Theor Appl Genet (2006) 113:619–629
DOI 10.1007/s00122-006-0326-y

ORIGINAL PAPER

Fine mapping of a quantitative trait locus for grain number per panicle from wild rice (*Oryza rufipogon* Griff.)

Feng Tian · Zuofeng Zhu · Boshen Zhang · Lubin Tan · Yongcai Fu · Xiangkun Wang · Chuan Qing Sun

Received: 19 February 2006 / Accepted: 9 May 2006 / Published online: 13 June 2006

Abstract SIL040, an introgression line (IL) developed by introgressing chromosomal segments from an accession of *Oryza rufipogon* into an *indica* cultivar Guichao 2, showed significantly less grains per panicle than the recurrent parent Guichao 2. Quantitative trait locus (QTL) analysis in F_2 and F_3 generations derived from the cross between SIL040 and Guichao 2 revealed that *gpa7*, a QTL located on the short arm of chromosome 7, was responsible of this variation. Alleles from *O. rufipogon* decreased grains per panicle. To fine mapping of *gpa7*, a high-resolution map with 1,966 F_2 plants derived from the cross between SIL040 and Guichao 2 using markers flanking *gpa7* was constructed, and detailed quantitative evaluation of the structure of main panicle of each of F_3 families derived from recombinants screened was performed. By two-step substitution mapping, *gpa7* was finally narrowed down to a 35-kb region that contains five predicted genes in cultivated rice. The fact that QTLs for five panicle traits (length of panicle, primary branches per panicle, secondary branches per panicle, grains on primary branches and grains on secondary branches) were all mapped in the same interval as that for *gpa7* suggested that this locus was associated with panicle structure, showing pleiotropic effects. The characterizing of panicle structure of IL SIL040 further revealed that, during the domestication from common wild allele to cultivated rice one at *gpa7*, not only the number of branches and grains per panicle increased significantly, more importantly, but also the ratio of secondary branches per panicle to total branches per panicle and the ratio of grains on secondary branches per panicle to total grains per panicle increased significantly. All these results reinforced the idea that *gpa7* might play an important role in the regulation of grain number per panicle and the ratio of secondary branches per panicle during the domestication of rice panicle.

Communicated by T. Sasaki

Feng Tian and Zuo Feng Zhu contributed equally to this work.

F. Tian · Z. Zhu · B. Zhang · L. Tan · Y. Fu · X. Wang · C. Q. Sun (✉)
Department of Plant Genetic and Breeding and State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100094, People's Republic of China
e-mail: suncq@cau.edu.cn

F. Tian · Z. Zhu · B. Zhang · L. Tan · Y. Fu · X. Wang · C. Q. Sun
Key Laboratory of Crop Heterosis and Utilization of Ministry of Education, Beijing 100094, People's Republic of China

F. Tian · Z. Zhu · B. Zhang · L. Tan · Y. Fu · X. Wang · C. Q. Sun
Beijing Key Laboratory of Crop Genetic Improvement, Beijing 100094, People's Republic of China

F. Tian · Z. Zhu · B. Zhang · L. Tan · Y. Fu · X. Wang · C. Q. Sun
Key Laboratory of Crop Genetic Improvement and Genome of Ministry of Agriculture, Beijing 100094, People's Republic of China

Introduction

Rice (*Oryza sativa* L.) grain yield was determined by three yield components, panicles per plant, grain weight and grain number. Of rice grain yield components, grain numbers showed the largest range of variation and was the major objective of improvement in rice

high yield breeding (Li et al. 1998; Yamagishi et al. 2002). *O. rufipogon* Griff. is the wild ancestor of cultivated rice (Second 1982; Oka 1988; Wang et al. 1992). During the course of domestication from wild rice to cultivated rice, profound changes of agronomic traits and genetic diversity occurred (Sun et al. 2001). One of the most important hallmarks of rice domestication is dramatic increases in grain number, as evidenced by the fact that most cultivated rice showed more grain number than wild rice. Typical strains of *O. rufipogon* usually showed less branch number and grain number (H.W. Cai and C.Q. Sun, unpublished data). Domestication of wild rice led to the production of panicle with more branches and grain number. To understand the molecular mechanism of changes occurred in panicle structure during domestication, we need to characterize these changes and identify the genes underling them. Thus, the molecular basis of grain number might not only set theory basis for the improvement of grain number, but also offer general insights into the rice domestication.

Rice grain number is quantitatively inherited and a great deal of quantitative trait locus (QTL) mapping for grain number have been conducted using various mapping populations derived from inter-specific crosses (Xiao et al. 1998; Xiong et al. 1999; Moncada et al. 2001; Thomson et al. 2003; Li et al. 2006), *indica–japonica* inter-subspecific crosses (Lu et al. 1996; Xiao et al. 1996; Yu et al. 1997; Redona and Mackill 1998; Sasahara et al. 1999; Yagi et al. 2001; Xing et al. 2002; Mei et al. 2003, 2005), *indica–indica* crosses (Lin et al. 1996; Zhuang et al. 1997) and *japonica–janonica* cross (Yamagishi et al. 2002). These QTLs detected were distributed throughout all rice chromosomes and created a firm basis to investigate the genetic control of grain number. However, QTL cloning projects for grain yield components were rarely reported. Li et al. (2004) fine mapped a grain-weight QTL, *gw3.1*, in the pericentromeric region of rice chromosome 3 and narrowed down the location of the gene underlying that QTL to a 93.8-kb region. More recently, Ashikari et al. (2005) cloned *Gn1a*, a QTL located on chromosome 1 that increases grain number in rice, and elucidated the molecular mechanism of this gene.

An introgression line (IL), SIL040, a member of 159 introgression lines derived from the BC_4F_4 population of *O. rufipogon* and cultivated rice (Tian et al. 2006), showed significant less grains per panicle than the recurrent parent Guichao 2 based on phenotypic evaluations at two sites over 2 years. In this study, a QTL mapped to the short arm of chromosome 7 was found responsible for this variation. This QTL was referred to as *gpa7* (abbreviation for grains per panicle). Then we fine mapped the *gpa7* and revealed that it resides within a 35-kb target region. We also further characterized the phenotypic effects of *gpa7* and the potential importance of *gpa7* for rice panicle domestication was also discussed.

