

# Genetic Control of a Transition from Black to Straw-White Seed Hull in Rice Domestication<sup>1[CI][W][OA]</sup>

Bo-Feng Zhu, Lizhen Si, Zixuan Wang, Yan Zhou, Jingjie Zhu, Yingying Shangguan, Danfeng Lu, Danlin Fan, Canyang Li, Hongxuan Lin, Qian Qian, Tao Sang, Bo Zhou, Yuzo Minobe, and Bin Han\*

National Center for Gene Research and Institute of Plant Physiology and Ecology (B.-F.Z., L.S., Y.Z., J.Z., Y.S., D.L., D.F., C.L., B.H.) and State Key Laboratory of Plant Molecular Genetics and Institute of Plant Physiology and Ecology (H.L.), Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200233, China; Plant Genome Center, Ibaraki 305-0856, Japan (Z.W., Y.M.); State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310006, China (Q.Q.); Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824 (T.S.); Biotechnology Institute, Zhejiang University, Hangzhou 310029, China (B.Z.); and Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China (B.H.)

The genetic mechanism involved in a transition from the black-colored seed hull of the ancestral wild rice (*Oryza rufipogon* and *Oryza nivara*) to the straw-white seed hull of cultivated rice (*Oryza sativa*) during grain ripening remains unknown. We report that the black hull of *O. rufipogon* was controlled by the *Black hull4* (*Bh4*) gene, which was fine-mapped to an 8.8-kb region on rice chromosome 4 using a cross between *O. rufipogon* W1943 (black hull) and *O. sativa indica* cv Guangluai 4 (straw-white hull). *Bh4* encodes an amino acid transporter. A 22-bp deletion within exon 3 of the *bh4* variant disrupted the *Bh4* function, leading to the straw-white hull in cultivated rice. Transgenic study indicated that *Bh4* could restore the black pigment on hulls in cv Guangluai 4 and Kasalath. *Bh4* sequence alignment of all taxa with the outgroup *Oryza barthii* showed that the wild rice maintained comparable levels of nucleotide diversity that were about 70 times higher than those in the cultivated rice. The results from the maximum likelihood Hudson-Kreitman-Aguade test suggested that the significant reduction in nucleotide diversity in rice cultivars could be caused by artificial selection. We propose that the straw-white hull was selected as an important visual phenotype of nonshattered grains during rice domestication.

Cultivated rice (*Oryza sativa*) is considered to be domesticated from its progenitor of the wild rice species *Oryza rufipogon* and *Oryza nivara* (Khush, 1997; Cheng et al., 2003; Kovach et al., 2007; Sang and Ge, 2007). The wild rice species exhibit seed-shattering habit, prostrate growth, and black-colored seed hull. Over the long time of domestication, large arrays of morphological traits have changed as a result of artificial selection, including the loss of seed shattering, changing of grain coloration, enlarging of the seed

size, and changing from prostrate to erect growth habit. There are a number of genes known to control seed-shattering habit (Konishi et al., 2006; Li et al., 2006), red grain pericarp (Sweeney et al., 2006), grain discoloration (Yu et al., 2008), grain size (Shomura et al., 2008), shoot gravitropism (Li et al., 2007), and prostrate growth (Jin et al., 2008; Tan et al., 2008). Examination of these genes will help us understand the forces driving the evolution of phenotypes during domestication.

It is now a natural phenomenon that the color of the immature seed hull of cultivated rice is green due to the color of chlorophyll and then turns into the so-called straw-white color (golden colored or yellowish) after full ripening. However, the seed hull of the wild ancestor of cultivated rice is colored black at maturity. The black seed hull is ubiquitous among the ancestors of cultivated rice. The black hull seems always associated with seed-shattering habit in wild rice. The seed-shattering habit is believed to be a key functional trait for wild rice to survive efficiently, and the black hull is thus reasonably thought to be a natural color for shattered wild rice seeds in a dark mud land. In contrast, the nonshattering seed of cultivated rice is always associated with a straw-white hull, which is accompanied by a withering of stems and leaves of a similar color at the end of the growing season.

<sup>1</sup> This work was supported by the Ministry of Science and Technology of China (grant no. 2006AA10A102), the Chinese Academy of Sciences (grant no. KSCX2-YW-N-024), the National Natural Science Foundation of China (grant no. 30821004), and the Ministry of Agriculture of China (grant nos. 2008ZX08009-002 and 2008ZX08012-002).

\* Corresponding author; e-mail bhan@ncgr.ac.cn.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors ([www.plantphysiol.org](http://www.plantphysiol.org)) is: Bin Han (bhan@ncgr.ac.cn).

[C] Some figures in this article are displayed in color online but in black and white in the print edition.

[W] The online version of this article contains Web-only data.

[OA] Open Access articles can be viewed online without a subscription.

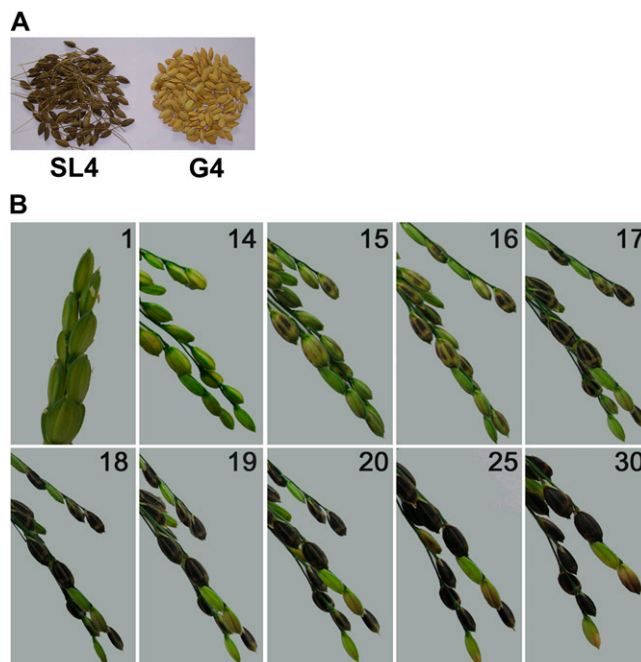
[www.plantphysiol.org/cgi/doi/10.1104/pp.110.168500](http://www.plantphysiol.org/cgi/doi/10.1104/pp.110.168500)

However, the genetic basis for the transition from the black hull of the wild rice to the straw-white hull of cultivated rice remains unknown. The black hull coloration has been identified to be controlled by two or three complementary genes (Chao, 1928; Nagao and Takahashi, 1954; Kuriyama and Kudo, 1967). Maekawa (1984) reports that three complementary genes, symbolized as *Bh-a*, *Bh-b*, and *Bh-c*, control black hull color. The *Ph* gene for phenol reaction is also found to be responsible for black hull color (Kuriyama and Kudo, 1967). The *Ph* gene (*Phr1*) corresponding to *Bh-c* has been cloned and characterized. *Phr1* localizes on the long arm of chromosome 4 and encodes a polyphenol oxidase. The frequency of *Ph* is high in the *indica* type and low in the *japonica* type (Yu et al., 2008). In another study, using an F2 population developed from the cross between the weedy strain SS18-2 and the breeding line EM93-1, the two loci *qHC4* and *qHC7* that explained a major and a minor proportion of the phenotypic variation in black pigmentations on the hull were identified (Gu et al., 2005). But no candidate gene has been cloned for the transition from the black hull of the ancestral wild rice to the straw-white seed hull of cultivated rice, and the genetic basis for this kind of hull color alteration remains unknown. The functional variations resulting from this transition are also to be investigated. In this study, we cloned the *Bh4* gene, a member of an amino acid transporter family conditioning for black hull on chromosome 4. A 22-bp deletion within *Bh4* caused a frame shift and truncated the BH4 protein that led to the straw-white hull phenotype in cultivated rice. Sequence analysis of *O. sativa* varieties and wild rice accessions revealed that *Bh4* has been the selection target and undergoes strong selection.

