Chinese Science Bulletin

Fine mapping and candidate gene analysis of purple pericarp gene *Pb* in rice (*Oryza sativa* L.)

WANG Caixia & SHU Qingyao†

IAEA-Zhejiang University Collaborating Center, National Key Laboratory of Rice Biology and Key Laboratory of Chinese Ministry of Agriculture for Nuclear-Agricultural Sciences, Institute of Nuclear Agricultural Sciences, Zhejiang University, Hangzhou 310029, China

Purple rice is a type of rice with anthocyanins deposited in its grain pericarp. The rice *Pb* gene controlling purple pericarp character is known to be on chromosome 4, and the purple color is dominant over white color. In this study, we fine mapped the *Pb* gene using two F_2 segregating populations, i.e. Pei'ai 64S (white) × Yunanheixiannuo (purple) and Pei'ai 64S × Chuanheinuo (purple). In the first-pass mapping, the *Pb* gene was located in the region downstream the SSR marker RM3820. In the fine mapping, the candidate region was saturated with InDel and CAPS markers developed specifically for this study. Eventually, the *Pb* gene was mapped within the 25-kb region delimited by the upstream marker RID3 and the downstream marker RID4. The delimited region contained two annotated genes, *Ra* and *bhlh16* (TIGR Rice Genome, R.5). The former is a homologue of the *Myc* transcription factor *Lc* controlling anthocyanin biosynthesis in maize, and the latter is a homologue of the *TT8* gene, which is also an *Myc* transcription factor gene controlling the pericarp pigmentation in *Arabidopsis thaliana*. Sequence analysis showed that the exon 7 of the *Ra* gene of Yunanheixiannuo and Chuanheinuo had a 2-bp (GT) deletion compared with those of the white rice varieties Pei'ai 64S, 9311 and Nipponbare. A CAPS marker, CAPSRa, was developed according to the GT deletion for analysis of the two F_2 segregating populations and 106 rice lines. The results showed that all F_2 plants with white pericarp, and all non-purple rice lines (63 white and 22 red) contained no GT deletion, but all 20 purple rice lines contained the GT deletion. These results suggested that the *Ra* gene may be the *Pb* gene and the purple pericarp characteristic of rice is caused by the GT deletion within exon 7 of the *Ra* gene.

Oryza sativa, purple pericarp, *Pb* gene, fine mapping, *Ra* gene

Most rice (*Oryza sativa*) varieties that are now planted and consumed all over the world are white rice; colored rice, mainly red and purple rice, is only grown in limited areas or as specialty rice. Cyanidin-3-glucoside and peonidin-3-glucoside are the two main pigments deposited in grain pericarp of purple rice[1,2]. Previous genetic investigations have demonstrated that anthocyanin pigmentation in rice involves at least the chromogen gene *C*, activator gene *A*, and tissue-specific regulator gene *Pl* for *C* and *A*[3]. Classical genetic analysis indicated that two loci, *Pb* (*Prp-b*) and *Pp* (*Prp-a*), located on chromosome 4 and 1, respectively[4,5], are required for the pericarp pigmentation with anthocyanins of purple rice. Rice grains appear brown in the presence of the *Pb* gene but without the *Pp* gene, whereas no pigmentation happens in the absence of the *Pb* gene.

When probed with the cDNA of the *Lc* gene that regulates anthocyanin biosynthesis in maize, Hu et al.[6] identified two bHLH homologs, *Ra* and *Rb* in the leaf cDNA library of the purple leaf rice line (Purple 522). The *Ra* and *Rb* genes were mapped on chromosome 4 and chromosome 1, respectively[6]. They further found that the *Ra* locus contains two genes, *Ra1* and *Ra2*[7].

Received July 18, 2007; accepted September 18, 2007
doi: 10.1007/s11434-007-0472-x
†Corresponding author (email: qyshu@zju.edu.cn)

Using a similar approach, Sakamoto et al.[8] probed the cDNA library prepared from purple leaves of T65-Plw using the maize *B-Peru* cDNA, identified the cDNA sequence of the Pl^w locus, which activates the anthocyanin biosynthesis pathway in most of the aerial tissues except the stem node in rice, and demonstrated that the Pl^w locus composed of at least two genes, *OSB1* and *OSB2*. Sakamoto et al.[8] further suggested that the *OSB1* gene was the *Ra* gene in nature, although Hu et al.[6] indicated that the *Ra* gene was not allelic to the *Pl* locus. Zhuang et al.[9] mapped the *Pb* gene in chromosome 4 close to RFLP markers RG214 and RG329 at distances of 26.3 and 18.9 cM, respectively, but no further analysis was reported. With the aim to more precisely locate the *Pb* gene so as to clone this gene and elucidate its function in anthocyanin biosynthesis, we fine-mapped the *Pb* gene using two F_2 populations and one of the two candidate genes was further investigated in this study.

