Haplotype variation at *Badh2*, the gene determining fragrance in rice

Gaoneng Shao [a,1], Shaoqing Tang [a,1], Mingliang Chen [b,1], Xiangjin Wei [a], Jiwai He [a], Ju Luo [a], Guiai Jiao [a], Yichao Hu [c], Lihong Xie [a], Peisong Hu [a,*]

[a] *State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou 310006, China*
[b] *Jiangxi academy of agricultural sciences, Nanchang 330200, China*
[c] *Hunan agricultural university, Changsha 410128, China*

A B S T R A C T

Fragrance is an important component of end-use quality in rice. A set of 516 fragrant rice accessions were genotyped and over 80% of them carried the *badh2.7* allele. A subset of 144 mostly fragrant accessions, including nine of *Oryza rufipogon*, was then subjected to a detailed diversity and haplotype analysis. The level of linkage disequilibrium in the *Badh2* region was higher among the fragrant accessions. Re-sequencing in the *Badh2* region showed that *badh2.7*, *badh2.2* and *badh2.4–5* all arose in the *japonica* genepool, and spread later into the *indica* genepool as a result of deliberate crossing. However, loss-of-function alleles of *Badh2* are also found in the *indica* genepools, and then transferred into *japonica*. Evidence for three new possible FNPs was obtained from the *Badh2* sequence of 62 fragrant accessions. Based on these data, we have elaborated a model for the evolution of *Badh2* and its participation in the rice domestication process.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Rice feeds about one third of the world's population. The two domesticated forms of rice are *Oryza sativa* and *Oryza glaberrima*; the former was domesticated from its wild ancestor *O. rufipogon* ~9000 years ago [1,2]. Genomic analysis has demonstrated that the two major subspecies of *O. sativa* (*japonica* and *indica*) arose from distinct *O. rufipogon* populations [3–7]. A small number of genes have been associated with the domestication process in the major crop species [8–12], while in rice, the domestication syndrome genes *GS3*, *rc*, *wx*, *sh4*, *qSH1* and *sd1* have all been isolated and extensively characterized [13,14]. The aroma of the rice grain ("fragrance") is an important consumer trait, and is largely controlled by allelic variation at *Badh2*, a gene which comprises 15 exons [15]. The full length BADH2 protein is associated with non-fragrance [16]. *Badh2* sequence variants associated with fragrance include an 8 bp deletion and three single nucleotide polymorphisms (SNPs) in exon 7, and a 7 bp deletion in exon 2 [15,17]; these polymorphisms are referred to as "functional nucleotide polymorphisms" or FNPs. The wild type BADH2 protein catalyzes the oxidation of a 2-acetyl-1-pyrroline-aminobutyraldehyde (2AP) precursor, so that a non-functional allele results in the enhanced synthesis of 2AP [18]. Fragrant accessions have been identified in various subspecies of rice, including notably the "Basmati" types, and within both the *indica* and the *japonica* genepools, and the fragrant allele has long been thought to have arisen from within the *indica* genepool.

An increasing number of fragrant rice cultivars lacking any of the known *Badh2* FNPs have been identified, suggesting the possibility that the trait may be in some cases controlled by gene(s) other than, or in addition to *Badh2*. Here, we have analyzed the *Badh2* genotype of a panel of 144 rice accessions (including a representation of *O. rufipogon*), some of which are fragrant and others not. Our objective was to catalogue the range of haplotype variation present at *Badh2*, with a view to shedding new light on the domestication of *indica* and *japonica* rice.

2. Materials and methods

2.1. Germplasm and the determination of fragrance

The entries making up the full collection of 549 accessions originated from 15 countries, and comprised 516 (405 listed in Table S1 and 111 in Table S2) classified previously as fragrant and 33 as non-fragrant (Table S2). Nine of the accessions were *O. rufipogon* (Table S2). The full set of material was field-grown at Hangzhou (Zhejiang Province) during 2009. The leaf fragrance of each entry was determined following [19]. About 2 g of green leaf harvested from plants at the tillering stage was sliced and immersed for 10 min in 10 mL 1.7% KOH at room temperature, after which the fragrance was graded independently by three operators. The determination of grain fragrance followed the protocol described in ref. [20]. Out of a random sample of 16 mature grains per line, if none were fragrant, the entry was deemed to be non-fragrant; if five consecutive sampled grains were all fragrant, the entry was deemed to be

* Corresponding author. Fax: +86 571 63370080.
  E-mail addresses: hupeisong@yahoo.com.cn, riceh@caas.net.cn (P. Hu).
[1] These authors contributed equally to this paper.

fragrant; and if some of the 16 grains were fragrant and others not, then the entry was deemed to be heterozygous.