## RESULTS

### Characterization of an Introgression Line with Black Hull

To characterize the gene responsible for black seed hull coloration and phenotypic variations resulting from the transition from black to straw-white hull, we constructed a set of chromosome segment substitution lines from backcross progeny derived from a cross between cultivated rice variety *O. sativa indica* cv Guangluai 4 (straw-white hull) as the recurrent parent and wild rice accession *O. rufipogon* Griff W1943 (black hull) as the donor parent. A chromosome segment substitution line was found to carry the entire wild rice chromosome 4 (designated as SL4). The mature seeds of SL4 showed black hull, long awn (Fig. 1A), and SL4 also exhibited later heading date and strong seed shattering. Therefore, we designated the black hull locus as *Black hull4* (*Bh4*) on chromosome 4. To separate the *Shattering4* (*Sh4*) locus, which controls the seed shattering in wild rice (Li et al., 2006), we developed a near isogenic line (NIL8) that contained a short segment from wild rice in Guangluai 4 (Supplemental Fig.



**Figure 1.** Phenotypes of mature rice grains of SL4 and Guangluai 4 and the NIL8 seed hull color changes. A, Mature grains of the introgression line SL4 and the recipient parent *indica* Guangluai 4 (G4). The SL4 seed hulls are black, and the Guangluai 4 seed hulls are straw white (yellowish). B, The course of the NIL8 seed hull color change on maturing NIL8 panicles. Days after heading are indicated. From the 15th d after heading, black spot appears on the hulls covering developing seeds.

S1). NIL8 showed black seed hull but did not shatter after ripening (Supplemental Fig. S2). NIL8 was awnless and had the same heading date as Guangluai 4.

All seed hulls of the NIL8 panicles kept green from the 1st to 14th d after heading. Black spots appeared on the hulls of the first developing grains on the 15th d. Most hulls of the panicle turned black on the 19th or 20th d. The hulls became completely black on about day 25 after flowering (Fig. 1B). We also observed that the grain hulls would not turn black if the flowers were not pollinated.

### Cloning and Confirmation of the *Bh4* Gene

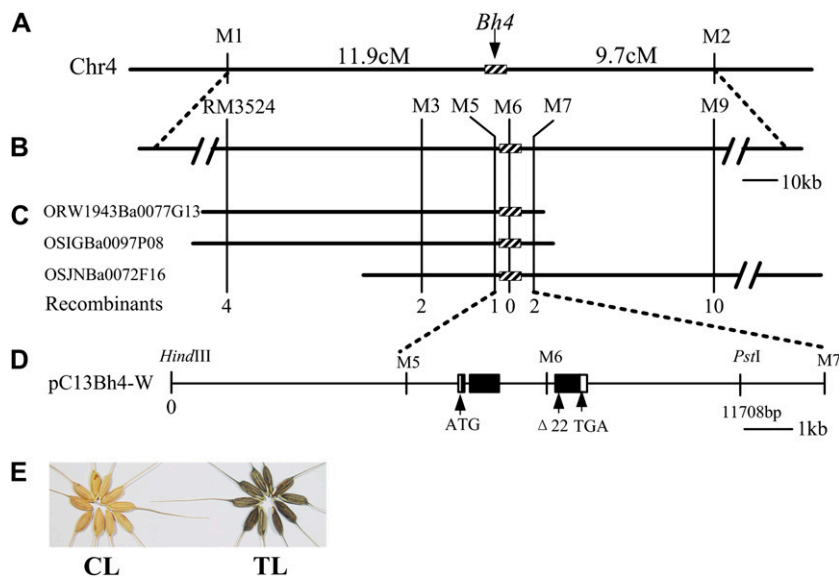
An F2 population was derived from the cross between SL4 and Guangluai 4, and a total of 300 individual F2 plants were genotyped using 28 insertion or deletion (InDel) markers along chromosome 4. All seeds of these 300 plants were harvested, and the F3 population, therefore, was derived from the F2 individuals. From the segregation of the F3 population, the genotype of each F2 individual for the *Bh4* locus was determined. With the combination of the genotype of the entire chromosome 4 and *Bh4*, *Bh4* was primarily mapped to the region between markers M1 and M2 and cosegregated with simple sequence repeat (SSR) marker RM3524 (Fig. 2A; Supplemental Fig. S3). Sub-

sequently, a larger F2 population was generated, and a total of 3,276 plants were genotyped in the seedling stage using the flanking markers M1 and M2. The genotype of the *Bh4* locus of the individual recombinants with B (Guangluai 4 homozygous) and H (heterozygous) was determined in this population, and the seeds of individual recombinants with A (SL4 homozygous) and H (heterozygous) between the two markers were collected and sown to determine the genotype of *Bh4*. By progressively examining SSR and single nucleotide polymorphism (SNP) markers between M1 and M2, we finally delimited the gene responsible for black hull in wild rice to an 8.8-kb region between the markers M5 and M7 (Fig. 2B).

A bacterial artificial chromosome (BAC) library of *O. rufipogon* W1943 was screened using the method of pool-PCR to identify the BAC clones that contained the 8.8-kb target region. Finally, two overlapped BAC clones were selected, and one of them (ORW1943-Ba0077G13) was fully sequenced (Fig. 2C). Only one open reading frame (ORF) was identified in the 8.8-kb DNA fragment using FGENESH gene prediction analysis (<http://linux1.softberry.com/berry.phtml>; Fig. 2D). Two BACs corresponding to the *Bh4* locus from cultivated rice *indica* Guangluai 4 (OSIGBa0097P08) and *japonica* Nipponbare (OSJNBa0072F16) were also identified and sequenced. In the Nipponbare annotations at the The Rice Annotation Project Database ([\[rapdb.dna.affrc.go.jp/viewer/gbrowse/build4/\]\(http://rapdb.dna.affrc.go.jp/viewer/gbrowse/build4/\)\), there was only one hypothetical gene within this region \(Os04g0460000\). To test whether the predicted gene was the \*Bh4\* gene, a complementation test was conducted. An 11,708-kb genomic DNA fragment containing the entire predicted \*Bh4\* coding region from W1943 including the 5,962-bp upstream sequence and the 3,282-bp downstream sequence was cloned into the \*Hind\*III-\*Pst\*I site of the binary vector pCAMBIA1301 to generate the transformation plasmid pC13Bh4-W for the complementation test \(Fig. 2D\). The binary plasmids were introduced into the \*Agrobacterium tumefaciens\* EHA105 strain by electroporation, and the \*indica\* cv Kasalath was transformed using \*Agrobacterium\*-mediated transformation. The vector pCAMBIA1301 was used as a control. We found that eight independent transgenic lines for pC13Bh4-W showed complementation of the black hull phenotype in mature seed hull of cv Kasalath \(Fig. 2E\). These results confirmed that Os04g0460000 in \*O. rufipogon\* was the \*Bh4\* gene controlling the synthesis of black hull pigment.](http://</a></p>
</div>
<div data-bbox=)

#### Sequence Comparison of *Bh4* in Wild Rice and Cultivated Rice

To investigate sequence variation between wild rice and cultivated rice, we compared the *Bh4* sequences of *japonica* Nipponbare and *indica* Guangluai 4 with that

