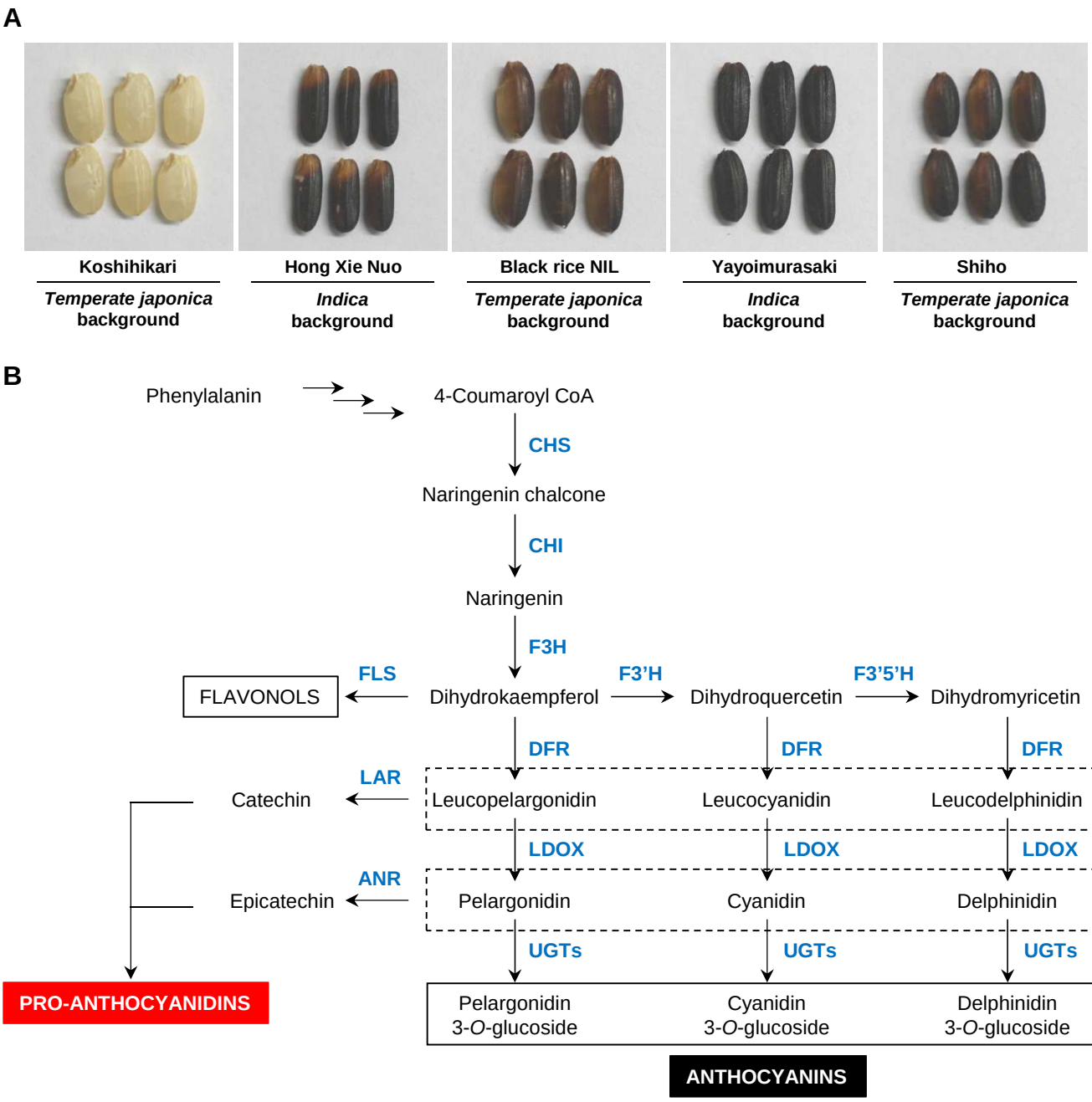
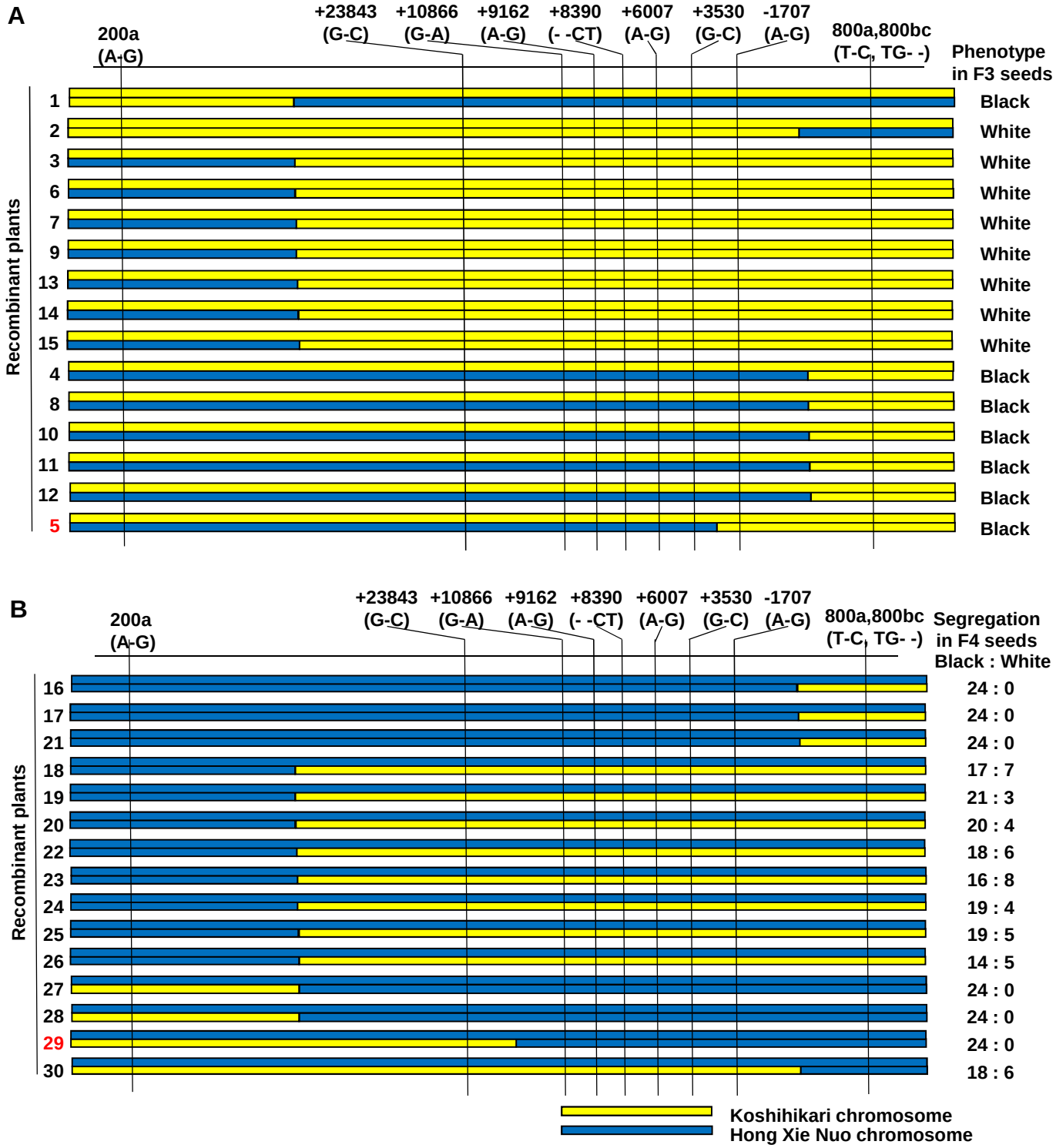


Supplemental Figure 1. Black rice cultivars and anthocyanin biosynthesis.

(A) Grain color of white and black rice cultivars. Photos show grains of one white and four black rice varieties.

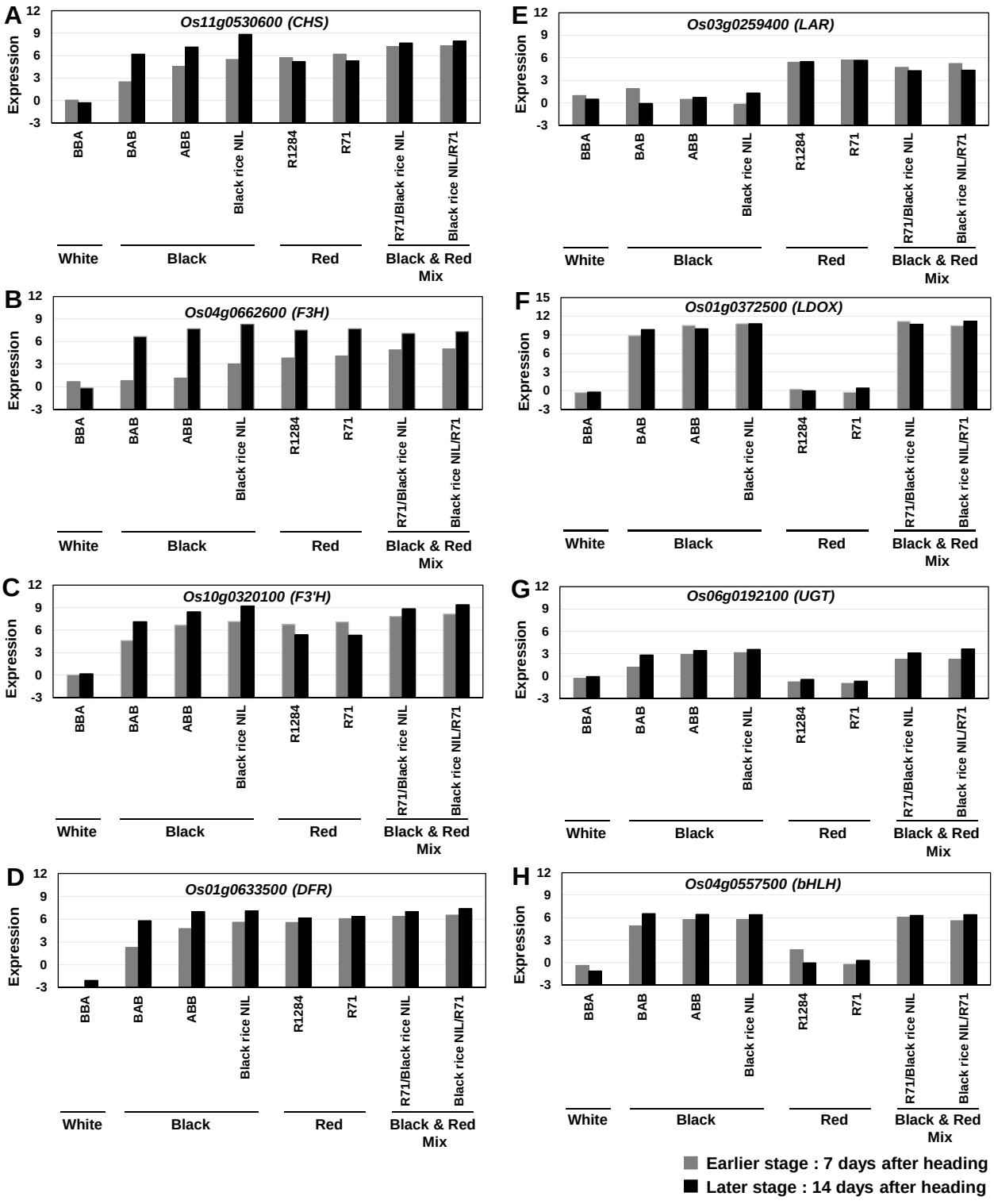
(B) Schematic diagram of the anthocyanin and pro-anthocyanidin biosynthesis pathway in plants. Blue characters indicate biosynthetic enzymes of each steps of the pathway. Enzyme names are abbreviated as follows: CHS, chalcone synthase; CHI, chalcone isomerase; F3H, Flavanone 3- hydroxylase; F3'H, Flavanone 3'-hydroxylase; F3'5'H, Flavanone 3'5'-hydroxylase; FLS, Flavonol synthase; DFR, Dihydroflavonol reductase; LAR, leucoanthocyanidin reductase; LDOX, leucoanthocyanidin dioxygenase; ANR, anthocyanidin reductase; UGTs, UDP-glucosyl transferases. This figure was redrawn from data described by Petroni and Tonelli (2011).



Supplemental Figure 2. Genotypes and phenotypes of thirty recombinant plants obtained by fine mapping.

