13	151 ± 28	137 ± 29
TNSP	101 ± 19	104 ± 19	184 ± 27	185 ± 28
SF	72.28 ± 5.96	69.23 ± 5.90	80.89 ± 4.80	72.97 ± 7.12

RHL-D and RHL-H: derivatives of RHL carrying, respectively, the cv. D50 and the cv. HB277 alleles. NIL-D and NIL-H: NILs carrying, respectively, the cv. D50 and the cv. HB277 alleles

GL grain length (mm), *GW* grain width (mm), *L/W* length to width ratio, *ED* equivalent diameter (mm), *TGW* thousand-grain weight (g), *PH* plant height (cm), *HD* heading date, *PL* panicle length (cm), *PN* panicle number per plant, *NFGP* number of filled grains per panicle, *TNSP* total number of spikelets per panicle, *SF* spikelet fertility (%)

Asterisks indicate significant differences between RHL-D and RHL-H, or between NIL-D and NIL-H, as determined by a Student's *t* test. *$P < 0.05$; **$P < 0.01$ ($n = 8$)

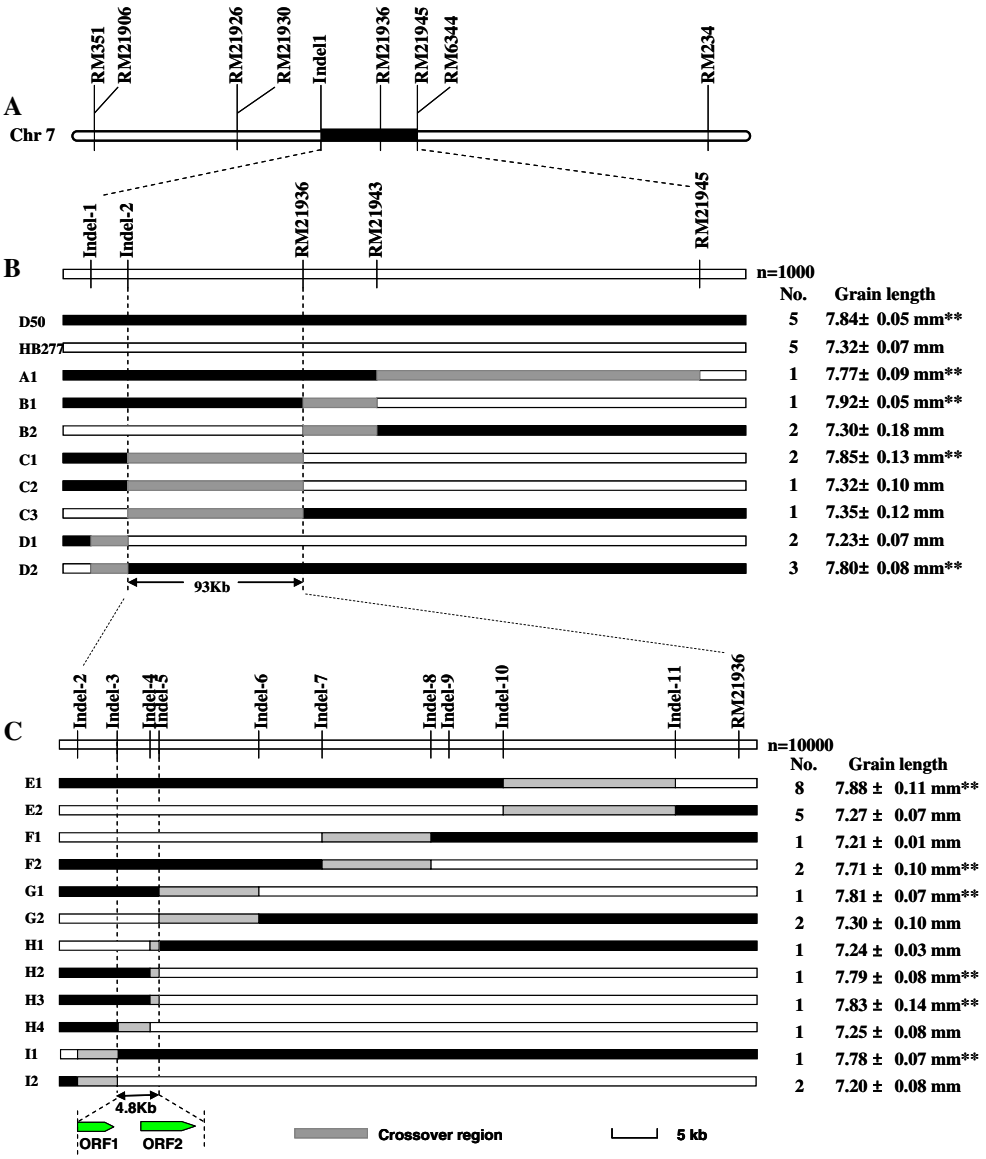

Fig. 3 Fine mapping of *GS7* and candidate gene analysis. **a** *GS7* lies between Indel1 and RM21945 on chromosome 7 (Shao et al. 2010). **b** Genotyping of RHL progeny placed *GS7* within a 93-kb region flanked by Indel2 and RM21936. The number of recombinants between adjacent markers indicated on the *right*. The 13 recombinants were arranged into eight genotypic groups, and the grain length associated with each is shown on the *right*. **c** Fine mapping based on 10,000 RHL progenies placed *GS7* within a 4.8-kb region flanked by Indel3 and Indel5. The 26 recombinants between Indel2 and RM21936 produced 12 genotypic groups. Grain lengths differing significantly ($P < 0.01$, $n = 8$) from that of RHL-H marked by *double asterisks*. RiceGAAS predicted the presence of *ORF1* and *ORF2* in the target region. *No.* number of recombinants