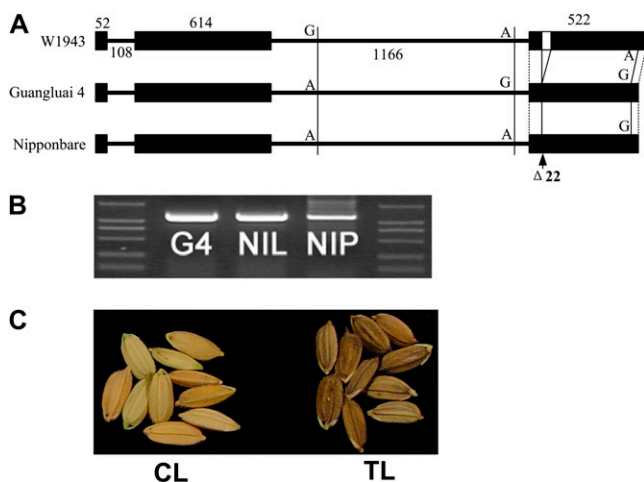


**Figure 2.** Map-based cloning and identification of the *Bh4* gene. A, The *Bh4* locus was mapped in the interval of markers M1 and M2 on the long arm of chromosome 4 (Chr4). B, The *Bh4* locus was further delimited to an 8.8-kb region between markers M5 and M7. C, Three BAC clones corresponding to the *Bh4* locus from *O. rufipogon* W1943 (ORW1943Ba0077G13), *O. sativa indica* Guangluai 4 (OSIGBa0097P08), and *O. sativa japonica* Nipponbare (OSJNBa0072F16) were identified and sequenced. The number of recombinants is indicated. D, Predicted ORF of *Bh4* and a construct of a genomic DNA fragment from W1943. The start codon (ATG) and the stop codon (TGA) are indicated. Black boxes indicate exons, white boxes indicate 5' and 3' untranslated regions, and horizontal lines between boxes indicate introns. The 22-bp deletion within the third exon of *bh4* in Guangluai 4 is indicated. The *Hind*III-*Pst*I genomic fragment from W1943 was cloned into the plant transformation vector pCAMBIA1301 to construct a plasmid, pC13Bh4-W. E, Complementation test. The straw-white seed hull of Kasalath turns black after transformation with plasmid pC13Bh4-W (TL), but it remains straw white when transformed with the control vector (CL).

of W1943. There were sequence differences in the promoter region. We designed a new marker, M10 (Supplemental Table S1), close to the site of the start codon to test the recombinant individuals and eliminated polymorphisms at the promoter as the source of the hull color change. So we focused on the mutations that could affect the protein sequence. The *Bh4* coding sequence of Guangluai 4 was identical to that of Nipponbare. Both varieties have straw-white seed hull. Sequence comparison of the *Bh4* locus between W1943 and Guangluai 4 revealed a 22-bp deletion and a SNP in the third exon in Guangluai 4 (Fig. 3A). The deletion induced a frame shift in the sequence, resulting in a premature stop codon. The stop codon truncated the protein by 150 amino acids. These results showed that the different alleles of *Bh4* were consistent with the observed phenotypic differences between Guangluai 4 and W1943, suggesting that the 22-bp deletion was the only apparent reason for the loss of the black hull phenotype in Guangluai 4 (Fig. 3A).

#### Overexpression Analysis of *Bh4* in Guangluai 4

To confirm the accuracy of the different gene models, we cloned the *Bh4* full-length cDNAs from W1943, Guangluai 4, and Nipponbare on the basis of the *Bh4* gene sequence derived from the W1943 gene model. The *Bh4* sequences could be amplified successfully from the three varieties (Fig. 3B). Sequencing of



**Figure 3.** Sequence comparison of *Bh4* and overexpression of *Bh4* in Guangluai 4. A, Comparison of the *Bh4* locus sequences from W1943, Guangluai 4, and Nipponbare. Sequence comparison of the *Bh4* locus between W1943 and Guangluai 4 reveals a 22-bp deletion in the third exon in Guangluai 4 and a SNP. Black boxes indicate exons, horizontal lines between the black boxes indicate introns, and the white box indicates a 22-bp deletion within the third exon of *bh4* in Guangluai 4. The lengths (bp) of exons and introns of *Bh4* with SNP sites are indicated. B, Full-length cDNAs of *Bh4* were detected through PCR analysis in Guangluai 4 (G4), NIL8 (NIL), and Nipponbare (NIP). C, A construct of the full-length cDNA of *Bh4* from W1943 was transformed into Guangluai 4 and turned the hulls of Guangluai 4 black. CL, Control line; TL, transgenic line overexpressing the *Bh4* full-length cDNA.

these cDNAs confirmed the gene model and the splice sites predicted from W1943. We then constructed an overexpression plasmid using the full-length cDNA of *Bh4* from W1943. Transformation of the *Bh4* overexpression plasmid succeeded in restoring the black hull phenotype in Guangluai 4 (Fig. 3C).

In addition, comparison of the *Bh4* cDNA sequence and its genomic DNA sequence in Guangluai 4 showed that there was one SNP-located coding region in the third exon. This SNP-located coding region exactly located the SNP site in the third exon between Guangluai 4 and W1943 genomic sequences. It is an adenine in W1943 and a guanine in Guangluai 4 (Fig. 3A). But the guanine turned to the same base of adenine as the wild-type W1943 allele after transcription in Guangluai 4. We also detected this transcriptional modification in Nipponbare. The function of this transition (CAT → CGT/His → Lys) might be involved in a protection mechanism that the plant adapted to random mutation in the evolutionary process.

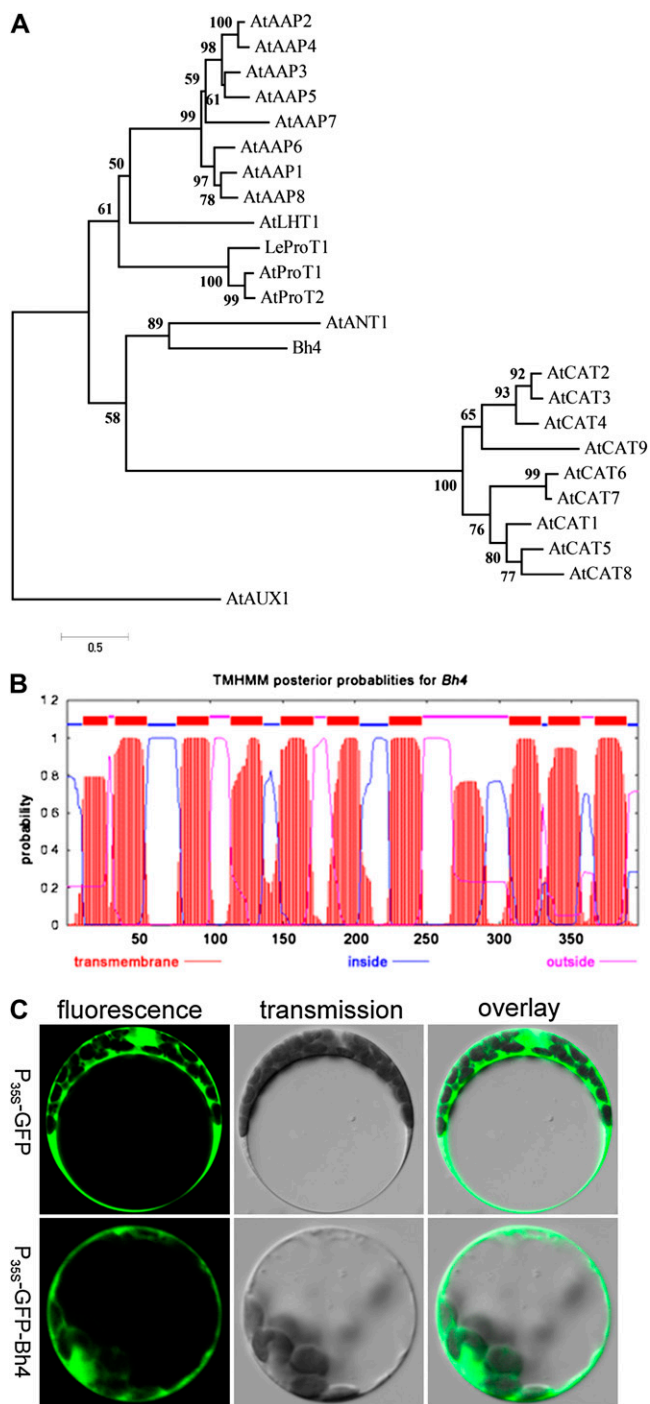
#### *Bh4* Encodes an Amino Acid Transporter Protein and Is Specifically Expressed in the Seed Hull