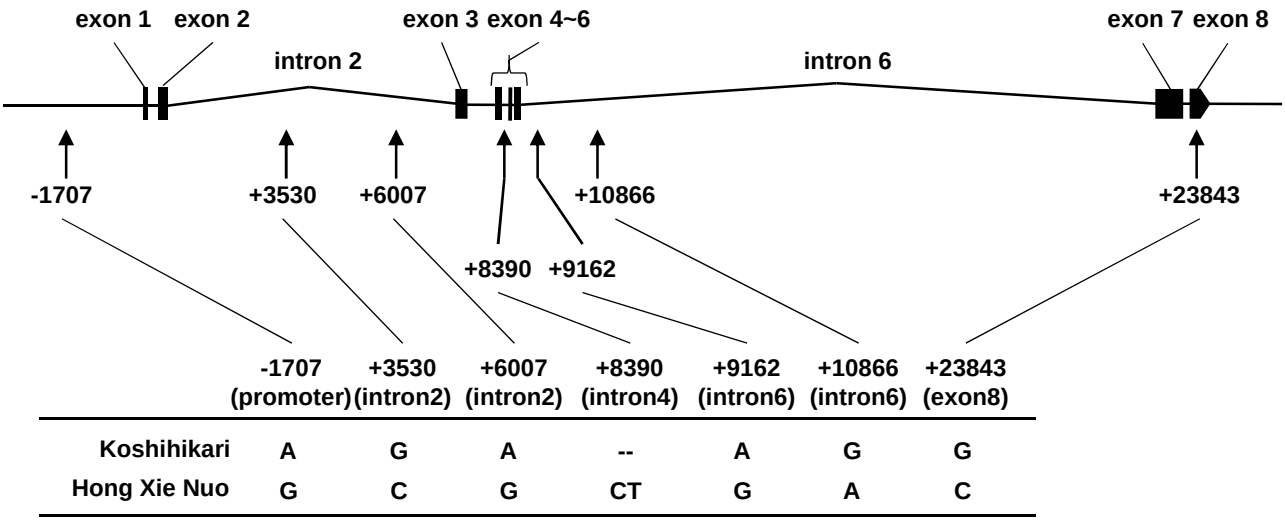
(A) These 15 recombinant plants are homozygous for the 'Koshihikari' allele in either 200a or 800a and 800bc markers. These plants were checked for grain color in F3 seeds. Plant No. 5 showed recombination between two sequence markers +3530 and -1707, and the grain color of F3 seeds were black. Thus, the candidate region was narrowed to downstream of the -1707 marker.

(B) These 15 plants showed the 'Hong Xie Nuo' homozygous genotypes in either "200a" or "800a and 800bc" markers. These plants were checked for the segregation ratio of the grain color in F4 seeds. Plant No. 29 showed recombination between two sequence markers +23843 and +10866, and the segregation ratio of grain color was 24:0 (black:white). Thus, the candidate region was narrowed to the region upstream of the +23843 marker.

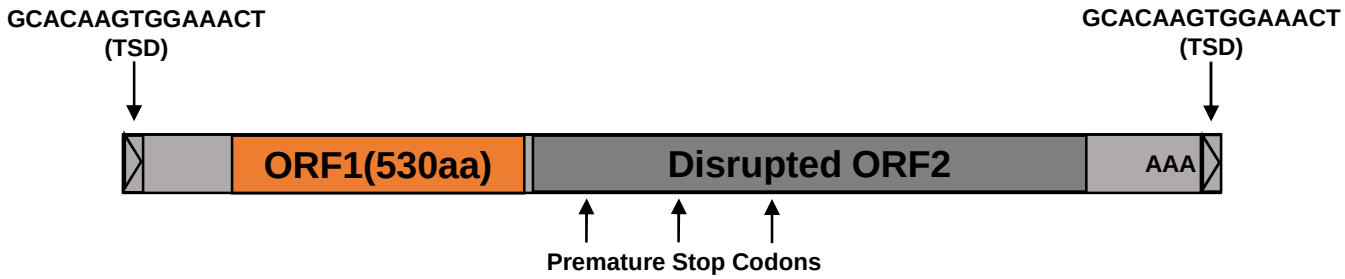


Supplemental Figure 3. Expression of genes related to anthocyanin and pro-anthocyanidin biosynthesis in the pericarps of white, black, and red rice lines.

Microarray data were subjected to 75 percentile normalization. The y axis presents the log2 ratio of test (black or red rice NILs) versus control (Koshihikari). The rice 4x44 K array (Agilent) was used in this analysis. Total RNAs were extracted from pericarps of the white, black, and red rice. BBA, BAB and ABB indicate genotypes of the NILs for *Kala1* (on chr.1), *Kala3* (on chr.3) and *Kala4* (on chr.4) loci in order from left to right. Genotype B and A are homozygous for the 'Hong Xie Nuo' allele and 'Koshihikari' allele, respectively. R1284 and R71 NILs are red rice that possess the functional *Rc* allele. R71/Black rice NIL and Black rice NIL/R71 indicate black and red mixed lines (F1 plants). Grey and black bars in graphs indicate expression levels in the pericarps at an earlier stage (7 d after heading) and later stage (14 d after heading), respectively. Probe names for all investigated genes are listed in Supplemental Table 7.

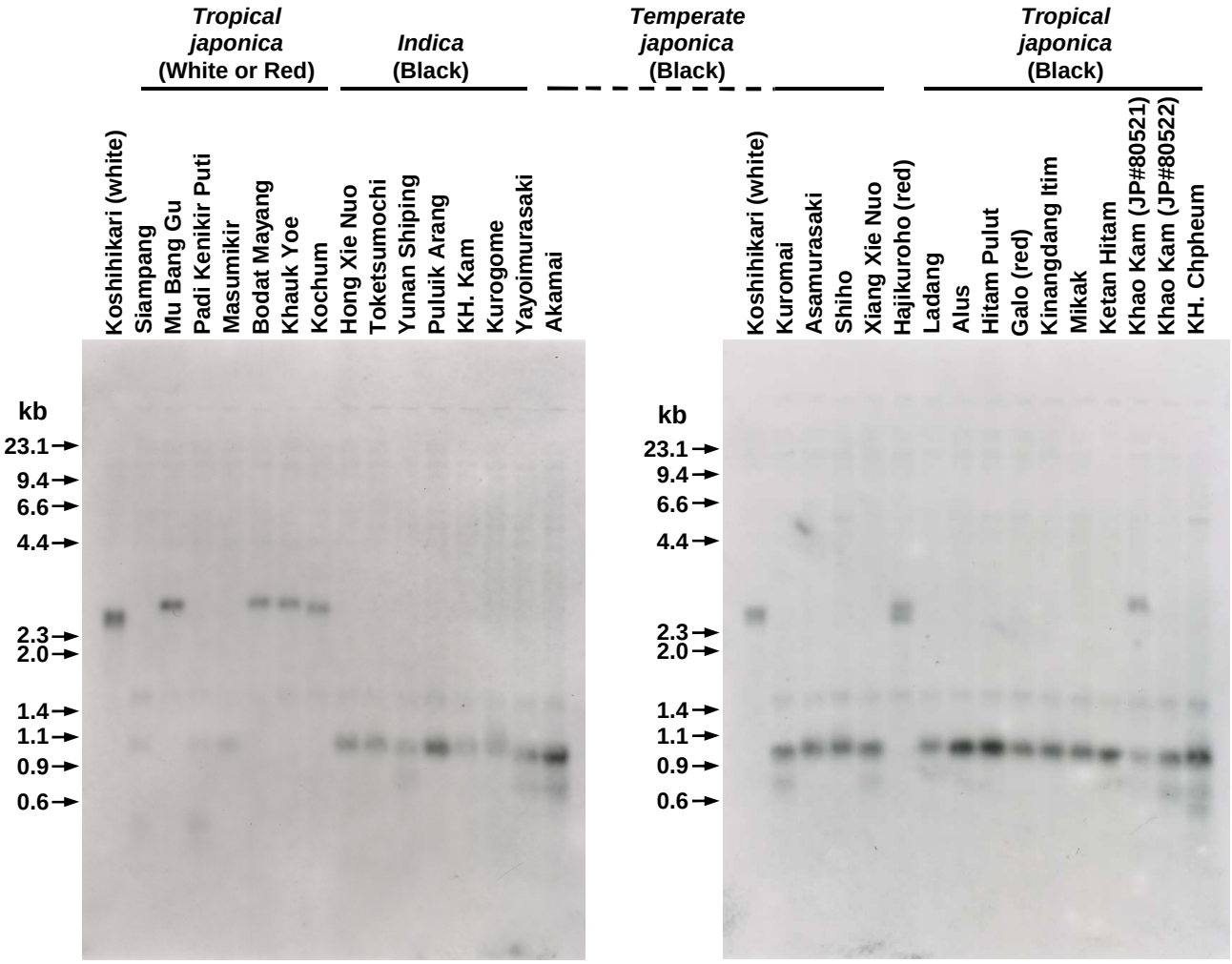


Supplemental Figure 4. Exon-intron structure of *Os04g0557500*.
The *Os04g0557500* gene is composed of 8 exons and 7 introns. Six SNPs and one in/del were observed in the approximately 25.6-kb candidate region. Black boxes and lines indicate exons and introns, respectively.



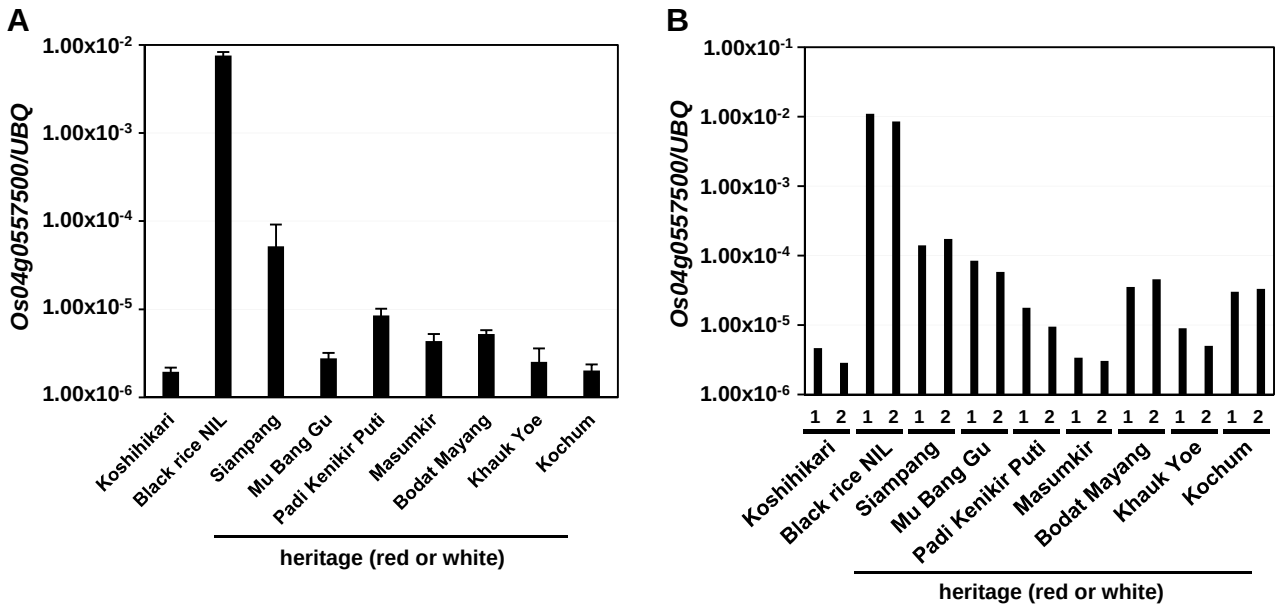
Supplemental Figure 5. Structure of the LINE1 element inserted into intron 2 of *Os04g0557500*.

The LINE1 element inserted into intron 2 of *Os04g0557500* has 15 bp of target site duplication (TSD) and an intact ORF1 sequence (orange box). ORF2 (grey box) is disrupted by nucleic acid changes causing premature stop codons.

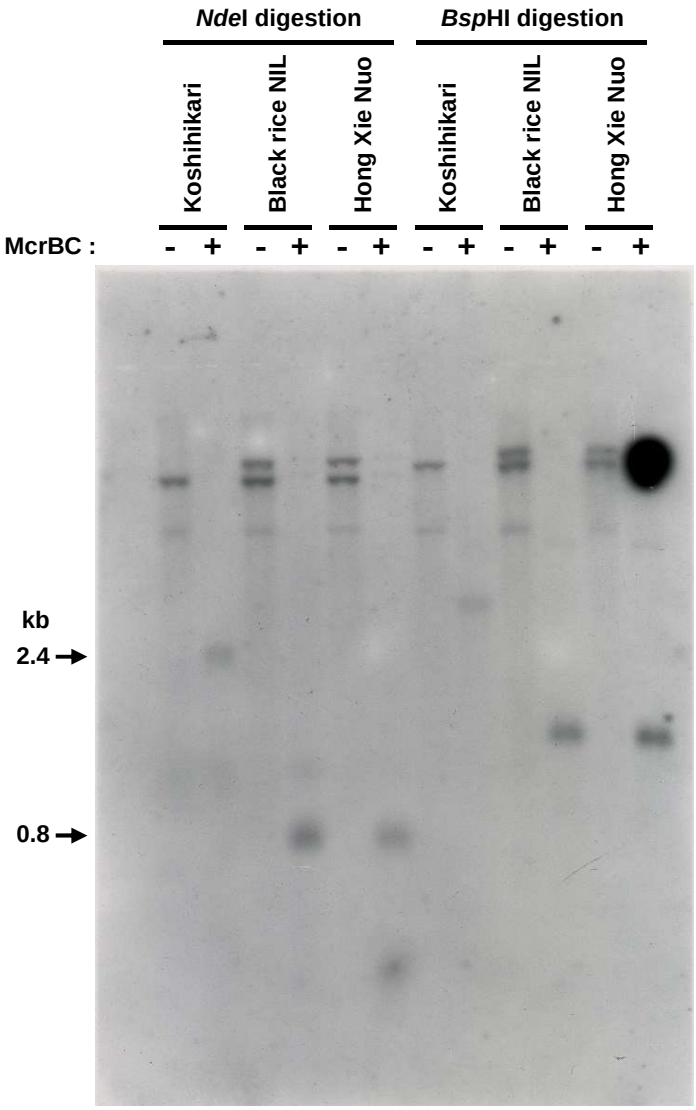


Supplemental Figure 6. Analysis of DNA methylation status of the twenty-one black rice varieties and seven heritage landraces by McrBC-DNA gel blot.

