Theor Appl Genet (2005) 111: 1495–1503
DOI 10.1007/s00122-005-0078-0

ORIGINAL PAPER

Qing Ji · Jufei Lu · Qing Chao · Minghong Gu
Mingliang Xu

Delimiting a rice wide-compatibility gene S_5^n to a 50 kb region

Received: 20 April 2005 / Accepted: 7 August 2005 / Published online: 30 August 2005

Abstract Wide-compatibility (WC) is one of the most important traits in rice, which can overcome the fertility barrier in the *indica*/*japonica* hybrids, and hence to make it possible to utilize the higher yield potential of inter-subspecific hybrids. The S_5^n gene located on chromosome 6 has been previously reported to be responsible for the wide-compatibility in rice. Here we report the precise location of the S_5^n gene. In the first-pass mapping, the S_5^n gene was restricted within a 200 kb region by using a population of 242 isogenic lines in combination with high-density markers developed in the S_5 region. In the fine mapping, the S_5 region was further saturated with newly developed markers and more isogenic lines (549 in total) were investigated. Eventually, the S_5^n gene was mapped within a 50 kb region delimited by the left marker J13 and the right marker J17. One BAC clone screened from the BAC library of the WC rice variety 02428 covered the whole S_5 region. Sequence analysis of the 50 kb region revealed two candidate genes, coding an aspartyl protease and a hypothetical protein. This result would greatly accelerate both cloning and marker-assisted selection of this important S_5^n gene.

Communicated by D.J. Mackill

Qing Ji and Jufei Lu have contributed equally to this paper.

Q. Ji · Q. Chao
College of Bioscience and Biotechnology,
Yangzhou University, 12 East Wenhui Road, Yangzhou,
Jiangsu, 225009 People's Republic of China

M. Xu (✉) · Q. Ji
National Maize Improvement Center,
China Agriculture University, 2 west Yuanmingyuan Road,
Beijing, 100094 People's Republic of China
E-mail: mxu@cau.edu.cn

J. Lu · M. Gu
Key lab for plant biotechnology of Jiangsu province,
Yangzhou University, 12 East Wenhui Road, Yangzhou,
Jiangsu, 225009 People's Republic of China

Introduction

Rice (*Oryza sativa* L.) is one of the most important crops in the world. It provides the food supply for almost half of the world's population. Although many modern technologies have been adopted to improve the yield potential in rice-breeding programs, the yield potential is still too low to meet an ever-increasing demand for rice (Cai 2001). Cultivated rice can be classified into two distinct subspecies—*indica* and *japonica* (Kato 1930). The hybrids between *indica* and *japonica* rice varieties show the higher yield potential compared with *indica*/*indica* or *japonica*/*japonica* hybrids due to greater genetic divergence (Yuan 1987). However, utilization of inter-subspecific heterosis has encountered many obstacles, and the primary one is the partial fertility of *indica*/*japonica* F_1 hybrids (Kato et al. 1928). Fortunately, wide compatibility varieties (WCVs) can produce normal fertility hybrids when crossing with either *indica* or *japonica* varieties (Ikehashi and Araki 1984). Based on this fact, Ikehashi and Araki (1986) proposed a genetic model named 'one-locus sporo-gametophytic interaction', the most prevalent model proposed so far, to explain this phenomenon. According to this model, there are three alleles at the S_5 locus, S_5^i, S_5^j, and S_5^n, present in *indica, japonica*, and WCVs, respectively. The heterozygote of *indica*/*japonica* (S_5^i/S_5^j) produces semi-sterile panicles, resulting from partial abortion of female gametes carrying S_5^j (Ikehashi and Araki 1986). However, this fertility barrier can be overcome by the S_5^n allele, resulting in normal fertile panicles in either *indica*/WCV (S_5^i/S_5^n) or *japonica*/WCV (S_5^j/S_5^n) hybrids. Apart from S_5^n, other loci with minor effects are also involved in improving hybrid fertility, including *S-7, S-8*, and *S-9* identified by Ikehashi and Wan (1996), *f1, f3*, and *f8* located on respective chromosome 1, 3, and 8 reported by Wang et al. (1998), two loci on chromosome 2 and 12 revealed by Liu et al. (1997), two minor loci $S_{d1(t)}^n$ and $S_{d2(t)}^n$ on chromosome 12 found by Yan et al. (2000), and a *S-p(t)* gene on chromosome 11 characterized by

Zhang et al. (1998). However, the S_5^n gene is the major contributor to the full fertility of *indica*/*japonica* hybrids (Liu et al. 1997; Yan et al. 2000; Liang et al. 2001; Yi et al. 2001).

Because of its important role in restoring panicle fertility, many studies have been conducted to determine the location of the S_5^n gene. Using morphological markers, Ikehashi and Araki (1986) roughly mapped the S_5^n gene on chromosome 6. This result has been confirmed by using isozyme markers (Li et al. 1991), as well as RFLP (restriction fragment length polymorphism) markers (Liu et al. 1992; Zheng et al. 1992; Yanagihara et al. 1995). A fine mapping of the S_5^n gene was also conducted with more markers available. Yan et al. (2000) revealed a major QTL locus linked with RG213 on chromosome 6. Liu et al. (1997) mapped the S_5^n gene within an interval of RZ450/R2349 with genetic distances of 1.0 cM right to R2349 and 13.4 cM left to RZ450, respectively. Wang et al. (1998) placed a possible wide compatibility gene in a region of 6.4 cM between R2349 and the *C* gene; the latter controls the production of pigment chromogen. However, the precise location of the S_5^n gene is still not determined due to several obstacles, such as lack of markers, permanent (fixed line) mapping populations, and efficient phenotyping approaches.

The availability of the entire genome sequences of the *indica* rice variety 93-11 and the *japonica* rice variety Nipponbare opens the door to develop high-density markers in the region of interest based on sequence divergence. The high-density markers would therefore accelerate the fine-mapping process, finally leading to cloning of the target gene.