2.2. DNA extraction, SSR analysis, primer synthesis and re-sequencing

Template DNA for the PCRs was extracted from leaf tissue using the CTAB method. Each 20 μL reaction contained 2 μL 10× PCR buffer (25 mM MgCl$_2$), 1.6 μL 2 mM dNTP, 2 μL of each SSR primer (5 μM), 1 U Taq DNA polymerase and 2 μL genomic DNA. A set of 48 SSR assays was applied to a subset of 144 accessions (Table S3). For primer sequences, refer to www.gramene.org. The amplification regime consisted of an initial denaturation of 94 °C/2 min, followed by 35 cycles of 94 °C/45 s, 55 °C/45 s, 72 °C/60s, and ending with a 72 °C/8 min final extension. Amplicons were electrophoresed through non-denaturing polyacrylamide gels, and visualized by silver staining. Details of the 12 primer pairs used to amplify segments of Badh2 for the purpose of re-sequencing [17,18], along with those for the three pairs each targeting a functional region of the gene, as described in ref. [21]. The amplicons in this case were electrophoresed through 1.2% w/v agarose gels, recovered using a TIAN gel Midi Purification Kit (TIANGEN), inserted into pGEM-T Easy Vector (Promega), and subsequently transformed into competent DH5α Escherichia coli cells. DNA sequencing was performed by a commercial company (Invitrogen, Shanghai).

2.3. Determination of genotypic diversity and population structure

The SSR allelic data for the 144 accessions (Table S2) was analyzed using MEGA v4.1 software in combination with PowerMarker v3.25 to generate an UPGMA dendrogram based on neighbor-joining [22]. The admixture model and the observed allele frequencies were adopted to obtain a representation of the population structure present using STRUCTURE v2.2 software [23]. Both the "length of burn-in period" and the "number of MCMC reps after burn-in" were set to 10,000.

2.4. Blast, haplotype and genetic diversity analysis of the Badh2 gene

Various regions of the Badh2 were re-sequenced for the 144 accessions, including a segment ~75 K bp upstream of the gene, the promoter, the 3′-UTR and a segment ~46 K bp downstream of the gene. The sequences were spliced by software contig and the resulting contigs were used as BLAST queries using ClustalW v1.8. The aligned sequences were then imported into TASSEL software to identify the polymorphisms present at a frequency >3%, leading to the recognition of Badh2 haplotypes [24]. The polymorphism data were used to estimate linkage disequilibrium (LD) using the software package DnaSP v4.52 [25].

3. Results

3.1. Allelic variation in Badh2

When the 516 fragrant rice accessions were classified at each of the three known Badh2 FNPs [17,21], four alleles were recognized: badh2.7 carries an 8 bp deletion in exon 7 (similar to badh2.1 reported in the ref. [18]), badh2.2 a 7 bp deletion in exon2 (the same to badh2.2 reported in the ref. [18]), badh2.4–5 an 806 bp deletion involving intron 4 and parts of exons 4–5, and the wild type (badh-wt) (Table 1). The bulk (~80%) of the fragrant rice varieties were of the badh2.7 type.

3.2. Determination of and allelism test for fragrance

516 accessions classified as fragrant (Tables 1, S1–2) were verified as having fragrance by means of the leaf and grain tests. Of these, 62 possessed none of the three known Badh2 FNPs, and so were crossed

Table 1
Descriptors of 549 rice accessions. Fragrance determination, allelism test and functional marker analysis are integrated to assess the Badh2/badh2 of those materials.

Allele	No.	Frg(+)/non-frg(−)
badh2.7	416	+
badh2.2	30	+
badh2.4–5	8	+
badh2-wt	62	+
Badh2	24	−
O. rufipogon	9	−

to both cv. Zhong2A and cv. Zhong3A to allow for an allelism test for fragrance; both these lines are male sterile, but while the former is fragrant, the latter is not. The F$_1$ seedling leaves and the F$_1$ hybrid involving cv. Zhong2A as the parent were all fragrant (data not shown), but those having cv. Zhong3A as the parent were all non-fragrant (data not shown). This result was consistent with the fragrant trait in all of the 62 accessions being determined by a Badh2 allele.

3.3. SSR diversity and genetic structure analysis

The phylogeny of the 144 accessions based on SSR genotype demonstrated the presence of three major groups, corresponding to indica types, japonica types and O. rufipogon accessions. Some admixture between indica and japonica has clearly occurred (Fig. 1). The badh2.7 allele was distributed among both the indica and the japonica groups, while both badh2.2 and badh2.4–5 were restricted to the japonica group. Thus both badh2.2 and badh2.4–5 acted as the mutant genes for Badh2, and fixed in the fragrant rice varieties. The structure-based analysis provided evidence for a significant population structure in the germplasm set, with three being the most likely number of distinct groups (the K parameter). The three clusters corresponded to the indica and the japonica types, and the O. rufipogon accessions (Fig. 2).

3.4. LD around the Badh2 locus

When the squared allele frequency correlations were related to the position of the various SNPs and indels, the resulting regressions demonstrated that the rate of LD decay among the non-fragrant accessions was less than among the fragrant ones, while the wild types were more stable across the whole gene Badh2 (Fig. 3). The expected value of r^2 for the non-fragrant rice varieties declined to 0.1 at a distance of >5 K bp, whereas the level remained >0.1 for at least 7.5 K bp in the fragrant varieties. This difference in LD can arise as a result of variable genome organization and/or of population structure. These results also indicated that the Badh2 gene was an evolution gene and differential between the fragrance and non-fragrance rice varieties.