1 Materials and methods

1.1 Plant materials

The two-line male sterile line Pei'ai 64S (white pericarp) was crossed to a *indica* purple rice Yunanheixiannuo and a *japonica* purple rice Chuanheinuo in the spring of 2005, respectively; and the F_2 seeds were produced in the summer of the same year. In the summer of 2006, F_2 plants were grown and F_3 seeds were harvested on individual F_2 plant basis. All experimental materials were transplanted at a single seedling per hill and grown at the university's experimental farms either in Hangzhou, Zhejiang for the summer crop or in Lingshui, Hainan for the winter/spring crop. The second top leaves of F_2 plants were collected at heading stage for DNA extraction. The pericarp color was observed after dehulling of mature and dry seeds.

1.2 DNA extraction

Genomic DNA was extracted using leaf tissues following the protocol of Lu and Zheng[10]. DNA samples were quantified using a Unican UV300 (Thermo Electron Corporation, Cambridge, UK) and adjusted to a final concentration of about 25 ng/uL.

1.3 Molecular makers

Twenty-seven SSR markers evenly distributing on chromosome 4[11] were used for roughly mapping the *Pb* gene. New SSR markers were synthesized according to information from Gramene (www.gramenne.org) for further mapping. For fine mapping the gene, InDel markers were developed according to the rice DNA polymorphism database (http://shenghuan.shnu.edu.cn/genefunction/ricemarker.htm)[12]. InDel markers, which differed at least for 7 bp based on genomic sequence alignment between 9311 and Nipponbare using BioEdit, were amplified for confirmation of polymorphism between Pei'ai 64S and the two purple varieties. Short InDels and SNPs were used to develop CAPS markers by identifying restriction sites between Nipponbare and Pei'ai 64S using Primer Premier 5. Genomic sequences of Pei'ai 64S in candidate region were provided by the Beijing Genomics Institute. Primers for candidate InDel markers and CAPS markers were designed based on the flanking sequences using Primer Premier 5 and were compared with whole genomic sequences (http://tigrblast.tigr.org/euk-blast/index.cgi?project=osa1) to check specification. All primers were synthesized by Shanghai Sangon Biological Engineering Technology & Services Co., Ltd.

PCRs were performed in 20 μL volumes containing approximately 50 ng genomic DNA, 1×PCR buffer, 400 nmol/L each primer, 200 μmol/L each dNTP, 2 mmol/L $MgCl_2$ and 1 U *Taq* enzyme. Amplification program of SSR markers were set according to Temnykh et al.[11]. For InDel markers and CAPS markers, the reaction was performed as follows: denaturing at 94℃ for 5 min, followed by 37 cycles of 94℃ for 45 s, 55℃ for 45 s, 72℃ for 90 s, and with a final extension step at 72℃ for 10 min. For some special markers, the annealing temperature and extension time were adjusted. For CAPS markers, the PCR product was digested with a corresponding restriction enzyme according the manufacturer's instructions.

The PCR products of SSR markers and InDel markers were run either on 3% agarose gels or on 8% polyacrylamide gels, depending on the actual need to differentiate polymorphic bands between the parents of each cross. The PCR products and the digested PCR products of CAPS markers were run on 1.2% agarose gels.

1.4 Preliminary mapping

DNA pools of 10 F_2 plants with white pericarp of each cross together with the three parent lines, i.e. Pei'ai 64S, Yunanheixiannuo and Chuanheinuo, were analyzed for 27 SSR markers that covered whole chromosome 4 to

identify SSR loci that are linked to the *Pb* gene. The SSR markers that appeared potentially linked to the *Pb* gene were further analyzed for all individual F_2 plants with white pericarp in each cross. To obtain SSR markers linked as closely as possible to the gene, new SSR markers in the neighborhood were synthesized and tested for polymorphism between the parent lines, and polymorphic SSR markers were further analyzed for individual plants of the two populations.

1.5 Fine mapping

Based on the result of preliminary mapping, InDel markers and CAPS markers were developed and tested for their polymorphism between the parent lines, and polymorphic markers were used for analysis of white pericarp F_2 plants. A band with the same size as that of purple rice line (Yunanheixiannuo or Chuanheinuo) was marked as '1', while a band showing the same size as that of white rice line (Pei'ai 64S) was labeled as '0'. For F_2 plants with white pericarp, there are three possible genotypes for these markers, namely non-recombinants with '0/0', single recombinants with '0/1', and double recombinants with '1/1'.