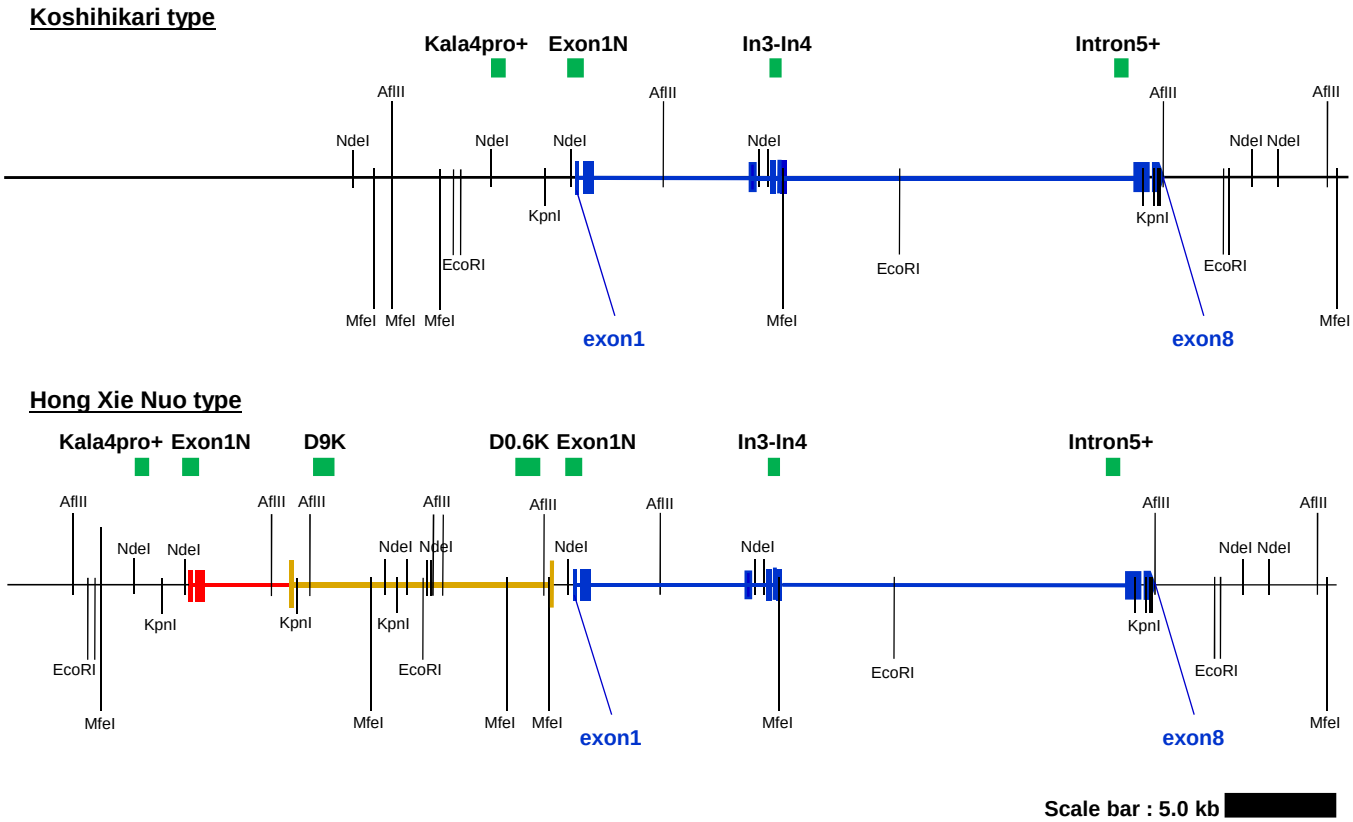
The genomic DNAs were digested with *Nde*I followed by McrBC treatment. To detect the DNA methylation of intron 2, the blotted membranes were hybridized with the exon1N probe. In the case of the 'Koshihikari'-type DNA methylation status of intron 2, the band was approximately 2.4 kb, whereas in 'Hong Xie Nuo', an approximately 0.8-kb band was detected. 'Khao Kam' (JP#80521) showed both sizes of band due to crossing with other white rice varieties in the paddy field. Hajikuroho was originally thought of as a black rice landrace, but was later found to be a red rice line.



Supplemental Figure 7. Expression of *Os04g0557500* in seven heritage landraces.
(A) The amount of transcript in the leaves. Total RNAs were extracted from mature leaves. Three independent total RNAs extractions and RT-qPCRs were performed. Average value and s.e. are shown.
(B) The amount of the transcript in the pericarps. Two bulk samples of pericarps consisting of approximately 20 seeds at 14 d after heading were subjected to RNA extraction. ‘Koshihikari’ and ‘Black rice NIL’ served as controls.

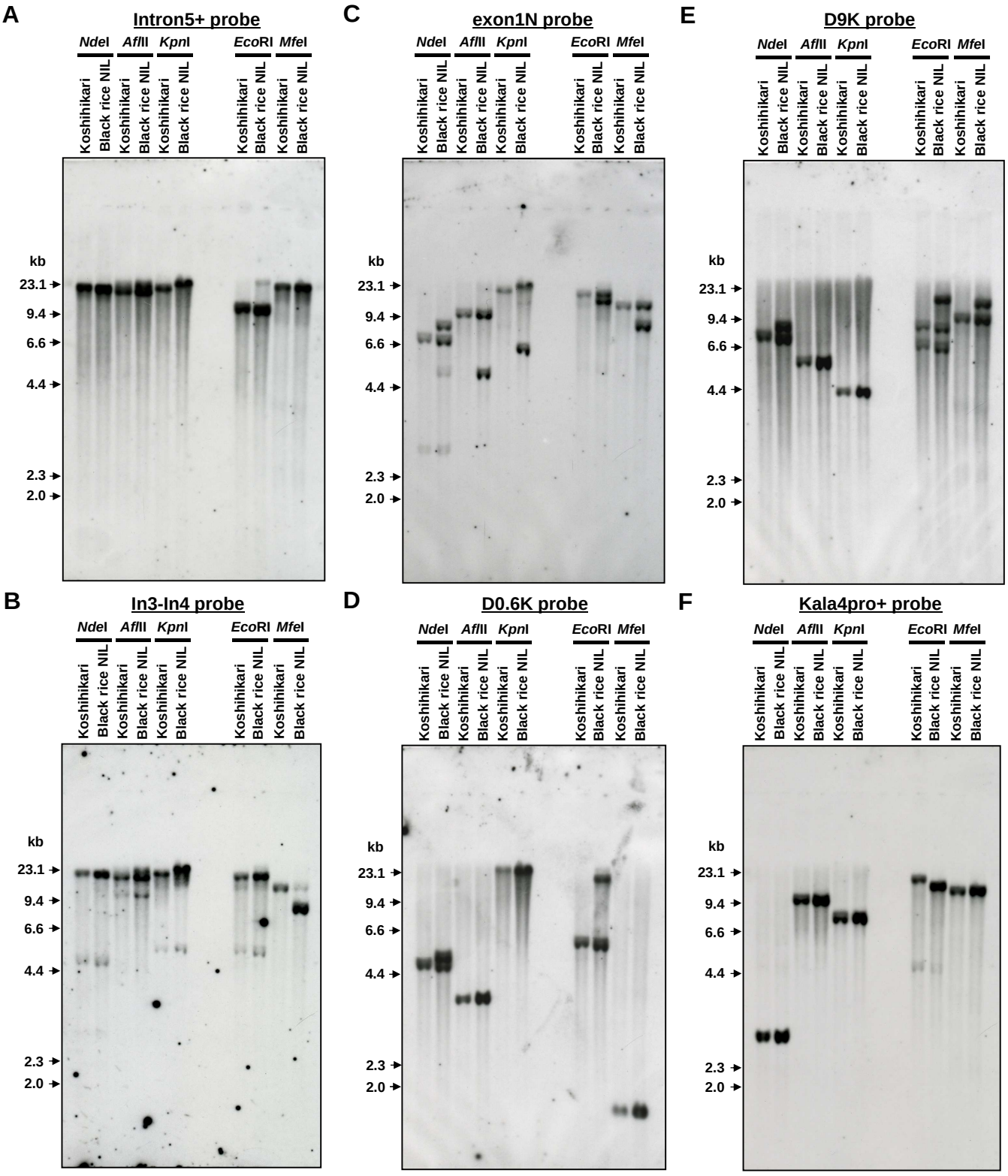


Supplemental Figure 8. Discovery of duplicated fragment related to Os04g0557500 ORF by McrBC-DNA gel blot analysis. Genomic DNAs were digested with *Nde*I or *Bsp*HI restriction enzymes, and after phenol/chloroform extraction and ethanol precipitation, these digested DNAs were further treated with McrBC-containing (+) or -lacking (-) solution. Electrophoresis was performed with 0.8 % agarose gels, and then transferred to positively charged nylon membrane. Hybridization was performed using the exon1N DNA probe. In the case of no treatment with McrBC, only single bands were detected in 'Koshihikari' genomic DNA. However, in 'Black rice NIL' and 'Hong Xie Nuo' genomic DNA, an extra band was observed in both *Nde*I- and *Bsp*HI-digested DNAs.



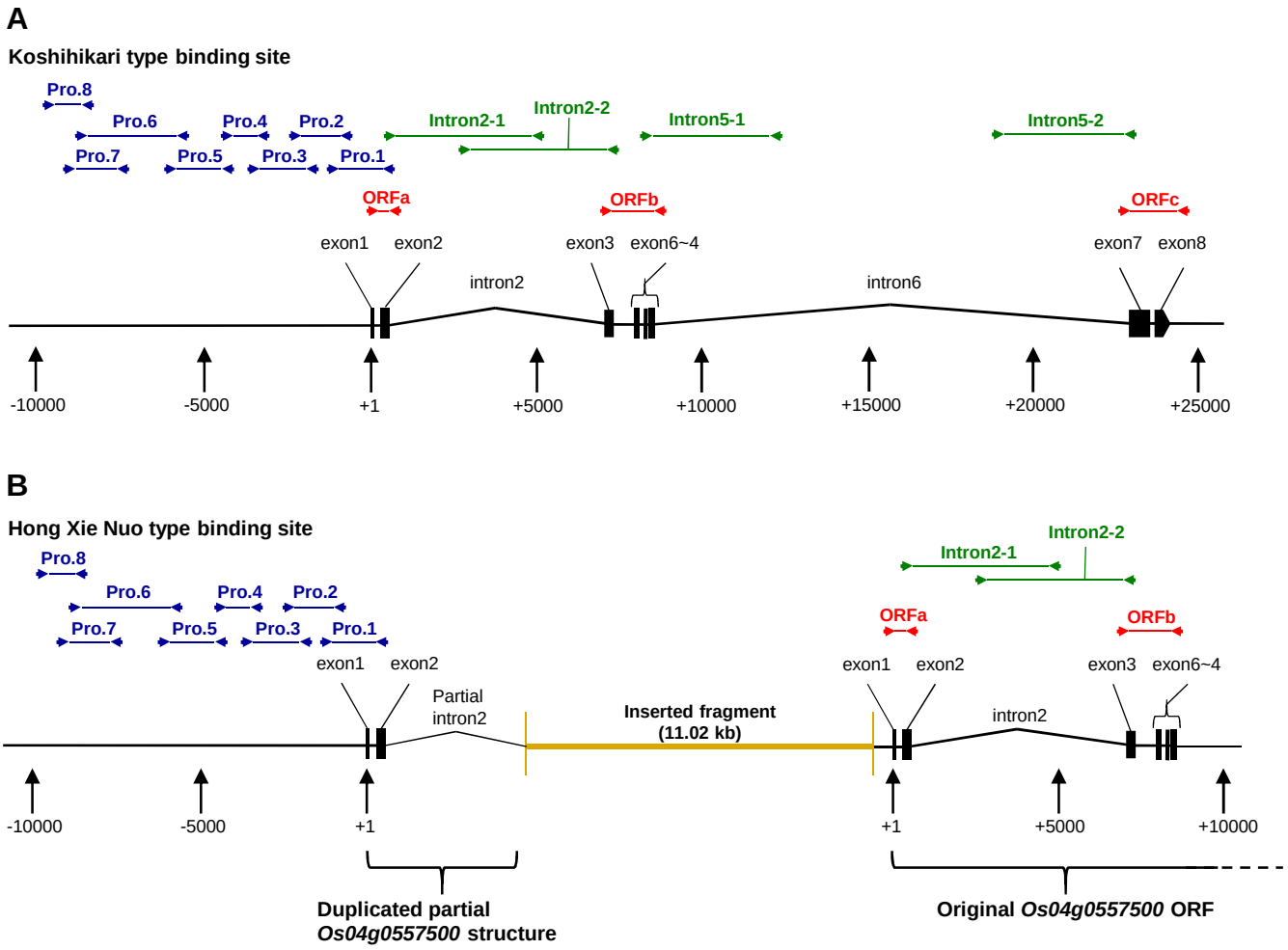
Supplemental Figure 9. Restriction maps of the 'Koshihikari' and the 'Hong Xie Nuo' types of the *Os04g0557500* gene.