Sequence analysis of the reverse transcription (RT)-PCR products and RACE-PCR indicated that the *Bh4* cDNA was 1,411 bp long, with an ORF of 1,191 bp, a 65-bp 5' untranslated region, and a 155-bp 3' untranslated region. BLAST analysis revealed that the rice *Bh4* encoded an amino acid transporter. Phylogenetic comparison of BH4 protein and other amino acid transporters showed that BH4 was closely related to AtANT1, an aromatic and neutral amino acid transporter in Arabidopsis (*Arabidopsis thaliana*; Chen et al., 2001; Fig. 4A; Supplemental Fig. S4). Given the precursor for the synthesis of pigment in plants, BH4 is most likely to transport aromatic amino acids, such as Phe, Tyr, and Trp (Grotewold, 2006). We then assayed the concentration of free amino acids in straw-white hull and black hull. The result showed that the concentration of most free amino acids and the total concentration of free amino acids decreased in black hull, but the concentration of Tyr increased 3- to 6-fold (Supplemental Table S2). On the basis of this analysis, we propose that BH4 is most likely a functional Tyr transporter.

Using the TMHMM program (<http://www.cbs.dtu.dk/services/TMHMM/>), BH4 was detected to have 10 transmembrane motifs, which are specific to transporter proteins in plants (Fig. 4B). Because amino acids are transported not only across the plasma membrane but essentially in and out of all cellular compartments, the cellular localization of the transporter was analyzed. Fluorescence microscopy revealed N-terminal P35S-GFP-BH4 fusions transiently expressed in Arabidopsis protoplasts. These results indicated that BH4 was located not only at the plasma membrane but also at other cellular compartments (Fig. 4C).

To profile *Bh4* transcription, we used RT-PCR to amplify mRNA from root, leaf, stem, hull, and the corresponding dehulled seed at about 18 d after head-



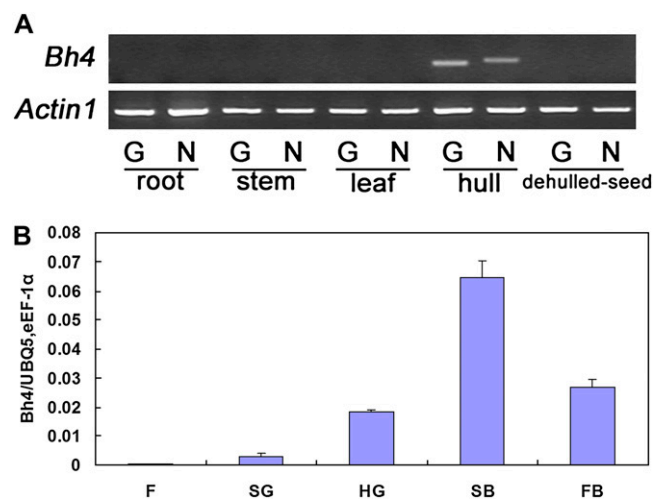


**Figure 4.** Function prediction of *Bh4*. A, Phylogenetic comparison of *Bh4* with other amino acid transporters. A phylogenetic tree of *Bh4* with other plant transporters was generated by MEGA 4.0. B, Prediction of the transmembrane domains of *Bh4* using the prediction program TMHMM ([www.cbs.dtu.dk/services/TMHMM](http://www.cbs.dtu.dk/services/TMHMM)). C, Subcellular localization of the *P<sub>35S</sub>-GFP-Bh4* fusion protein. *Bh4* cDNA was translationally fused with GFP and expressed transiently in *Arabidopsis* protoplasts (*P<sub>35S</sub>-GFP-Bh4*). GFP fluorescence was detected at the plasma membrane and some cellular compartments. GFP fluorescence in control (*P<sub>35S</sub>-GFP*) protoplasts was detected throughout the cytosol. The fluorescence (left) and bright-field (middle) images are overlaid at right.

ing. The total RNA was isolated from Guangluai 4 and NIL8 plants. RT-PCR showed no expression of *Bh4* in root, leaf, stem, or dehulled seed, as expected for *Bh4* associated with seed hull phenotype. Because the promoter of the *Bh4* gene had been eliminated as the source of polymorphism based on the recombination data, we anticipated that similar expression levels of *Bh4* would be detected in straw-white and black hull seeds. Our results confirmed this expectation (Fig. 5A), and the RNA transcripts from Guangluai 4 contained the 22-bp deletion predicted by the sequence information. We also carried out quantitative real-time RT-PCR and found that the expression level of *Bh4* reached the highest point when black spot appeared on the hull (Fig. 5B).

#### Effect of Expression of *Bh4* in Transgenic Rice

To identify functional variations resulting from the black-to-straw-white hull transition, we compared grain number per panicle, fertility rate, thousand grain weight, tiller number per plant, amylose content, gel consistency, gelatinization temperature, and grain chalkness degree between NIL8 and Guangluai 4. Only thousand grain weight, tiller number, and AC exhibited significant changes (Supplemental Fig. S5). Consequently, we further compared the three traits in *Bh4* transgenic and control lines of Kasalath in order to make sure that the changes were caused by *Bh4*. But there were no significant changes in these traits between transgenic and control lines (Supplemental Fig. S6). This indicated that *Bh4* did not associate with the yield or any grain quality.



**Figure 5.** Expression analysis of *Bh4*. A, Transcripts of *Bh4* and *Actin1* detected by RT-PCR in root, stem, leaf, hull about 18 d after heading, and the corresponding dehulled seeds from Guangluai 4 (G) and NIL8 (N). B, qRT-PCR analysis of *Bh4* transcripts in the hull of wild-type NIL8. RNA was extracted from early flowerings (F), soft green grains (SG), hard green grains (HG), spot-black grains (SB), and full-black grains (FB). The abundance of *Bh4* transcripts was normalized to that of *UBQ5* and *eEF-1α* transcripts. Data represent means with SD of three independent experiments. [See online article for color version of this figure.]

**Table 1.** Plant materials sampled in this study and information of geographic origin and functional nucleotide polymorphisms