1.6 Sequence analysis of the *Ra* gene

The *Ra* gene of the two purple rice parents was sequenced using PCR fragments. PCR primers were designed according to the genome sequence of Nipponbare and 9311 to amplify overlapping fragments of 700 and 1000 bp in length. PCR was carried out using TaqPlus polymerase (Sangon); amplified fragments were separated on 1.2% agarose gels, purified using DNA Gel Extraction Kit (AP-GX-250, Axygen), and sequenced at Shanghai Invitrogen Biotechnology Co. Sequences were assembled using ContigExpress and aligned using public programs ClustalX and BioEdit.

1.7 Analysis of CAPSRa

A CAPS marker within the exon 7 of the *Ra* gene, CAPSRa, was designed according to the genomic sequences, with forward primer 5′-CGTCCATTCACAGGGTA-3′, and reverse primer 5′-CAGCAGATGAGGCAAACT-3′. The length of CAPSRa fragment before digestion was about 858 bp; the 2-bp (GT) deletion generated a *Bam*H I restriction site, resulting in two digested products of 653 and 203 bp.

CAPSRa analysis was performed for Pei'ai 64S, Yunanheixiannuo and Chuanheinuo, and all F_2 plants with white pericarp of each cross, and 106 rice germplasm materials including 63 lines with white pericarp, 20 lines with purple pericarp and 23 lines with red pericarp.

2 Results and analysis

2.1 Preliminary mapping

Among the 27 SSR markers of chromosome 4, there were 6 and 9 SSR markers, respectively, showing polymorphism between Pei'ai 64S and Yunanheixiannuo, and Pei'ai 64S and Chuanheinuo. Analysis of 192 white pericarp F_2 plants of Pei'ai 64S × Yunanheixiannuo showed that the *Pb* gene was linked to RM252 at a genetic distance of 6.51 cM. Similarly, the *Pb* gene also appeared to be linked to RM252 in the F_2 population (n = 190) of Pei'ai 64S × Chuanheinuo, at a genetic distance 8.68 cM.

Because there was no information about the physical location of RM252 on chromosome 4, we blasted Gramene (http://filetta.cshl.org/multi/blastview) using the sequence of RM252 (http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=AF344072), and identified the physical position of RM252 to be 25144688－25145368 bp (TIGR, psuedomolecule v4) on Chromosome 4 in the Gramene.

New SSR markers were synthesized according to information available in the Gramene database. Although none of them is polymorphic between Pei'ai 64S and Yunanheixiannuo, four markers downstream RM252, i.e. RM17231, RM1354, RM3288 and RM3820, were polymorphic between Pei'ai 64S and Chuanheinuo, and linked with the *Pb* gene at genetic distances of 7.11, 4.74, 1.58, 0.79 cM, respectively. Based on these results, the *Pb* gene was rough restricted to the downstream region of RM3820 (TIGR, pseudomolecule v4: 27589809－27589903).

2.2 Development of InDel markers and CAPS markers

Attempts were made to find more polymorphic SSR markers in the downstream region of RM3820. However, it ended without success, because there are limited number of SSR markers available in this region, and none of which is polymorphic between Pei'ai 64S and two purple rice lines. Therefore, we opted to develop InDel and CAPS markers to narrow down the *Pb* locus. A total of 22 InDel markers and five CAPS markers were developed (Table 1).