Materials and methods

Plant materials

SIL040 is an introgression line derived from a set of 159 introgression lines developed in a previous study by introgressing chromosomal segments from an accession of Chinese common wild rice (*O. rufipogon* Griff.), collected from Dongxiang county, Jiangxi Province, China, into an *indica* cultivar (*O. sativa* L.), Guichao 2 background based on four generations of backcrossing and four generations of selfing (Tian et al. 2006), and showed significantly less grains per panicle than the recurrent parent Guichao 2 based on phenotypic evaluations at two sites over 2 years. To uncover the genetic basis of this variation, an F_2 population consisting of 400 plants was constructed by selfing the F_1 plant of SIL040 as female and Guichao 2 as male parent and grown in Experiment Station of China Agricultural University, Beijing (39°N, 116°E), in the summer of 2004. Then F_3 family derived from each 400 F_2 plant and a larger F_2 population containing 1,566 plants derived from the cross between SIL040 and Guichao 2 were planted at Sanya (18°N, 109°E), Hainan Province, China, in the winter of 2004. The 400 F_3 families were laid out in a single plot without replication, four rows per plot, 12 plants per row, 13.3 cm between plants within each row and 26.4 cm between rows. F_3 families derived from all recombinants screened from 1,966 plants along with parental controls were grown at the same site in Beijing in the summer of 2005. The field planting for progeny testing of recombinants and parental controls followed a complete randomized block design with two replications. The planting manner of each plot was the same as that of the 400 F_3 families described above. The field management was similar to that under normal rice production conditions. In addition, another F_2 population containing 100 plants derived from the cross between SIL040 and *indica* cultivar 93-11 was constructed and grown at the same site in Beijing in the summer of 2005.

Phenotypic evaluation

The 400 F_2 plants were individually evaluated for panicles per plant and grains per plant. Grains per panicle were obtained by dividing grains per plant by panicles

per plant. Ten plants were harvested from each of 400 F_3 families. The phenotypic evaluations of 4,000 F_3 plants were the same as those for F_2 plants described above. The phenotypic evaluations for 100 F_2 individuals derived from the cross between SIL040 and 93-11 were also the same as those described above. Detailed quantitative analysis of panicle structure for each of recombinant derived F_3 families and the parents SIL040 and Guichao 2 and their F_1 hybrid were performed using 40 main panicles each. The traits measured for panicle structure included length of panicle, primary branches per panicle, secondary branches per panicle, grains on primary branches per panicle, grains on secondary branches per panicle and total grains per panicle.

DNA extraction and molecular marker analysis

DNA was extracted from fresh leaves according to the CTAB method (Murray and Thompson 1980) with minor modifications. The molecular markers analyzed in the target region were published SSRs (McCouch et al. 2002), to construct high density map of the target region, new SSR and Indel markers were developed in this study. Four new SSR primers (63130, 3617, 3683 and 5452) were designed according to available public rice genome sequence (http://www.rgp.dna.affrc.go.jp/). In addition, two Indel markers (ID52 and ID68.3) were developed based on partial sequencing of target region for the parent types. PCR fragments were amplified with primers designed to match rice genomic sequence and directly sequenced in both directions. The PCR reaction mixture and PCR conditions were the same as those described by Tian et al. (2006). Primer information of six newly developed markers used in this study was listed in Table 1. In each of recombinant families, the genotypes for all markers in the target region were determined and homozygous recombinants were identified.

Data analysis

Linkage map construction was performed using Mapmaker/Exp 3.0 (Lander et al. 1987). QTL analysis was carried out by interval analysis with Map Manager QTXb17 (Manly et al. 2001).

In each recombinant family, the mean phenotypic value of each of the six traits of main panicle, including length of panicle, primary branches per panicle, secondary branches per panicle, grains on primary branches per panicle, grains on secondary branches per panicle and total grains per panicle, for the homozygous recombinants was compared with that of the controls SIL040 and Guichao 2 using the SAS statistical software package (SAS 1990, SAS Institute, Cary, NC, USA) at significant level $P < 0.001$. Recombinant families were grouped based on the genotypes of homozygous recombinants they contained and the mean phenotypic value of each of above six traits for each of the groups was calculated. In the fine mapping of the position of *gpa7*, the substitution mapping strategy described by Paterson et al. (1990) was used.

Results

Characterizing of IL SIL040

Phenotypic evaluations of grains per panicle for the introgression line SIL040 and the recurrent parent Guichao 2 were conducted at two sites, Beijing (39°N, 116°E) and Sanya (18°N, 109°E), over 2 years (2003 and 2004) (Fig. 1). The results demonstrated that SIL040 showed consistent phenotype in all trials. The mean grains per panicle at two sites over 2 years for SIL040 and Guichao 2 were 44 and 114, respectively. SIL040 showed nearly threefold less than Guichao 2 for mean grains per panicle. Table 2 showed comparison of some important agronomic traits between SIL040 and Guichao 2 obtained in the trial conducted at Sanya in 2004. The results indicated that there were significant differences ($P < 0.01$) in plant height, panicles per plant, grains per plant, 1,000-grain weight and grain yield per plant between them, while there were no significant difference in seed set rate and days to heading between them.

Table 1 New molecular markers developed in this study to fine map *gpa7*

Marker	Marker type	Product size in Nipponbare (bp)	Forward primer (5′–3′)	Reverse primer (5′–3′)
63130	SSR	159	GGATCTAGCTAGGGTTCGGG	TATCCGCGTCCCTAGCTTAG
3683	SSR	229	GGTAGGGATCAAACTATTGC	GCTGAGTCTCATCTACATAG
5452	SSR	186	TAGTGACCGAGGGTGGGAATT	GCCGCGTCAAAGATAGTCGTT
3617	SSR	119	GGATGGATTTGAAGGATTTG	AACCACTTCATTCACCACCC
ID52	Indel	140	GTTTGGTGGTGTTCATGGTCT	GATCAGCTCTCACCAATCCAG
ID68.3	Indel	158	CAGGTTGAGATTAAGGGAGGA	AGACGAATGGTCAAACAGTGC