Taxon	Accession <sup>a</sup>	Code <sup>b</sup>	Country of Origin	Cultivar Subgroup <sup>c</sup>	Phenotype <sup>d</sup>	Nucleotide Position in Genomic DNA (ATG–TGA)							
						610 <sup>e</sup>	991	1,893	1,968–1,989 <sup>f</sup>		1,089	2,457	
<i>O. sativa indica</i>	GLA4	ind-CHN1	China	IND	Y	G	A	G	DEL		C	G	
	MH63	ind-CHN2	China	IND	Y	G	A	G	DEL		C	G	
	RA4966	ind-BRA	Brazil	IND	Y	G	A	A	DEL		C	G	
	RA4921	ind-THA	Thailand	IND	Y	G	A	A	DEL		C	G	
	RA4892	ind-MDG	Madagascar	IND	Y	G	A	A	DEL		C	G	
	RA4890	ind-IRN	Iran	AUS	Y	G	A	A	DEL		C	G	
	RA4976	ind-AFG	Afghanistan	AUS	Y	G	A	A	DEL		C	G	
	RA4893	ind-IND1	India	AUS	Y	G	A	A	DEL		C	G	
	Kasalath	ind-IND2	India	AUS	Y	G	G	A	TGTGCACGCTCAACTACGGTGC	A	A	G	
	HP-256	ind-CHN3	China	NA	Y	DEL	A	A	TGTGCACGCTCAACTACGGTGC	C	G	G	
	HP-381	ind-CHN4	China	NA	Y	DEL	A	A	TGTGCACGCTCAACTACGGTGC	C	G	G	
	<i>O. sativa japonica</i>	RA4950	jap-BRA	Brazil	ARO	Y	G	A	A	DEL		C	G
		RA4913	jap-IND	India	ARO	Y	G	A	A	DEL		C	G
		RA5004	jap-MDG	Madagascar	ARO	Y	G	A	A	DEL		C	G
RA4952		jap-IRN	Iran	ARO	Y	G	A	A	DEL		C	G	
RA5363		jap-AUS	Australia	TEJ	Y	G	A	A	DEL		C	G	
RA5123		jap-USA1	United States	TEJ	Y	G	A	A	DEL		C	G	
RA5295		jap-CHN	China	TEJ	Y	G	A	A	DEL		C	G	
Nipponbare		jap-JPN	Japan	TEJ	Y	G	A	A	DEL		C	G	
RA5364		jap-NRA	Nigeria	TRJ	Y	G	A	A	DEL		C	G	
RA4875		jap-USA2	United States	TRJ	Y	G	A	A	DEL		C	G	
Lansheng		jap-CHN1	China	NA	Y	DEL	A	A	TGTGCACGCTCAACTACGGTGC	C	G	G	
Suyunuo		jap-CHN2	China	NA	Y	DEL	A	A	TGTGCACGCTCAACTACGGTGC	C	G	G	
HP-4		jap-CHN3	China	NA	Y	DEL	A	A	TGTGCACGCTCAACTACGGTGC	C	G	G	
HP-73		jap-CHN4	China	NA	Y	DEL	A	A	TGTGCACGCTCAACTACGGTGC	C	G	G	
HP-117	jap-CHN5	China	NA	Y	DEL	A	A	TGTGCACGCTCAACTACGGTGC	C	G	G		
HP-123	jap-CHN6	China	NA	Y	DEL	A	A	TGTGCACGCTCAACTACGGTGC	C	G	G		
Lijiang	jap-CHN7	China	NA	B	G	A	A	TGTGCACGCTCAACTACGGTGC	C	G	G		
<i>O. rufipogon</i>	W1943	ruf-CHN	China	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
	80505	ruf-IND1	India	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
	80529	ruf-IND2	India	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
	80472	ruf-IND3	India	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
	80528	ruf-IND4	India	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
	80542	ruf-IND5	India	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
	106122	ruf-IND6	India	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
	104311	ruf-THA1	Thailand	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
	81969	ruf-THA2	Thailand	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
	105953	ruf-IDN	Indonesia	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
	105494	ruf-MMR	Myanmar	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
	<i>O. nivara</i>	101967	niv-IND1	India	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A
		104697	niv-IND2	India	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A
		104658	niv-THA1	Thailand	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A
105801		niv-THA2	Thailand	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
105867		niv-THA3	Thailand	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
103422		niv-LKA1	Sri Lanka	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
105417		niv-LKA2	Sri Lanka	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
105454		niv-LKA3	Sri Lanka	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
105702		niv-NPL1	Nepal	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
105703		niv-NPL2	Nepal	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A	
105705	niv-NPL3	Nepal	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A		
105706	niv-NPL4	Nepal	/	B	G	G	A	TGTGCACGCTCAACTACGGTGC	C	A	A		
<i>O. barthii</i>	101196	niv-CMR	Cameroon	/	B	G	T	A	TGTGCACGCTCAACTACGGTGC	C	A	A	

<sup>a</sup>Accession numbers preceded by RA refer to Garris et al. (2005); GLA4, MH63, Nipponbare, Kasalath, Lansheng, Suyunuo, Lijiang, W1943, and HP series were collected by other authors; all other accessions were obtained from the Genetic Resources Center of the International Rice Research Institute at Los Banos, Philippines. <sup>b</sup>Individual accession is abbreviated by first three letters of the taxon name followed by the code of its origin country. <sup>c</sup>The cultivar subgroup classification followed Garris et al. (2005). NA indicates that the classification is not available for an accession. <sup>d</sup>Y indicates straw-white (yellowish) hull phenotype and B indicates black hull phenotype. <sup>e</sup>DEL represents a 1-bp deletion.

<sup>f</sup>DEL represents a 22-bp deletion.

### Multiple Independent Mutations of *Bh4* in Rice Cultivars

To investigate the occurrence of *Bh4* deficiency among rice cultivars, we first detected a 22-bp deletion among 16 *japonica* and 11 *indica* varieties using molec-

ular marker M22 (Supplemental Table S1). All cultivars examined were with straw-white seed hull. Eighteen of them had the 22-bp deletion in the site as identified in Guangluai 4; the remainder had no such deletion. We then sequenced the *Bh4* alleles of the 27

cultivars. Among the nine varieties without the 22-bp deletion, eight had a 1-bp deletion in the second exon, which also resulted in a frame shift (produced a truncated protein of 213 amino acids). One *indica* var Kasalath had a single base change from C to A (TCG → TAG), leading to a premature stop codon in the third exon (producing a truncated protein of 359 amino acids; Table I). Taken together, at least three mutations in the *Bh4* coding region of cultivated rice were associated with the straw-white hull color, suggesting multiple origins of the phenotype.

We further surveyed the distribution of the three mutations in a large panel of 433 rice landraces with straw-white hull collected in China. Among them, 94.9% (411) contained a 22-bp deletion, 3.7% (16) contained a 1-bp deletion, and the remaining 1.4% (six) had none of the three mutations. Among the six varieties, three of them had the same ORF as W1943, and another three accessions had the same ORF as *O. sativa japonica* cv Lijiang with black hull color. These results indicated that some other genes may be involved in the synthesis of the rice black hull pigment. Remarkably, the C-to-A transversion, which caused a premature stop codon, was not found in other varieties except for Kasalath (Table I).

#### Phylogenetic Analysis of *Bh4*

We obtained *Bh4* sequences from 28 cultivated rice (27 of them are with straw-white hull and one of them is with black hull) and 24 wild rice varieties with black hull (Table I). The length of aligned sequences for each taxon varied from 2,442 to 2,458 bp (Table II). All of the four taxa included both coding and noncoding regions from ATG to TGA. These sequences were aligned with an outgroup sequence of *Oryza barthii* (Table I).

Phylogenetic analysis of the *Bh4* sequences showed that all of the cultivars were grouped together with 98% bootstrap support except for *indica* var Kasalath (code ind-IND2 in Fig. 6), which was nested within a wild rice group from Thailand. Within this group, *indica* and *japonica* cultivars are intermixed. Given

their sequence differences, except Guangluai 4 and Minghui 63, which have a SNP in the second intron, they differ only in a 22-bp or a 1-bp deletion (Table I). These results suggested that those cultivars might originate from a common progenitor. Over the long time of domestication, different mutations have been selected on the *Bh4* gene. Outside of this clade, four wild rice accessions (one *O. nivara* accession and three *O. rufipogon* accessions) formed a sister group to the cultivated alleles. All of the accessions in this group came from India except W1943, and the bootstrap support came up to 83%. This result suggested that the *bh4* mutation might originate from India.