occurred in the same interval were sorted into the same group, thus this analysis produced 12 distinct genotypic groups (Fig. 3c). There are 13 recombinants between Indel10 and Indel11 (E1-2), 3 between Indel7 and Indel8 (F1-2), 3 between Indel5 and Indel6 (G1-2), 3 between Indel4 and Indel5 (H1-3), 1 between Indel3 and Indel4 (H4) and 3 between Indel2 and Indel3 (I1-2). The relevant grain shape phenotypes were obtained from derivatives in which the recombined segments had been fixed by selfing. As a result, an important recombinant (H4) was found and allowed the size of the segment harboring *GS7* to be narrowed to a 4.8-kb region flanked by Indel3 and Indel5 by a comparison with the other three recombinants, H1–3 (Fig. 3c), with the locus co-segregating with Indel4. This region is present on rice BAC clone OSJNBb0018H10.

A candidate gene for GS7

No annotated genes lay within the critical 4.8-kb genomic region of the cv. Nipponbare (*japonica*) genome (http://rice.plantbiology.msu.edu/). The target regions were therefore re-sequenced in cvs. D50 and HB277 and were shown by RiceGAAS analysis to contain two open reading frames named *ORF1* and *ORF2*. The predicted *ORF1* translation products of the two cultivars differed from one another by only one residue (data not shown), but the *ORF2* sequences were rather divergent (Fig. S1). As a result, the latter gene was considered to be the more likely candidate for *GS7*. Its sequence, however, gave no clue as to its function, since it shared no similarity to any currently annotated gene.