Blue boxes and lines indicate the *Os04g0557500* exons and introns, respectively. Orange lines indicate 11.02-kb inserted fragment originating from approximately 83-kb upstream of the *Os04g0557500* ORF. Red boxes and lines indicate duplicated partial sequence of the *Os04g0557500* ORF. Green boxes indicate the binding position of the 6 probes used in Supplemental Figure 10. Scale bar: 5.0 kb.



Supplemental Figure 10. Confirmation of the *Os04g0557500* structure in 'Black rice NIL' by DNA gel blot analysis.

Approximately 10-µg samples of genomic DNAs were digested with 5 kinds of restriction enzymes. Hybridizations was performed with six probes termed Intron5+ (A), In3-In4 (B), exon1N (C), D0.6K (D), D9K (E), and Kala4pro+ (F), respectively.

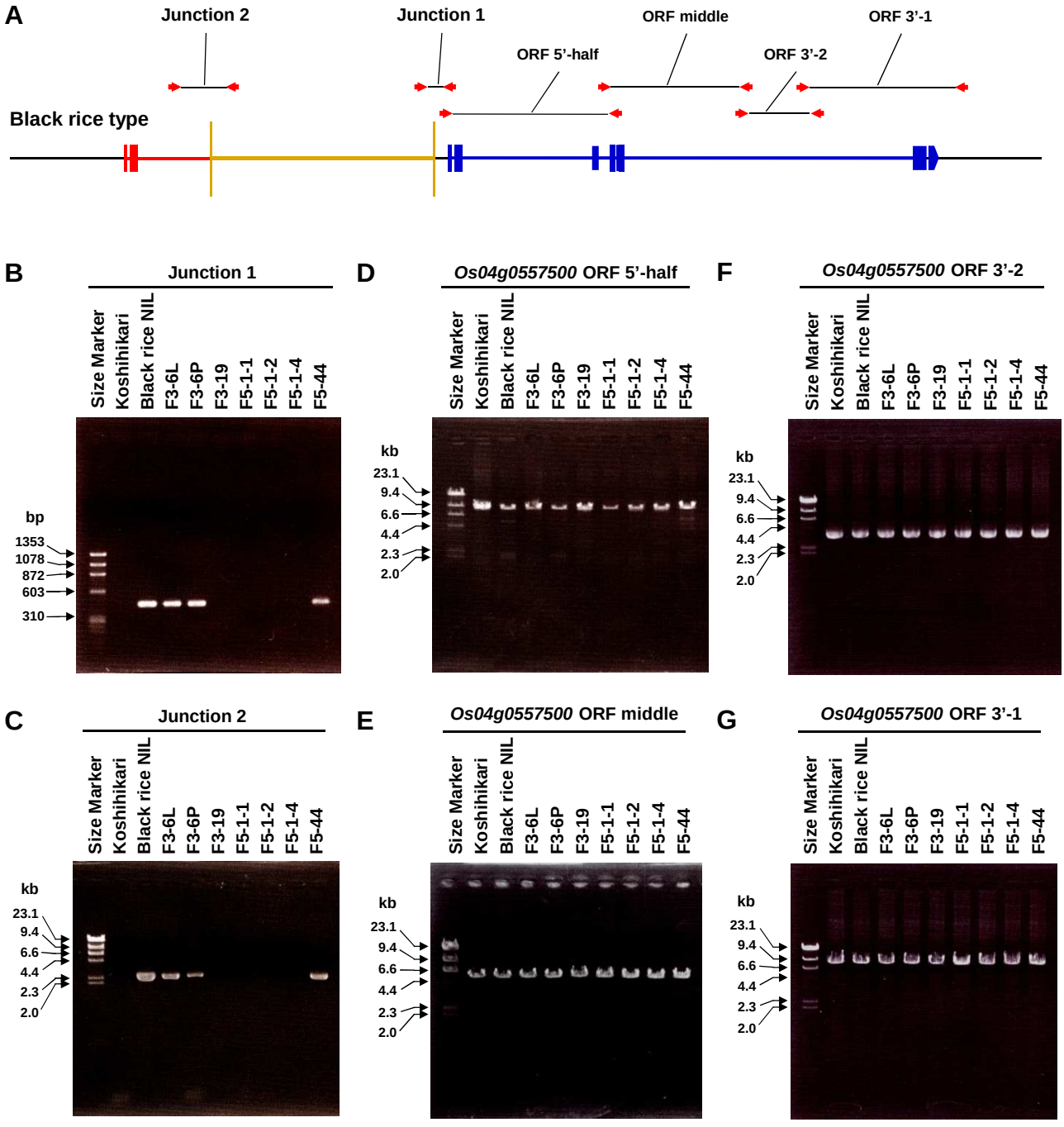


Supplemental Figure 11. Binding position of the primers used for sequencing of *Os04g0557500* and its promoter.

(A) The 'Koshihikari'-type *Os04g0557500* gene structure, the binding positions of the primers, and the corresponding amplified PCR products.

(B) The 'Homg Xie Nuo'-type *Os04g0557500* structure, the binding positions of the primers, and the corresponding amplified PCR products.

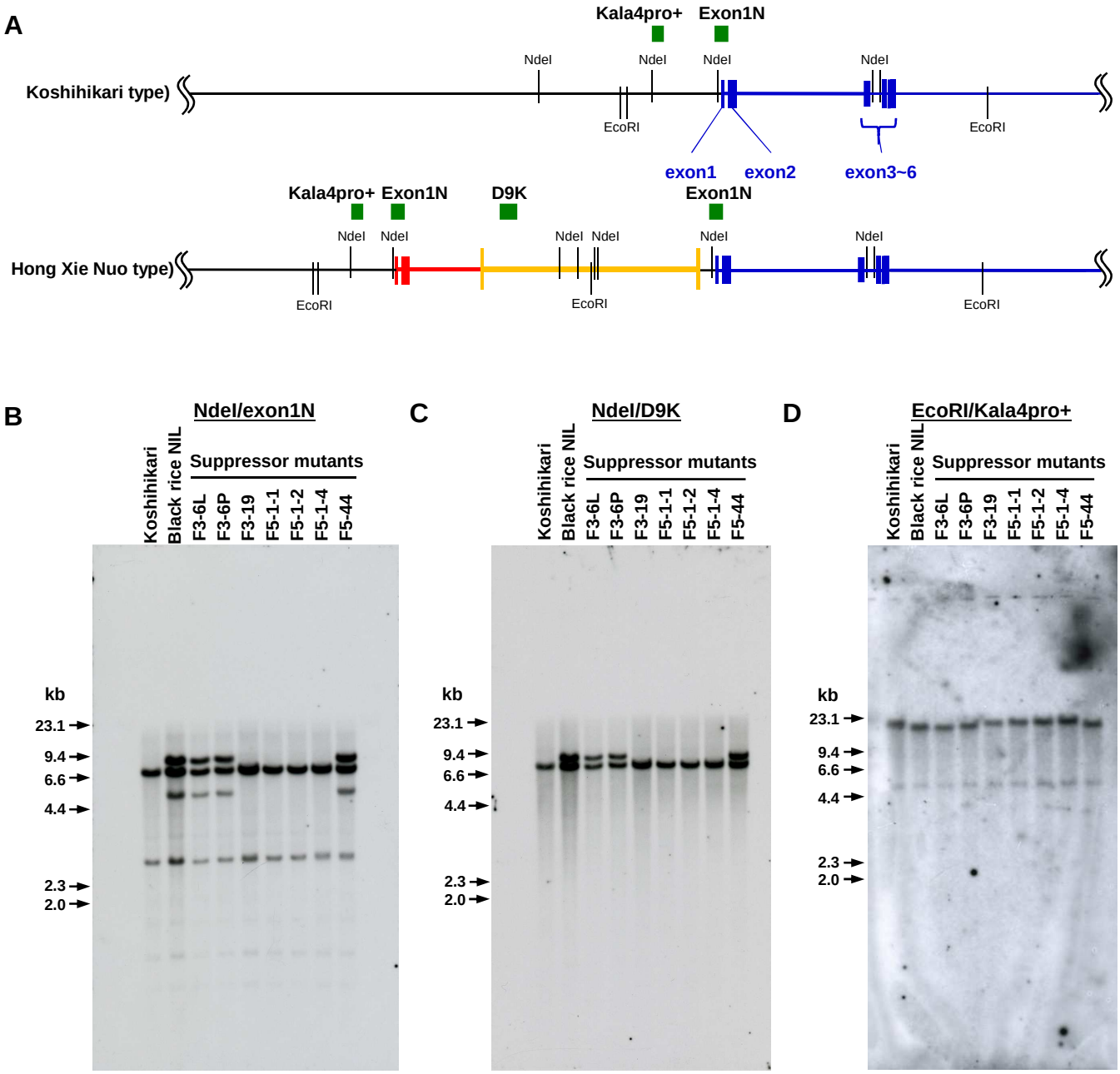
Black boxes indicate the *Os04g0557500* exons. Red, blue and green arrows indicate the binding sites of the primers used for amplification of the *Os04g0557500* ORF, promoter, and intron regions, respectively. The amplified fragments were used for the direct sequencing. The numbers under the *Os04g0557500* structures indicate the positions relative to the transcription start site +1.



Supplemental Figure 12. Confirmation of the structure of the *Os04g0557500* promoter and ORF in the seven suppressor mutants by PCR.

(A) Primer design and *Os04g0557500* structure in 'Black rice NIL'. Red arrows indicate binding positions of the designed primer pairs. Blue boxes and blue lines indicate the *Os04g0557500* exons and introns, respectively. Orange line indicates approximately 11.02-kb inserted fragment. Red boxes and lines indicate duplicated partial *Kala4* ORF.

(B), (C), (D), (E), (F) and (G) Gel images of the PCR products amplified using junction 1 (B), junction 2 (C), 5'-half of the ORF (D), middle of the ORF (E), 3'-2 of the ORF (F) and 3'-1 of the ORF (G) primer pairs. 'Koshuhikari' and 'Black rice NIL' served as controls.

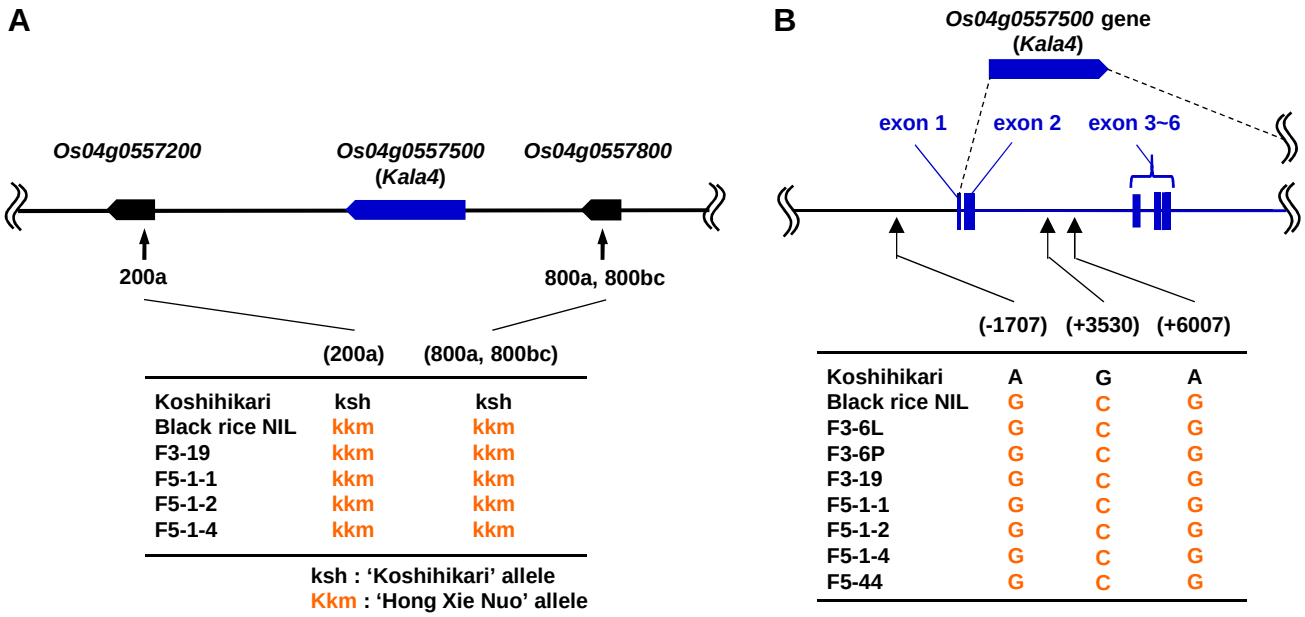


Supplemental Figure 13. Confirmation of the *Os04g0557500* promoter structure in the suppressor mutants by DNA gel blot analysis.