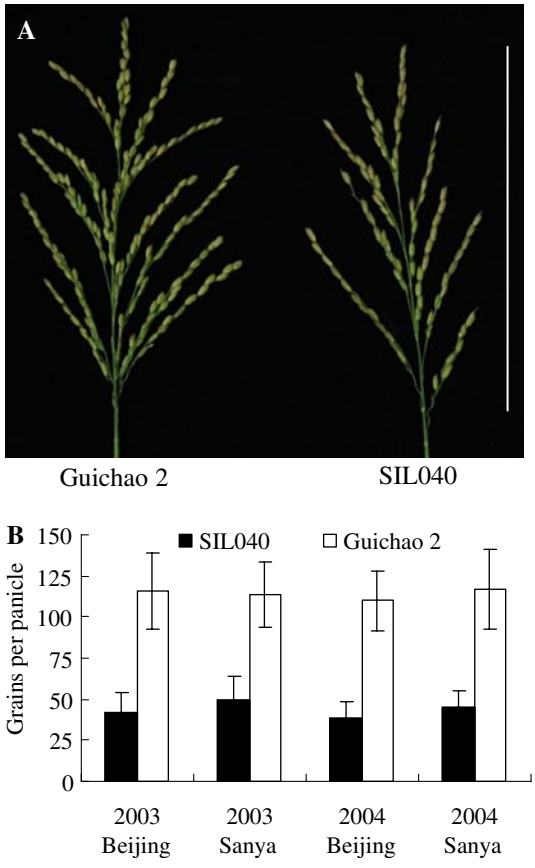


Fig. 1 Comparison of grain number per panicle of introgression line SIL040 and Guichao 2. **a** Main panicle structure of Guichao 2 and SIL040. *Scale bar* 20 cm. **b** Mean grain number per panicle of introgression line SIL040 and Guichao 2 at two sites, Beijing (39°N, 116°E) and Sanya (18°N, 109°E), over 2 years (2003 and 2004). In each trial, ten plants for SIL040 and Guichao 2 were individually harvested and evaluated for panicles per plant and total grains per plant, and grains per panicle were obtained by dividing total grains per plant by panicles per plant. Data presented are means with SD ($n = 10$ plants)

Detailed quantitative observations of the structure of main panicle for the parents SIL040 and Guichao 2 and their F_1 hybrid were conducted (Table 3), the results revealed that SIL040 showed significant difference ($P < 0.001$) from Guichao 2 and the F_1 for all panicle traits measured. The F_1 also showed significant difference ($P < 0.001$) from Guichao 2 for all panicle traits measured. In SIL040, the mean grains on secondary branches per panicle were significantly ($P < 0.001$) less than the mean grains on primary branches per panicle. However, the case in Guichao 2 was completely converse, which was that the mean grains on secondary branches per panicle were significantly ($P < 0.001$) more than the mean grains on primary branches per panicle. In the F_1, there was no significant difference between the mean grains on primary branches and the mean grains on secondary branches.

Figure 2 showed the ratio of primary branches per panicle and secondary branches per panicle to total branches per panicle and the ratio of grains on primary branches per panicle and grains on secondary branches per panicle to total grains per panicle in SIL040, F_1 and Guichao 2. In analysis of branches per panicle, a common tendency was obvious for SIL040, F_1 and Guichao 2, which was that the secondary branches per panicle contributed most to the total branches per panicle. In SIL040, the primary and secondary branches per panicle contributed 38.4 and 61.6%, respectively, to the total branches per panicle, while in F_1 and Guichao 2, the proportions were 33.3 and 66.7%, 28.4 and 71.6%, respectively. The ratio of secondary branches per

Table 2 Comparison of some important agronomic traits between SIL040 and Guichao 2

Trait	SIL040	Guichao 2	Significance
Plant height (cm)	61.1	75.1	*
Panicles per plant	13.8	9.1	*
Grains per plant	619.2	1060.6	*
Seed set rate	88%	86%	NS
1,000-grain weight (g)	25.3	26.5	*
Grain yield per plant	13.7	24.2	*
Days to heading	106.1	104.2	NS

The above data was from the trial conducted at Sanya in 2004. Ten plants for SIL040 and Guichao 2 were individually harvested and evaluated for above seven traits. The plant height was scored as the length of the tallest tiller from the ground to the tip of the panicle. Days to heading was recorded as the number of days from sowing to the appearance of the first panicle. For each trait, the mean phenotypic values of ten plants were compared between SIL040 and Guichao 2

NS Not significant

*Significant at $P < 0.01$

Table 3 Quantitative observations of panicle structure of SIL040, Guichao 2 and their F_1 hybrid

Line	Length of panicle	No. of primary branches per panicle	No. of secondary branches per panicle	No. of grains on primary branches	No. of grains on secondary branches	Total grains per panicle
SIL040	20.0 ± 1.5	8.7 ± 1.0	14.8 ± 4.9	51.1 ± 6.9	42.9 ± 16.2	94.0 ± 22.0
F_1	21.2 ± 1.2	10.0 ± 1.1	20.7 ± 5.0	60.8 ± 7.8	62.4 ± 17.05	123.2 ± 22.9
Guichao 2	22.3 ± 1.0	11.4 ± 0.9	29.5 ± 5.9	71.3 ± 7.1	93.5 ± 21.9	164.8 ± 26.6

The main panicle of each of 40 plants from the parents SIL040 and Guichao 2 and their F_1 hybrid was individually harvested and measured for above panicle traits. Data presented are the means with SD ($n = 40$ plants)