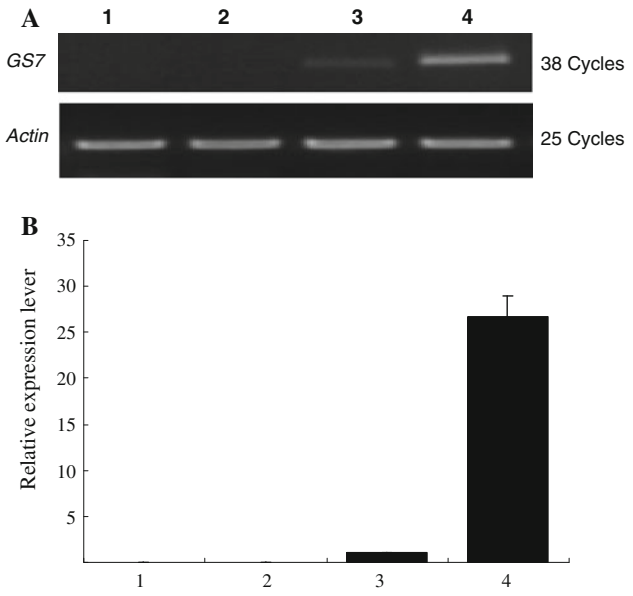

Fig. 4 Transcription of *ORF2* in RHL controls. **a** Semi-quantitative RT-PCR; **b** real-time quantitative RT-PCR. Values shown represent the mean ± SD ($n = 3$). *1* RHL-HB277 leaf; *2* RHL-HB277 panicle; *3* RHL-D50 leaf; *4* RHL-D50 panicle

Transcription analysis

Semi-quantitative RT-PCR was used to characterize the transcription of *ORF1* and *ORF2* in the leaf and panicle at the heading stages of RHL plants. *ORF1* was not transcribed in either organ (data not shown); however, with respect to *ORF2*, although no transcription was identified in the presence of the cv. HB277 allele, in the presence of the cv. D50 allele, the gene was clearly, if only rather weakly, transcribed (Fig. 4a; Fig. S1). When the *ORF2* sequence was subjected to real-time quantitative RT-PCR, a similar result was obtained (Fig. 4b).

Haplotype analysis and genetic variation for GS7

When the cvs. Nipponbare, 9311, D50 and HB277 sequences lying upstream of *ORF2* were aligned (Fig. S1), it was observed that compared to the cv. D50 sequence, those of cvs. 9311 and HB277 shared the same 142-bp insertion, while that of cv. Nipponbare had an 88-bp insertion. An analysis of a further 48 rice accessions identified six *ORF2* haplotypes, with a functional nucleotide polymorphism (FNP) present at position 672 (Fig. 5). Most of the type 1 (H1-2) cultivars were *indica* varieties, most of the type 2 (H3) and type 4 (H5-6) ones were *temperate japonica* varieties, and most of the type 3 (H4) ones were *tropical japonica* varieties (Figs. 6, 7; Table 2).

The FNP in *GS7* was exploited to develop the marker FMGS7 (Table S2; Fig. S1). Using this marker, the 52 entries could be organized into four types (Table 2; Figs. 6, S2); type 1 varieties (such as cvs. HB277 and 9311) produced intermediate length grains, type 2 (cv. Nipponbare) and type 4 (cv. Longjingchanglixiang) ones short grains, while type 3 (cv. D50) ones produced long grains. In addition, allelic variation at *GS3* suggested that a single nucleotide polymorphism can explain a major proportion of the phenotypic variance for grain length (Fan et al. 2009), so this was targeted by developing the CAPS marker FMGS3 (Table S2). Of the 52 entries analyzed using FMGS3, 23 carried the A allele and 29 the C allele. Combining the genotyping outcomes of FMGS7 and