(A) Strategy for the DNA gel blot analysis. The 'Koshihikari' and 'Hong Xie Nuo' type *Os04g0557500* promoters are shown. Blue boxes and lines indicate the exons and the introns of *Os04g0557500*. Orange lines indicate the 11.02-kb inserted fragment. Red boxes and lines indicate duplicated partial *Kala4* ORF. Green boxes indicate hybridizing regions of the three probes.

(B) and (C) DNA gel blot analysis using *Nde*I restriction enzyme. Hybridizations were performed with exon1N (B) and D9K (C) probes, respectively. In the four suppressor mutant lines, i.e. F3-19, F5-1-1, F5-1-2 and F5-1-4, extra bands were not detected.

(D) DNA gel blot analysis with *EcoR*I digestion. Hybridization was performed with Kala4pro+ probe. In these four suppressor mutants, the structures of the *Os04g0557500* promoter were reverted to the 'Koshihikari' type.



	(200a)	(800a, 800bc)
Koshihikari	ksh	ksh
Black rice NIL	kkm	kkm
F3-19	kkm	kkm
F5-1-1	kkm	kkm
F5-1-2	kkm	kkm
F5-1-4	kkm	kkm

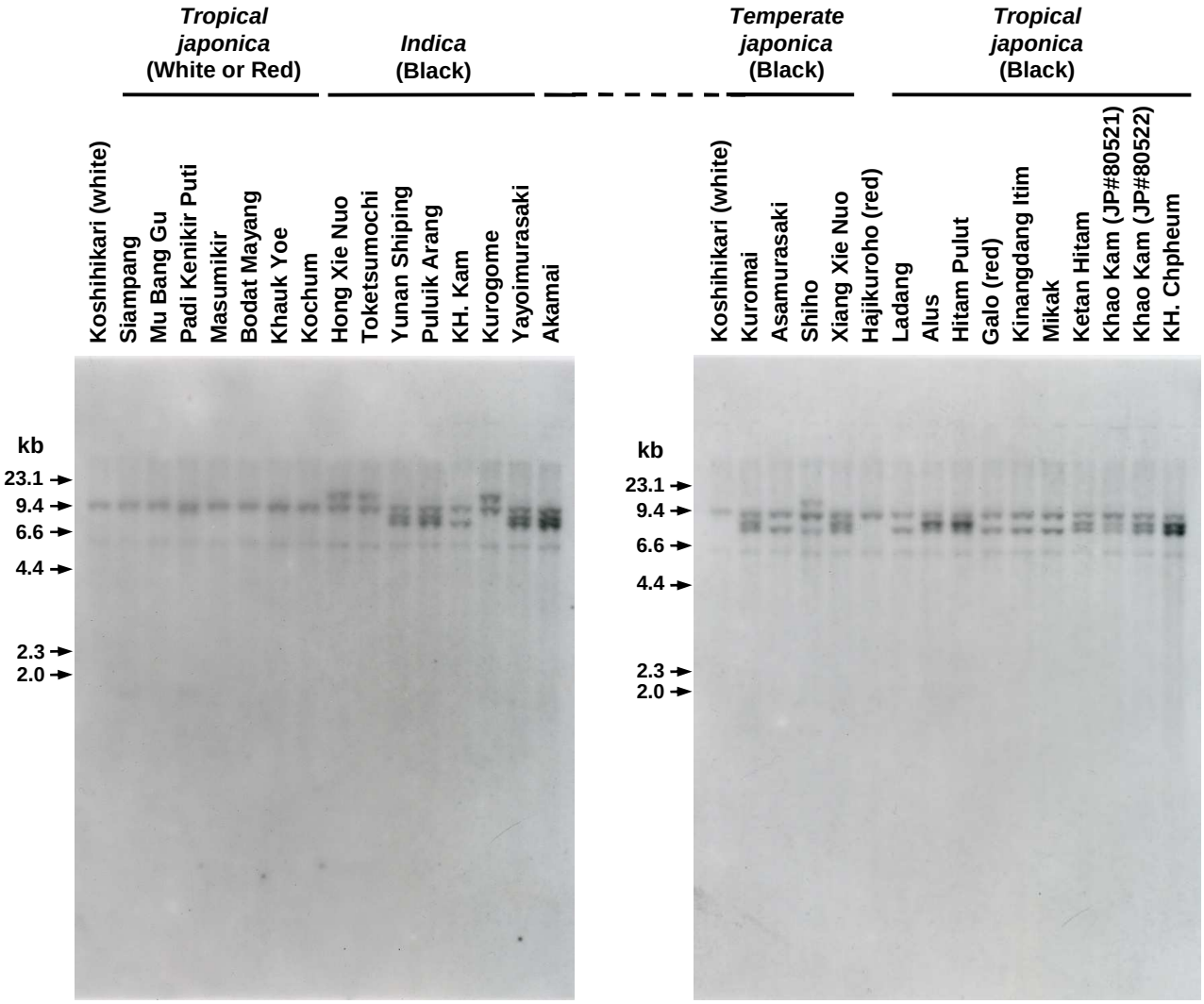
ksh : 'Koshihikari' allele
Kkm : 'Hong Xie Nuo' allele

	A	G	A
Koshihikari	A	G	A
Black rice NIL	G	C	G
F3-6L	G	C	G
F3-6P	G	C	G
F3-19	G	C	G
F5-1-1	G	C	G
F5-1-2	G	C	G
F5-1-4	G	C	G
F5-44	G	C	G

Supplemental Figure 14. Genotypes of *Os04g0557500* and the two PCR markers in the seven suppressor mutants.

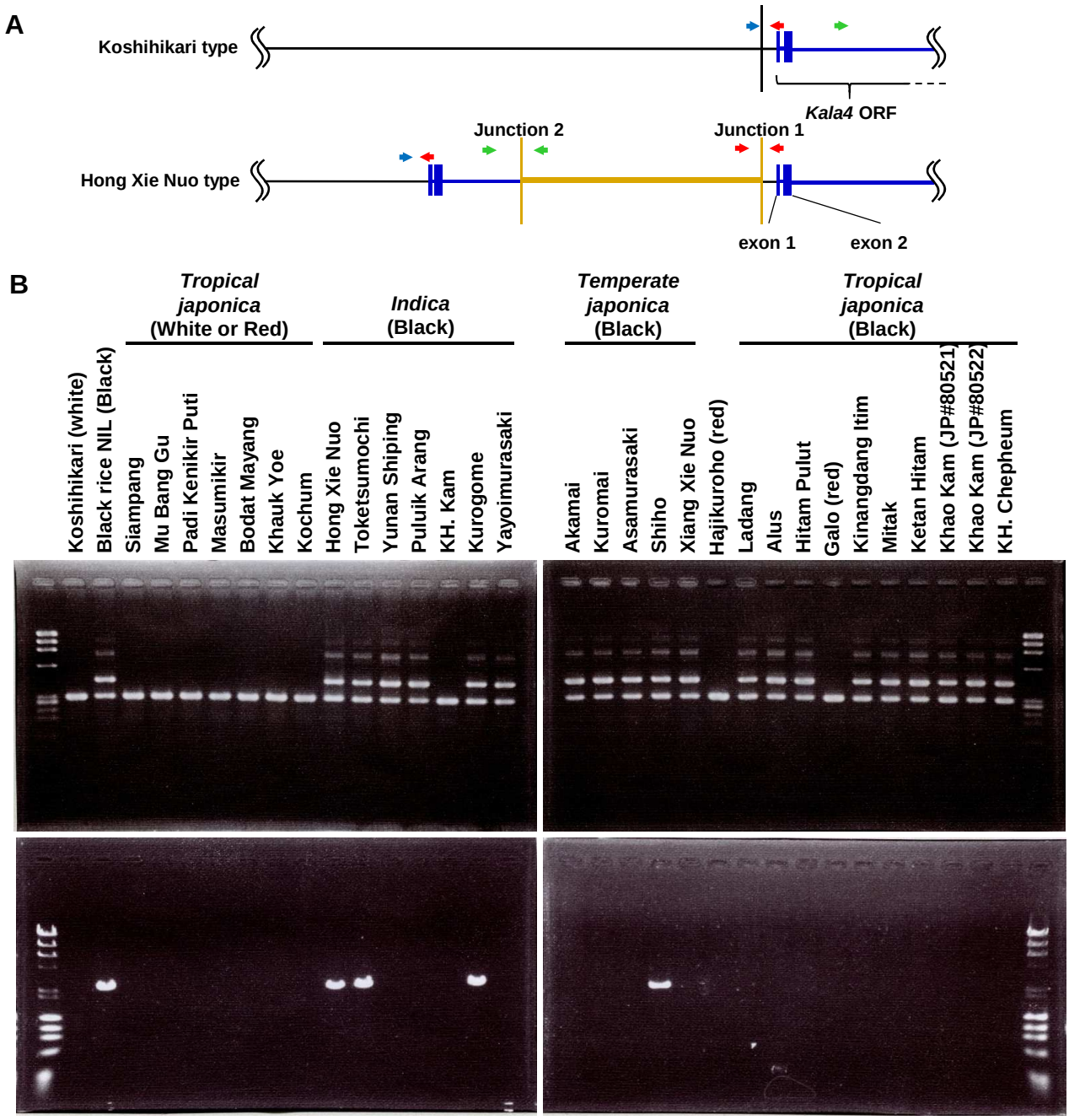
(A) Genotype of PCR markers at the two neighboring genes, *Os04g0557200* and *Os04g0557800*, in the suppressor mutants. The upper line and boxes indicate the position of three bHLH genes. The lower table indicates the polymorphisms of the four suppressor mutants that have low expression levels for *Os04g0557500*.

(B) Genotype of *Os04g0557500* in the suppressor mutants. The upper blue box indicates direction of the *Os04g0557500* gene. Blue line and boxes (middle) indicate the *Os04g0557500* gene partial sequence and the positions of three sequence markers. The lower table indicates the polymorphisms of three SNPs in *Os04g0557500*.



Supplemental Figure 15. DNA gel blot analysis of the *Kala4* promoter in twenty-one black rice varieties.

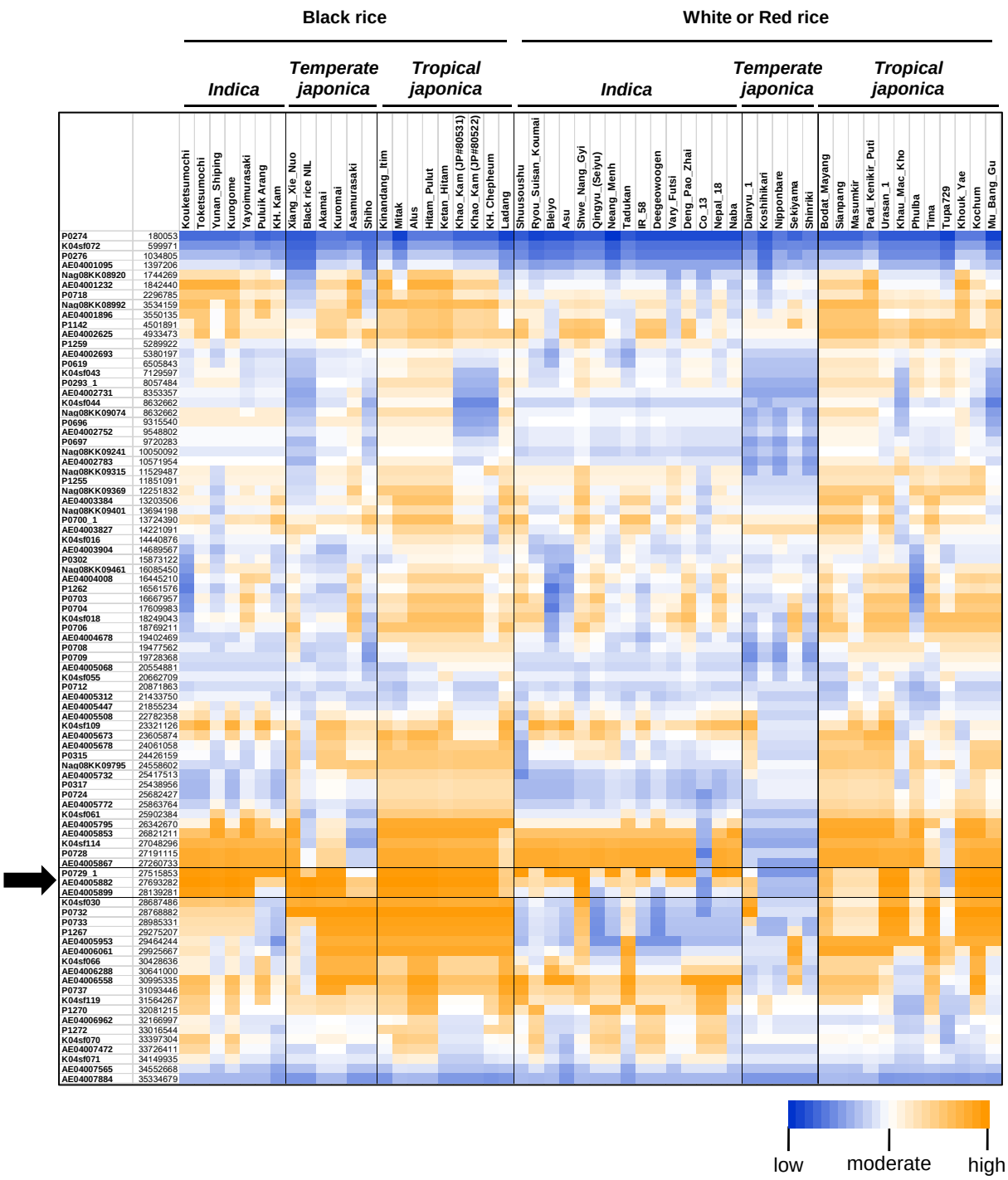
Ten-μg samples of genomic DNAs were digested with *Nde*I restriction enzyme. Hybridization was performed with the exon1N probe to detect duplicated sequence of the *Kala4* gene. 'Koshihikari' and seven heritage landraces served as controls.