The overall objective of this research was to delimit the S_5^n gene in a narrow interval by using high-density molecular markers in the S_5 region in combination with a population consisting of 549 isogenic lines. The result would therefore lay the basis for the map-based cloning of the S_5^n gene and the marker-assisted selection of this gene in rice-breeding programs.

Materials and methods

Assay of wide-compatibility

The rice spikelet fertility is a ratio of fertile spikelets to total spikelets by scoring two or three panicles in a single plant. The S_5^n allele can enhance the spikelet fertility by at least 15% in *indica*/*japonica* hybrids, thus exhibiting wide-compatible capability. Since the spikelet fertility itself varies widely under different *indica*/*japonica* backgrounds (Gu et al. 1993a, b; Liu et al. 1996); this in turn makes it difficult to determine the wide-compatibility trait, especially in F_2 or early backcross populations. Fortunately, the S_5^n allele is closely linked to the *C* gene that controls the production of pigment chromogen, showing purple apiculus at the top of the glume (Ikehashi and Araki 1986). Therefore, the purple apiculus trait can be used as a morphological indicator for the presence of the S_5^n allele, and hence be very helpful in distinguishing an individual with or without the S_5^n allele.

In the process of introduction of the S_5^n gene from a WC variety into non-WC rice varieties, individuals from each backcross generation were divided into two groups, with/without purple apiculus. The plants from each group were individually crossed to the test variety, and assayed for their spikelet fertilities in the F_1 hybrids. If there was no crossover between *C* and S_5^n alleles in the parental line, offspring with purple apiculus must show 30% higher fertility on average than that of offspring without purple apiculus. Conversely, if there was no difference in fertility between these two offspring groups, this indicated that a crossover between *C* and S_5^n alleles occurred in the parental line. With the aid of the morphological marker, purple apiculus at top of glume, the S_5^n allele would generally not be lost during long-term backcrossing.

Plant materials

Four rice cultivars, including a typical *indica* variety Nanjing11, a typical *japonica* variety Balilla, and two wide-compatibility varieties 02428 (*japonica*) and Dular (*indica*), were used to prepare isogenic lines in the present study. In the initial crossing, each wide-compatibility variety was crossed to both non-WCVs 'Nanjing 11' and 'Balilla', producing four crosses: 'Dular'×'Nanjing 11', 'Dular'×'Balilla', '02428'×'Nanjing 11', and '02428'×'Balilla'. The Non-WC variety ('Nanjing 11' or 'Balilla') in each cross was used as the recurrent parent in the continued backcross to introgress the S_5^n gene from the WC variety into the typical *indica* or *japonica* variety. In each backcross generation, the WC trait in every individual using the 'Balilla' as the recurrent parent was assayed by crossing with the tester 'Nanjing 11', while, the WC trait in every individual using the 'Nanjing 11' as the recurrent parent was assayed by crossing with the tester 'Balilla'. Only those individuals showing WC traits were selected for continuous backcrossing. After backcrossing for 10–13 generations, followed by self-pollination for two or three generations, 549 isogenic lines were obtained. The isogenic lines could be divided into three phenotypes at the purple apiculus and WC traits: phenotype I, both purple apiculus and WC were present; phenotype II, only purple apiculus remained; phenotype III, only the WC trait remained. The phenotype I accounted for most of the developed isogenic lines. The phenotypes II and III were derived from crossovers occurring between the *C* and S_5^n loci in the selfing step. For some important isogenic lines, such as recombinants within the S_5^n region, their phenotypes were assayed for more than 2 years, so as to definitely confirm the WC phenotype.

DNA extraction

Leaves of young seedlings were harvested and pounded into pieces with a bamboo skewer in liquid nitrogen, and ground to fine powder with steel beads on a high-speed vortex. The DNA extraction was conducted according to the method described by Murray and Thompson (1980).

Primer design

Taking all previous mapping results into account, it could be concluded that the S_5^n gene was located in an interval between markers R1954 and R2349 with a genetic distance of ~13.6 cM. The BLAST (basic local alignment search tool) search was used to find sequence divergence between the *indica* rice 93-11 and the *japonica* rice Nipponbare in the R1954/R2349 region. The genome sequences of these two rice varieties were available in the websites (http://www.btn.genomic.org.cn/rice) and (http://rgp.dna.affrc.go.jp/), respectively. The InDels (inserts and deletions) were amenable for developing sequence-tagged site markers (STS), while those single nucleotide polymorphisms (SNPs) embedded in some restriction recognition sites were useful for developing cleaved-amplified polymorphic sequence (CAPS) markers.

Primers for candidate STS or CAPS markers were designed based on the flanking regions of the InDels or SNPs using the PRIMER0.5 software (http://www-genome.wi.mit.edu/ftp/pub/software/primer.0.5). The criteria setup for primer design were as follows: (1) length of 20–25 bp; (2) GC content of 50–60%; (3) Tm of 60–75°C; (4) no secondary structure and no run of three identical nucleotides; (5) preference for either G or C at the 3′-end of the primer.

Development of markers

The primers for candidate markers were first used to amplify parental lines, and only those primers that gave unambiguous and expected PCR products were retained. For STS markers, polymorphic PCR bands should appear after separation on an agarose or a polyacrylamide gel, while, for those CAPS markers, polymorphic bands could be observed only after digestion with a certain restriction enzyme.

The PCR reaction mixture was adjusted to a total volume of 10 μl containing 30 ng of template DNA, 10 mM Tris–HCl, 50 mM KCl, 2.5 mM $MgCl_2$, 0.2 mM dNTP, 2 μM primers, and 0.5 U Taq polymerase. The reaction was performed as follows: denaturation at 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 2 min, and with a final extension step at 72°C for 10 min. For CAPS marker, the PCR product was then subjected to digestion with a corresponding restriction enzyme. The reaction was carried out in a volume of 15 μl, consisting of 1.5 μl 10x buffer, 5U restriction endonuclease, and 4 μl PCR product. The mixture was incubated at 37°C for 2 h.