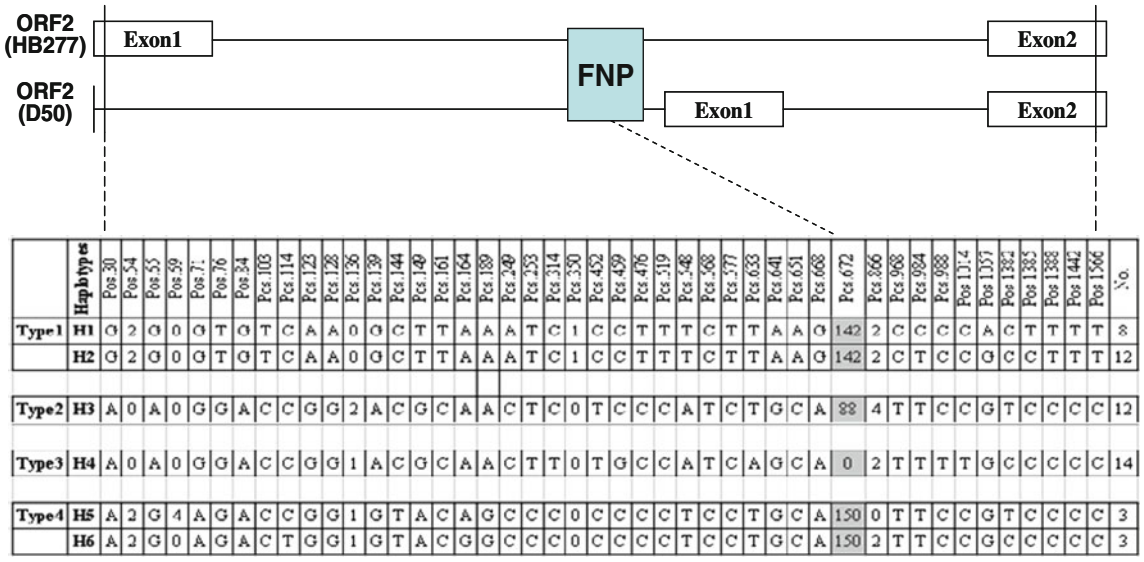

Fig. 5 Schematic position of nucleotide polymorphisms at *ORF2* and haplotypes across the *ORF2* in 52 rice accessions. The classification of haplotype based on the FNP in position 672

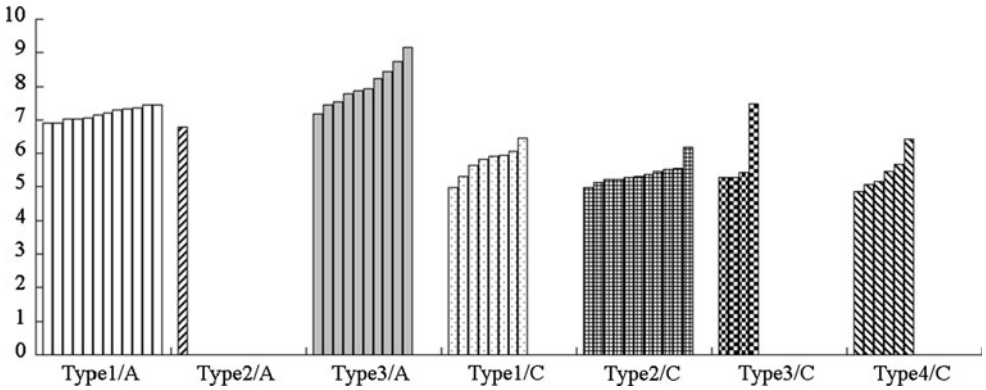

Fig. 6 Variation in grain length among 52 rice accessions related to genotype as revealed by markers FMGS7 and FMGS3

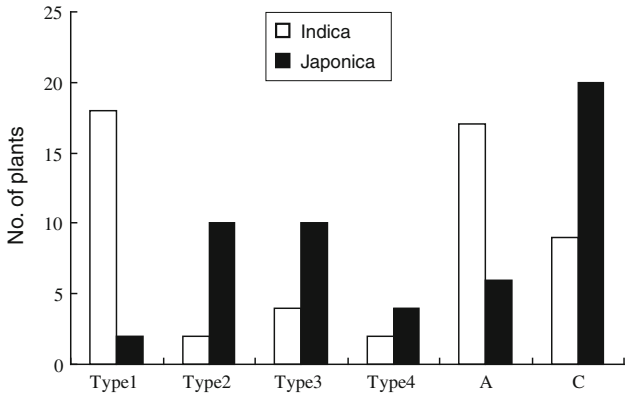

Fig. 7 The frequency of *indica* and *japonica* types related to genotype as revealed by markers FMGS7 and FMGS3

FMGS3 analysis revealed that the majority of the *indica* varieties were either type 1/A or type 1/C, while the *japonica* ones fell into three classes: 2/C, 3/C and 4/C. Type 3/A entries were *tropical japonica*, while type 2/A was only present in one of the 52 entries (Table 2; Fig. 6). Cultivars with the A allele tended to produce significantly longer grains than those produced by C allele carriers (Table 2; Fig. 6), and there was a significant difference with respect to grain length between type1/A and type3/A cultivars (Table 3).