Supplemental Figure 16. Classification of twenty-one black rice varieties based on PCR analysis.

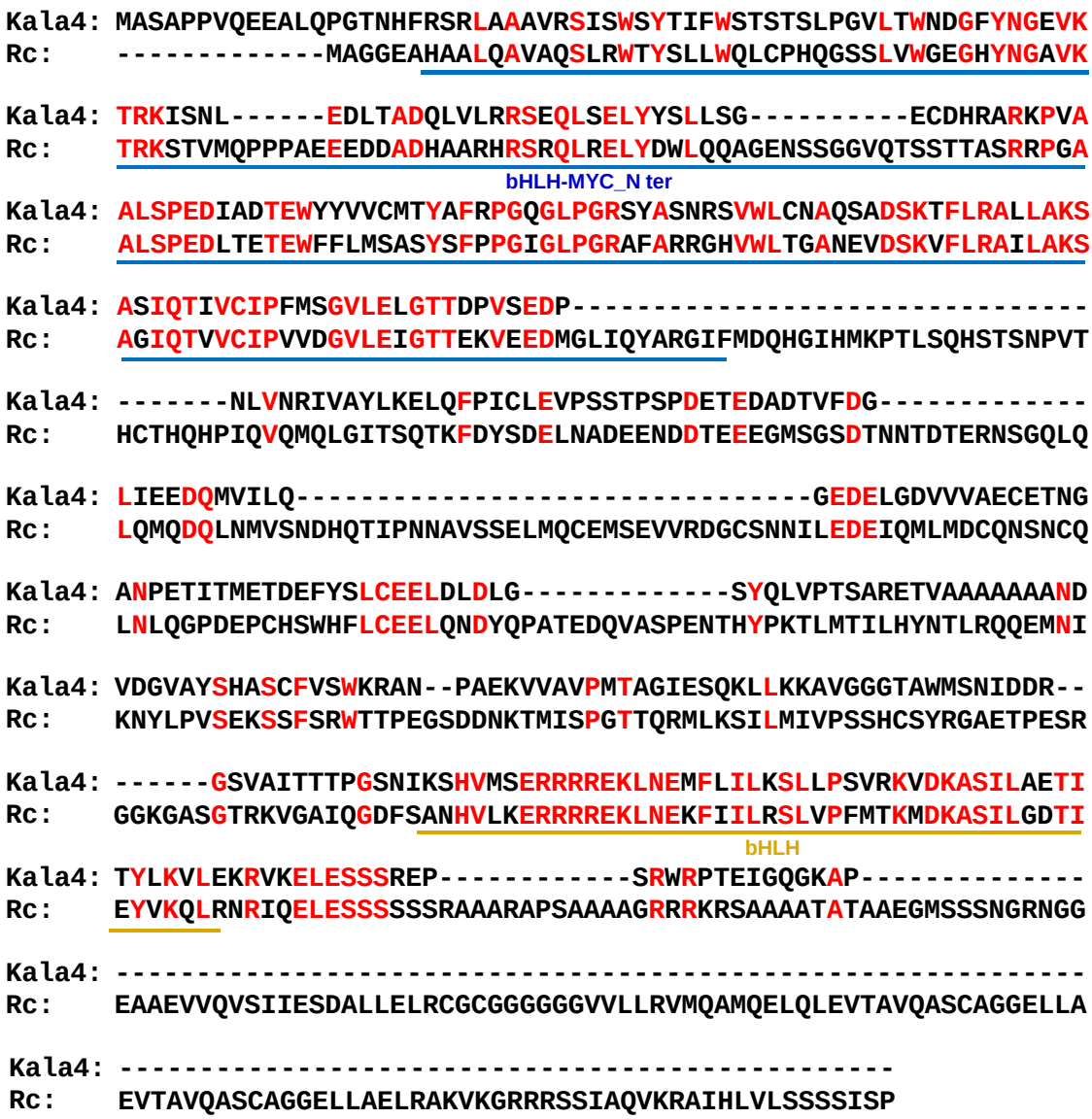
(A) The 'Koshihikari' and the 'Hong Xie Nuo' types promoter structure and the binding positions of the designed primer pairs. Two set of primer pairs were designed from the 'Hong Xie Nuo'-type *Kala4* promoter across junction 1 (red arrows) and junction 2 (green arrows). When PCR analysis checking the junction 1 structure was performed, the control primer (blue arrows), which can amplify the both 'Koshihikari' and 'Hong Xie Nuo' type promoter fragments, was added to the reaction. Blue lines and boxes indicate the original *Kala4* and duplicated *Kala4* sequences.

(B) Gel images of the PCR products. Upper 2 panels show the PCR products from the junction 1 primer pair and control. Lower 2 panels show the PCR products from the junction 2 primer pair. In the PCR experiments, the *Kala4* promoter structures in these 21 black rice were divided into 3 types, type I, which possess the same sequence at the both ends (the 'Hong Xie Nuo' type), type II, which possess the same junction 1 but not the proper junction 2, and type III, in which no bands were amplified for either end.



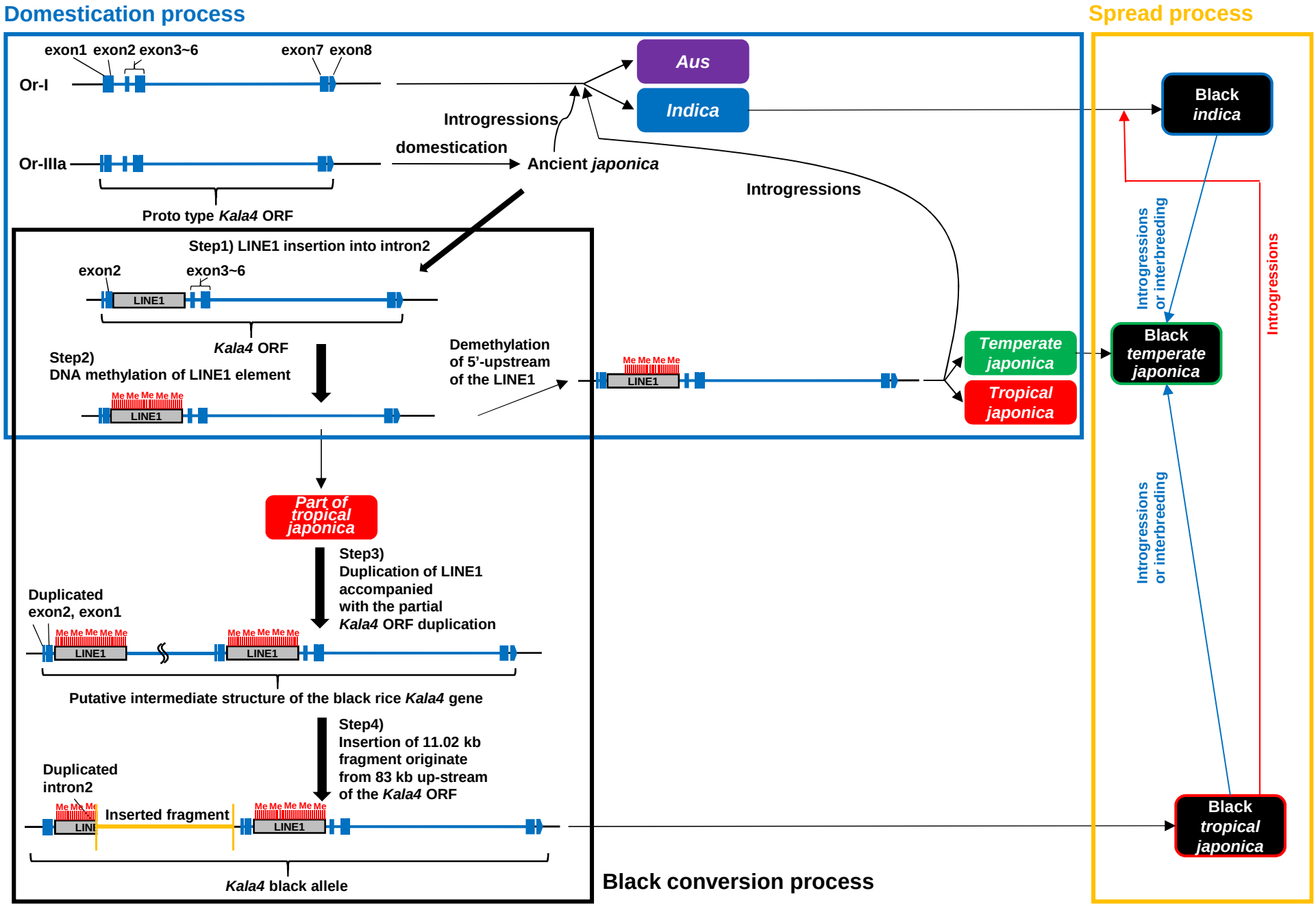
Supplemental Figure 17. Local genetic distances of the black rice varieties on chromosome 4.

Genetic distances from *tropical japonica* are represented in this figure. Information for each polymorphism was obtained from Golden-Gate assay, and the correlation coefficients were calculated. Orange and blue colors indicate higher and lower correlation with *tropical japonica* subspecies, respectively. Black arrow indicates the peripheral region of the *Kala4* gene.



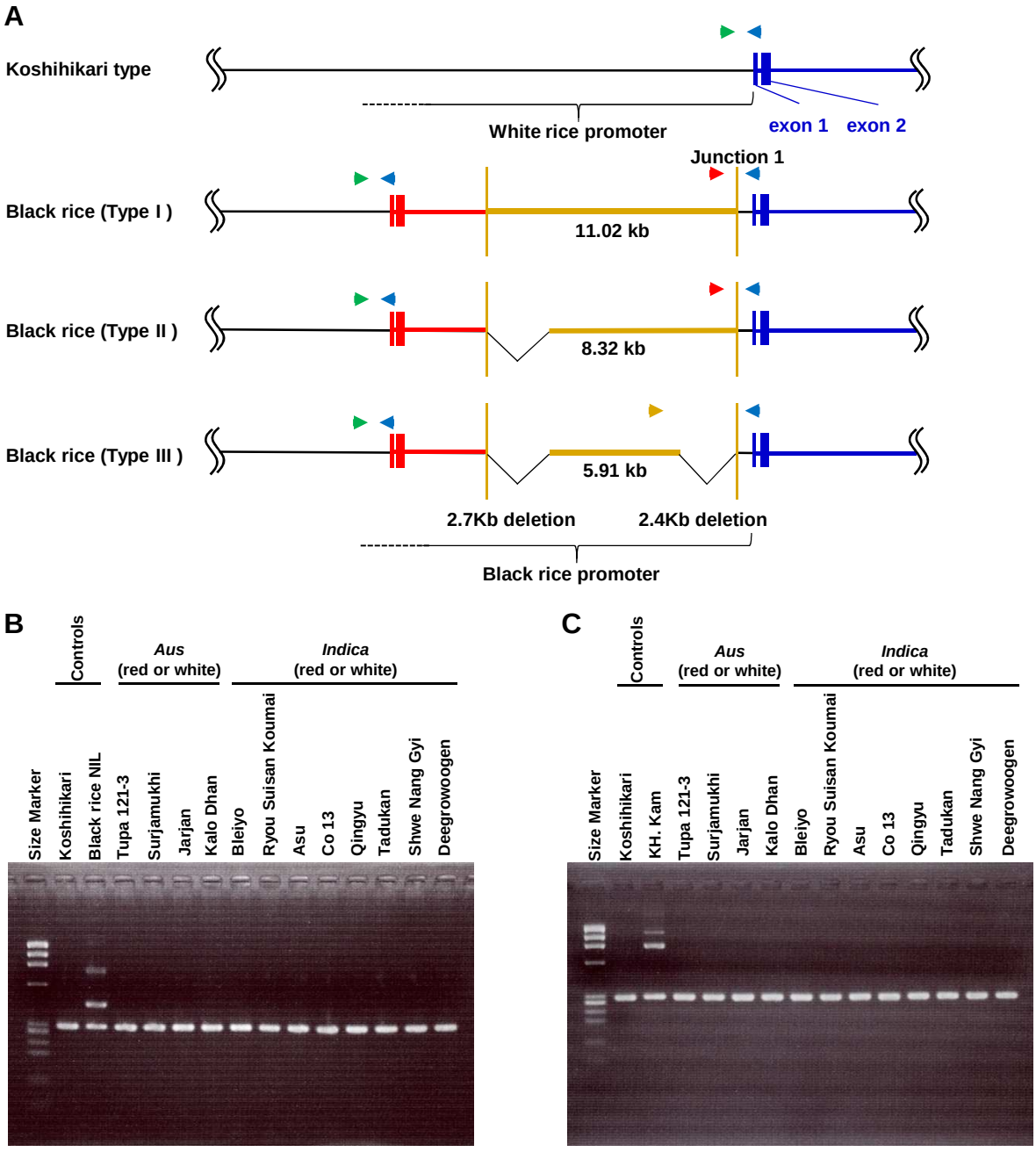
Supplemental Figure 18. Sequence alignment of two bHLH proteins.