3.5. Haplotype analysis of Badh2

The re-sequenced segments of Badh2 covered 5.86 K b of the reference cv. Nipponbare sequence. The analysis revealed the presence of >100 variable sites, including SNPs, indels and a TA microsatellite; of these, 37 present at a frequency of >3% were used to define 27 haplotypes of Badh2 (Fig. 4). Ten of the haplotypes were restricted to a group of accessions dominated by japonica types, and the other 17 to a group composed of indica types and O. rufipogon accessions. Haplotype H1 was identical to badh2.7 defined the presence of an 8 bp deletion and three SNPs in exon 7, and represented ~20% of the accessions (in 144 rice accessions). It is found in both japonica and indica types, although is thought to have originated in the japonica genepool [18], a notion confirmed by the observation that the polymorphic sites in the key exon 7 region were mostly of japonica origin (Figs. 4–5; Tables S4–5). Twelve fragrant rice varieties
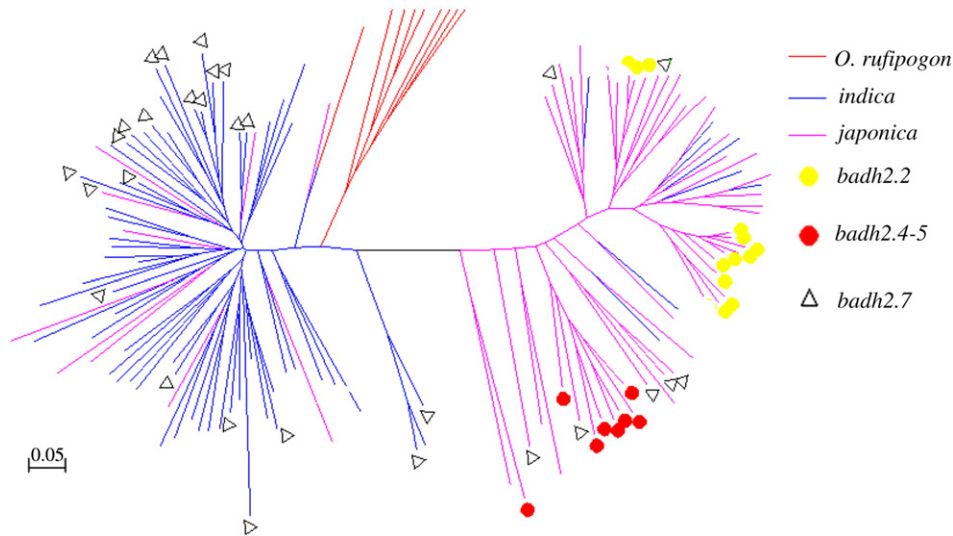
Fig 1. Diversity analysis of 144 rice accessions based on 48 SSR markers. The three clusters generated corresponded to the *indica* and *japonica* types, and *O. rufipogon*. The *badh2.7* allele (H1) was present both in the *indica* and *japonica* groups, while *badh2.2* (H2) and *badh2.4–5* (H3) were restricted to the *japonica* group.

carried the H3 haplotype (*badh2.2*), and eight H2 (*badh2.4–5*); both H2 and H3 appear to have arisen within the *japonica* genepool (Figs. 2, 4), as confirmed from their sequence in four other regions (Fig. 5). H4 had no variance in exons while have a consistent allele across the *badh2* from *japonica* group. Unacceptable, they all appeared in the *indica* fragrant rice varieties. It can be possible that this haplotype may have originated from *japonica* and later entered and fixed in the *indica* genepool, since a segment of the downstream ~45 K bp sequence of H4 was of the *japonica* type (Fig. 5). H5 and H9 both had the same phenomenon with H4, and the recombinant site may have taken place far from the *Badh2* locus. The sequence associated with haplotypes H6, H7 and H10 suggested that these all arose in the *japonica* genepool.