The PCR products (STS markers) or digested PCR products (CAPS markers) were subjected to electrophoresis on either 1% agarose gel or 6% polyacrylamide gel, depending on the sizes and difference of two polymorphic bands. If the difference between any two polymorphic bands was larger than 50 bp, then a 1% agarose gel was used; whereas, if the difference was smaller than 50 bp, then a 6% polyacrylamide gel was used instead. Following electrophoresis, agarose gels were stained with ethidium bromide, while polyacrylamide gels were sliver-stained for visualization.

Data collection and analysis

A band with the same size as that of a WCV ('02428' or 'Dular') was marked as '1', while a band showing the same size as that of a non-WCV ('Nanjing 11' or 'Ballila') was labeled as '0'. For each marker, three banding patterns were observed in the tested isogenic lines, two homozygotes '0/0' and '1/1', and one heterozygote '0/1'. The physical distance between any two markers could be easily calculated according to their locations on the *japonica* rice Nipponbare genome.

Results

Development of saturated markers in the S_5 region

Generally, three kinds of sequence divergences were observed in the BLASTn searches between 'Nipponbare' and '93-11', including large InDels (inserts/deletions), small InDels, and SNPs. Theoretically, large InDels are feasible to be developed into STS markers, and PCR products can easily be resolved on an agarose gel. However, out of ten candidate STS markers based on large InDels, only one was finally developed into a STS marker. The remaining nine candidate STS markers showed no differences in the PCR products between two parents, and all amplicons were the same sizes as those expected from the *japonica* rice variety Nipponbare. The result indicated that most large InDels obtained from the BLASTn search were not present between the *indica* and the *japonica* parents in our mapping populations, and were therefore unlikely to be developed into STS markers. Compared with large InDels, small InDels were more likely to be developed into STS markers. In the present study, 41 candidate markers were designed based on small InDels, and 24 of them were successfully developed into STS markers. The SNPs are the most frequent sequence polymorphisms, which are widely distributed in the rice genome. It is estimated that there is one SNP in every 268 bp (Feng et al. 2002). However, most SNPs detected between '93-11' (*indica*) and 'Nipponbare' (*japonica*) seemed not present between our

Table 1 Primer sequences and sizes of polymorphic bands for the developed STS and CAPS markers in the S_5 region

Primer	Type	Sequence(5′–3′)		Size (bp)			
		Forward	Reverse	N11	02428	Balilla	Dular
J1	STS	AGCTAGGCACCACAAGACTG	CTTTGTGGTAGCTTCATGGG	730	1240	1240	1240
J2	STS	TGAAGGCCAGAGCATTACAG	TGGTATGGAGATCGATGGAC	226	222	226	226
J3	STS	GATCAACATGCTGTCAGCTC	GGAGATAGTTGCAGTGGATG	197	197	202	202
J4	STS	TGACGATGTCGAGGACGAAC	CTCCAGCTCACCAACGACAG	139	142	139	139
J5	STS	CAGGAAGCCACATCGAAGTG	ATATCCGCGTCTTCTGGCTC	283	283	289	289
J6	STS	ATCGGTGATCGAGCATATCG	GGTAGCTAATCGGCAATGTG	316	316	306	306
J7	STS	TCATAAGCTCGATCTCGTGC	TGCAGCACCATGAATTCGTC	127	127	130	127
J8	STS	TGTAGCCAACAGCCACAGTG	ACATCAGTAGCAGCACAAGG	262	262	272	272
J9	STS	CTGAGGAGGTAGCAAGGAGG	CGGTCCTCGAATGAATTCAG	400	400	390	400
J10	STS	TTATATTATGGGACGGAGGGAG	ACTCTATCAAGGCAGCATGC	163	158	158	163
J11	STS	GATCCGTCGTCCAATCCGTC	CCGATGCGTGTCTGAGTCTG	153	155	153	153
J12	STS	TCGCATAGCTTCAGCTCTAG	CCACAACTCCACTCAAGACT	194	200	200	200
J13	STS	GTAAGGAATGTGCAATCGCT	GGAACTAAAGAGGGCCTAGG	224	224	228	228
J14	STS	GTGTTCTACATAGACATTCT	CCTACTCAAAATAAAGACTA	223	225	227	227
J15	STS	GACCAGCACCATCTTCGATC	CCTTGAGAGTTCACAGCGTG	803	569	803	569
J16	STS	TACACTTCGCCAGCTTCAAG	GCCAGCCAATCAATAGTGTC	172	174	172	172
J17	STS	GATCTGGCGGACTTGAAGTC	GGACATTCTCGACGGATTAG	353	382	353	382
J18	STS	GTCGTTGGCTTAGGATCGAG	TTGGACAGATGTGAGACGTC	264	260	264	260
J19	STS	ATGTTCTCGGCCTGCTCCAG	CCGCAAGGTCTGACGAGTAC	232	237	232	237
J20	STS	TGAGCGGCTTAATGATCACG	TTGGTGTGGAGGTACATTCG	263	260	263	260
J21	STS	CCACGTTGCAACCACTTATG	GTGAGATGATGTGGAACACC	346	353	346	353
J22	STS	ATTACTACGGTCAGTTTGGC	TCAGACCTTCAAGTTCCAGTAC	107	100	107	100
J23	STS	TCCTTGAACCACTCGTCCTC	ATCGCCTACAAGACGTCGAG	174	178	174	178
J24	STS	ACTTGTGCAGGAGAAGGTTG	TTGCTGTACAGGTGTTCTCC	334	329	334	329
J25	STS	GCCATTCAGACCAAAAGAAC	CCGGCTACGATGTATACGTG	172	189	172	189
J26	CAPS(*Mse* I)	CGAAGACAGCAGCATCAACG	CCGGAATCGACGATCATCTC	706	706	506/200	706

mapping parents, since only one out of the nine SNPs was developed into a CAPS marker. In total, 26 markers were developed in the S_5 region (Table 1), and each marker showed polymorphism in at least one pair of the parental lines. These high-density markers were indispensable in the fine mapping of the S_5^n gene.