Kala4 (upper sequence) and Rc (lower sequence) proteins are shown. Red characters indicate identical amino acids between Kala4 and Rc proteins. Blue and orange underlines indicate conserved bHLH-MYC N-terminal domain and bHLH DNA-binding domain, respectively.



Supplemental Figure 19. Schematic diagram of a model showing the processes in the origin of black rice and the spread of the *Kala4* black allele.

Supplemental Figure 19. (continued)
The processes involved in the creation of black rice and its spread can be divided into three categories, i.e. domestication (in blue frame), black rice conversion (in black frame) and spread (in orange frame).
Domestication process: Or-IIIa type *O. rufipogon* is the most closely related ancestral varieties of *japonica* cultivars and landraces (Lu et al., 2008; Huang et al., 2012). The two *O. rufipogon* accessions, W1943 and W1963 which belong to Or-IIIa, do not have LINE1 insertion at intron 2 of the *Kala4* gene (data from Gramene database, http://www.gramene.org/ and our NGS). This observation suggests that the LINE1 insertion occurred after domestication. Moreover, the LINE1 insertion at this position was observed in both *tropical japonica* and *temperate japonica*. Therefore, the LINE1 insertion probably occurred in the common ancestor before the differentiation to the two *japonica* subspecies (step1) (however, in *temperate japonica* subspecies, the LINE1 insertions were surveyed in only two cultivars). There are relatively few LINE1-inserted cultivars in *indica* and *aus* subspecies, suggesting that the LINE1-inserted allele of the *Kala4* locus was transferred from ancient *japonica* or *tropical japonica* to *indica* and *aus* subspecies.
Black conversion process: The difference of the DNA methylation status at intron 2 of the *Kala4* gene is observed in only few varieties of *tropical japonica* subspecies (however, only 2 white *temperate japonica* and no *indica* varieties were investigated). Higher DNA methylation of the LINE1 element has been thought to be necessary for inactivation of LINE1 transposition. Therefore, DNA methylation of the LINE1 element likely occurred immediately after LINE1 insertion (step 2), and then after the nonsense mutation arose in ORF2, the DNA methylation status of the 5'-upstream region of LINE1 was decreased in the common ancestral varieties of most *temperate japonica* and *tropical japonica*. However, the higher DNA methylation status 5'-upstream of LINE1 remained in some *tropical japonica*, and the variety was the origin of black rice. In this variety, the duplication of the LINE1 sequence accompanied by part of *Kala4* structure likely occurred (step 3), followed by insertion of the 11.02-kb fragment originated from approximately 83-kb upstream of the *Kala4* ORF (step4). As a consequence of these duplication and insertion events, the *Kala4* expression was upregulated in both transcription and translation, resulting in the creation of the black *Kala4* allele.
Spread process: The black allele of the *Kala4* locus, which was generated in *tropical japonica* subspecies, was transferred into *indica* subspecies by introgression (red arrow). The identical alteration of the *Kala4* promoter structure is also observed in black rice of *temperate japonica* subspecies. The black allele of the *Kala4* locus likely was transferred from black *indica* or black *jropical japonica* subspecies by interbreeding (blue arrows).



Supplemental Figure 20. PCR analysis of the promoter in white and red rice varieties with LINE1 element insertion in intron 2 of the *Kala4* gene.

(A) Binding positions of the primers. 'Koshihikari' type and three types of black rice promoters are shown in this figure. Blue and green arrowheads indicate the primers that are able to amplify the fragment of the Koshihikari-type promoter and 5'-upstream of the black rice promoter. Red arrowheads indicate the primers that bind to the type I and type II inserted fragment specifically. Orange arrowhead indicates the primer that binds to type III inserted fragment and the primer can amplify the type III junction 1 region.

(B) Gel image of the analysis of type I and type II insertion in the white and red varieties. PCR experiments were performed using 3 primers, blue, green and red arrowheads in (A). In addition to the common band, an extra upper band was observed in the 'Black rice NIL'.

(C) Gel image of the analysis of type III insertion in the white and red varieties. PCR experiments were performed using 3 kinds of primers, blue, green and orange arrowheads in (A). In addition to the common band, an extra band was observed in the 'KH. Kam'.

Supplemental Table 1. Anthocyanin biosynthesis-related genes in rice and maize.

Categories	Proteins	Species	Gene names	Locus IDs	References
Structural genes	Chalcone Synthase (CHS)	maize	*Colored aleurone2* (*C2*)[a]	GRMZM2G151227 (Chr.2)	Wienand et al. (1986)
		rice	*OsCHS1*[b]	Os11g0530600	Reddy et al. (1996)
			OsCHS2[b]	Os07g0214900	Shih et al. (2008)
	Chalcone isomerase (CHI)	maize	*CHI1*[b]	GRMZM2G155329 (Chr.1)	Grotewold and Peterson (1994)
		rice	*OsCHI*[b]	Os03g0819600	Druka et al. (2002)
	Flavanone 3-hydroxylase (F3H)	maize	*F3H*[b]	GRMZM2G062396 (Chr.2)	Deboo et al. (1995)
		rice	*OsF3H-1*[b]	Os04g0662600	Kim et al. (2007)
			OsF3H-2[b]	Os10g0536400	
			OsF3H-3[b]	Os04g0667200	
	Flavanone 3'-hydroxylase (F3'H)	maize	*Purple aleurone1* (*Pr1*)[a]	GRMZM2G025832 (Chr.5)	Sharma et al. (2011)
		rice	*OsF3'H*[b]	Os10g0320100	Shih et al. (2008)
	Dihydroflavonol reductase (DFR)	maize	*Anthocyaninless1* (*A1*)[a]	GRMZM2G026930 (Chr.3)	O'Reilly et al. (1985)
		rice	*Rd*[a]	Os01g0633500	Furukawa et al. (2006)
	Leucoanthocyanidin dioxygenase (LDOX)	maize	*Anthocyaninless2* (*A2*)[a]	GRMZM2G345717 (Chr.5)	Menssen et al. (1990)
		rice	*OsANS1*[b]	Os01g0372500	Shih et al. (2008)
			OsANS2[b]	Os06g0626700	
	UDP-glycosyl transferase (UGT)	maize	*Bronze1* (*Bz1*)[a]	GRMZM2G165390 (Chr.9)	Dooner et al. (1977)
		rice	*OsUGT*[b]	Os06g0192100	Tanaka et al. (2008)
Regulator genes	basic helix-loop-helix transcriptional factor (bHLH)	maize	*Red* (*R*)[a]	GRMZM5G822829 (Chr.10)	Styles et al. (1973)
			Booster (*B*)[a]	GRMZM2G172795 (Chr.2)	Chandler et al. (1989)
		rice	*OSB1*[b]	Os04g0557800	Sakamoto et al. (2001)
			OSB2[b]	Os04g0557500	
	R2R3-Myb transcriptional factor (Myb)	maize	*Colored aleurone1* (*C1*)[a]	GRMZM2G005066 (Chr.9)	Paz-Ares et al. (1978)
			Purple plant (*Pl*)[a]	GRMZM2G701063 (Chr.6)	Cone et al. (1993)
		rice	*OsC1*[b]	Os06g0205100	Saitoh et al. (2004)
	WD40 repeat	maize	*Pale aleurone color1 (PAC1)*[a]	GRMZM2G058292 (Chr.5)	Carey et al. (2004)
		rice	unidentified		

[a]These genes were genetically identified in maize or rice.
[b]These genes were identified based on sequence similarity with other plant species.
Although there are many homologous genes annotated in both genomic sequences, only the genes that have been reported as cloned are described in this table.

Supplemental Table 2. Polymorphisms in the 25.6-kb candidate region of the *Kala4* locus.

			Position/Polymorphism						
Subspecies	Varieties	Seed color	+23843 (exon8)	+10866 (intron6)	+9162 (intron6)	+8390 (intron4)	+6007 (intron2)	+3530 (intron2)	-1707 (promoter)
Temperate japonica	Nipponbare	white	G/G	G/G	A/A	--/--	A/A	G/G	A/A
	Koshihikari	white	G/G	G/G	A/A	--/--	A/A	G/G	A/A
Tropical japonica (heritages)	Siampang	red	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Mu Bang Gu	red	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Padi Kenikir Puti	white	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Masmikir	white	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Bodat Mayang	red	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Khauk Yae	white	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Ko chum	red	C/C	G/G	G/G	--/--	G/G	C/C	G/G
Indica	Hong Xie Nuo	black	C/C	A/A	G/G	CT/CT	G/G	C/C	G/G
	Toketsumochi	black	C/C	A/A	G/G	CT/CT	G/G	C/C	G/G
	Yunan Shiping	black	C/C	G/G	G/G	CT/CT	G/G	C/C	G/G
	Puluik Arang	black	C/C	G/G	G/G	CT/CT	G/G	C/C	G/G
	KH. Kam	black	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Kurogome	black	C/C	A/A	G/G	CT/CT	G/G	C/C	G/G
	Yayoimurasaki	black	C/C	G/G	G/G	CT/CT	G/G	C/C	G/G
Temperate japonica	Akamai	black	C/C	G/G	G/G	CT/CT	G/G	C/C	G/G
	Kuromai	black	C/C	G/G	G/G	CT/CT	G/G	C/C	G/G
	Asamurasaki	black	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Shiho	black	C/C	G/A	G/G	--/CT	G/G	C/C	G/G
	Xiang Xie Nuo	black	C/C	G/G	G/G	CT/CT	G/G	C/C	G/G
Tropical japonica	Ladang	black	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Alus	black	C/C	G/G	G/G	CT/CT	G/G	C/C	G/G
	Hitam Pulut	black	C/C	G/G	G/G	CT/CT	G/G	C/C	G/G
	Kinangdang Itim	black	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Mitak	black	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Ketan Hitam	black	C/C	G/G	G/G	--/--	G/G	C/C	G/G
	Khao Kam (JP#80521)	black	C/C	G/G	G/G	CT/CT	G/G	C/C	G/G
	Khao Kam (JP#80522)	black	C/C	G/G	G/G	CT/CT	G/G	C/C	G/G
	KH. Chepheum	black	C/C	G/G	G/G	CT/CT	G/G	C/C	G/G

x/x : 'Koshihikari' Type y/y : 'Hong Xie Nuo' Type

Supplemental Table 3. Deduced band sizes for DNA gel blot analysis.

		Restriction enzyme/Band size				
Probe name	Cultivar name	*Nde* I	*Afl* II	*Kpn* I	*Eco* RI	*Mfe* I
Intron5+	Koshihikari	19.927	20.142	24.351	12.425	23.222
	Black rice NIL	19.927	20.142	30.731	12.435	23.222
In3-In4	Koshihikari	19.927	20.142	24.351	19.349	13.634
	Black rice NIL	19.927	20.142	30.731	20.477	9.575
Exon1N	Koshihikari	7.609	11.656	24.351	19.349	13.634
	Black rice NIL	9.133	11.656	30.731	20.477	13.634
		7.609	5.153	6.616	16.008	9.575
D0.6K	Koshihikari	4.979	3.670	35.362	6.095	1.754
	Black rice NIL	5.380	3.670	35.362	20.477	1.754
		4.979	3.670	30.731	6.095	1.754
D9K	Koshihikari	7.779	5.730	4.140	6.837	10.195
	Black rice NIL	9.133	5.730	4.140	16.008	13.874
		7.779	5.730	4.140	6.837	10.195
Kala4pro+	Koshihikari	2.917	11.656	8.320	19.349	13.634
	Black rice NIL	2.917	11.656	8.320	16.001	13.874

The band sizes (kb) when digested by each enzyme were estimated from the RAP-DBs complete genome sequence of the ‘Nipponbare’ (build 5). For probes prepared from the duplicated region, two bands or one overlapping band were detected.