H15, with no variance in exons, was a significant allele of *badh2* that appeared in *indica* types. Nine non-fragrant *indica* and two fragrant *japonica* varieties shared haplotype of H15. The sequence context suggested that the H15 haplotype was of the *indica* type, which clarified that the allele *badh2* in *japonica* may from the *indica* type with the trace of two recombinants happened in the upstream of *badh2* (Figs. 4–5; Tables S4–5). Haplotypes H11–12, 14 and 16–18 might perform the same results with H15. H19 were verified that the allele *badh2* moved from *indica* type to the *japonica* type. H21–27 were only found in *O. rufipogon* accessions with rich allele variances in *badh2*, which demonstrated that the allele *badh2* are mainly similar to the *indica* types, while the around sequence are *japonica* type. That may due to the differential of *indica/japonica* from the *O. rufipogon*.
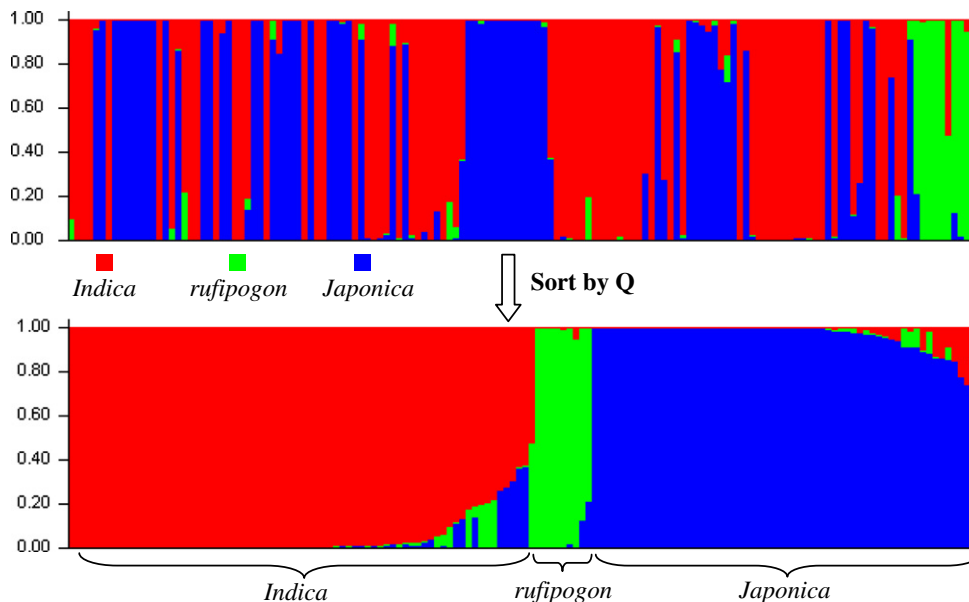
Fig. 2. Structure analysis of 144 rice accessions. Three groups were obtained, corresponding to the *indica* and *japonica* types, and *O. rufipogon*. The analysis provides evidence of some inter-group introgression.
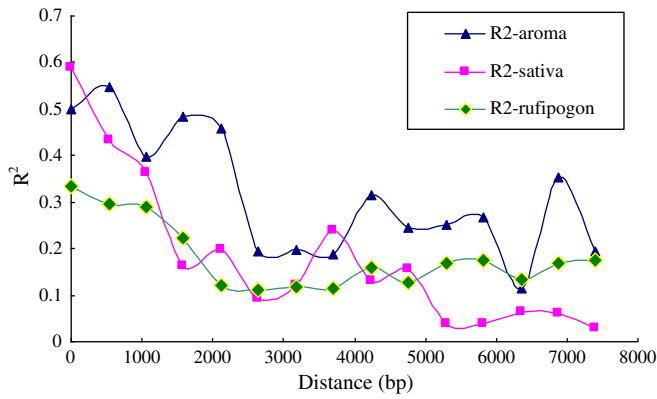
Fig. 3. LD analysis at *Badh2*. The regression curves showed that the decay in LD is more rapid among the non-fragrant varieties.

The overall conclusion is that the origin of fragrance among *indica* types is associated with sequence variation at *Badh2*.

3.6. Novel deletions in Badh2 associated with fragrance

Certain deletions in *Badh2* result in the enhanced synthesis of 2AP and therefore fragrance. The common H1 haplotype involves an 8 bp deletion in exon 7, while the less common ones H2 and H3 consist of, respectively, a 7 bp deletion in exon 2 and an 806 bp deletion involving segments of exons 4 and 5. The re-sequencing of the 62 fragrant accessions which lacked any of the three prior known FNPs revealed a 75 bp deletion in exon 2 and a SNP in each of exons 10 and 13 (Table 2). Each of these variants was present in two accessions, but the causative sequence variants for the other 56 accessions could not be identified.

4. Discussion

4.1. Distinguishing indica from japonica types

Identifying a diagnostic to distinguish between *indica* and *japonica* types has occupied a deal of research effort [26]. Early attempts based on isoenzyme alleles were partially successful, but more recently SSR markers have been developed which are highly diagnostic and easy to apply. Thus, for example, genotyping on the basis of 169 nuclear SSRs and two chloroplast loci was able to effectively sort a large germplasm panel into the five recognized types, i.e., *indica*, *aus*, *aromatic*, *temperate japonica* and *tropical japonica* [7]. SSRs have also been proven useful as a means of assessing the genetic diversity represented in germplasm panels, as exemplified by the analysis of 428 accessions (including *O. rufipogon*, *nivara*, *spontanea* and *sativa*) using 36 SSRs [27]. SSR genotyping has shown that fragrant rice accessions can be either of the *indica* or the *japonica* type [28]. Here, we applied a set of 48 SSRs to classify 144 accessions, and were able to readily recognize the three groups *indica*, *japonica* and *O. rufipogon*. Although these findings suggested that the divergence between *indica* and *japonica* attributed to the differentiation of the same ancestral *O. rufipogon* population in different locations and at different times, different genetic backgrounds have been stayed in the *indica* or *japonica* groups that may genotype flowing with each other. In a word, SSR cluster and structure analysis not only can explain the molecular difference between *indica* and *japonica* rice, but also be helpful for offering useful genetic information to breeders and improving the selective effect of useful germplasm resource.

