Independent Losses of Function in a Polyphenol Oxidase in Rice: Differentiation in Grain Discoloration between Subspecies and the Role of Positive Selection under Domestication

Yanchun Yu,a,1 Tian Tang,b,1 Qian Qian,c Yonghong Wang,a Meixian Yan,c Dali Zeng,c Bin Han,d,e Chung-I Wu,b,e,f,2 Suhua Shi,b,2 and Jiayang Li,a,2

Asian rice (Oryza sativa) cultivars originated from wild rice and can be divided into two subspecies by several criteria, one of which is the phenol reaction (PHR) phenotype. Grains of indica cultivars turn brown in a phenol solution that accelerates a similar process that occurs during prolonged storage. By contrast, the grains of japonica do not discolor. This distinction may reflect the divergent domestication of these two subspecies. The PHR is controlled by a single gene, Phr1; here, we report the cloning of Phr1, which encodes a polyphenol oxidase. The Phr1 gene is indeed responsible for the PHR phenotype, as transformation with a functional Phr1 can complement a PHR negative cultivar. Phr1 is defective in all japonica lines but functional in nearly all indica and wild strains. Phylogenetic analysis showed that the defects in Phr1 arose independently three times. The multiple recent origins and rapid spread of phr1 in japonica suggest the action of positive selection, which is further supported by several population genetic tests. This case may hence represent an example of artificial selection driving the differentiation among domesticated varieties.

INTRODUCTION

Morphological and physiological changes during domestication have intrigued generations of geneticists and evolutionists. There are now a number of genes known to control phenotypic variation among domesticated cultivars or between cultivars and their wild progenitors (Purugganan et al., 2000; Nesbitt and Tanksley, 2002; Olsen and Purugganan, 2002; Clark et al., 2004; Wang et al., 2005; Li et al., 2006; Sweeney et al., 2006). In addition to illuminating the physiological mechanisms, which are often highly relevant to agriculture, examination of these genes also sheds light on the underlying forces driving the evolution of phenotypes during domestication (Wang et al., 2005).

Among domesticated plants and animals, Asian cultivated rice (Oryza sativa) is of particular interest as it comprises two subspecies, indica and japonica, which differ in various morphological, physiological, and life history traits (Oka, 1988). These two subspecies are also partially reproductively isolated and provide a unique opportunity to study divergence associated with speciation. Races and subspecies are attractive subjects for speculation studies as they are at the incipient stage (Ting et al., 2004).

Although the two subspecies of rice differ broadly in >40 characters (Kato et al., 1928; Oka, 1958), the indica- and japonica-type cultivars are generally distinguished by four traits: resistance to KClO3, tolerance to cold, hair length of glume tips, and the phenol reaction (PHR) (Oka, 1953; Morishima and Oka, 1960). These characteristics have been suggested to be germane to rice domestication (Oka and Chang, 1962; Chang, 1976). Of particular interest to us is PHR, which reflects grain reaction to phenol treatment (Oka, 1953; Morishima and Oka, 1960). The grains, especially hulls, of the indica-type cultivars show positive PHR by turning brown after being soaked in phenol solution, whereas those of the japonica type are PHR-negative and their color remains unchanged (Oka, 1953). Phenol treatment presumably accelerates the browning of rice grains that happens under normal storage conditions.

While both indica and japonica are generally golden-hulled at harvest, the bran and unpolished or coarse grains of indica rice darken gradually during storage. Such a process may resemble
the discoloration of wheat (*Triticum aestivum*), which is catalyzed by polyphenol oxidases (PPOs) (Anderson and Morris, 2001; Simeone et al., 2002). As the white color of grains is preferred by consumers, wheat breeding has aimed to reduce or eliminate the grain discoloration. In rice, previous studies have shown that PHR is controlled by a single Mendelian gene, Phenol reaction 1 (*Phr1*; historically *Ph*), located on the long arm of chromosome 4 (McCouch et al., 1988; Saito et al., 1991; Tanksley et al., 1993; Lin et al., 1994). However, the nature of *Phr1* and the mechanism underlying PHR remain to be elucidated. In this study, we report the cloning, molecular characterization, transformation rescue, and evolutionary analysis of *Phr1* in rice.