Discussion

Increasing standards of living are gradually shifting rice breeders' priorities away from just grain yield toward the simultaneous improvement of end-use quality and yield. Grain shape (length, width and their ratio) are important quality criteria. Consumer preference, however, varies from region to region. In most of Asia and elsewhere, long, slender grains are preferred, but in Japan, South Korea and Sri Lanka, short bold ones predominate (Unnevehr et al. 1992; Juliano and Villareal 1993). A number of grain length QTLs have been described (Xu et al. 2002; Zheng et al. 2007; Shao et al. 2009). One of these, *qGL7-2*, was detected among RILs bred from the cross cv. D50 × cv. HB277 and was shown to lie within a segment of rice chromosome 7 (Shao et al. 2010). *GS7* is a robust QTL, as its effect was visible in comparisons involving either the RHL derivatives or the NIL pairs (Table 1). It is also associated with variation in grain width and the ratio of length to width (Tables 1, 4). Given that it affects not just grain length, but also grain circumference, area, roundness and equivalent diameter (Table 1), we have renamed this QTL *GS7*. *GS7* had no influence over grain weight (probably because long grains tend to be slender). Overall, the ratio between the grain's length and its width explained a greater proportion of the variance than either its length or width on their own (Tables 1, S1). The implication of this result is that *GS7* acts rather differently from any of the other related QTLs described to date (*GW2*, *GS3*, *qGW5* and *GS5*). Moreover, the alignment of the cv. D50 and cv. HB277 *ORF2* sequences suggests that its nucleotide variation is responsible for variation in grain shape (Fig. S1). The *ORF2* sequences present in cvs. 9311 and Nipponbare have been also screened. The cvs. 9311 and HB277 are both *indica* rice varieties and own the same genomic sequence in *ORF2*, so the gene presumably encodes an identical product (data not shown). However, although cvs. D50 and Nipponbare share the same start and end codes, they have different exon/intron composition (data not shown).

Grain shape has been heavily selected during the domestication of rice. At *GS3*, a single nucleotide polymorphism produces a clear difference in grain length (Fan et al. 2009). It is possible that a deletion in *GW5* was also subjected to positive selection during domestication, since genotypic analysis has revealed that its presence is strongly correlated with the wide grain type (Shomura et al. 2008). At *GS5*, a recently cloned QTL responsible for grain size, three variants in the promoter region have been correlated with grain size classes (Li et al. 2011). The present research has shown that while most type 1 cultivars were *indica* varieties, type 2 ones were dominated by *japonica*

Table 2 Genotype of 52 rice accessions as revealed by the FMGS7 and FMGS3 markers