4.2. Allelic diversity and fragrance variance at Badh2/badh2

Fragrance is an important aspect of the end-use quality of rice. It has been noted previously that the biochemical basis of fragrance

Fig. 4. *Badh2* haplotyping of 144 rice accessions.

Fig. 5. Haplotype analysis of the sequence covering four genomic regions in the vicinity of *Badh2*.

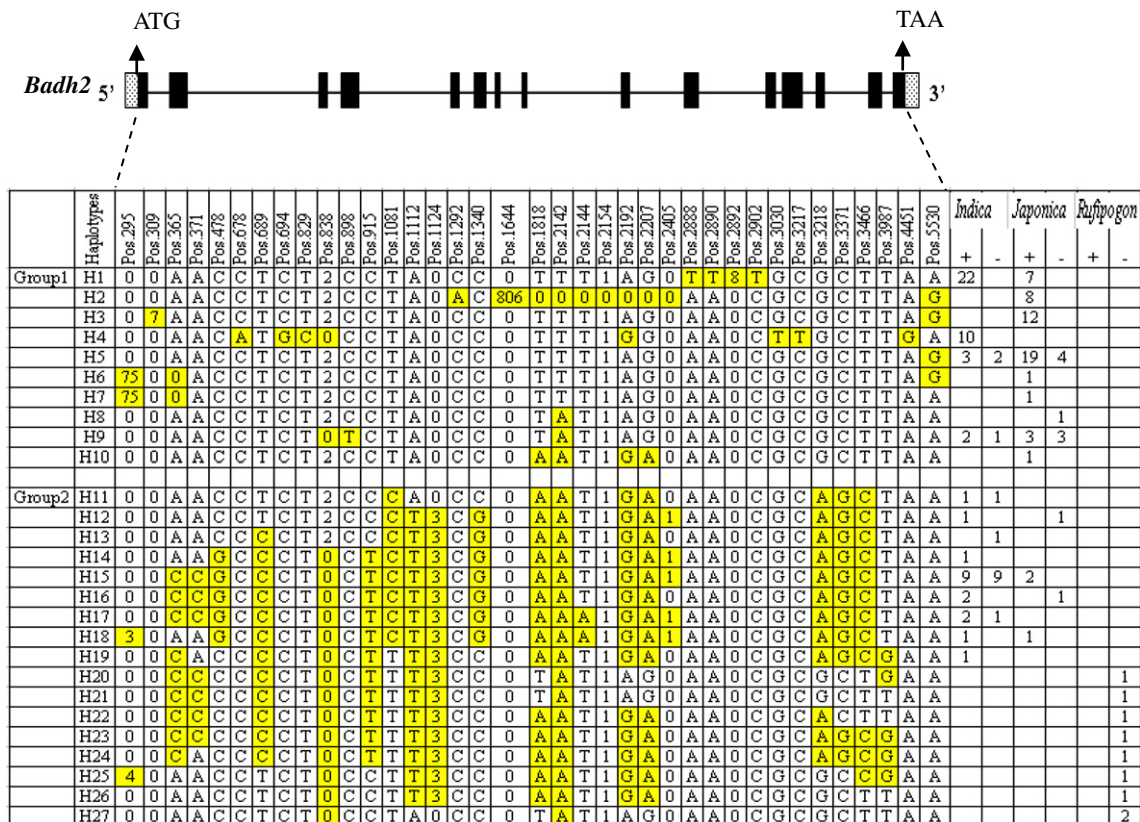		Haplotypes	Up-Badh2			Promoter and 5'UTR of Badh2																																										3'UTR	Down-Badh2		
		Pos.-95996	Pos.-95970	Pos.-95785	Pos.-1010	Pos.-988	Pos.-965	Pos.-962	Pos.-957	Pos.-937	Pos.-925	Pos.-876	Pos.-843	Pos.-842	Pos.-841	Pos.-833	Pos.-817	Pos.-794	Pos.-764	Pos.-738	Pos.-733	Pos.-726	Pos.-695	Pos.-670	Pos.-633	Pos.-605	Pos.-599	Pos.-562	Pos.-554	Pos.-536	Pos.-510	Pos.-493	Pos.-489	Pos.-454	Pos.-448	Pos.-421	Pos.-387	Pos.-357	Pos.-350	Pos.-325	Pos.-324	Pos.-323	Pos.-315	Pos.-309	Pos.-304	Pos.-181	Pos.-167	Pos.6081	Pos.45202	Pos.45209	Pos.45258
Group1 H1	H1	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	T	T	A
	H2	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	T	T	A
	H3	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	T	T	A
	H4	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	**C**	C	G	T	G	C	T	T	C	G	T	0	**A**	T	A	A	T	T	T	A	G	28	T	G	0	**C**	**C**	**G**
	H5	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	T	T	A
	H6	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	T	T	A
	H7	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	**43**	T	T	A
	H8	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	T	T	A
	H9	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	T	T	A
	H10	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	T	T	A
Group2 H11	H11	A	A	A	C	G	C	G	A	C	A	C	1	A	A	G	T	3	T	A	C	A	22	C	T	T	C	A	A	C	C	T	A	C	1	A	C	T	T	A	A	A	G	A	0	T	G	**43**	**C**	**C**	A
	H12	A	A	A	C	G	C	G	A	C	A	C	1	A	A	G	T	3	T	A	C	A	22	C	T	T	C	A	A	C	C	T	A	C	1	A	C	T	T	A	A	A	G	A	0	G	A	0	**C**	**C**	A
	H13	A	A	A	C	G	C	G	A	C	A	C	1	A	A	G	T	3	T	A	C	A	22	C	T	T	C	A	A	C	C	T	A	C	1	A	C	T	T	A	A	A	G	A	0	G	A	0	**C**	**C**	A
	H14	A	A	A	C	G	C	G	A	C	A	C	1	A	A	G	T	3	T	A	C	A	22	C	T	T	C	A	A	C	C	T	A	C	1	A	C	T	T	A	A	A	G	A	0	G	A	**43**	**C**	**C**	A
	H15	A	A	A	C	G	C	G	A	C	A	C	1	A	A	G	T	3	T	A	C	A	22	C	T	T	C	A	A	C	C	T	A	C	1	A	C	T	T	A	A	A	G	A	0	G	A	**43**	**C**	**C**	A
	H16	A	A	A	C	G	C	G	A	C	A	C	1	A	A	G	T	3	T	A	C	A	22	C	T	T	C	A	A	C	C	T	A	C	1	A	C	T	T	A	A	A	G	A	0	G	A	0	**C**	**C**	A
	H17	A	A	A	C	G	C	G	A	C	A	C	1	A	A	G	T	3	T	A	C	A	22	C	T	T	C	A	A	C	C	T	A	C	1	A	C	T	T	A	A	A	G	A	0	G	A	**43**	T	T	A
	H18	A	A	A	C	G	C	G	A	C	A	C	1	A	A	G	T	3	T	A	C	A	22	C	T	T	C	A	A	C	C	T	A	C	1	A	C	T	T	A	A	A	G	A	0	G	A	**43**	T	T	A
	H19	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	T	T	A
	H20	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	**C**	G	**C**	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	**C**	**C**	A
	H21	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	**C**	G	**C**	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	**43**	**C**	**C**	A
	H22	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	**C**	**C**	A
	H23	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	**C**	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	**G**	**G**	0	**C**	**C**	A
	H24	G	G	G	C	A	C	G	A	C	A	A	1	A	A	G	T	3	T	A	C	A	22	C	T	T	C	A	A	C	C	C	A	C	1	A	C	T	T	A	A	A	G	A	0	G	A	**43**	T	T	A
	H25	G	G	G	T	A	C	G	A	C	A	A	0	T	G	G	C	3	C	G	C	A	0	C	T	G	C	G	A	C	C	C	G	C	1	G	T	T	T	A	A	A	G	0	G	A	0	T	T	A
	H26	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	**C**	G	0	**C**	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	T	T	A
	H27	G	G	G	T	A	T	A	G	T	G	A	0	T	G	A	C	0	C	G	T	G	0	T	C	G	T	G	C	T	T	C	G	T	0	G	T	A	A	T	T	T	A	G	28	T	G	0	T	T	A