**RESULTS**

**Cloning and Confirmation of **<i>*Phr1*</i>*

To map and clone *Phr1*, we constructed a large mapping population derived from cross between PHR-positive *indica* cv MH63 and PHR-negative *japonica* cv CJ06 (Figure 1). The cross produced a total of 5589 F2 plants (PHR positive: 4203; PHR negative: 1386) with a segregation ratio of 4:1 (χ² = 0.06; P > 0.75), consistent with the previous conclusion that a single nuclear recessive gene controls the negative PHR. Based on the previous mapping results that *Phr1* was on chromosome 4 (McCouch et al., 1988), we developed two new PCR-based molecular markers, S100 and S115, located on both sides of *Phr1* in the rice genetic linkage map (Chen et al., 2002; see Supplemental Table 1 online). We subsequently screened all PHR-negative F2 plants with them. The *Phr1* gene was mapped to the interval between markers S100 and S115 with genetic distances of 9.3 and 8.5 centimorgan (cM) to *Phr1*, respectively (Figure 2A). To fine-map *Phr1*, we obtained 66 recombinants between S100 and *Phr1* and 40 recombinants between *Phr1* and S115. With six more markers developed for this study, the breakpoints can be finely delineated (Figure 2B; see Supplemental Table 1 online). The *Phr1* locus was pinpointed to an 88-kb interval between markers P60 and P168 on a single BAC clone (OSJN-Ba0053K19; Figure 2B). Within this DNA segment, 14 open reading frames (ORFs) have been predicted (Figure 2C) (Feng et al., 2002). Among them, OSJN-Ba0053K19.18 (referred to as *K18* hereafter) is highly similar to plant PPO genes (Cary et al., 1992; Chevalier et al., 1999; Constabel et al., 2000; Gooding et al., 2001; DeMeke and Morris, 2002). The similarity in protein sequence ranges from 43 to 68%, and the highest is found with wheat PPO.

In higher plants, PPOs have been proposed to be responsible for the browning of damaged kernels, fruits, or vegetables, which may be response for disease resistance (Nicolas et al., 1994; Thipyapong et al., 1995; Gooding et al., 2001; DeMeke and Morris, 2002; Li and Steffens, 2002). Therefore, we PCR amplified and sequenced the corresponding *K18* sequences from MH63 and CJ06 varieties with primers designed based on the sequence of AK108237 given in GenBank (see Supplemental Table 1 online). DNA sequence comparison revealed an 18-bp deletion (Δ18) in this ORF in CJ06 compared with MH63 (Figure 2D), and the same deletion was also found in Nipponbare (PHR-negative, *japonica* type) compared with GLA (PHR-positive, *indica* type) (see Supplemental Figure 1 online). Moreover, the relative PPO activity is high in MH63 grains but nearly undetectable in CJ06 (Figure 2E). These results strongly suggested that *K18* is very likely to be the *Phr1* gene.

To test the prediction that *k18* is the *Phr1* gene, a complementation test was conducted. Transformation of the plasmid pC13Phr1 that contains the entire 93-11 *Phr1* gene, including 1411-bp 5’ upstream and 454-bp downstream sequences (Figure 2D), succeeded in rescuing the negative PHR phenotype of Nipponbare, whereas the control vector pC13p (Figure 2D) containing a truncated *Phr1* failed (Figure 2F). The authenticity of the transgenic plants was assured by the specific amplification of the transgene (Figure 2G). These results confirmed that the *K18* is indeed the *Phr1* gene and the 18-bp deletion (Δ18) is responsible for the loss-of-function phenotype of the *japonica* rice Nipponbare.

**Phr1** Encodes a **PPO**

The steady state level of *Phr1* mRNA was not detectable by RNA gel blot analysis; nevertheless, by RT-PCR, we could detect *Phr1* expression in grains at the early flowering stage (G1) and at the mature stage (G2) (Figure 3A). An earlier study had also reported *Phr1* expression in leaf tissue when the plants were under stress (Nobuta et al., 2007).