Code	Alleles		Accession name	Grain length	*Indica/Japonica*	Origin	Sub-group
	FMGS7	FMGS3					
1	Type 2	C	Fuxiang1	5.00	*Japonica*	China	Group 2
2	Type 3	A	DXBC	7.17	*Japonica*	America	Group 3
3	Type 3	A	Kaybonnet	7.44	*Japonica*	America	Group 3
4	Type 2	C	Chujiangxiang2	5.15	*Japonica*	China	Group 2
5	Type 1	A	Minghui86	6.91	*Indica*	China	Group 1
6	Type 1	A	Xinxiang1	6.92	*Indica*	China	Group 1
7	Type 1	A	Xiangxiang2	7.02	*Indica*	China	Group 1
8	Type 1	A	Ganhui319	7.03	*Indica*	China	Group 1
9	Type 2	A	Duoxi1	6.80	*Indica*	China	Group 1
10	Type 1	A	HB277	7.05	*Indica*	Chian	Group 1
11	Type 1	A	Huahangxinzhan	7.16	*Indica*	China	Group 1
12	Type 2	C	Xiangjing111	5.22	*Japonica*	China	Group 2
13	Type 1	A	Peiai64	7.21	*Indica*	China	Group 1
14	Type 1	A	9311	7.30	*Indica*	China	Group 1
15	Type 2	C	Yueguang	5.24	*Japonica*	Japan	Group 2
16	Type 1	A	Jasmine85	7.33	*Indica*	Philippines	Group 1
17	Type 1	A	Baxiang308	7.35	*Indica*	China	Group 1
18	Type 1	A	Basmati370	7.44	*Indica*	India	Group 1
19	Type 1	A	IR74053	7.46	*Indica*	Philippines	Group 1
20	Type 2	C	Xiangbao1	5.29	*Japonica*	China	Group 2
21	Type 1	C	Xiangjingdao	4.97	*Japonica*	China	Group 2
22	Type 4	C	Baimaoxiangnuo	5.47	*Indica*	China	Group 2
23	Type 2	C	Wagwag	5.31	*Japonica*	Philippines	Group 2
24	Type 2	C	Wuqinxiangsidao	5.38	*Japonica*	China	Group 2
25	Type 4	C	Longjingchanglixiang	5.68	*Japonica*	China	Group 2
26	Type 1	C	Teqing	5.84	*Indica*	China	Group 1
27	Type 2	C	Huadixiangdao	5.46	*Japonica*	China	Group 2
28	Type 2	C	Gongxiang	5.53	*Japonica*	Japan	Group 2
29	Type 1	C	Nanjing11	5.96	*Indica*	China	Group 1
30	Type 1	C	Mianhui501	6.47	*Indica*	China	Group 1
31	Type 3	A	Changlipinzhong	7.54	*Indica*	America	Group 3
32	Type 3	A	WAB56-104	7.79	*Japonica*	Philippines	Group 3
33	Type 3	A	Naire1	7.86	*Japonica*	Russia	Group 3
34	Type 3	A	D50	7.93	*Japonica*	America	Group 3
35	Type 3	A	HEP-77	8.23	*Japonica*	America	Group 3
36	Type 2	C	Nipponbare	5.57	*Japonica*	Japan	Group 2
37	Type 3	A	AD95035	8.44	*Indica*	Philippines	Group 1
38	Type 2	C	Miyang46	6.19	*Indica*	Korea	Group 1
39	Type 3	A	Luxiang90	8.73	*Indica*	China	Group 3
40	Type 3	A	IR73002-146-2-3-3	9.17	*Indica*	Philippines	Group 1
41	Type 3	C	CPSLO17	7.49	*Japonica*	America	Group 3
42	Type 3	C	Baikexiangdao	5.29	*Japonica*	China	Group 2
43	Type 1	C	Xiangguiyouzhan	5.31	*Indica*	China	Group 1
44	Type 1	C	Wuxiangjing9	6.08	*Japonica*	China	Group 1
45	Type 1	C	Xianghenuo	5.66	*Indica*	China	Group 1
46	Type 4	C	Daxiangnuogu	6.43	*Indica*	China	Group 1
47	Type 4	C	Xiangdanuo	4.86	*Japonica*	China	Group 2

Table 2 continued

Code	Alleles		Accession name	Grain length	*Indica/Japonica*	Origin	Sub-group
	FMGS7	FMGS3					
48	Type 4	C	Xianghongnuo	5.07	*Japonica*	China	Group 2
49	Type 1	C	Yizhixiang	5.92	*Indica*	China	Group 1
50	Type 3	C	Suxiangjing1	5.44	*Japonica*	China	Group 2
51	Type 4	C	Wuxiangjing14	5.18	*Japonica*	China	Group 2
52	Type 3	C	Mixiangnuo	5.29	*Japonica*	China	Group 2

[a] Groups 1, 2 and 3 classifications derived from the data shown in Table 2 and Fig. S3