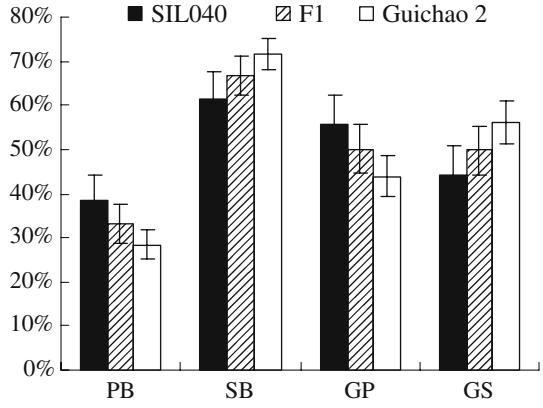


Fig. 2 The ratio of primary branches per panicle and secondary branches per panicle to total branches per panicle and the ratio of grains on primary branches and grains on secondary branches to total branches per panicle in SIL040, Guichao 2 and their F_1 hybrid. The main panicle of each of 40 plants from the parents SIL040 and Guichao 2 and their F_1 hybrid was individually harvested and calculated for the above ratioes. Data presented are the means with SD ($n = 40$ plants). *PB* primary branches per panicle, *SB* secondary branches per panicle, *GP* grains on primary branches per panicle, *GS* grains on secondary branches per panicle

panicle to total branches per panicle significantly ($P < 0.001$) increased. The analysis of grains per panicle revealed that, in SIL040, the grains on primary branches per panicle and the grains on secondary branches per panicle contributed 55.7 and 44.3%, respectively, to total grains per panicle, while in F_1 and Guichao 2, the proportions were 50.1 and 49.9%, 43.9 and 56.1%, respectively. The ratio of grains on secondary branches to total grains per panicle significantly ($P < 0.001$) increased.

Detecting of *gpa7*

Using 208 SSR markers showing polymorphism between *O. rufipogon* and Guichao 2, distributed evenly throughout 12 chromosomes, SIL040 was scanned and three *O. rufipogon* introgression segments were detected, located on chromosome 1, 7 and 8, respectively. To dissect the genetic factors underlying the less grain number per panicle in SIL040, an F_2 population containing 400 plants was constructed by using SIL040 as female and Guichao 2 as male parent. The 400 F_2 individuals were genotyped using 24 SSR markers distributed on three *O. rufipogon* introgression regions, and evaluated for grains per panicle. QTL analysis for grains per panicle revealed that there was a significant peak between markers RM3325 and 3683 on the short arm of chromosome 7 with a LRS score of 130.0 (LOD = 28.3) and an % (phenotypic variance explained by the QTL) of 37%. The *O. rufipogon* derived allele contributed a decreasing effect on grains per panicle. We referred to this locus as *gpa7*. No QTL was detected in the other two introgression regions. QTL analysis in F_3 families derived from selfing each F_2 plant further confirmed the unique QTL peak between markers RM3325 and 3683 on the short arm of chromosome 7 with a LRS score of 300.4 (LOD = 65.3) and an % of 54% (Table 4), and also no QTL was detected in the other two introgressions, consistent with the F_2 mapping results. All above results demonstrated that *gpa7* was responsible for the less grains per panicle in SIL040, and thus SIL040 was a nearly isogenic line to Guichao 2, suggesting that using the F_2 population derived from the cross between SIL040 and Guichao 2 to fine map *gpa7* was feasible. In addition, to examine the expression of *gpa7* in different genetic background, another F_2 population containing 100 plants derived from the cross between SIL040 and *indica* cultivar 93-11 was constructed and were evaluated for grains per panicle. QTL analysis in this population revealed that there was a significant peak between markers RM3325 and 3683 on the short arm of chromosome 7 with a LRS score of 28.0 (LOD = 6.1) and an % of 28% (Table 4).

Fine mapping of *gpa7*

To further refine the position of *gpa7*, a larger F_2 population containing 1,566 plants derived from the cross

Table 4 QTL analysis for grain number per panicle in the F_2 and F_3 generations derived from the cross between SIL040 and Guichao 2 and in the F_2 population derived from the cross between SIL040 and 93-11

Population	Interval	LRS	LOD[a]	%[b]	Add[c]	Dom[d]
SIL040/Guichao2 F_2	RM3325–3683	130.0	28.3	37	−19.3	3.0
SIL040/Guichao2 F_3	RM3325–3683	300.4	65.3	54	−25.8	8.2
SIL040/93–11 F_2	RM3325–3683	28.0	6.1	28	−31.2	15.2

[a] Likelihood ratio statistic (LRS) value was divided by 4.6 to obtain the equivalent logarithm of the odds (LOD) score (Manly et al. 2001)

[b] Phenotypic variance explained by the QTL

[c] Additive effect associated with *O. rufipogon*

[d] Dominance effect associated with *O. rufipogon*

between SIL040 and Guichao 2 was constructed. From 1,966 F_2 plants (plus the previous 400 plants), a total of 62 recombinants were detected between markers RM3325 and 3683. All recombinants screened were subjected to genotyping with four markers between the interval RM3325 and 3683. A high-resolution map with 1,966 plants using published and newly developed markers flanking *gpa7* was constructed (Fig 3b). In each of the 62 recombinants families, homozygous recombinants were identified with the appropriate markers and evaluated for six panicle traits described above. Using homozygous recombinants, the mean phenotypic value of each trait for each recombinant F_3 family was compared to that of the controls SIL040 and Guichao 2 at $P < 0.001$ level. According to their marker genotypes, the 62 recombinant families were grouped into 12 groups (Fig. 3c). Group A1 contained seven recombinant families between RM3325 and RM5055. All of the recombinants in group A1 showed significant difference ($P < 0.001$) from Guichao 2 and no significant difference from SIL040 for all six panicle traits evaluated. The reciprocal group A2 contained 13 recombinant families between RM3325 and RM5055. All of the recombinants in group A2 showed significant difference ($P < 0.001$) from SIL040 and no significant difference from Guichao 2. Thus, the A group placed the QTL for all six panicle traits to a region downstream of RM3325. Using the same procedure, the B and C group placed the QTL for all six panicle traits to a region downstream of RM5055 and RM427, respectively, and the F group placed the QTL to a region upstream of marker 3683. Of greatest important was the group D1 and D2, the genotypes of them were identical between markers 63130 and RM481, however, the differences of them from the controls SIL040 and Guichao 2 were completely reverse for six panicle traits evaluated. Group D1 showed significant difference ($P < 0.001$) from SIL040 while group D2 showed no significant difference from SIL040. Group D2 showed significant difference ($P < 0.001$) from Guichao 2 while group D1 showed no significant difference from Guichao 2. These results indicated that the QTL for all six panicle traits was located in the recombination region between markers 63130 and RM481. This conclusion was further conformed by the group E, where the similar case as that described in group D occurred between group E1 and E2.