The sequenced RT-PCR product, which was amplified from the *indica* rice MH63, confirmed the high similarity of cDNA sequence of *Phr1* with LOC _Os04g53300_ (The Institute for Genomic Research). The latter encodes a protein of polyphenol oxidase with 570 amino acid residues. Alignment of the *Phr1* cDNA with its genomic DNA revealed that *Phr1* contains two introns (Figure 3B; see Supplemental Figure 1 online). As a
member of the tyrosinase family, the deduced amino acid sequence of Phr1 contains two putative copper binding domains (Figure 3B) (Steffens et al., 1994; Klabunde et al., 1998). Phr1 also contains a putative thylakoid-targeting sequence at its NH₂-terminal, which is rich in hydroxyl amino acid residues as in other plant PPOs (Constabel et al., 2000). In addition, an NH₂-terminal transit peptide of 30 amino acid residues is predicted (Figure 3B) (Keegstra, 1989; De Boer et al., 1991).

Therefore, Phr1 appears to encode an ~62.6-kD precursor protein, which is processed into a mature PPO of ~56.3 kD after removal of the NH₂-terminal signal peptide. The deduced Phr1 amino acid sequence shows 68% identity to wheat PPO (GenBank AAS00454), indicating that Phr1 is a rice homolog of the wheat PPO (Figure 3B). BLASTX searches for Phr1 homologs identified three Phr1-like (Phr1L) proteins encoded in the rice genome. Sequence alignment analysis showed that Phr1L1 and Phr1L2 are truncated PPOs (see Supplemental Figure 2 online). It also showed that, among the 13 conserved amino acid residues in the copper binding domain B of the Phr1L3 protein, five residues have been changed (see Supplemental Figure 2 online). Most importantly, we did not detect PPO enzyme activity in the three phr1 mutation lines (Table 1). These results indicate that Phr1 is the sole source of PPO activity. Since none of the three paralogs can be functionally interchangeable with Phr1, we did not pursue these Phr1-like genes.

**Phr1 and Grain Discoloration during Storage**

We hypothesized that functional Phr1 contributes to the discoloration of hulls and coarse grains of indica-type cultivars during prolonged storage and that the nonfunctional Phr1 thus resulted in nondiscoloration of japonica grains. To determine whether this hypothesis is correct, we transferred the functional indica Phr1 gene into the Phr1-negative Nipponbare rice and compared the discoloration of their grains during storage. Although no color difference between transformed and nontransformed grains could be observed at the mature grain stage, the color of transgenic hulls after harvest became darker gradually during storage (Figures 4A to 4C), indicating that Phr1 is responsible for the discoloration of Phr1-positive hulls. It should be pointed out that the coarse grains of transgenic rice also show an apparent discoloration of PHR-positive hulls. It should be pointed out that the coarse grains of transgenic rice also show an apparent discoloration during storage (Figures 4D to 4F) as well as a strong PHR-positive reaction (Figure 4G). We further suggest that the Δ18 is the predominant, if not the sole, lesion responsible for the loss of PPO activity in Nipponbare. In Figures 5 to 7 (see below) and Supplemental Figure 3 online, four sets of observations are given to support this suggestion.

**Multiple Independent Mutations of Phr1 in japonica Cultivars**

To measure the occurrence of Phr1 deficiency among rice cultivars, we first genotyped 35 japonica and 20 indica lines using the molecular marker pSTS18, which specifically detects Δ18 (see Supplemental Table 1 online). All 20 indica lines examined are PHR-positive, although one line appears to be heterozygous for both the wild-type and Δ18 alleles. Curiously, although all 35 japonica lines examined are negative in PHR, only
Figure 3. Phr1 Expression Patterns and Alignment of Phr1 Protein Sequences of MH63 and Nipponbare with Wheat PPO.

(A) Phr1 expression pattern revealed by RT-PCR using total RNA isolated from roots (R), culms (C), leaves (L), and panicles 7 and 21 d after flowering (G1 and G2, respectively). Amplification of actin cDNA was used to ensure that approximately equal amounts of cDNA were loaded.