Table 1 The InDel and CAPS markers used for fine mapping of the *Pb* gene

Marker	Position[a)]	Sequence of primer (5′→3′)		Size[b)] (bp)		
		Forward	Reverse	Yunanheixiannuo	Chuanheinuo	Pei'ai 64S
RID1	27623211	GTGGCTCTTTCCGTTTC	TGTGCTCCTCATACAAAC	305	305	285
RID2	27654641	CATCCCAAGATCATATTACA	TCTCCCAGTCCCAGGTC	271	271	254
RID3	27712299	GCTCAAGCACGCCACTC	CGTCCCTTATGCCTCCA	360	360	403
RID4	27737750	ATTGAAACGGAGGGAGT	CTGTACGTGATGGCAAA	180	180	158
RID5	27761483	GTGTATTTAAGGTTTTCGAA	TTGTTCATCTGTAGCTTTGT	153	153	145
RID6	27764264	GGCTGAGTCAAAGAAGGG	TTTTCTACGTGGCGACC	143	143	136
RID7	27772442	CAGCATCTGACACTGGA	CTACTGCAACCTAATCTAA	160	160	171
RID8	27787666	CCGCCAAATACGTTCTG	TTCCTCCCGTGCTTCTG	131	131	117
RID9	27793566	TCGGGCAACTCACCTCC	GCGGGCCTCCTCCTTTA	179	179	170
RID10	27821277	ATGGGCTTGATCGGTAG	TTGGGTTATTTGAGTGCTAA	198	198	210
RID11	27827101	CACATCTCGCTAAGGAAA	CCTATGCAATTCGGTCA	299	299	279
RID12	27851717	CCGCTGAGCTTGAAAGT	CGGAGGGAGTAGAGTAATG	290	290	266
RID13	27879300	GAGATAGGCTCACTTCG	ACATACATACGTGTATAAACT	295	295	279
RID14	27883558	CAAGTTACATAGGGCTCG	GGTGGAGGAGTTCATTCT	488	488	424
RID15	27932108	GAGTGAACAGAAAGGGAC	CAACTACAAGCGTACAAAC	280	280	302
RID16	27947572	CAGGGTTGCGTCGAGAT	TTTCAAAGGTGCTTATAGAG	393	393	435
RID17	28020277	GTTCAAGACCTTCCCATCA	CCTACAAGAGCCCAGCA	283	283	256
CAPS1 (*Sma* I)	27722111	CGCATCGTCAATTTCGT	GGCACAAGGTAGGCACT	399	399	151/248
CAPS2 (*Xba* I)	27728667	AGCATTATCCCTGAGTGT	GACTGTAGAGGGCAAGAC	840	840	562/278

a) The physical position of the first base of InDels or SNPs on chromosome 4 (TIGR, psuedomolecule v4); b) The band size was concluded on the basis of electrophoretic band and genomic sequence of 9311 and Nipponbare.

Among the 22 InDel markers, seventeen were polymorphic between Pei'ai 64S and the two purple lines, Yunanheixiannuo and Chuanheinuo (Table 1). Notably, electrophoresis experiments revealed that the latter two had completely the same genotype for these polymorphic InDel markers (Table 1). Furthermore, we developed five CAPS markers in the genome region 27.72—27.73 Mb of chromosome 4. However, only two markers, CAPS1 and CAPS2 were proved to be polymorphic between Pei'ai 64S and two purple lines, Yunanheixiannuo and Chuanheinuo. The PCR products of Pei'ai 64S could be digested by corresponding restricted enzymes, while those of Yunanheixiannuo and Chuanheinuo could not (Table 1).

2.3 Fine mapping

The polymorpic 17 InDel and 2 CAPS markers were used for fine mapping of the *Pb* gene. We first genotyped white pericarp F_2 plants of both crosses for the RID16 and RID17 markers. The two loci were at the furthest end of chromosome 4 among 19 polymorphic markers. Among the 192 white percarp F_2 plants of Pei'ai 64S × Yunanheixiannuo, four plants had a recombinant genotype for both RID16 and RID17 loci, while among the 190 F_2 plants of Pei'ai 64S × Chuanheinuo, seven plants were found to be recombinant for the RID17 locus and six for the RID16 locus. Accordingly, we narrowed down the *Pb* gene to the region between RM3820 and RID 16 (Figure 1(a)).

Subsequently, we analyzed the genotypes of white pericarp F_2 plants of both crosses for the other 15 InDel loci between RM3820 and RID16. Among the 192 F_2 plants of Pei'ai 64S × Yunanheixiannuo, one plant (Y628) was recombinant on the RID1 locus, and 4 plants (Y214, Y508, Y586, Y712) on the RID15 locus (Table 2). Step by step, the plant Y508 was the last recombinant for the downstream locus RID10, but no recombinant loci were found to be closer than RID1 in upstream. Among the 190 F_2 plants of Pei'ai 64S × Chuanheinuo, two plants (C104, C176) were recombinant on the RID1 locus and six plants (C217, C271, C325, C513, C830, C913) on the RID15 locus. Closing to center, there was still one plant (C176) appeared to be recombinant on locus RID3 in the upstream and one plant (C325) on locus RID4 in the downstream (Table 2). Eventually, the *Pb* gene was delimited within a 25 kb region (TIGR, pseudomolecule v4: chr4: 27712299—27737750) by the upstream marker RID3 and the downstream marker RID4 (Figure 1(b), Table2).