specifically, the accumulation of 2AP is similar in rice, soybean and *Pandanus amaryllifolius* [16,29,30]. The sequence of *Badh2* homologues has proven to be quite variable. A SNP in exon 10 of the soybean gene GmBADH2 is responsible for a glycine to aspartic acid polymorphism, associated with contrasting levels of fragrance [29]. In rice, the major genetic bases of fragrance lie in the 8 bp deletion along with three SNPs in *Badh2* exon 7 [15] and in the 7 bp deletion in exon 2 [17]. At least eight functional alleles associated with fragrance have also been identified [18]. Of these, four are frameshift-inducing indels, one is a SNP which generates a premature stop codon, two comprise SNPs in exon sequence and the last is a 3 bp insertion. Here, we have shown that 62 accessions, although showing fragrance, do not carry any of the known *Badh2* FNPs. Nevertheless, an allelism test indicated that the gene underlying the phenotype was *Badh2*. Re-sequencing through the *Badh2* region has identified a few novel polymorphisms which may represent functional sites (Table 2), but there still remain 56 accessions where the exon sequence matched that of the non-fragrant allele. This suggests that fragrance might also be determined by sequence polymorphism in non-exonic DNA, such as in the promoter or the introns.

In addition, fragrance is mainly controlled by a simply inherited recessive gene mutating from the functional gene *Badh2*. More evidences have revealed that the fragrance is closely related to the RNA expression content of *Badh2*. Compared with the non-fragrant rice cv Nanjing11, a significant reduction in its transcription levels was detected in the fragrant rice cv. Wuxiangjing in all tissues except for the roots [16]; to ensure that the fragrance is correlated with reduced endogenous transcript, semi-quantitative RT-PCR and real-time quantitative RT-PCR revealed a considerable reduction in endogenous *OsBADH2* transcript levels in OsBADH2-RNAi repression lines compared to that of wild-type plants [31,32]. In a word, the alteration of expression lever within *Badh2* was mainly attributed to the mutant of *Badh2*, which finally affects the emerging of fragrance.