The first-pass mapping

Two markers, J1 (a developed STS marker) and RM6701 (a SSR marker), were used as starting points to screen all isogenic lines. The J1/RM6701 region with a genetic distance of ~1534 kb overlapped with the

Table 2 The important isogenic lines used in the first-pass mapping

Phenotype	I											II	III
Marker	F19[a] (31.86)[c]	F2031[a] (30.96)	F2045[a] (34.55)	F543-1-5[b] (31.22)	F2005[b] (28.50)	F2010[b] (31.34)	H204[b] (28.33)	F0043[b] (34.41)	F2001[b] (30.97)	F0116[b] (27.47)	F2009[b] (33.28)	F0123[b] (0.85)	F2115[b] (37.31)
J1	0/0	1/1	1/1	0/0	1/1	1/1	0/0	0/0	1/1	0/0	1/1	0/0	0/0
RM5754	/	/	/	0/1	1/1	1/1	0/1	0/1	1/1	0/1	1/1	0/1	0/0
J2	/	/	/	0/1	1/1	1/1	0/1	0/1	1/1	0/1	1/1	0/1	0/1
J14	1/1	1/1	1/1	0/1	1/1	1/1	0/1	0/1	1/1	0/1	1/1	0/0	0/1
J17	1/1	1/1	1/1	0/1	1/1	1/1	0/1	0/1	1/1	0/1	1/1	0/0	0/1
J18	1/1	1/1	1/1	0/0	1/1	1/1	0/1	0/1	1/1	0/1	1/1	0/0	0/1
J19	1/1	1/1	1/1	0/0	0/0	1/1	0/1	0/1	1/1	0/1	1/1	0/0	0/1
J20	1/1	0/0	0/0	0/0	0/0	1/1	0/1	0/1	1/1	0/1	1/1	0/0	0/1
RM8274	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/1	1/1	0/1	1/1	0/0	0/1
RM6701	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	1/1	0/0	0/1

0/0: The banding pattern was the same as that of a non-WC variety,
1/1: The banding pattern was the same as that of a WC variety,
0/1: One band was the same as that of a non-WC variety, and the other band was the same as that of a WC variety,
/: no polymorphism was present between two parents

[a]Isogenic lines from the cross Dular'×Nanjing11'

[b]Isogenic lines from the cross 02428'×Nanjing11'

[c]Spikelet fertility (%) in parenthesis is an effective value of the S_5^n gene that was calculated by subtracting the spikelet fertility of the cross recurrent parent'×tester line' from that of the respective cross isogenic line'×tester line'. The spikelet fertility showed significant difference between these two crosses for all phenotype I and III isogenic lines, but not for phenotype II isogenic lines

R1954/R2349 region, and was therefore assumed to encompass the S_5 locus. Within the J1/RM6701 region, a total of 17 markers, including 12 developed STS and five SSR markers, were used in the first-pass mapping of the S_5^n gene with a population consisting of 242 isogenic lines. Most isogenic lines belonged to phenotype I, only several isogenic lines fell into the other two phenotypes. Since the *C* allele is located upstream of the S_5^n allele, mapping with phenotype I isogenic lines will effectively delimit the right border, but not the left border of the S_5^n gene. However, both phenotype II (the purple apiculus only) and III (the WC trait only) can be used to delimit the left border of the S_5 locus. In mapping of the S_5 locus present in the WCV 'Dular', one isogenic line F19, derived from the cross 'Dular'×'Nanjing 11', revealed the closest crossover upstream of the S_5 locus; while, F2031 and F2045 from the same cross showed the closest crossovers downstream of the S_5 locus. These three recombinants allowed delimitation of the S_5 locus between markers J1 and J20 in 'Dular'. In mapping of the S_5 locus in the WCV '02428', six recombinants were identified downstream of the S_5 locus, and all of them were derived from the cross '02428'×'Nanjing 11'. These six recombinants were separately derived from crossovers that occurred in intervals of J17/J18 (F543–1-5), J18/J19 (F2005), J20/RM8274 (F2010), and RM8274/RM6701 (H204, F0043, and F2001) (Table 2). It was obvious that the nearest crossover occurred between markers J17 and J18, hence delimiting the right border of the S_5 locus to the marker J18. There were three recombinants (all from the cross '02428'×'Nanjing 11') upstream of the S_5 locus, derived from crossovers in intervals J1/RM5754 (F0116 and F0123) and RM5754/J2 (F2115, phenotype III) (Table 2). Interestingly, F0123 was a phenotype II isogenic line, carrying an insert with its right border up to marker J2 (Table 2). This result convinced us that the left border of the S_5 locus should not extend beyond the marker J2. In summary, the S_5 locus in the WCV '02428' has been roughly restricted within an interval between J2 and J18 with a physical distance of 200 kb. The individual F2009 was a representative for most of phenotype I isogenic lines that had both purple apiculus and wide compatibility traits, and showed "1/1" genotypes for all markers in the S_5 region (Table 2). A number of non-WC isogenic lines were also investigated with these 17 markers, and no DNA segment was found to derive from WC varieties. This, on the other hand, supported the fact that the S_5^n gene should be located within the J1/RM6701 region (Figs. 1, 2).

The fine mapping of the S_5^n gene

With an aim to further narrow down the S_5 locus, especially from the left border, more phenotype II and III isogenic lines were employed. Accordingly, an additional 307 isogenic lines were selected, 25 of which belonged to phenotypes II or III. Meanwhile, 14 more markers were developed within the 200 kb of the S_5 region. For the S_5^n gene in the WCV 'Dular', two recombinants from the cross 'Dular' ×'Nanjing11' (F1509 and F1518) which belonged to phenotype III were identified. These two recombinants were derived from a crossover in intervals of J13/J14, delimiting the shortest left border of the S_5 locus to the marker J13. On the right side of the S_5 locus, two recombinants (F115 and F309, phenotype I) were found in the cross 'Dular'×'Balilla', which could further shorten the right border to the marker J17 (Table 3). It could be concluded that the S_5^n gene was located between markers J13 and J17 in the WCV 'Dular'. Likewise, in mapping of the S_5^n gene for the WCV '02428', three phenotype III recombinants (F404, F405, and F406) from the cross '02428'×'Nanjing 11' were derived from crossovers in the interval of J11/J14; while, one phenotype I recombinant F133 from the cross '02428'×'Balilla' was derived from a crossover that occurred between J17 and J26. These recombinants allowed restriction of the S_5^n gene into an interval of J11/J17 in the WCV '02428'.