According to the TIGR Rice Genome, R.5, there are two annotated genes within the about 25 kb region. Both are Myc transcription factor genes, one is the *bhlh16* gene

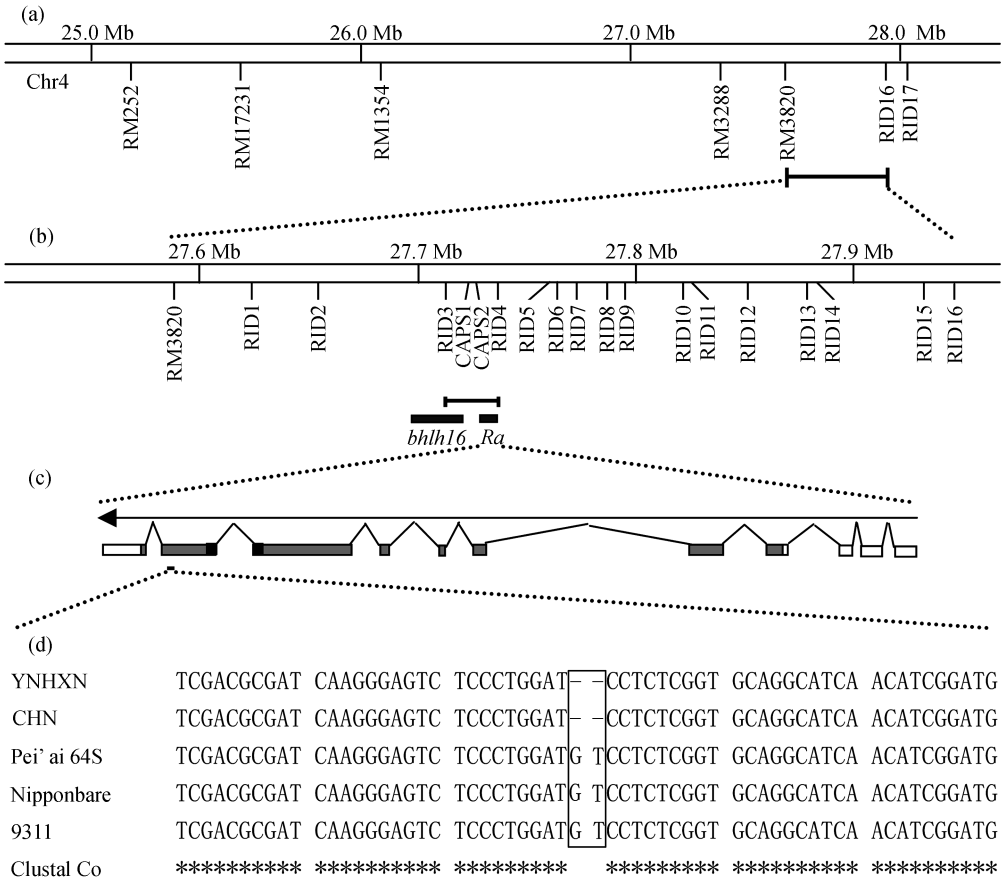

Figure 1 Fine mapping of the *Pb* gene and sequence analysis of candidate gene *Ra*. (a) The *Pb* gene was restricted to the region between markers RM3820 and RID16; (b) the *Pb* gene was further narrowed down to the region between markers RID3 and RID4; (c) the gene model of candidate gene *Ra*, exons were showed in grey boxes, and the bHLH domain was indicated in black boxes at the end of exon 6 and the beginning of exon 7; (d) sequence analysis of the *Ra* gene. The purple rice lines Yunanheixiannuo (YNHXN) and Chuanheinuo (CHN) contained 2 bp (GT) deletion at the end of exon 7 compared with white rice lines Pei'ai 64S, Nipponbare and 9311.

Table 2 The InDel and CAPS genotype of white pericarp F_2 plants of two populations used for fine mapping of the *Pb* gene[a)]

Marker	Pei'ai 64S × Yunanheixiannuo[b)]					Pei'ai 64S × Chuanheinuo[b)]							
	Y628	Y214	Y508	Y586	Y712	C104	C176	C217	C271	C325	C513	C830	C913
RID1	0/1	0/0	0/0	0/0	0/0	0/1	0/1	0/0	0/0	0/0	0/0	0/0	0/0
RID2	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0	0/0
RID3	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0	0/0
CAPS1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPS2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
RID4	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/0	0/0
RID5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/0	0/0
RID6	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/0	0/0
RID7	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/1	0/0	0/0	0/0
RID8	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/1	0/1	0/0	0/0
RID9	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/1	0/1	0/0	0/0
RID10	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/1	0/0	0/1	0/1	0/1	0/0
RID11	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/1	0/0	0/1	0/1	0/1	0/1
RID12	0/0	0/1	0/1	0/1	0/0	0/0	0/0	0/1	0/1	0/1	0/1	0/1	0/1
RID13	0/0	0/1	0/1	0/1	0/1	0/0	0/0	0/1	0/1	0/1	0/1	0/1	0/1
RID14	0/0	0/1	0/1	0/1	0/1	0/0	0/0	0/1	0/1	0/1	0/1	0/1	0/1
RID15	0/0	0/1	0/1	0/1	0/1	0/0	0/0	0/1	0/1	0/1	0/1	0/1	0/1

a) 0, The banding pattern was the same as that of the white rice line Pei'ai 64S; 1, the banding pattern was the same as that of purple rice line Yunanheixiannuo or Chuanheinuo. b) The total number of white pericarp F_2 plants genotyped was 192 and 190, respectively, for Pei'ai 64S × Yunanheixiannuo and Pei'ai 64S × Chuanheinuo.