#### Neutrality Tests

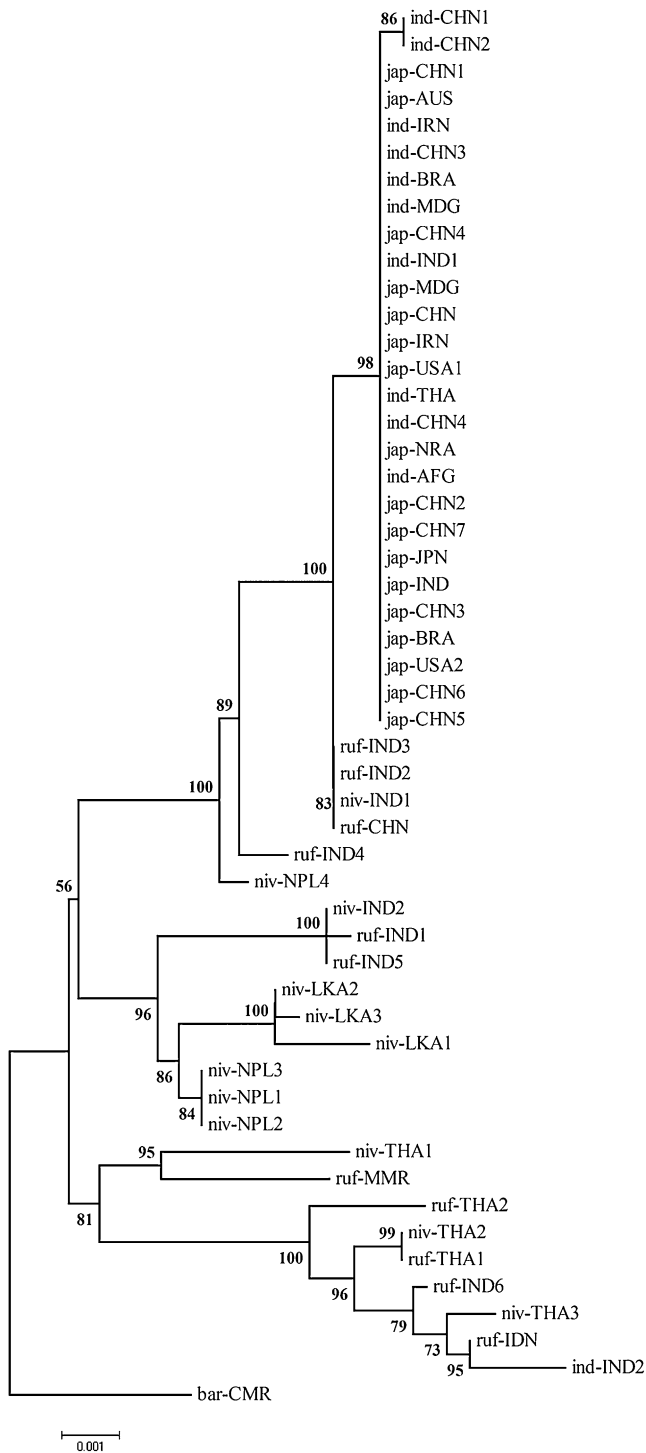
At the taxon level, polymorphisms in wild rice (Watterson's estimator of  $\theta$  per base pair calculated based on silent sites [ $\theta_{\text{sil}}$ ] = 0.00787 for *O. rufipogon* and  $\theta_{\text{sil}}$  = 0.00876 for *O. nivara*) were significantly higher than those in cultivated rice ( $\theta_{\text{sil}}$  = 0.00021 for *indica* excluding Kasalath and  $\theta_{\text{sil}}$  = 0 for *japonica*). With respect to both the entire region and silent sites, the wild rice maintained comparable levels of nucleotide diversity, which were about 70 times higher than those in the cultivated rice (Table II). A neutrality test was conducted to determine whether the reduction in nucleotide diversity in rice cultivars could be caused by artificial selection. The maximum likelihood Hudson-Kreitman-Aguade (MLHKA) test, performed in reference to seven neutral genes (Zhu et al., 2007), indicated that the test of neutrality is borderline significant in *indica* varieties ( $P < 0.025$ ) but not significant for *O. nivara* and *O. rufipogon*. Similar results were obtained for cultivated rice and wild species. The test of neutrality was significant in cultivated rice ( $P < 0.005$ ) but not significant for the wild species (Table III).

Unlike the MLHKA test, Tajima's *D* did not detect significant selection in any of four taxa (Tajima, 1989; Table III). However, those nonsignificant statistics are likely to be an artifact of population demographics (Tajima, 1989; Nielsen, 2001) and do not necessarily

**Table II.** Estimates of nucleotide variations

Taxon	N <sup>a</sup>	L <sub>T</sub> <sup>b</sup>	H <sup>c</sup>	S <sup>d</sup>	$\pi$ <sup>e</sup>	$\theta_T$ <sup>f</sup>	$\theta_{\text{sil}}^g$	$\theta_{\text{sil}}^h$
<i>O. sativa indica</i>	20	2,442	2	1	0.00014	0.00012	0.00021	0.00017
<i>O. sativa japonica</i>	34	2,442	0	0	0	0	0	0
<i>O. rufipogon</i>	22	2,450	9	52	0.00738	0.00582	0.01005	0.00783
<i>O. nivara</i>	24	2,458	10	59	0.00608	0.00643	0.00848	0.00876
Cultivated rice	54	2,442	2	1	0.00006	0.00009	0.00009	0.00014
Wild rice species	46	2,449	16	68	0.00697	0.00632	0.00963	0.00861

<sup>a</sup>Total number of sequences. <sup>b</sup>Average length (bp) of the sequences from each species or subspecies. <sup>c</sup>Total number of haplotypes at an individual locus in each taxon. <sup>d</sup>Total number of polymorphic sites. <sup>e</sup>Average number of pairwise nucleotide differences per site calculated based on the total number of polymorphic sites. <sup>f</sup>Watterson's estimator of  $\theta$  per base pair calculated based on the total number of polymorphic sites. <sup>g</sup>Average number of pairwise nucleotide differences per site calculated based on silent sites. <sup>h</sup>Watterson's estimator of  $\theta$  per base pair calculated based on silent sites.



**Figure 6.** *Bh4* phylogeny of rice cultivars and wild relatives.

indicate the absence of selection. The previous analysis of rice domestication gene *Sh4* also showed that the multilocus HKA test is more powerful in detecting artificial selection than the single-locus Tajima's *D* when polymorphisms are much reduced (Zhang et al., 2009).

## DISCUSSION

Straw-white panicles filled with ripened grains are a characteristic feature of mature cultivated rice plants. In this study, we cloned a *Bh4* gene encoding an amino acid transporter protein and identified that a 22-bp deletion within the third exon caused the loss of function of *Bh4*, resulting in the transition from black hull in wild rice to straw-white hull in cultivated rice. The 22-bp deletion accounted for 94.9% of the straw-white hull color in the screened cultivated rice varieties. Transgenic study confirmed that the *Bh4* gene controlled the synthesis of black hull pigment during seed maturation.

Although the prevalent mutation of *Bh4* was the 22-bp deletion within the third exon, there were other mutations in this gene that led to the loss of black hull pigment. This suggested that different mutations occurred in the *Bh4* gene during rice domestication, suggesting that this phenotype of cultivated rice might have multiple origins.

Why did modern rice cultivars discard the black hull trait? Did *Bh4* undergo some artificial or nature selection? MLHKA test showed that the *Bh4* gene was fixed in cultivated rice by artificial selection. It is now a natural phenomenon that the nonshattering seed of cultivated rice is always associated with straw-white hull, which is accompanied by a withering of stems and leaves near the harvest. To identify functional variations resulting from the transition of black to straw-white hull, we compared a number of yield- and grain quality-related traits between NIL8 and cv Guangluai 4 as well as transgenic lines. Since there were no significant functional variations detected to be associated with the transition from black to straw-white seed hull, we proposed that the straw-white hull for non-shattering grains was the favored visible phenotype and therefore artificially selected in cultivated rice.

Previous studies reported that *Phr1* participates in the formation of black hull (Kuriyama and Kudo, 1967) and that *Phr1* in *japonica* is a nonfunctional allele (Yu et al., 2008). We transferred the *Bh4* gene into the two subspecies of cultivated rice, *indica* (Guangluai 4, Kasalath) and *japonica* (Nipponbare). We found that only

**Table III.** Tests for neutrality and selection

Taxon	Tajima's <i>D</i>	$\chi^2$ (MLHKA) <sup>a</sup>	<i>P</i> (MLHKA) <sup>b</sup>
<i>O. sativa indica</i>	0.35195	6.3218	<i>P</i> < 0.025
<i>O. sativa japonica</i>	n.t.	n.t.	n.t.
<i>O. rufipogon</i>	-0.05211	0.0604	<i>P</i> > 0.9
<i>O. nivara</i>	-1.02514	0.095	<i>P</i> > 0.9
Cultivated rice	-0.47833	19.653	<i>P</i> < 0.005
Wild rice species	-0.69702	0.0314	<i>P</i> > 0.9

<sup>a</sup>MLHKA test was performed with *O. barthii* as the outgroup and seven neutrally evolving rice genes as the reference. The  $\chi^2$  distributed with degrees of freedom = 1 according to MLHKA. <sup>b</sup>Significance of the MLHKA test.