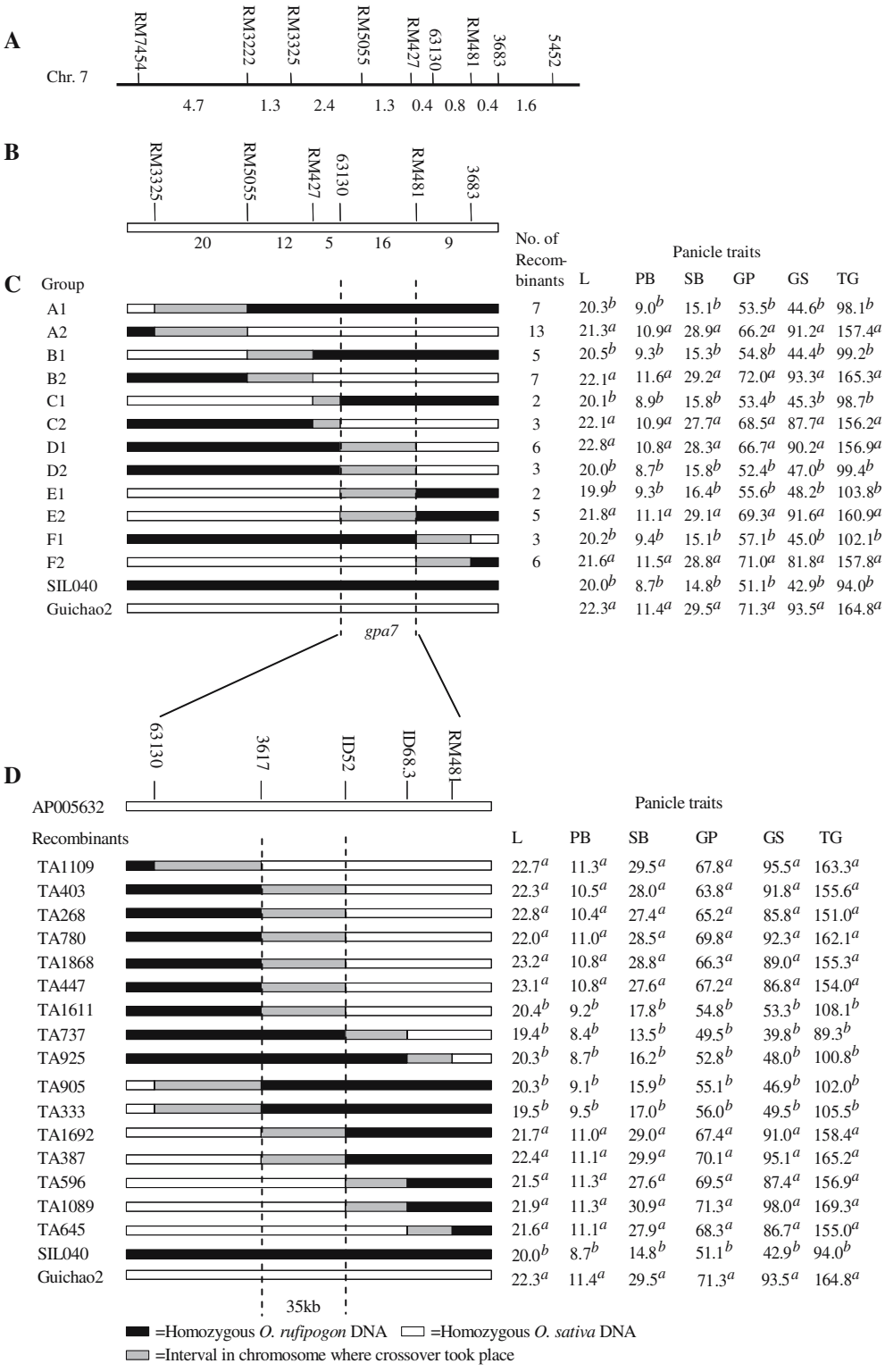
To more precisely determine the critical recombination point, three new markers were developed to further subdivide the interval between markers 63130 and RM481. The 16 recombinants carrying crossover between markers 63130 and RM481 were further genotyped with the new markers. Similarly using homozygous recombinants, the mean phenotypic value of each of six panicle traits for each of 16 recombinant families was compared to that of the controls SIL040 and Guichao 2. By pairwise comparisons between each recombinant family and the two parental controls, the recombinants TA1109, TA905 and TA333 placed *gpa7* to a region downstream of marker 63130, the recombinants TA387 and TA1692 placed *gpa7* to a region downstream of marker 3617, the recombinants TA925 and TA645 placed *gpa7* to a region upstream of RM481, and the recombinants TA596 and TA1089 placed *gpa7* to a region upstream of marker ID68.3. The most informative recombinants were recombinants TA403, TA268, TA780, TA1868, TA447 and TA1611 with identical genotype between markers 3617 and ID52. The recombinants TA403, TA268, TA780, TA1868 and TA447 all showed significant difference ($P < 0.001$) from SIL040 and no significant difference from Guichao 2 for all six panicle traits, however, the case in the recombinant TA1611 was completely reverse, which was that TA1611 showed significant difference ($P < 0.001$) from Guichao 2 and no significant difference from SIL040 for all six panicle traits. These results demonstrated that *gpa7* must reside in the recombination region between markers 3617 and ID52. Thus, by further substitution mapping, we finally narrowed down the location of *gpa7* to a 35-kb region between markers 3617 and ID52 (Fig. 3d).

Candidate genes in the 35-kb target region

Based on available sequence annotation database (http://www.rgp.dna.affrc.go.jp; http://www.tigr.org), there are five predicted genes (LOC_Os07g05870, LOC_Os07g05880, LOC_Os07g05890, LOC_Os07g05900 and LOC_Os07g05910) in the 35-kb target region of cultivated rice genome. LOC_Os07g05870 is expressed protein and has corresponding full-length cDNA (AK066146) in GenBank. LOC_Os07g05880 belongs to Kelch repeat containing F-box protein family, and has corresponding full-length cDNA (AK069618). LOC_Os07g05890, LOC_Os07g05900 and LOC_Os07g05910 are all hypothetical protein, and there are no full-length cDNAs or ESTs corresponding to any of them. LOC_Os07g05900 contained a C2H2 type of Zinc finger domain.