Supplemental Table 4. Classification of the twenty-one black rice varieties according to three types of insertions.

Insertion type	Varieties	Subspecies
Type I (11.02kb)	Black rice NIL	*Temperate japonica* backgrand
	Hong Xie Nuo	*Indica*
	Toketsumochi	*Indica*
	Kurogome	*Indica*
	Shiho	*Temperate japonica*
Type II (8.32kb)	Yunan Shiping	*Indica*
	Puluik Arang	*Indica*
	Yayoimurasaki	*Indica*
	Akamai	*Temperate japonica*
	Kuromai	*Temperate japonica*
	Asamurasaki	*Temperate japonica*
	Xiang Xie Nuo	*Temperate japonica*
	Ladang	*Tropical japonica*
	Alus	*Tropical japonica*
	Hitam Pulut	*Tropical japonica*
	Kinangdang Itim	*Tropical japonica*
	Mitak	*Tropical japonica*
	Ketan Hitam	*Tropical japonica*
	Khao Kam (JP#80521)	*Tropical japonica*
	Khao Kam (JP#80522)	*Tropical japonica*
	KH. Chepheum	*Tropical japonica*
Type III (5.91kb)	KH. Kam	*Indica*

Supplemental Table 5. White and red rice cultivars used for Golden-Gate assay.

Subspecies	Line names	Origin
Aus	Chinsurah Boro 2	Java
	Kasalath	India
	Shoni	Bangladesh
	ARC 7047	India
	Anjena Dhan	Nepal
	Surjamukhi	India
	Tupa 121-3	Bangladesh
	Muha	India
	Jhona 2	India
	Kinandang Puti	Philippines
	Local Basmati	India
	Jena 035	Nepal
	Jarjan	Nepal
	Kalo Dhan	Nepal
	Shwe Nang Gyi	Myanmar
Indica	Tadukan	Philippines
	Qingyu (Seiyu)	Taiwan
	Shouusoushu	China
	Deng Pao Zhai (Toufutsusai)	China
	Deegeowoogen	Taiwan
	Nepal 18	Nepal
	Asu	Bhutan
	Naba	India
	Co 13	India
	Neang Menh	Cambodia
	Ryou Suisan Koumai	China
	IR 58	Philippines
	Bleiyo	Thailand
	Vary Futsi	Madagascar
Tropical japonica	Bodat Mayang	Indonesia
	Siampang	Indonesia
	Urasan 1	Japan
	Masumikir	Indonesia
	Padi Kenikir Puti	Indonesia
	Tima	Bhutan
	Tupa 729	Bangladesh
	Kochum	Bhutan
	Khouk Yoe	Myanmar
	Mu Bang Gu	Vietnam
	Khau Mac Kho	Bhutan
	Phulba	India
Temperate japonica	Koshihikari	Japan
	Dianyu 1	China
	Nipponbare	Japan
	Shinriki	Japan
	Sekiyama	Japan

Supplemental Table 6. Analysis of LINE1 insertion at intron 2 of the *Kala4* gene in rice varieties.

Subspecies	Varieties	Local origin	Grain color	LINE1 insertion in the *Kala4* locus
Aus	Kasalath	India	Red	-
	Jena 035	Napal	Red	-
	Shoni	Bangladesh	ND	-
	Tupa 121-3	Bangladesh	white	+
	ARC 7047	India	red	-
	Muha	India	red	-
	Jhona 2	India	ND	-
	Surjamukhi	India	red	+
	Jarjan	Bhutan	ND	+
	Kalo Dhan	Nepal	ND	+
	Anjana Dhan	Nepal	ND	-
	Local Basmati	India	ND	-
	Chinsurah Boro 2	India	red	-
	Kinandang Puti	Philipines	white	-
Indica	Neang Menh	Cambodia	ND	-
	Naba	India	red	-
	Bleiyo	Thailand	red	+
	Ryou Suisan Koumai	China	red	+
	Shuusoushu	China	white	-
	Aus	India	red	+
	Co 13	India	white	hetero
	Vary Futsi	Madagascar	purplish brown (bot)	-
	IR 58	Philipines	ND	-
	Qingyu (Seiyu)	Taiwan	white	+
	Deng Pao Zhai (Toufutsu)	China	white	-
	Tadukan	Philipines	white	+
	Shwe Nang Gyi	Myanmer	ND	+
	Deegeowoogen	Taiwan	ND	+
	Nepal 18	Nepal	ND	-
Temperate japonica	Nipponbare	Japan	white	+
	Koshihikari	Japan	white	+
Tropical japonica (heritage)	Siampang	Indonesia	red	+
	Mu Bang Gu	Vietnam	red	+
	Padi Kenikir Puti	Indonesia	white	+
	Masumikir	Indonesia	white	+
	Bodat Mayang	Indonesia	red	+
	Khauk Yoe	Myanmer	white	+
	Kochum	Bhutan	red	+
Indica	Hong Xie Nuo	China	black	+
	Toketsumochi	China	black	+
	Yunan Shiping	Chaina	black	+
	Puluik Arang	Indonesia	black	+
	KH. Kam	Laos	black	+
	Kurogome	Japan	black	+
	Yayoimurasaki	Japan	black	+
Temperate japonica	Akamai	Japan	black	+
	Kurogome	Japan	black	+
	Asamurasaki	Japan	black	+
	Shiho	Japan	black	+
	Xiang Xie Nuo	Chaina	black	+
Tropical japonica	Ladang	Indonesia	black	+
	Alus	Malaysia	black	+
	Hitam Pulut	Malaysia	black	+
	Galo	Philipines	black	+
	Kinangdang Itim	Philipines	black	+
	Mitak	Indonesia	black	+
	Ketan Hitam	Indonesia	black	+
	Khao Kam (JP#80521)	Laos	black	+
	Khao Kam (JP#80522)	Laos	black	+
	KH. Chepheum	Laos	black	+

Supplemental Table 7. Information on probes in the rice 4x44 K array (Agilent) related to anthocyanin and pro-anthocyanidin biosynthesis genes.

Gene name	Feature No.	Probe name	Locus ID	Accession No.
CHS	41547	Os11g0530600|mRNA|AB000801|CDS+3'UTR	Os11g0530600	AB000801
	36195	Os11g0530600|mRNA|AK067810|CDS+3'UTR	Os11g0530600	AK067810
	36457	Os11g0530600|mRNA|D29697|5'UTR+CDS	Os11g0530600	D29697
	40810	Os11g0530600|mRNA|X89859|CDS+3'UTR	Os11g0530600	X89859
F3H	39072	Os04g0662600|COMBINER_EST|CI252891|6	Os04g0662600	CI252891
F3'H	7139	Os10g0320100|mRNA|AK064736|CDS+3'UTR	Os10g0320100	AK064736
DFR	8526	Os01g0633500|mRNA|AB003496|CDS+3'UTR	Os01g0633500	AB003496
	16439	Os01g0633500|mRNA|Y07956|CDS+3'UTR	Os01g0633500	Y07956
LDOX	37661	Os01g0372500|mRNA|Y07955|CDS+3'UTR	Os01g0372500	Y07955
LAR	3141	Os03g0259400|COMBINER_EST|Os03g0259400|8	Os03g0259400	
UGT	4657	Os06g0192100|COMBINER_EST|Os06g0192100|8	Os06g0192100	
Kala4	26232	Os04g0557500|mRNA|AB021080|CDS+3'UTR	Os04g0557500	AB021080

Supplemental References

1. **Chandler, V., Radicalla, J.P., Robbins, T.P., Chen, J., and Turks, D.** (1989). Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of *B* utilizing *R* genomic sequence. Plant Cell **1:** 1175-1183.
2. **Carey, C.C., Strahle, J.T., Selinger, D.A., and Chandler, V.L.** (2004). Mutation in the *pale aleurone color1* regulatory gene of the *Zea mays* anthocyanin pathway have distinct phenotypes relative to the functionally similar *TRANSPARENT TESTA GLABRA1* gene in *Arabidopsis thaliana*. Plant Cell **16:** 450-464.
3. **Cone, K.C., Cocciolone, S.M., Burr, F.A., and Burr, B.** (1993). Maize anthocyanin regulatory gene *pl* is a duplicate of *c1* that functions in the plant. Plant Cell **5:** 1795-1805.
4. **Deboo, G.B., Albertsen, M.C., and Taylor, L.P.** (1995). Flavanone 3-hydroxylase transcripts and flavonol accumulation are temporally coordinate in maize anthers. Plant J. **7:** 703-713.
5. **Dooner, H.K., and Nelson, O.E.** (1977). Genetic Control of UDP glucose: flavonol 3-*O*-glucosyltransferase in the endosperm of maize. Biochem. Genet. **15:** 509-519.
6. **Druka, A., Kudurna, D., Rostoks, N., Brueggeman, R., Wettstein, D.V., Kleinhofs, A.** (2003). Chalcone isomerase gene from rice (*Oryza sativa*) and barley (*Hordeum vulgare*): physical, genetic and mutation mapping. Gene **302:** 171-178.
7. **Furukawa, T., Maekawa, M., Oki, T., Suda, I., Iida, S., Shimada, H., Takamure, I., and Kadowaki, K.** (2006). The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. Plant J. **49:** 91-102.
8. **Grotewold, E., and Peterson, T.** (1994). Isolation and characterization of a maize gene encoding chalcone flavonone isomerase. Mol. Gen. Genet. **242:** 1-8.
9. **Kim, J.H., Lee, Y.J., Kim, B.G., Lim, Y., and Ahn, J-H.** (2007). Flavanone 3β-hydroxylases from rice: key enzymes for flavonol and anthocyanin biosynthesis. Molecules and Cells. **25:** 312-316.
10. **Lu, T., Yu, S., Fan, D., Mu, J., Shangguan, Y., Wang, Z., Minobe, Y., Lin, Z., and Han, B.** (2008). Collection and comparative analysis of 1888 full-length cDNAs from wild rice *Oryza rufipogon* griff. W1943. DNA Res. **15:** 285-295.
11. **Menssen, A., Hohmann, S., Martin, W., Schnable, P.S., Peterson, P.A., Saedler, H., and Gierl, A.** (1990). The En/Spm transposable element of *Zea mays* contains splice sites at the termini generating a novel intron from a dSpm element in the *A2* gene. EMBO J. **9:** 3051-3057.
12. **O'Reilly, C., Shepherd, N.S., Pereira, A., Schwarz-Sommer, Z., Bertram, I., Robertson D.S., Peterson P.A., and Saedler, H.** (1985). Molecular cloning of the *a1* locus of *Zea mays* using the transposable elements *En* and *Mu1*. EMBO J. **4:** 877-882.
13. **Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A., and Saedler, H.** (1987). The regulatory *c1* locus of *Zea mays* encodes a protein with homology to *myb* proto-oncogene products and with structural similarities to transcriptional activators. EMBO J. **6:** 3553-3558.
14. **Reddy, A.R., Scheffler, B., Madhuri, G., Srivastava, M.N., Kumar, A., Sathyanarayanan, P.V., Nair, S., and Mohan, M.** (1996). Chalcone synthase in rice (*Oryza sativa* L): Detection of the CHS protein in seedlings and molecular mapping of the *chs* locus. Plant Mol. Biol. **32:** 735-743.
15. **Saitoh, K., Onishi, K., Mikami, I., Thidar, K., and Sano, Y.** (2004). Allelic diversification at the *C* (*OsC1*) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. Genetics **168:** 997-1007.

Supplemental References (continued)

16. **Sakamoto, W., Ohmori, T., Kageyama, K., Miyazaki, C., Saito, A., Murata, M., Noda, K., and Maekawa, M.** (2001). The *purple leaf* (*Pl*) locus of rice: the Pl^{w} allele has a complex organization and induces two genes encoding basic helix-loop-helix proteins involved in anthocyanin biosynthesis. Plant Cell Physiol. **42:** 982-991.
17. **Sharma, M., Cortes-Cruz, M., Ahern, K.R., McMullen, M., Brutnell T.P., and Chopra, S.** (2011). Identification of the *Pr1* gene product completes the anthocyanin biosynthesis pathway of maize. Genetics **188:** 69-79.
18. **Shih, C.H., Chu, H., Tang, L.K., Sakamoto, W., Maekawa, M., Chu, I.K., Wang, M., and Lo, C.** (2008). Functional characterization of key structural genes in rice flavonoid biosynthesis. **228:** 1043-1054.
19. **Styles, E.D., Ceska, O., and Seah, K.T.** (1973). Developmental difference in action of *R* and *B* alleles in maize. Can. J. Genet. Cytol. **15:** 59-72.
20. **Tanaka, T. et al.** (2008). The rice annotation project database (RAP-DB): 2008 update. Nucleic Acid Res. **36:** D1028-1033.
21. **Wienand, U., Weydemann, U., Niesbach-Klosgen, U., Peterson, P.A., and Saedler, H.** (1986). Molecular cloning of the *c2* locus of *Zea mays*, the gene coding for chalcone synthase. Mol. Gen. Genet. **203:** 202-207.