*indica* varieties can be rescued with black hull phenotype, which demonstrated that some other genes in this path might mutate in *japonica* varieties during domestication. These results are consistent with previous genetic analyses. As multiple mutations occurred in *Rc* (Sweeney et al., 2006) and *Phr1* (Yu et al., 2008) genes, *Bh4* also had multiple mutations. All three of these genes are related to color changes in grains and seeds.

We tested whether birds played a vital role in the selection of straw-white seed hull in rice. We proposed that if the loss of function of *Sh4* (Li et al., 2006) was selected and fixed first, nonshattered black-hull seeds in the background of yellowish stems and leaves at maturity would be easily targeted by birds. In wild rice, black-hull seeds fall off easily from plants at maturation, and the black hull color can protect them from being targeted by birds. However, nonshattered black-hull seeds of panicles were presumably easily targeted by birds in the background of the withering of stems and leaves at maturity. When a black-hull seed mutated to straw white, it would have been protected from bird predation and the allele frequency would increase over generations of cultivation. However, we found that birds almost equally favored eating straw-white hull and black hull rice grains while they were ripening in the field. It is possible that the birds around the rice field might have adapted to eating straw-white hull grains. Perhaps more rigorous field experiments have to be designed in order to replicate the process of rice domestication thousands of years ago. If this works, it might also explain why the majority of cereal crops have straw-white hulls while their wild relatives have dark-colored hulls.

We observed that the seeds began to turn black about 15 d after pollination and that seeds did not turn black if they were not pollinated. From the temporal and spatial expression of *Bh4*, the phenotype is in accordance with the expression pattern of *Bh4*. We think that there must be something transported from grain to hull for the black hull formation. Further study will be required to investigate which substance leads to black hull at maturity.

## MATERIALS AND METHODS

### Plant Materials

In the cloning of *Bh4* and expression analysis, SL4 and NIL8 were used. SL4 is an introgression line developed by introgressing chromosomal segments from an accession of common wild rice (*Oryza rufipogon*) W1943 into an *indica* cultivar (*Oryza sativa*), Guangluai 4, based on four generations of backcrossing and four generations of self-fertilization. All the genetic background of SL4 comes from Guangluai 4 except for the entire chromosome 4. NIL8 is derived from the cross between SL4 and Guangluai 4 (Supplemental Fig. S1).

In phylogenetic analysis and neutrality test, we sampled 28 accessions of cultivated rice, including 11 accessions of *O. sativa indica* and 17 accessions of *O. sativa japonica*. This sample represents all five recognized subgroups within *O. sativa* (Garris et al., 2005). A total of 24 accessions of the wild rice progenitors were sampled, including 11 accessions of *O. rufipogon*, 12 accessions of *Oryza nivara*, and one accession of *Oryza barthii* (Table I). These collections covered a wide distributional range of the wild rice from southeastern Asia to India, including all potential areas for rice domestication. *O. barthii* was used as an outgroup for phylogenetic analyses.

### Measurement of Grain Quality Traits

Mature rice grains were milled after being harvested, air dried, and stored at room temperature for 3 months. Amylose content, gel consistency, and gelatinization temperature were measured according to methods reported by Tan et al. (1999). Gelatinization temperature was evaluated as the alkali spreading value.

For chalkness degree analysis, we adopted the image autorecognize chalkness system. Briefly, first, rice seeds were dehulled and the seed coat removed. Then, the polished seeds (intact) were imaged through a scanner, and using specific software, the chalkness region was recognized automatically by color difference between chalkness and endosperm. The chalkness degree was calculated by dividing the area of chalkness by the area of intact seed. All this work was done automatically by computer. Each sample was repeated four times, and each time we analyzed 50 seeds and obtained the average chalkness degree.

### DNA Extraction and Molecular Marker Analysis

DNA was extracted from fresh leaves according to the cetyl-trimethylammonium bromide method (Murray and Thompson, 1980) with minor modifications. The molecular markers analyzed in the target region contained InDels, SNPs, cleaved-amplified polymorphic sequences, and one SSR marker, RM3524; they were all newly designed in this study except RM3524. For InDels and SSR, PCR products were loaded on agarose gels to assay polymorphism. For SNPs, PCR fragments were directly sequenced in both directions by the ABI 3730 Sequence instrument after purification by agarose gels. For cleaved-amplified polymorphic sequences, PCR fragments were digested with specific restriction endonuclease enzyme and then loaded on agarose gels for polymorphism. The information for newly developed markers is given in Supplemental Table S1.

### Positive BAC Clone Screening

The BAC library comprises 80 384-well microtiter plates. The coverage of the library is about eight times the rice genome. To get the specific clone, three PCR rounds were adopted. In the first round, the 80 384-well microtiter plates were 80 superpools, and DNA was extracted from each superpool by the alkaline lysis method for PCR screening. Each superpool was screened with the two markers flanking the target region. The superpools that were positively screened by two markers were selected for the next PCR screening round. Each superpool was divided into four 96-well plates, which formed four subpools. The subpools that were also positively screened by two markers were selected for the third PCR screening round. In the third round, each positive pool was divided into eight-row pools and 12-column pools. The target clone was specifically detected after three screening rounds.

### Subcellular Localization of BH4

Transient transformation of the protoplasts with polyethylene glycol was performed according to the protocol of Negruțiu et al. (1992). For confocal laser scanning microscopy, protoplasts were incubated overnight at 24°C in the dark after transformation and observed with an Olympus microscope.

### Quantitative Real-Time PCR

Total RNA was extracted from rice tissues using TRIZOL reagent (Invitrogen) as described by the supplier. Two micrograms of RNA was reverse transcribed with oligo(dT)<sub>18</sub> primer using the PrimeScriptRT reagent kit (Takara). For quantitative real-time RT-PCR, first-strand cDNAs were used as templates in real-time PCR using the SYBR Green PCR Master Mix (Takara) according to the manufacturer's instructions. The amplification of the target genes was analyzed using the ABI Prism 7500 Sequence Detection System and Software (PE Applied Biosystems). *UBQ5* (AK061988) and *eEF-1α* (AK061464; Jain et al., 2006) were used as controls to normalize all data.

### *Bh4* Allele Sequencing

A 2,465-bp region of the *Bh4* gene was amplified using PCR primers Bh4aF (5'-CTCCAAGATGACCCTGCAIT-3') and Bh4eR (5'-AGGCCACCACAC-TAATCGAC-3'). This region includes the entire coding and noncoding sequences of *Bh4* from ATG to TGA. The PCR and internal sequencing primers were designed based on W1943 genome sequences. The *Bh4* region

was amplified with Takara LA Taq DNA polymerase with GC buffer I (Takara). The PCR products were purified with a purification kit from Tiangen. All PCR products were sequenced directly on both strands with the primers Bh4a, Bh4b, Bh4c, Bh4d, and Bh4e (Supplemental Table S1).

## Phylogenetic Analyses of DNA Sequences

A phylogenetic tree of the sequenced accessions (excluding heterozygote accessions) was reconstructed by the neighbor-joining (NJ) method (Saitou and Nei, 1987) based on the two-parameter distances of Kimura (1980). MEGA version 4.0 (Tamura et al., 2007) was used to perform the phylogenetic reconstruction. Bootstrap values were estimated (with 1,000 replicates) to assess the relative support for each branch. All positions containing alignment gaps were eliminated in pairwise sequence comparisons in NJ analyses. The NJ tree is shown rooted by the midpoint to improve clarity.