Discussion

In natural populations, most phenotypic variation is continuous and affected by alleles at multiple loci



(Alpert and Tanksley 1996; Frary et al. 2000). And the genetic resolution of quantitative traits in populations that segregate simultaneously for different QTL scattered throughout the genome was low compared with QTL analysis in lines that segregate for a single region (Zamir 2001). Introgression lines were identical for the entire genome except for few introgression segments compared with the recurrent parent, and all the phenotypic variation in the ILs was associated with the introduced segment (Fridman et al. 2004). In tomato, a total of 76 segmental ILs that are composed of marker-defined genomic regions of the wild species *L. pennellii*,

Fig. 3 Substitution mapping of *gpa7*. **a** The genetic linkage map (in cM) of *gpa7* region on chromosome 7 based on 400 F_2 plants. *Numbers below the line* indicate genetic distance between adjacent markers. **b** High-resolution linkage map of the *gpa7* region produced with 1,966 F_2 plants. The number of recombinants between adjacent markers is indicated under the linkage map. **c** Progeny testing of homozygous recombinants delimited the *gpa7* locus to the region between markers 63130 and RM481. The 62 recombinants were grouped into 12 groups based on their genotypes. On the right, the number of recombinants contained in each group and phenotypic difference of each group from the controls SIL040 and Guichao 2 for six panicle traits were indicated. **d** Fine mapping of *gpa7*. The 16 recombinants between markers 63130 and RM481 are shown on the left. The *open bar* shows a part of the PAC clone AP005632. To the right are the phenotypic differences of each recombinant family from the controls SIL040 and Guichao 2 for six panicle traits. An "a" following the phenotypic value indicates that the mean phenotypic value of recombinant is significantly different from that of the control SIL040 at $P < 0.001$; a "b" indicates that the mean phenotypic value of recombinant is significantly from that of the control Guichao 2 at $P < 0.001$. *L* length of panicle, *PB* primary branches per panicle, *SB* secondary branches per panicle, *GP* grains on primary branches per panicle, *GS* grains on secondary branches per panicle, *TG* total grains per panicle

substituting for the homologous intervals of the cultivated variety M82, was developed and a number of phenotypic traits in these lines were quantified and QTLs were identified (Eshed and Zamir 1994, 1995). These tomato introgression lines have been used successfully to clone genes underlying quantitative traits (Frary et al. 2000; Fridman et al. 2000, 2004). In rice, Sobrizal et al. (1999) and Kurakazu et al. (2001) constructed serials of introgression lines carrying *Oryza glaberrima* and *Oryza meridionalis* introgressed segments, respectively, in cultivated rice, *O. sativa* L.. To attempt to gain an insight into the genetic factors underlying differences between common wild rice and cultivated rice, in our laboratory, a set of 159 introgression lines derived from a cross between Guichao 2, a high-yielding commercial *indica* cultivar (*O. sativa*), as the recurrent parent and an accession of common wild rice collected from Dongxiang county, Jiangxi Province, China, as the donor, was developed by using marker-assisted selection (Tian et al. 2006). The assay of yield-related traits for this set of ILs has been conducted. Because each introgression line harbors a single or fewer segments in the near-isogenic background of the recurrent parent, high-resolution mapping of QTLs as Mendelian factors would be feasible in many ILs by constructing secondary F_2 population derived from a cross between an IL and the recurrent parent. In this study, using SIL040, a member of the 159 ILs, we detected a QTL *gpa7* for grain number per panicle with large effect and this QTL was successfully delimited to a 35-kb genome region. Our results revealed that this set of introgression lines was powerful tools in identifying genes underlying complex quantitative traits.

Yu et al. (1997), Xiong et al. (1999), Xing et al. (2002) and Yamagishi et al. (2002) have detected QTLs associated with grains or spikelets per panicle on chromosome 7, however, by comparison of the physical location, *gpa7* was located in the different location of chromosome 7 from the QTLs reported above. Notably, Li et al. (2006) detected an interval between markers RC7-20 and RC7-31 on chromosome 7 associated with four panicle structure traits (panicle length, primary branch, secondary branch and spikelet number). By comparing, this locus is in the same interval as *gpa7*. This locus might be an important domesticated locus during rice domestication. During the construction of introgression lines (Tian et al. 2006), there existed a large gap of about 50 cM around *gpa7* left uncovered on the chromosome 7, calculating according to the newly developed genetic map (McCouch et al. 2002). Thus, *gpa7* was not detected in previous QTL analysis using introgression lines (Tian et al. 2006). In this study, a total of 208 polymorphic SSR markers, distributed evenly throughout 12 chromosomes, were first used to scan SIL040. *gpa7* was found located on the newly detected *O. rufipogon* introgression on chromosome 7. Using the markers flanking *gpa7* to genotype ILs and phenotypic data of ILs deposited in our laboratory, QTL analysis for grains per panicle in ILs conformed that there was a significant peak near RM481 on the short arm of chromosome 7. The phenotypic variance explained by this QTL in ILs was up to 17%. This result has been included in another article submitted. QTL analysis in another F_2 population derived from the cross between SIL040 and *indica* cultivar 93-11 demonstrated that a major QTL for grains per panicle was located in the same interval as *gpa7*. These results, in combination with the phenotypic evaluations of multiple sites and years for SIL040, revealed that the effect of the wild allele at *gpa7* was consistent across different genetic backgrounds and environments. From the data of F_1 in Table 3 and QTL parameters in Table 4, *gpa7* behaved incomplete dominant and the allele from *O. rufipogon* decreased grains per panicle.