As shown, the location of the S_5^n gene in the 'Dular' perfectly overlapped with that in the '02428', indicating that the S_5^n gene responsible for the wide-compatibility trait was located at the same chromosomal site in both 'Dular' and '02428'. Eventually, the S_5^n gene was mapped within a 50 kb region delimited by the left marker J13 and the right marker J17.

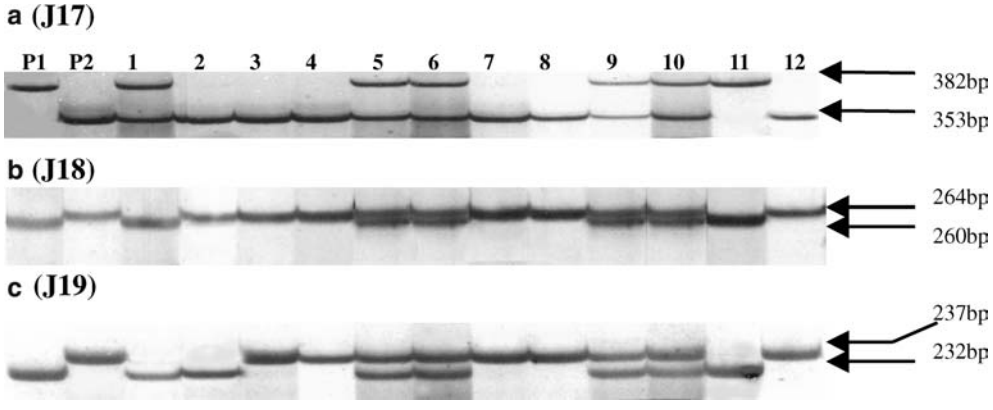

Fig. 1 Banding patterns of three developed markers (J17, J18, and J19) downstream of the S_5^n locus in the first pass mapping. P1—Nanjing11, P2—02428 *Lanes 1* to *12* represent isogenic lines: F543-1-5, F2005, F2031, F2045, H204, F0043, F2010, F2001, F0116, F2115, F0123, and F2009, respectively

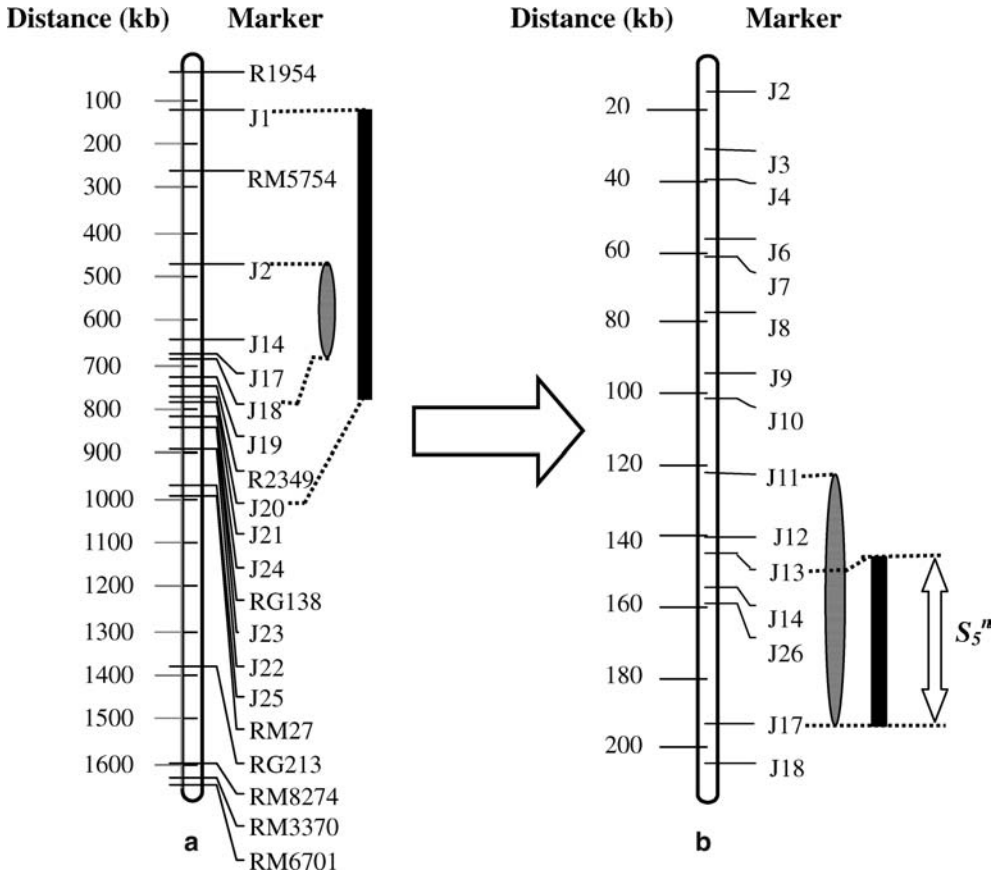

Fig. 2 Delimitation of the S_5^n gene to a 50 kb region. **a** In the first-pass mapping, the S_5^n gene was restricted within the region between markers J1 and J20 in the cross 'Dular'×'Nanjing11' (*rectangle, black*); while, the S_5^n gene was delimited to a 200 kb region with the left marker J2 and the right marker J18 in the cross '02428'×'Nanjing11' (*ellipse, gray*). **b** In the fine mapping process, the S_5^n gene was further narrowed down into a region between markers J13 and J17 (in the crosses 'Dular'×'Nanjing11' and 'Dular'×'Balilla', *rectangle, black*) and a region between markers J11 and J17 (in the crosses '02428'×'Nanjing11' and '02428'×'Balilla', *ellipse, gray*). Obviously, the S_5^n regions from both WCVs 'Dular' and '02428' are overlapped, it could be concluded that a tentative region encompassing the S_5^n gene was bordered by the left marker J13 and the right marker J17 with a physical distance of 50 kb (*hollow arrow*)

Confirmation of the S_5 location

All recombinants in the J2/J18 region, 11 on the left and 11 on the right to the S_5 locus, were genotyped at all 26 markers again to ensure that no mistake had occurred in the mapping process. Meanwhile, all recombinants were assayed for their purple apiculus and WC phenotypes by crossing to respective test varieties and investigating the spikelet fertilities in the offspring. All data obtained in the first-pass mapping and the fine mapping was confirmed, indicating the reliability of both phenotypes and genotypes for all recombinants.