## Neutrality Tests

Sequence alignments were performed using ClustalX (Thompson et al., 1997) and were refined manually. Genetic variation was estimated with average pairwise differences per base pair between sequences ( $\pi$ ; Nei and Li, 1979) and Watterson's estimate ( $\theta$ ; Watterson, 1975) based on the number of segregating sites using DnaSP version 4.10 (Rozas et al., 2003). Under the standard neutral model for an autosomal gene from a random-mating population of a constant size, both  $\pi$  and  $\theta$  are expected to be equal to  $4Ne\mu$ , where  $Ne$  represents the effective population size and  $\mu$  represents the mutation rate per generation per site (Watterson, 1975).

Selection on *Bh4* was tested using two methods. Departure from neutrality was tested at segregating nucleotide sites with Tajima's  $D$  using the program DnaSP (Rozas et al., 2003). We also performed the MLHKA test (Wright and Charlesworth, 2004). The maximum likelihood ratio test assesses departure from neutrality at a focal locus compared with neutral standards. Seven genes (*Adh1*, *GBSSII*, *Ks1*, *Lhs1*, *Os0053*, *SSIII*, and *TFIIA*; Zhu et al., 2007) were used as reference sequences, and *O. barthii* was used as the outgroup for the MLHKA test. The test was performed using the program MLHKA provided by S.I. Wright ([http://labs.eeb.utoronto.ca/wright/Stephen\\_I.\\_Wright/MLHKA/](http://labs.eeb.utoronto.ca/wright/Stephen_I._Wright/MLHKA/)). With different random numbers of seeds, Markov chain lengths of 100,000 were run in the MLHKA tests to specify each gene in each accession, and at least three independent runs per model were performed to assess the maximum likelihood values.

Sequence data from this article can be found in the GenBank/EMBL data libraries under the following accession numbers: cDNA from W1943 (FQ377518), cDNA from Guangluai 4 (FQ377519), ORW1943Ba0077G13 (FQ377585), and *Bh4* genomic sequences of 52 cultivated and wild rice varieties (FQ377520–FQ377566 and FQ377579–FQ377583).

## Supplemental Data

The following materials are available in the online version of this article.

**Supplemental Figure S1.** The size of the W1943 segment in NIL8.

**Supplemental Figure S2.** Plant and panicle architecture of NIL8 and Guangluai 4.

**Supplemental Figure S3.** Some of the F2 individuals used for *Bh4* primary mapping.

**Supplemental Figure S4.** Multiple sequence alignment of the rice BH4 protein and the amino acid transporter proteins from Arabidopsis and tomato (*Solanum lycopersicum*).

**Supplemental Figure S5.** Comparison of grain number per panicle (GNP), fertility rate, thousand grain weight (TGW), tiller number per plant (TN), amylose content (AC), gel consistency (GC), gelatinization temperature (GT), and chalkness degree (CD) between NIL8 and Guangluai 4 (GLA4).

**Supplemental Figure S6.** Comparison of tiller number (TN; A), thousand grain weight (TGW; B), and amylose content (AC; C) between control (Ka-1301) and transgenic (Ka-Bh4) lines.

**Supplemental Table S1.** Primers used in this study.

**Supplemental Table S2.** Concentration of free amino acids in seed hull.

## ACKNOWLEDGMENTS

We thank the China National Rice Research Institute for helping assay rice grain quality and Linbin Zhang for help with population genetic analyses.

Received October 28, 2010; accepted January 21, 2011; published January 24, 2011.

## LITERATURE CITED

- Chao LF (1928) Linkage studies in rice. *Genetics* **13**: 133–169
- Chen L, Ortiz-Lopez A, Jung A, Bush DR (2001) ANTI1, an aromatic and neutral amino acid transporter in Arabidopsis. *Plant Physiol* **125**: 1813–1820
- Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohtsubo E (2003) Polyphyletic origin of cultivated rice: based on the interspersed pattern of SINEs. *Mol Biol Evol* **20**: 67–75
- Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169**: 1631–1638
- Grotewold E (2006) The genetics and biochemistry of floral pigments. *Annu Rev Plant Biol* **57**: 761–780
- Gu XY, Kianian SF, Hareland GA, Hoffer BL, Foley ME (2005) Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). *Theor Appl Genet* **110**: 1108–1118
- Jain M, Nijhawan A, Tyagi AK, Khurana JP (2006) Validation of house-keeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochem Biophys Res Commun* **345**: 646–651
- Jin J, Huang W, Gao JP, Yang J, Shi M, Zhu MZ, Luo D, Lin HX (2008) Genetic control of rice plant architecture under domestication. *Nat Genet* **40**: 1365–1369
- Khush GS (1997) Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol* **35**: 25–34
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**: 111–120
- Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication. *Science* **312**: 1392–1396
- Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. *Trends Genet* **23**: 578–587
- Kuriyama H, Kudo M (1967) Complementary genes Ph and Bh controlling ripening black coloration of rice hulls and their geographical distribution. *Jpn J Breed* **17**: 13–19
- Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. *Science* **311**: 1936–1939
- Li P, Wang Y, Qian Q, Fu Z, Wang M, Zeng D, Li B, Wang X, Li J (2007) LAZY1 controls rice shoot gravitropism through regulating polar auxin transport. *Cell Res* **17**: 402–410
- Maekawa M (1984) Geographical distribution of the genes for black hull coloration. *Rice Genet Newsl* **1**: 104–105
- Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* **8**: 4321–4325
- Nagao S, Takahashi M (1954) Genetical studies on rice plant. XVI. Some genes responsible for yellow, brown and black color of glume. *Jpn J Breed* **4**: 25–30
- Negrutiu I, Hinnisdaels S, Cammaerts D, Cherdshewasart W, Gharti-Chhetri G, Jacobs M (1992) Plant protoplasts as genetic tool: selectable markers for developmental studies. *Int J Dev Biol* **36**: 73–84
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* **76**: 5269–5273
- Nielsen R (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity* **86**: 641–647
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425
- Sang T, Ge S (2007) The puzzle of rice domestication. *J Integr Plant Biol* **49**: 760–768
- Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M

- (2008) Deletion in a gene associated with grain size increased yields during rice domestication. *Nat Genet* **40**: 1023–1028
- Sweeney MT, Thomson MJ, Pfeil BE, McCouch S** (2006) Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* **18**: 283–294
- Tajima F** (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595
- Tamura K, Dudley J, Nei M, Kumar S** (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599
- Tan L, Li X, Liu F, Sun X, Li C, Zhu Z, Fu Y, Cai H, Wang X, Xie D, et al** (2008) Control of a key transition from prostrate to erect growth in rice domestication. *Nat Genet* **40**: 1360–1364
- Tan YF, Li JX, Yu SB, Xing YZ, Xu CG, Zhang Q** (1999) The three important traits for cooking and eating quality of rice grains are controlled by a single locus in an elite rice hybrid, Shanyou 63. *Theor Appl Genet* **99**: 642–648
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG** (1997) The CLUSTAL\_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882
- Watterson GA** (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**: 256–276
- Wright SI, Charlesworth B** (2004) The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* **168**: 1071–1076
- Yu Y, Tang T, Qian Q, Wang Y, Yan M, Zeng D, Han B, Wu CI, Shi S, Li J** (2008) Independent losses of function in a polyphenol oxidase in rice: differentiation in grain discoloration between subspecies and the role of positive selection under domestication. *Plant Cell* **20**: 2946–2959
- Zhang LB, Zhu Q, Wu ZQ, Ross-Ibarra J, Gaut BS, Ge S, Sang T** (2009) Selection on grain shattering genes and rates of rice domestication. *New Phytol* **184**: 708–720
- Zhu Q, Zheng X, Luo J, Gaut BS, Ge S** (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol Biol Evol* **24**: 875–888