Table 3 Analysis of variance for grain length among the seven haplotypes revealed by the markers FMGS7 and FMGS3

Haplotype	Type 1/A	Type 2/A	Type 3/A	Type 1/C	Type 2/C	Type 3/C	Type 4/C
Type 1/A	–	–	<0.001***	<0.001***	<0.001***	<0.05*	<0.001***
Type 2/A		–	–	–	–	–	–
Type 3/A			–	<0.001***	<0.001***	<0.01**	<0.001***
Type 1/C				–	<0.05*	0.434 n.s.	0.137 n.s.
Type 2/C					–	0.219 n.s.	0.417 n.s.
Type 3/C						–	0.252 n.s.
Type 4/C							–

Haplotypes as illustrated in Table 2

n.s. not significant

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$

Table 4 The identity of the *GS7* allele present affects both grain width (GW) and the ratio between grain length and grain width (L/W)

Groups	No.	GW (mm)	RL/W
RHL-D	5	2.21 ± 0.03**[a]	3.57 ± 0.06**
RHL-H	5	2.34 ± 0.05	3.13 ± 0.08
E1	8	2.22 ± 0.03**	3.55 ± 0.07**
E2	5	2.37 ± 0.03	3.08 ± 0.05
F1	1	2.35 ± 0.04	3.07 ± 0.05
F2	2	2.16 ± 0.03**	3.58 ± 0.05**
G1	1	2.20 ± 0.04**	3.56 ± 0.08**
G2	2	2.39 ± 0.02	3.06 ± 0.06
H1	1	2.34 ± 0.03	3.11 ± 0.05
H2	1	2.20 ± 0.03**	3.56 ± 0.07**
H3	1	2.22 ± 0.04**	3.53 ± 0.07**
H4	1	2.37 ± 0.03	3.07 ± 0.06
I1	1	2.17 ± 0.03**	3.59 ± 0.06**
I2	2	2.36 ± 0.04	3.06 ± 0.05

The phenotypic value of each of the 12 recombinants derived from 10,000 progenies of RHL is shown. GW and L/W differ significantly (**·[a]$P < 0.01$; $n = 8$) from that of RHL-H

varieties, and type 3 entries were all *tropical japonica* types (Tables 2, S3; Fig. S3). Four *ORF2* haplotypes emerged from a re-sequencing exercise of 52 entries, suggesting that the *GS7* allele present in cv. D50 might have originated from the *tropical japonica* genepool.

Phenotype/genotype associations between the allelic state at FMGS3 and grain length have suggested that carriers of the A allele produce significantly longer grains than carriers of the C allele (Fan et al. 2009; Takano-Kai et al. 2009), as confirmed in the present study (Table 3). Takano-Kai et al. (2009) further proposed that *GS3* had evolutionary importance over grain size, and that there were no significant differences among *indica*, *japonica* and *tropical japonica* types which carried the A allele, a result which was not confirmed by the analysis of the present materials. The discrepancy may reflect differences in the size of the populations analyzed and/or the ignoring of the effects of allelic variation at other genes underlying grain shape, in particular, *GS7*. Although the effect of *GS7* was expressed in the presence of the *GS3* A allele, it was ineffective when combined with the *GS3* C allele (Fig. 6; Table 3). This phenomenon may be introduced by the negative regulation of C allele in *GS3*. In addition, *tropical japonica* types have been widely used in various Chinese rice improvement programs, since they contribute stronger stems, longer grains, less grain chalkiness, better end-use quality and are photosynthetically more efficient than *indica* types (Hu et al. 1999, 2000). A significant level of heterosis is also expressed when members of these two genepools are

inter-crossed (Liu et al. 2006; Zhang et al. 2006). With the development of the functional FMGS7 and FMGS3 markers, cv. D50, along with other *tropical japonica* rice varieties, could be readily used as a source of positive alleles for improving not just grain shape, but also other important agricultural traits.