4.3. The contribution of Badh2 to rice domestication

The domestication process applies a strong selection pressure on the crop genome. In particular, with respect to genes which are part of the domestication syndrome, fixation is typically rapid for certain alleles which may well be rare in wild populations. In some cases, gene flow can still occur between wild accessions and cultivars, at least where these types are sympatric [33,34]. In rice, deliberate hybrids between *indica* and *japonica* types have contributed to the spread of domestication-related genes. Most of these introgression events have involved transfer from *japonica* to *indica* [14,18,35,36]. The *Badh2* haplotype H1 is found among both *indica* and *japonica* types, although its origin was likely to have been in the *japonica* genepool [18]. Haplotypes H2 and H3 appear to be fixed among *japonica* types (Figs. 1, 4; Table S4). H2 is largely confined to glutinous and Basmati type cultivars, while H3 is associated with accessions

Table 2

Exon polymorphisms in *Badh2*. The exon sequence of 62 fragrant accessions which lack any of the known FNPs identified three novel polymorphisms, leaving 56 accessions with no putative FNP.

Alleles	Deletion site	FNPs	Varieties
Badh2	None	No	89
badh2.2(1)	Exon 2	7 bp deletion	12
badh2.2(2)	Exon 2	75 deletion	2
badh2.4–5	Exons 4–5	806 deletion	8
badh2.7	Exon 7	8 deletion	29
badh2.10	Exon 10	G→A	2
badh2.13	Exon 13	C→T	2

Codes 25 and 26 with *badh2.2(2)*.
Codes 2 and 56 with *badh2.10*.
Codes 9 and 14 with *badh2.13*.

O. rufipogon — Badh2

Domestication — mutant — Domestication

O. sativa

badh2-I1, badh2-I2, badh2-I3, badh2-I4, badh2-I5, ……… Indica

badh2-J1, badh2-J2, badh2-J3, badh2-J4, badh2-J5, ……… Japonica

Fig. 6. A model for the evolution of *Badh2*, and its role in domestication.

originating in Jiangsu Province, suggesting that this was where the haplotype first arose (Fig. 1; Tables S1–2). Sequence analysis also indicates that haplotypes H4–7 and H9–10 all originated in the *japonica* genepool, while H11–12 and H14–18 originated in the *indica* genepool, and were only later transferred into *japonica* (Figs. 4–5). Thus, the mutant gene *badh2* both existed in *japonica* and *indica*, and was strongly affected by artificial selection during later stages of domestication (Figs. 4–5; Table 2).

A remaining interesting question relates to whether or not fragrance was present in wild populations prior to the domestication of *O. sativa*. Although no fragrance had been detected in wild rice varieties, one accession of *O. rufipogon* has been shown to be heterozygous for the H1 haplotype [18]. Its overall morphology, however, suggests that this represents a back introgression from a cultivated type, an event which has been considered the most likely source of fragrance in wild rice [37]. Our working model for the evolution at *Badh2* is shown in Fig. 6. Most of the 62 fragrant accessions lacking known *Badh2* FNPs lie to local rice varieties with long history of cultivation and owned their mutant *Badh2* genes (Tables S4–5). Thus we can conclude that original mutants both happened in the *indica* and *japonica* in the process of rice domestication, and then *badh2* genes have transferred to each other with the help of the artificial selection.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2012.11.010.