(LOC_Os04g47059, 27696505 − 27720731), and the other is the *Ra* gene (LOC_Os04g47080, 27730292-27736548). The *bhlh16* gene putatively encodes the protein homologue of *TRANSPARENT TESTA 8* (*TT8*), which regulates biosynthesis of flavonoid in the pericarp of Arabidopsis[13]. The *Ra* gene encodes a protein homologue of *Lc*, which is involved in the biosynthesis of anthocyanin in maize[14]. Attempts were made to exclude one of the two candidate genes through analysis of CAPS markers CAPS1 and CAPS2, but resulted in no success because no recombinants were identified in the two F_2 populations.

2.4 Sequence analysis of Ra

We tried to clone and sequence both candidate genes to identify the causative DNA lesion(s) leading to mutation of the *Pb* gene. However, there are dozen regions across the rice genome homologous to the *bhlh16* gene, no suitable PCR primers could be designed to specifically amply this gene. In addition, the *bhlh16* gene is involved in proanthocyanidin synthesis while the *Ra* gene in anthocynanin synthesis, we assumed that the *Ra* gene is more likely to be the candidate gene. Therefore, the *Ra* gene of Yunanheixiannuo and Chuanheinuo was sequenced (GenBank accession number: Yunanheixiannuo, EU095985; Chuanheinuo, EU095986) and compared with those of white rice lines, Pei'ai 64S, 9311 and Nipponbare. Although numerous variations exist among those five varieties (data not shown), only the 2 bp (GT) insert/deletion within the exon 7 coincided the white/purple difference. The GT was absent in both purple rice lines (Yunanheixiannuo and Chuanheinuo) but present in all three white rice lines (Pei'ai 64S, 9311 and Nipponbare) (Figure 1(c), (d)).

2.5 Analysis of marker CAPSRa

The GT deletion resulted in a *Bam*H I restriction site, accordingly, a CAPS marker CAPSRa was designed to confirm whether the GT indel correlates with the pericarp color difference. The PCR products of all materials were about 858 bp in length, but those of purple rice, Yunanheixiannuo and Chuanheinuo, could be digested by *Bam*H I into two fragments (about 653 and 203 bp), while that of Pei'ai 64S couldn't (Figure 2(a)). In each cross, the PCR products of homologous or heterozygous F_2 plants with purple pericarp could be digested, while that of F_2 plants with white pericarp could not (Figure 2(b)).

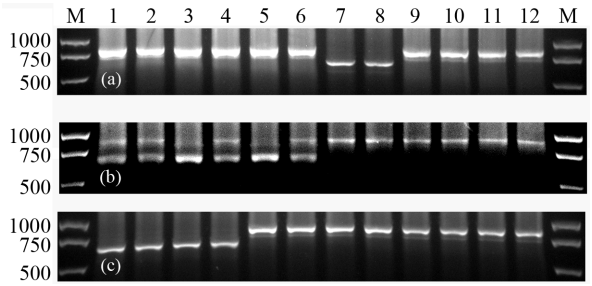

Figure 2 Analysis of the CAPSRa marker. M, DNA ladder. (a) Lanes 1 and 7, Yunanheixiannuo; 2 and 8, Chuanheinuo; 3 and 9, Pei'ai 64S; 4 and 10, Nipponbare; 5 and 11, Zhoushanhongmi (red rice); 6 and 12, Honghuami (red rice); 1−6 were PCR products, and 7−12 were digested products by *Bam*H I. (b) *Bam*H I digested PCR products of F_2 plants. Lanes 1−6, homologous or heterozygous plants with purple pericarp; lanes 7−12, plants with white pericarp. (c) *Bam*H I digested PCR products. Lanes 1−4, purple rice lines; lanes 5−8, white rice lines; lanes 9−12, red rice lines.

In addition, 106 rice lines with different pericarp color were also analyzed for CAPSRa. The PCR products of all materials were about 858 bp, and the PCR products of all 20 purple rice lines could be digested resulting in two fragments with about 653 and 203 bp, while that of 63 white lines and 23 red rice lines could not be digested (Figure 2(c)).