Acknowledgments This research was financially supported by grants from the Chinese '863' program (Grant No. 2009AA101101), the Natural Science Foundation (Grant No. 31161140348), the Department of Agriculture (Grant Nos. 2011ZX08001-001, 2011ZX08001-002 and 2011ZX08001-006), and the Zhejiang Province Science and Technology Project (Grant No. 2010C32061).

References

Bai XF, Luo LJ, Yan WH, Rao KM, 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 Genetic 11:16

Fan CC, Xing YZ, Mao HL, Lu TT, Han B, Xu CG, Li XH, Zhang QF (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

Fan CC, Yu SB, Wang CR, Xing YZ (2009) A causal C-A mutation in the second exon of GS3 highly associated with rice grain length and validated as a functional marker. Theor Appl Genet 118:465–472

Frary A, Nesbitt TC, Frary A, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–88

Hu PS, Tang SQ, Luo J, Huang FS (1999) Utilization of American glabrous rice and breeding of super high yielding varieties. Acta Agron Sin 1:32–38 (in Chinese with English abstract)

Hu PS, Luo J, Tang SQ (2000) Utilization of American glabrous rice and breeding of good-quality varieties. CNRR 8:13–15

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

Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A, Lundqvist U, Fujimura T, Matsuoka M, Matsumoto T, Yano M (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc Natl Acad Sci 104:1424–1429

Konishi S, Ebana K, Izawa T (2008) Inference of the *japonica* rice domestication process from the distribution of six functional nucleotide polymorphisms of domestication-related genes in various landraces and modern cultivars. Plant Cell Physiol 49(9):1283–1293

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

Liu K, Muse S (2004) PowerMarker: new genetic data analysis software, Version 2.7. http://www.powermarker.net

Liu WB, Zhang JX, Luo WL, Lin Q, Chi XW, Cai WM (2006) Progress on glabrous hybrid rice breeding on hybrid rice. Hybrid Rice 21(6):11–13 (in Chinese with English abstract)

Lu YJ, Zheng KL (1992) A simple method for isolation of rice DNA. Chin Rice Sci 6:47–48 (in Chinese with English abstract)

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(45):19579–19584

Paterson AH, Deverna J, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, from an interspecies cross of tomato. Genetics 124:735–742

Shao GN, Tang SQ, Luo J, Jiao GA, Tang A, Hu PS (2009) QTL analysis for flag leaf and grain shape and populations construction derived from related residual heterozygous lines in rice. Plant Mol Breed 7:16–22 (in Chinese with English abstract)

Shao GN, Tang SQ, Luo J, Jiao GA, 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

Shi YF, Ying JZ, Wang L, Zhu ZW, Zhuang JY (2005) Screening SSR markers for rice variety identification. Chin Rice Sci 19:195–201 (in Chinese with English abstract)

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

Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai YS, Gill BS, Faris JD (2006) Molecular characterization of the major wheat domestication gene Q. Genetics 172:547–555

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

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(4):1323–1334

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

Unnevehr LJ, Duff B, Juliano BO (1992) Consumer demand for rice grain quality. International Rice Research Institute, Manila, and International Development Research Center, Ottawa

Wang H, Wagler TN, Li BL, Zhao Q, Vigouroux Y, Faller M, Bomblies K, Lukens L, Doebley JF (2005) The origin of the naked grains of maize. Nature 436:714–719

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

Xu JL, Xue QZ, Luo LJ, Li ZK (2002) Genetic dissection of grain weight and its related traits in rice (*Oryza sativa* L.). Chin Rice Sci 16:6–10

Zhang JX, Huang JH, Rao MD, Tang XD, Cheng TJ, Hong W (2006) A preliminary study on heterosis of the hybrid rice of indica/nuda and indica/indica I. The performance of agronomic traits. Subtrop Agric Res 2(3):161–164 (in Chinese with English abstract)

Zheng TQ, Xu JL, Li ZK, Zhai HQ, Wan JM (2007) Genomic regions associated with milling quality and grain shape identified in a set of random introgression lines of rice (*Oryza sativa* L.). Plant Breed 126:158–163