References

[1] T.T. Chang, The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices, Euphytica 25 (1976) 425–441.
[2] O. Panaud, The molecular bases of cereal domestication and the history of rice, Biologies CR 332 (2009) 267–272.
[3] T. Sang, S. Ge, Genetics and phylogenetics of rice domestication, Curr. Opin. Genet. Dev. 17 (2007) 533–538.
[4] C.C. Yang, Y. Kawahara, H. Mizuno, J. Wu, T. Matsumoto, T. Itoh, Independent domestication of Asian rice followed by gene flow from japonica to indica, Mol. Biol. Evol. (2011), http://dx.doi.org/10.1093/molbev/msr315.
[5] J. Molina, et al., Molecular evidence for a single evolutionary origin of domesticated rice, Proc. Natl. Acad. Sci. 108 (2011) 8351–8356.
[6] B.R. Lu, K.L. Zheng, H.R. Qian, J.Y. Zhuang, Genetic differentiation of wild relatives of rice as assessed by RFLP analysis, Theor. Appl. Genet. 106 (2002) 101–106.
[7] A.J. Garris, T.H. Tai, J. Coburn, S. Kresovich, S.R. McCouch, Genetic structure and diversity in *Oryza sativa* L. Genetics 169 (3) (2005) 1631–1638.
[8] A. Frary, et al., fw2.2: a quantitative trait locus key to the evolution of tomato fruit size, Science 289 (2000) 85–88.
[9] H. Wang, et al., The origin of the naked grains of maize, Nature 436 (2005) 714–719.
[10] R.T. Clark, T.N. Wagler, P. Quijada, J. Doebley, A distant upstream enhancer at the maize domestication gene, tb1, has pleiotropic effects on plant and inflorescent architecture, Nat. Genet. 38 (2006) 594–597.
[11] K.J. Simons, et al., Molecular characterization of the major wheat domestication gene Q, Genetics 172 (2006) 547–555.
[12] T. Komatsuda, et al., Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene, Proc. Natl. Acad. Sci. 104 (2007) 1424–1429.
[13] S. Konishi, K. Ebana, T. Izawa, Inference of the japonica rice domestication process from the distribution of six functional nucleotide polymorphisms of domestication-related genes in various landraces and modern cultivars, Plant Cell Physiol. 49 (2008) 1283–1293.
[14] K. Asanoa, et al., Artificial selection for a green revolution gene during japonica rice domestication, Proc. Natl. Acad. Sci. 108 (2011) 11034–11039.
[15] L.M.T. Bradbury, T.L. Fitzgerald, R.J. Henry, Q.S. Jin, D.L.E. Waters, The gene for fragrance in rice, Plant Biotechnol. J. 3 (2005) 363–370.
[16] S.H. Chen, et al., Badh2, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-Acetyl-1-Pyrroline, a major component in rice fragrance, Plant Cell 20 (2008) 1850–1861.
[17] W.W. Shi, Y. Yang, S.H. Chen, M.L. Xu, Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties, Mol. Breed. 22 (2008) 185–192.
[18] M.J. Kovach, M.N. Calingacion, M.A. Fitzgerald, S.R. McCouch, The origin and evolution of fragrance in rice (*Oryza sativa* L.), Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 14444–14449.
[19] B.C. Sood, E.A. Sidiq, A rapid technique for scent determination in rice, Indian J Genet Plant Breed. 38 (1978) 268–271.
[20] D.K. Berner, B.J. Hoft, Inheritance of scent in American long grain rice, Crop. Sci. 26 (1986) 876–878.
[21] G.N. Shao, et al., A new deletion mutation of fragrant gene and the development of three molecular markers for fragrance in rice, Plant Breed. 130 (2) (2011) 172–176.
[22] K. Liu, S. Muse, PowerMarker: New genetic data analysis software, Version 2.7, http://www.powermarker.net, 2004.
[23] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, Genetics 164 (2003) 1567–1587.
[24] P.J. Bradbury, Z.W. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler, TASSEL: software for association mapping of complex traits in diverse samples, Bioinformatics 23 (2007) 2633–2635.
[25] P. Librado, J. Rozas, DnaSP v5: software for comprehensive analysis of DNA polymorphism data, Bioinformatics 25 (2009) 1451–1452.
[26] N. Saini, N. Jain, S. Jain, R.K. Jain, Assessment of genetic diversity within and among Basmati and non-Basmati Rice varieties using AFLP, ISSR and SSR markers, Euphytica 140 (3) (2004) 133–146.
[27] J.Z. Lu, et al., SSR analysis on diversity of AA genome *Oryza* species in the Southeast and South Asia, Rice Sci. 15 (4) (2008) 289–294.
[28] A. Tang, et al., Comparative assessment of SSR diversity in aromatic rice Germplasm, Hereditas 31 (4) (2009) 412–419.
[29] R. Juwattanasomran, et al., A SNP in GmBADH2 gene associates with fragrance in vegetable soybean variety "Kaori" and SNAP marker development for the fragrance, Theor. Appl. Genet. 122 (2011) 533–541.
[30] A.B. Nadaf, S. Krishnan, K.V. Wakte, Histochemical and biochemical analysis of major aroma compound (2-acetyl-1-pyrroline) in basmati and other scented rice (*Oryza sativa* L.), Curr. Sci. 91 (11) (2006) 1533–1536.
[31] X.L. Niu, et al., RNAi-directed downregulation of OsBADH2 results in aroma (2-acetyl-1-pyrroline) production in rice (*Oryza sativa* L.), BMC Plant Biol. 8 (2008) 100.
[32] M.L. Chen, Fragrance of the rice grain achieved via artificial microRNA-induced down-regulation of OsBADH2, Plant Breed. (2012), http://dx.doi.org/10.1111/j.1439-0523.2012.01989.x.
[33] Z. Lin, et al., Origin of seed shattering in rice (*Oryza sativa* L.), Planta 226 (2007) 11–20.
[34] H. Wang, et al., The origin of the naked grains of maize, Nature 436 (2005) 714–719.
[35] M.T. Sweeney, et al., Global dissemination of a single mutation conferring white pericarp in rice, PLoS Genet. 3 (2007) e133.
[36] N. Takano-Kai, et al., Evolutionary history of GS3, a gene conferring grain length in rice, Genetics 182 (2009) 1323–1334.
[37] P. Prathepha, The fragrance (fgr) gene in natural populations of wild rice (*Oryza rufipogon* Griff.), Genet. Resour. Crop. Evol. (2008), http://dx.doi.org/10.1007/s10722-008-9337-7.