Screening of the positive BAC clones

A bacterial artificial chromosome (BAC) library was constructed from the WC variety 02428 with 5× coverage of the rice haploid genome equivalents. Screening of this BAC library using the flanking markers (J14 and J17) of the S_5^n gene revealed one positive clone with an insert size of $\sim$ 100 kb. This positive BAC clone would be valuable towards cloning of the S_5^n gene.

Discussion

For the three kinds of sequence divergences between the *indica* rice 93-11 and the *japonica* rice Nipponbare, only short InDels are most likely to be developed into STS markers. This suggests that most short InDels present in the sequenced rice varieties 93-11 and Nipponbare are also conserved in other *indica* and *japonica* rice varieties. For large InDels and SNPs, only a few could be developed into markers. The result may be explained by two factors: (1) These sequence diversities present between '93-11' and 'Nipponbare' may not be present in other rice varieties; (2) the *indica* rice 93-11 may have many incorrect nucleotides resulting in the observation that all PCR products were the same as expected from the 'Nipponbare' sequences. Since the *indica* rice 93-11 only has the draft map derived from whole genome shotgun approach, it is unavoidable to incorporate false nucleotides in sequence assembly.

Many hypotheses have been proposed to explain the mechanism underlying wide compatibility since its discovery. The two influential models were one locus model and two loci model (Ikehashi and Araki 1986; Oka 1974). In addition, multiple loci, such as *S-7, S-8, S-9, f1, f3, f5, f8, S-p(t)* $S_{d1(t)}^n$, $S_{d2(t)}^n$, and etc, were reported to be involved in the wide-compatibility trait (Ikehashi and Wan 1996; Wang et al. 1998; Zhang et al. 1998; Yan et al. 2000). On the contrary, Lu et al. (1992) claimed that all hypotheses proposed so far couldn't fully explain wide-compatibility. Some researchers even doubted the presence of the S_5^n gene (Sano 1993). Our result, however, clearly showed the effect of the S_5^n allele and its

Table 3 The important recombinants used for the fine mapping of the S_5^n gene

Phenotype	I				III			
Marker	F115[c] (39.05)[e]	F309[c] (41.08)	F133[d] (39.51)	F1509[a] (32.34)	F1518[a] (29.40)	F404[b] (35.31)	F405[b] (30.54)	F406[b] (33.42)
J2	/	/	1/1	/	/	0/0	0/0	0/0
J3	/	/	1/1	0/0	0/0	/	/	/
J5	/	/	1/1	0/0	0/0	/	/	/
J4	/	/	1/1	/	/	0/0	0/0	0/0
J6	/	/	1/1	0/0	0/0	/	/	/
J7	1/1	1/1	1/1	/	/	/	/	/
J8	/	/	1/1	0/0	0/0	/	/	/
J9	1/1	1/1	1/1	/	/	/	/	/
J10	1/1	1/1	/	/	/	0/0	0/0	0/0
J11	/	/	1/1	/	/	0/0	0/0	0/0
J13	/	/	1/1	0/0	0/0	/	/	/
J14	/	/	1/1	0/1	1/1	0/1	0/1	1/1
J26	1/1	1/1	1/1	/	/	/	/	/
J17	0/0	0/0	0/0	0/1	1/1	0/1	0/1	1/1
J18	0/0	0/0	0/0	0/1	1/1	0/1	0/1	1/1

0/0: The banding pattern was the same as that of a non-WC variety,
1/1: The banding pattern was the same as that of a WC variety,
0/1: One band was the same as that of a non-WC variety, and the other band was the same as that of a WC variety,
/: no polymorphism was present between two parents
[a]Isogenic lines from the cross 'Dular'×'Nanjing11'
[b]Isogenic lines from the cross '02428'×'Nanjing11'
[c]Isogenic lines from the cross 'Dular'×'Balilla'
[d]Isogenic lines from the cross '02428'×'Balilla'
[e]Spikelet fertility (%) in parenthesis is an effective value of the S_5^n gene that was calculated by subtracting the spikelet fertility of 'recurrent parent'×'tester line' from that of 'isogenic line'×'tester line'

contribution to the spikelet fertilities of *indica*/*japonica* hybrids.