3 Discussion

Although extensive researches on anthocyanin biosynthesis in other plants were reported in the past years, little is known for the molecular genetic mechanism of pigmentation in grain pericarp of purple rice. In this study, we delimited the *Pb* gene to a 25 kb region, where only two genes were annotated; we further found that the *Pb* gene is likely to be the same gene as the *Ra1* gene (a homologue of the maize *Lc* gene)[7] and the *OSB1* gene which is located in the Pl^w locus[8]. Moreover, the purple characteristic of rice pericarp maybe results from a GT deletion within exon 7 of the *Ra* gene. Therefore, our findings are of importance to the understanding of anthocyanin biosynthesis in plants, and especially of the molecular and genetic mechanism of pericarp pigmentation in rice.

With the completion of genomic sequencing of the *indic*a rice 9311 and the *japonica* rice Nipponbare, a great number of InDel and SNP markers can now be developed through sequence comparison and used for mapping and eventually cloning genes[12,15,16]. Zhuang et al.[9] mapped the *Pb* gene controlling pericarp pigmenta-

tion in the purple rice 'Heizhenmi' to a position at a distance of 18.9 cM to the RFLP marker RG329. RG329 is now known to be at the physical position of 30819286—30819457 bp on chromosome 4. If 1 cM equals to about 244 kb of physical distance on average[17], the *Pb* gene should be roughly located around 26.21 Mb in the genome of 'Heizhenmi'. There are two possibilities that could lead to the about 1.5 Mb difference compared with our result. Firstly, 'Heizhenmi' resulted from a somaclonal variation[18], while two purple lines used in this study were conventional rice lines, so the genes controlling pigmentation in grain could be different and not allelic. Secondly, the difference resulted from the imprecision of mapping together with different ratio of physical to genetic distance due to hot and cold recombination spots in different genome region. For example, there was also a 2 cM difference between the distances of the *Pb* gene to RM252 in two populations according to the preliminary mapping in this study. Hence, the inconsistency of mapping results between our study and Zhuang et al.[9] not necessarily indicated the genes controlling pericarp color are different between 'Heizhenmi' and the purple rice lines used in our study.

Extensive studies on anthocyanin metabolism have been performed on maize, pentunia and snapdragon. Three types of regulatory gene, Myb, Myc and WD40, are involved in regulation of anthocyanin biosynthesis in plants. The Myc proteins contain a basic helix-loop-helix (bHLH) domain, which interacts with Myb protein in anthocyanin biosynthesis[19,20]. With the completion of genome sequence, once a gene is cloned in one plant species, comparative genomics tools or other techniques can now be used to identify their homologue genes in other plant species. In rice, Hu et al.[6,7] identified three rice genes, *Ra1*, *Ra2* and *Rb*, which are homologous to the *Lc* gene in maize and involved in rice anthocyanin biosynthesis using maize *Lc* cDNA; Sakamoto et al.[8] also found two rice genes, *OSB1* and *OSB2* homologous to the maize *B-Peru* gene. Both *Lc* and *B-Peru* are members of the *R*/*B* gene family and belong to Myc type of transcription factors in maize. In the meantime, Saitoh et al.[21] isolated the rice gene homologous to maize *C1* gene, *OsC1*, which belongs to Myb transcription factor. In this study, we delimited the *Pb* gene in a 25 kb region, in which two annotated Myc family genes, *Ra* and *bhlh16*, are located. The *Ra* gene is known to be involved in anthocyanin biosynthesis in maize and can activate the biosynthesis of anthocyanin in rice[6,8]. The homologoue gene of *bhlh16* in *Arabidopsis*, *TT8*, is involved in the biosynthesis of proanthocyanidins through cell specific accumulation of flavonoids[22]. Therefore, based on the general nature of the two genes, none of them could be excluded as a candidate of the *Pb* gene.

The *Pl* locus is one of the six factors that regulate pigmentation in rice leave issues; plants with the Pl^w allele usually have pigmented aerial tissues except internodes and produce seeds with purple pericarp[8]. The two purple rice lines used in this study had purple pericarp but green leaves, hence they are different from the purple rice T65-Plw with purple leaf and purple pericarp used in the study of Sakamoto et al.[8]. Because one of the two candidate genes of the *Pb* gene (the *Ra* gene) is identical to one of the two genes at the Pl^w locus, the *OSB1* gene, we inclined to assume that the *Pb*, *Ra* and *OSB1* genes are likely to be indeed identical. Hence, we sequenced the *Ra* gene and found a 2-bp (GT) difference between white rice and purple rice. We further developed a CAPS marker CAPSRa based on this 2 bp deletion. It is reasonable that the CAPSRa co-segregated with pericarp color in the two F_2 populations, because the maker is within the delimited region, hence no further information could be obtained from the results; however, the complete correlation of the 2 bp insertion/deletion with the non-purple/purple pericarp color in more than 100 varieties strongly supported that the *Ra* gene is more likely than the *bhlh16* gene to be the gene controlling purple pericarp characteristic in rice.