Based on available sequence annotation database (http://www.rgp.dna.affrc.go.jp/; http://www.tigr.org/), gene candidate for *gpa7* was searched. There are five predicated genes in the target region of cultivated rice genome. Of them, the more interesting candidates were LOC_Os07g05880 and LOC_Os07g05900. LOC_Os07g05880 belongs to Kelch repeat containing F-box protein family. The Kelch motif is an ancient and

evolutionarily-widespread sequence motif of 44–56 amino acids in length. It occurs as five to seven repeats that form a beta-propeller tertiary structure. In general, Kelch-repeat beta-propellers are involved in protein-protein interactions (Prag and Adams 2003). LOC_Os07g05900 contained a C2H2 type of Zinc finger domain, which contains two cysteines and two histidines that coordinate a zinc atom, creating a compact nucleic acid-binding domain. The majority of such proteins characterized to date are DNA-binding transcription factors, and many have been shown to play crucial roles in the development of plants. Ashikari et al. (2005) elucidated the molecular mechanism of *OsCKX2* gene underlying *Gn1a* for gains per panicle, and found that *OsCKX2* is a negative regulator and it is through expression of *OsCKX2* to regulate grain number. Reduced expression of *OsCKX2* causes cytokinin accumulation in inflorescence meristems and increases the number of reproductive organs, resulting in enhanced grain yield. They identified a total of 11 putative *CKX* genes (*OsCKX1* to *OsCKX11*) in rice genome. There is no *OsCKX* gene in the 35 kb target region *gpa7* located. Thus the gene underlying *gpa7* might regulate grain number through another unknown mechanism. However, based on current information, we could not determine the candidate gene underlying *gpa7*. It is because that the *gpa7* is derived from wild rice (*O. rufipogon*) and all of available annotation database was based on sequence of cultivated rice. The cultivated form and its related wild form may be divergent in sequence level. For example, in tomato, Van Der Knaap et al. (2004) revealed that the *sun*, a *L. pennellii* locus controlling tomato fruit shape, is located in a region that *L. pennellii* appears to be prone to DNA rearrangements, including an about 30-kb insertion, a paracentric inversion and a tandem duplication when compared with cultivated tomato. In maize, the gene lost is presented even between different cultivar, e.g., some gene lost in the cultivar B73 when compared with the sequence of cultivar Mo17 (Brunner et al. 2005), similar result is also reported in barley (Scherrer et al. 2005). Thus, the determination of candidate gene must be dependent on the complete sequencing of *O. rufipogon* in the target region. We are progressing with construction of BAC library of *O. rufipogon* to identify the gene underlying *gpa7*.

Domestication of foxtail millet led to the production of much larger inflorescences with a more complex branching patter (Doust et al. 2004). Some of the same changes during domestication are also seen in domestication of maize from its wild progenitor teosinte (Doebley and Stec 1991, 1993). In rice, typical strains of *O. rufipogon* usually have less total grain number per panicle, less primary and secondary branches when compared to cultivated rice (H.W. Cai and C.Q. Sun, unpublished data), and the greatly increased grain number in domestication process is a most important event. Years of domestication and selection for panicle structure have resulted in more branches, especial more secondary branches, and more grains, especial more grains on secondary branches. Many QTLs for grain numbers were reported, however, the changes occurred in panicle structure during domestication was not clearly indicated and little is known about the genes underling these changes. In this study, substitution mapping of panicle traits and detailed quantitative analysis of panicle structure in SIL040, Guichao 2 and their F_1 hybrid demonstrated that *gpa7* affected not only the grain number per panicle but also was responsible for the primary and the secondary branches per panicle, grains on the primary branches and grains on the secondary branches per panicle, showing pleiotropic effects. This case of pleiotropic effects was similar to that of *fw2.2*, a fruit weight QTL key to the evolution of domesticated tomatoes, which not only affects the size of developing tomato fruit, but also has effects on fruit number and photosynthate distribution (Nesbitt and Tanksley 2001). The results of detailed analysis of the ratio of primary branches per panicle and secondary branches per panicle to total branches per panicle and the ratio of grains on primary branches per panicle and grains on secondary branches per panicle to total grains per panicle in SIL040, F_1 and Guichao 2 demonstrated that, during the domestication from common wild allele to cultivated rice one at *gpa7*, not only the number of branches and grains per panicle increased significantly, more importantly, but also the ratio of secondary branches per panicle to total branches per panicle and the ratio of grains on secondary branches per panicle to total grains per panicle increased significantly. All these results suggested that *gpa7* might play an important role in the regulation of grain number per panicle and the ratio of secondary branches per panicle during the domestication of rice panicle. The results reported by Li et al. (2006) supported this conclusion. They identified a location on chromosome 7 that contains QTL of the largest effect on plant height, panicle length, primary branch, secondary branch and spikelet number in the QTL analysis of key domestication traits using an F_2 population derived from a cross between *O. sativa* and the annual wild species, *O. nivara*, and further proposed that fixation of mutation on chromosome 7 could have substantially improved plant architecture and panicle structure. By comparing, the locus is located in the same interval as *gpa7*. *O. nivara* is annual, while *O. rufipogon* is perennial, and both are

all the closest wild relatives of *O. sativa*. The phylogenetic relationships among the three species, *O. rufipogon*, *O. nivara* and *O. sativa*, has not yet been elucidated entirely. In our study, we used the perennial *O. rufipogon* as the initial parent, and revealed that *gpa7* was associated with six panicle traits by fine mapping. This two studies confirmed that the location on chromosome 7 *gpa7* resided in is an important target of rice panicle domestication selection. The elucidation of molecular mechanism underlying *gpa7* would provide direct clue about which wild species served as the direct progenitor.

Acknowledgements This work was funded by the "973" Project of the Ministry of Sciences and Technology of China (No. 2001CB108801), the Project of Conservation and Utilization of Agricultural Wild Plants of the Ministry of Agriculture of China, and a grant from China National High-Tech Research and Development ("863") Program (No. 2003AA207040).