In summary, four factors in the present study lead to the successful fine-mapping of the S_5^n gene: a high-density marker map, improved mapping populations, duplicate tests for recombinants, and a two-step mapping strategy. Firstly, the high-density marker map is indispensable for the fine-mapping efforts. In the previous reports, only SSR and RFLP markers were available for mapping, and not many SSR or RFLP markers were present in the S_5 region. With the advent of whole genome sequences for both *indica* and *japonica* rice, development of high-density markers in the region of interest becomes realistic by making use of sequence polymorphisms between the *indica* and *japonica* genomes. This in turn ensures a successful fine-scale mapping of the gene of interest. Secondly, a population composed of 549 isogenic lines was used to map the S_5^n gene, and these lines were generated in the breeding program after backcrossing for more than 10 generations, followed by two or three generations of selfing. Apart from the S_5 region, all isogenic lines from the same cross shared the same background as the recurrent parent. Moreover, our progeny testing of spikelet fertility in each backcross ensured an introgression of the S_5^n gene from a WC variety into a non-WC variety. It has been shown that genetic background has a large impact on spikelet fertility, leading to variation in spikelet fertility in different *indica*/*japonica* hybrids, therefore the wide compatibility trait could be accurately assessed only in the same genetic background. In the previous reports, the mapping populations were either F_2 or BC_1, and this made it very difficult to accurately determine wide-compatibility. As a consequence, the location of the S_5^n gene was not consistent among different researchers (Liu et al. 1997; Wang et al. 1998; Yan et al. 2000). Thirdly, repeated testing of some important recombinants was also a key requirement for the precise mapping of the S_5^n locus. Considering all difficulties in measuring the wide-compatibility trait, unambiguous phenotypes of the important recombinants were very necessary to get the precise location of the S_5^n gene. This required repeated testing of the wide-compatibility traits across several generations. Finally, a two-step mapping program adopted in the present study was an efficient and cost-effective approach towards precise location of a gene like S_5^n . In the first-pass mapping, isogenic lines with typical phenotypes were surveyed with the markers within the S_5 region as revealed by the previous reports. This effectively narrowed down the S_5^n gene to 200 kb, covered by two BAC clones. In the following fine-mapping step, all sequence divergences between *indica* and *japonica* within this 200 kb were exploited to develop as many markers as possible to saturate the S_5 region. The high-density markers (on average one marker every 14.3 kb) allowed detection of almost all crossovers within the 200 kb harboring the S_5^n gene, leading to delimitation of the S_5^n gene to 50 kb covered by one single BAC clone.

Purple apiculus in rice could be observed in the presence of the three complementary genes: chromogen *C*, and anthocyanin-activating genes *A* and *P*. For the four parental varieties, only the WC variety 'Dular' showed purple apiculus, therefore its genotype must be *CCAAPP*. Another WC variety '02428', together with the two test varieties ('Balilla' and 'Nanjing 11') did not

exhibit any purple apiculus. The genotype of 'Balilla' was reported to be *ccAAPP* (Gu et al. 1993a, 1993b; Lu et al. 1998). Interestingly, when '02428' was crossed with either 'Balilla' or 'Nanjing 11', the F_1 hybrids showed purple apicilus, suggesting that '02428' must carry the gene *C*. In the process of an introgression of the S_5^n locus from a WC variety into a non-WC variety, this closely linked indicator, the purple apiculus at top of glume, was employed to judge whether or not the parental line carried the S_5^n gene. This simple method has proved to be very effective, especially in early backcross generations. Since spikelet fertility was greatly affected by genetic background, testing of spikelet fertility for each individual was not enough to judge whether or not it carried the S_5^n allele. The morphological marker allowed rapid grouping of offspring in each backcross generation. If there was no crossover between the *C* and S_5^n genes, most individuals with purple apiculus (group I) should carry the S_5^n allele, and most individuals without purple apiculus (group II) did not carry the S_5^n allele. The spikelet fertility of group I genotypes should be 30% on average higher than that of group II. Conversely, if crossing-over occurred between the *C* and S_5^n genes, there would be no difference in spikelet fertility between the two groups. Although it is a very tedious work to test fertility for every individual for both groups, this method is very reliable to transfer the S_5^n gene. Our work clearly indicates the presence of the wide-compatibility gene S_5^n and its ability to enhance spikelet fertility in *indica*/*japonica* hybrids.

Besides the four main crosses, another four populations derived from the four crosses between WC varieties and non-WC varieties, including 'Woaiga'/'Balilla', 'Ketan Nangka'/'Balilla', 'Huanuo'/'Balilla', and 'Cpslo17'/'Balilla', were also investigated using the markers in the S_5 region (data not shown). The results confirmed the location of the S_5^n gene and its contribution to wide compatibility.

Sequence of the 50 kb region encompassing the S_5 locus is already available in both *indica* and *japonica* genome sequence databases. Two genes (coding an aspartyl protease and a hypothetical protein) predicted in this region seem more likely to be candidate S_5^n genes due to their possible function involved in organ development. The candidate S_5^n genes have been cloned and transferred into non-WC rice varieties.

The genetic distance between *C* and S_5^n is estimated to be 14.6 cM (Zheng et al. 1992), however, the physical distance between them is only about 500 kb as revealed by the genome sequence of 'Nipponbare'. The 34.2 kb/cM in the S_5 region is much lower than an average 244 kb/cM in the rice genome (Chen et al. 2002), suggesting frequent recombination in the S_5 region. Due to a high frequency of crossover between *C* and S_5^n, individuals with purple apiculus was frequently found to lose an ability to enhance the spikelet fertility both in backcross and self-crossing generations in the present study. As for the fine-mapping efforts, such individuals were valuable to delimit the left border of the S_5 locus. In other words, the fine mapping of the S_5^n gene was facilitated by the recombination hotspot in the S_5 region.

The closely linked markers would be very useful in marker-assisted selection of the S_5^n gene, thus to breed inter-subspecific hybrids with normal fertility. As shown, apart from poor spikelet fertility, heterosis of inter-subspecific hybrids is stronger than that of intra-subspecific hybrids (Yuan 1987). Therefore, an efficient introgression of the S_5^n gene into non-WC rice varieties with an aid of molecular markers would greatly accelerate the utilization of strong heterosis of inter-subspecific hybrids.

Acknowledgements We thank Bofeng Zhu for the technical assistance. This study was financially supported by the National High-Tech '863' Program of P.R. China, grant No. 2002AA224041, the talent people's foundation of China Agricultural University, and the natural science foundation of Jiangsu province, P.R. China, grant No. BK2001145.