The purple rice lines Yunanheixiannuo and Chuanheinuo were provided by Dr. J S Bao and Dr. D X Wu (Institute of Nuclear Agricultural Sciences, Zhejiang University), respectively. In CAPSRa analysis, the white rice materials were provided by Dr. S Q Shen, and the red rice and purple rice materials were supplied by National Rice Germplasm Bank, Chinese National Rice Research Institute. The sequence of Pei'ai 64S used in this study was provided by Dr. G Z Liu of Beijing Genomics Institute. We thank all of them for their kindly help.

1 Fossen T, Slimestad R, Ovstedal D O, et al. Anthocyanins of grasses. Biochem Syst Ecol, 2002, 30: 855—864

2 Abdel-aal E S M, Young J C, Rabalski I. Anthocyanin compositon in black, blue, pink, purple, and red cereal grains. J Agric Food Chem, 2006, 54: 4696—4704

3 Nagao S, Takahashi M. Genetical studies on rice plants. XXVII. Trial construction of twelve linkage groups in Japanese rice. J Fac Agri Hokkaido Univ, 1963, 53: 72—130

4 Causse M A, Fulton T M, Cho Y G, et al. Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics, 1994, 138: 1251－1274
5 Yoshimura A, Ideta O, Iwata N. Linkage map of phenotype and RFLP markers in rice. Plant Mol Biol, 1997, 35: 49-60
6 Hu J P, Anderson B, Wessler S R. Isolation and characterization of rice *R* genes: Evidence for distinct evolutionary paths in rice and maize. Genetics, 1996, 142: 1021－1031
7 Hu J P, Reddy V S, Wessler S R. The rice *R* gene family: Two distinct subfamilies containing several miniature inverted-repeat transposable elements. Plant Mol Biol, 2000, 42: 667－678
8 Sakamoto W, Ohmori T, Kageyama K et al. The purple leaf (*Pl*) locus of rice: Pl^{w} allele has a complex organization and includes two genes encoding basic helix-loop-helix proteins involved in anthocyanin biosynthesis. Plant Cell Physiol, 2001, 42(9): 982－991
9 Zhuang J Y, Yang C D, Qian H R, et al. Linkage analysis of RFLP markers and the gene for purple pericarp of rice (in chinese). Acta Genetica Sinica, 1996, 23(5): 372－375
10 Lu Y J, Zheng K L. A convenient method for isolating genomic DNA from rice (in chinese). Chin J Rice Sci, 1992, 6: 47－48
11 Temnykh S, DeClerck G, Lukashova A, et al. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Res, 2001, 11: 1441－1452
12 Shen Y J, Jiang H, Jin J P, et al. Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. Plant Physiol, 2004, 135: 1198－1205
13 Nesi N, Debeaujon I, Jond C, et al. The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* Siliques. Plant Cell, 2000, 12: 1863－1878
14 Ludwig S R, Habera L F, Dellaporta S L, et al. *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. Proc Natl Acad Sci USA, 1989, 86: 7092－7096
15 Ji Q, Lu J F, Chao Q et al. Delimiting a rice wide-compatibility gene s5n to a 50 kb region. Theor Appl Genet, 2005, 111: 1495－1503
16 Chen S H, Wu J, Yang Y et al. The *fgr* gene responsible for rice fragrance was restricted within 69 kb. Plant Sci, 2006, 171: 505－514
17 Chen M S, Presting G, Barkazuk W B et al. An integrated physical and genetic map of the rice genome. Plant Cell, 2002, 14: 537－545
18 Zhao C Z, Qi X F, Yang C D et al. Study of 'Black superior rice' selected by the use of somaclonal variation (in chinese). J Agri Biotech, 1993, 1(1): 100－105
19 Chandler V L, Radicella J P, Robbins T P, et al. Two regulatory genes of maize anthocyanin pathway are homologous: Isolation of *B* utilizing *R* genomic sequences. Plant Cell, 1989, 1: 1175－1183
20 Ludwig S R, Wessler S R. Maize *R* gene family: Tissue-specific helix-loop-helix protein. Cell, 1990, 62: 849－851
21 Saitoh K, Onishi K, Mikami I, et al. Allelic diversification at the *C* (OsC1) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. Genetics, 2004, 168: 997－1007
22 Baudry A, Caboche M, Lepiniec L. TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and *bHLHL* factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*. Plant J, 2006, 46: 768－779