References

Alpert KB, Tanksley SD (1996) High-resolution mapping and isolation of a yeast artificial chromosome contig containing *fw2.2*: a major fruit weight quantitative trait locus in tomato. Proc Natl Acad Sci USA 93:15503–15507

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. Science 309:741–745

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17:343–360

Doebley J, Stec A (1991) Genetic analysis of the morphological differences between maize and teosinte. Genetics 129:285–295

Doebley J, Stec A (1993) Inheritance of the morphological differences between maize and teosinte: comparison of results for two F_2 populations. Genetics 134:559–570

Doust AN, Devos KM, Gadberry MD, Gale MD, Kellogg EA (2004) Genetic control of branching in foxtail millet. Proc Natl Acad Sci USA 101:9045–9050

Eshed Y, Zamir D (1994) Introgressions from *Lycopersicon pennellii* can improve the soluble-solids yield of tomato hybrids. Theor Appl Genet 88:891–897

Eshed Y, Zamir D (1995) Introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield associated QTL. Genetics 141:1147–1162

Frary A, Nesbitt TC, 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

Fridman E, Pleban T, Zamir D (2000) A recombination hotspot delimits a wild species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc Natl Acad Sci USA 97:4718–4723

Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305:1786–1789

van der Knaap E, Sanyal A, Jacksom SA, Tanksley SD (2004) High-resolution fine mapping and fluorescence in situ hybridization analysis of *sun*, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. Genetics 168:2127–2140

Kurakazu T, Sobrizal, Ikeda K, Sanchez PL, Doi K, Angeles ER, Khush GS, Yoshimura A (2001) *Oryza* meridionalis chromosomal segment introgression lines in cultivated rice, *O. sativa* L. Rice Genet Newsl 18:81–82

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Li ZK, Pinson SRM, Stansel JW, Paterson AH (1998) Genetic dissection of the source–sink relationship affecting fecundity and yield in rice (*Oryza sativa* L.). Mol Breed 4:419–426

Li J, Thomson M, McCouch SR (2004) Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. Genetics 168:2187–2195

Li CB, Zhou AL, Sang T (2006) Gentetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. New Phytol 170:185–194

Lin HX, Qian HR, Zhuang JY, Lu J, Min SK, Xiong ZM, Huang N, Zheng KL (1996) RFLP mapping of QTLs for yield and related characters in rice (*Oryza sativa* L.). Theor Appl Genet 92:920–927

Lu C, Shen L, Tan Z, Xu Y, He P, Chen Y, Zhu L (1996) Comparative mapping of QTLs for agronomic traits of rice across environments using a doubled haploid population. Theor Appl Genet 93:1211–1217

Manly KF, Cudmore RH, Meer JM (2001) Map Manager QTX: cross-platform software for genetic mapping. Mamm Genome 12:930–932

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development of 2,240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

Mei HW, Luo LJ, Ying CS, Wang YP, Yu XQ, Guo LB, Paterson AH, Li ZK (2003) Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two testcross populations. Theor Appl Genet 107:89–101

Mei HW, Li ZK, Shu QY, Guo LB, Wang YP, Yu XQ, Ying CS, Luo LJ (2005) Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. Theor Appl Genet 110:649–659

Moncada P, Martinez CP, Borrero J, Chatel M, Gauch H, Guimaraes E, Tohme J, McCouch SR (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor Appl Genet 102:41–52

Murray MG., Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Nesbitt TC, Tanksley SD (2001) *fw2.2* directly affects the size of developing tomato fruit, with secondary effects on fruit number and photosynthate distribution. Plant Physiol 127:575–583

Oka HI (1988) Origin of cultivated rice. Developments in crop science, vol 14. Elsevier, Amsterdam

Paterson AH, De-Verna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping

recombinant chromosomes, in an interspecific cross of tomato. Genetics 124:735–742

Prag S, Adams JC (2003) Molecular phylogeny of the kelch-repeat superfamily reveals an expansion of BTB/kelch proteins in animals. BMC Bioinformatics 4:42

Redona ED, Mackill DJ (1998) Quantitative trait locus analysis for rice panicle and grain characteristics. Theor Appl Genet 96:957–963

Sasahara H, Fukuta Y, Fukuyama T (1999) Mapping of QTLs for vascular bundle system and spike morphology in rice, *Oryza sativa* L. Breed Sci 49:75–81

SAS Institute, Inc. (1990) SAS/STAT user's guide. SAS Institute, Inc., Cary

Scherrer B, Isidore E, Klein P, Kim JS, Bellec A, Chalhoub B, Keller B, Feuillet C (2005) Large intraspecific haplotype variablity at the *Rph7* locus results from rapid and recent divergence in the barley genome. Plant Cell 17:361–374

Second G (1982) Origin of the genetic diversity of cultivated rice (*Oryza* spp.), study of the polymorphism scored at 40 isozyme loci. Jpn J Genet 57:25–57

Sobrizal, Ikeda K, Sanchez PL, Doi K, Angeles ER, Khush GS, Yoshimura A (1999) Development of *Oryza glumaepatula* introgression lines in rice, *Oryza sativa* L. Rice Genet Newsl 16:107–108

Sun CQ, Wang XK, Yoshimura A, Iwata N (2001) Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. sativa* L.) using RFLP markers. Theor Appl Genet 102:157–162

Thomson MJ, Tai TH, McClung AM, Lai XH, Hinga ME, Lobos KB, Xu Y, Martinea 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

Tian F, Li DJ, Fu Q, Zhu ZF, Fu YC, Wang XK, Sun CQ (2006) Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*O. sativa* L.) background and characterization of introgressed segments associated with yield-related traits. Theor Appl Genet 112:570–580

Wang ZY, Second G, Tanksley SD, (1992) Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor Appl Genet 83:565–581

Xiao J, Li J, Yuan L, Tanksley SD (1996) Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a subspecific rice cross. Theor Appl Genet 92:230–244

Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899–909

Xing YZ, Tan YF, Hua JP, Sun XL, Xu CG, Zhang Q (2002) Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor Appl Genet 105:248–257

Xiong LZ, Liu KD, Dai XK, Xu CG, Zhang Q (1999) Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O. rufipogon*. Theor Appl Genet 98:243–251

Yagi T, Nagata K, Fukuta Y, Tamuta K, Ashikawa I, Terao T (2001) QTL mapping of spikelet number in rice (*Oryza sativa* L.). Breed Sci 51:53–56

Yamagishi M, Takeuchi Y, Kono I, Yano M (2002) QTL analysis for panicle characteristics in temperate japonica rice. Euphytica 128:219–224

Yu SB, Li JX, Xu CG, Tan YF, Cao YJ, Li XH, Zhang Q, Maroof MA (1997) Importance of epistasis as the genetic basis of heterosis in elite rice hybrid. Proc Natl Acad Sci USA 94:9226–9231

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983–989

Zhuang JY, Lin HX, Lu J, Qian HR, Hittalmani S, Huang N, Zheng KL (1997) Analysis of QTL × environment interaction for yield components and plant height in rice. Theor Appl Genet 95:799–808