References

Cai DT, Yuan LP, Lu XG (2001) A new strategy of rice breeding in the 21st century II. Searching a new pathway of rice breeding by utilization of double heterosis of wide cross and polyploidization. Acta Agronom Sin 27:110–116

Chen MS, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang GC, Kim H, Frisch D, Yu Y, Sun SH, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li P, Liu JD, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo MZ, Bancroft I, Salse J, Regad F, Mohapatra Trilochan, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14: 537–545

Feng Q, Zhang YJ, Hao P, Wang SY, Fu G, Huang YC, Li Y, Zhu JJ, Liu YL, Hu X, Jia PX, Zhang Y, Zhao Q, Ying K, Yu SL, Tang YS, Weng QJ, Zhang L, Lu Y, Mu J, Lu YQ, Zhang LS, Yu Z, Fan DL, Liu XH, Lu TT, Li C, Wu YR, Sun TG, Lei HY, Li T, Hu H, Guan JP, Wu M, Zhang RQ, Zhou B, Chen ZH, Chen L, Jin ZQ, Wang R, Yin HF, Cai Z, Ren SX, Lv G, Gu WY, Zhu GF, Tu YF, Jia J, Zhang Y, Chen J, Kang H, Chen XY, Shao CY, Sun Y, Hu QP, Zhang XL, Zhang W, Wang LJ, Ding CW, Sheng HH, Gu JL, Chen ST, Ni L, Zhu FH, Chen W, Lan LF, Lai Y, Chen ZK, Gu MH, Jiang JM, Li JY, Hong GF, Xue YB, Han B (2002) Sequence and analysis of rice chromosome 4. Nature 420: 316–320

Gu MH, You AQ, Pan XB (1993a) Analysis on allelic relationship of wide compatibility genes among several WCVs (*Oryza Sativa* L.) (in Chinese). Sci Agric Sin 26:13–21

Gu XY, Gu MH, Pan XB (1993b) A study on mapping of wide compatibility gene in rice variety Lunhui 422 (in Chinese). J Jiangsu Agric Coll 14:7–14

Ikehashi H, Araki H (1984) Variety screening of compatibility types revealed in F_1fertility of distant cross in rice. Jpn J Breed 34:304–313

Ikehashi H, Araki H (1986) Genetics of F_1 sterility in remote crosses of rice. In: Rice genetics: proceedings of the international rice genetics symposium, 27–31 May 1985. IRRI, P.O. Box 933, Manila, Philippines, pp 119–130

Ikehashi H, Wan J (1996) Differentiation of alleles at seven loci for hybrid sterility in cultivated rice (*Oryza Sativa* L.). In: Rice genetics: proceedings of the international rice genetics symposium, 16–20 October 1995, IRRI, Manila, pp 404–408

Kato S (1930) On the affinity of the cultivated rice varieties of rice plants, *Oryza sativa* L. J Dept Agr Kyushu Imp Univ 2:241–275

Kato S, Kosaka H, Hara K (1928) On the affinity of rice varieties as shown by fertility of hybrid plants. Bull Sci Fac Agric Kyushu Univ 3:132–147

Li H, Tang L, Zou J (1991) Marker-based analysis of wide compatibility in rice (in Chinese). Hybrid Rice 4:22–24

Liang GH, Yan CJ, Yi CD, Zhu LH, Gu MH (2001) Re-analysis of wide compatibility in rice (in Chinese). Acta Genet Sin 28:447–457

Liu AM, Zhang QF, Li HB (1992) Location of a gene for wide-compatibility in the RFLP linkage map. Rice Genet Newsl 9:134–136

Liu KD, Zhou ZQ, Xu CG, Zhang QF, Saghai Maroof MA (1996) An analysis of hybrid sterility in rice using a diallel cross of 21 parents involving *indica, japonica* and wide compatibility varieties. Euphytica 90:275–280

Liu KD, Wang J, Li HB, Xu CG, Liu AM, Li XH, Zhang QF (1997) A genome-wide analysis of wide compatibility in rice and the precise location of the *S5* locus in the molecular map. Theor Appl Genet 95:809–814

Lu YG, Zhang GQ (1992) Specific compatibility gene and the utilize hypothesis in hybrid rice breeding (in Chinese). Fujian Sci Technol Rice Wheat 10:1–4

Lu J, Yan CJ, Tang SZ, Zhu LH, Gu MH (1998) Genetic analysis of the wide compatibility of rice variety HUANUO from Yunnan province (in Chinese). J Yangzhou Univ (Nat Sci Ed) 4:31–35

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

Oka HI (1974) Analysis of genes controlling F_1 sterility in rice by the use of isogenic lines. Genetics 77:521–534

Sano Y (1993) Is an egg-killer present in rice? Theor Appl Genet 86:1038–1042

Wang J, Liu KD, Xu CG, Li XH, Zhang QF (1998) The high level of wide compatibility of variety 'Dular' has a complex genetic basis. Theor Appl Genet 97:407–412

Xu YB, Wang JJ, Shen ZT (1989) Study on intersubspecific compatibility in rice (*Oryza sativa*) (in Chinese). Chin J Rice Sci 3:113–118

Yan CJ, Liang GH, Zhu LH, Gu MH (2000) RFLP analysis on wide compatibility genes in rice variety 'Dular' of ecotype Aus (in Chinese). Acta Genet Sin 27:409–417

Yanagihara S, Mccouch SR, Ishikawa K, Ogi Y, Maruyama K, Kehashi H (1995) Molecular analysis of the inheritance of the S-5 Locus, conferring wide compatibility in *Indica*/*Japonica* hybrids of rice (*O. Sativa* L). Theor Appl Genet 90:182–188

Yi CD, Yan CJ, Liang GH, Zhu LH, Gu MH (2001) RFLP analysis of the effect of wide compatibility genes in Aus variety 'Dular' (in Chinese). Acta Genet Sin 28:540–549

Yuan LP (1987) Strategic hypotheses of breeding for hybrid rice (in Chinese). Hybrid Rice 1:1–3

Zhang XH, Zhu XD, Qian Q, Zhu LH, Zeng DL, Cao LY, Wang JL (1998) Mapping a new gene for wide compatibility of rice in relation to RFLP markers (in Chinese). Chin J Rice Sci 12:11–16

Zheng KL, Shen B, Qian HR, Wang JL (1992) Tagging genes for wide compatibility in rice via linkage to RFLP markers (in Chinese). Chin J Rice Sci 6:145–150