(B) Alignment of Phr1 protein sequences of MH63 and Nipponbare with wheat (*T. aestivum*, abbreviated as Tae) PPO (GenBank AAS00454). A dot stands for an identical amino acid residue and a dash for a gap. Putative N-terminal signal peptide and copper binding domains (A and B) are indicated in pink and orange, respectively. Black arrow indicates the predicted cleavage site of N-terminal signal peptide, and vertical lines show the intron positions. The amino acid residues encoded by the 18- and 29-bp deletion in *japonica* cultivars are highlighted in red and yellow, respectively. The blue triangle indicates the position of the 1-bp insertion in Tx36, a *japonica* cultivar.
Table 1. Summary of Phr1 Variations and PHR Phenotypes

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Genome</th>
<th>Genotype</th>
<th>No. of Lines Genotyped</th>
<th>No. of Lines Sequenced</th>
<th>PHR Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>japonica (n = 35)</td>
<td>AA</td>
<td>Δ18</td>
<td>32</td>
<td>11</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Δ29</td>
<td>2</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ins-1bp</td>
<td>1</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>indica (n = 20)</td>
<td>AA</td>
<td>+</td>
<td>19</td>
<td>6</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Δ18/+</td>
<td>1</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>O. rufipogon (n = 523)</td>
<td>AA</td>
<td>+</td>
<td>517</td>
<td>21</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Δ18 or 18/+</td>
<td>3</td>
<td>3</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Δ29 or 29/+</td>
<td>3</td>
<td>3</td>
<td>na</td>
</tr>
<tr>
<td>O. nivara</td>
<td>AA</td>
<td>+</td>
<td>10</td>
<td>2</td>
<td>+</td>
</tr>
<tr>
<td>O. barthii</td>
<td>AA</td>
<td>+</td>
<td>67</td>
<td>2</td>
<td>+b</td>
</tr>
<tr>
<td>O. logistaminata</td>
<td>AA</td>
<td>+</td>
<td>5</td>
<td>0</td>
<td>na</td>
</tr>
<tr>
<td>O. glaberrima (Africa cultivar)</td>
<td>AA</td>
<td>+</td>
<td>1</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>O. meridonalis</td>
<td>AA</td>
<td>+</td>
<td>16</td>
<td>0</td>
<td>na</td>
</tr>
<tr>
<td>O. glumaeaputula</td>
<td>AA</td>
<td>+</td>
<td>81</td>
<td>0</td>
<td>na</td>
</tr>
<tr>
<td>O. panictata</td>
<td>BB</td>
<td>+</td>
<td>2</td>
<td>0</td>
<td>na</td>
</tr>
<tr>
<td>O. officinalis</td>
<td>CC</td>
<td>+</td>
<td>2</td>
<td>0</td>
<td>na</td>
</tr>
<tr>
<td>O. alta</td>
<td>CCDD</td>
<td>+</td>
<td>2</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>O. latifolia</td>
<td>CCDD</td>
<td>+</td>
<td>1</td>
<td>0</td>
<td>na</td>
</tr>
<tr>
<td>O. granulata</td>
<td>GG</td>
<td>+</td>
<td>1</td>
<td>0</td>
<td>na</td>
</tr>
</tbody>
</table>

Genotypes are indicated with "+" for alleles without molecular lesions, "Δ18" for alleles bearing an 18-bp deletion, "Δ29" for alleles bearing a 29-bp deletion, "ins-1bp" for alleles bearing a 1-bp insertion, and "Δ18/+" and "Δ29/+" for heterozygous sites. PHR phenotypes with plus and minus signs for positive and negative phenol reactions, respectively. na, not available.

* Phenotype of sequenced lines.

b One of two lines shows positive PHR phenotype, and another is not available.

32 bear the Δ18 mutation. We therefore sequenced the Phr1 alleles from the three lines that do not carry Δ18. Interestingly, all three lines carry frame-shift mutations in the Phr1 coding region, of which two carry a 29-bp deletion (Δ29) and one has a 1-bp insertion (Table 1; see Supplemental Figure 1 online).

We further surveyed the distribution of the Δ18 and Δ29 alleles with PCR markers pSTS18 and pSTS29 (see Supplemental Table 1 online) in a large panel of wild rice lines and the African cultivar Oryza glaberrima (Table 1; see Supplemental Table 2 online). Outside of O. sativa, the deletions exist only in Oryza rufipogon, but at very low frequencies. Among 523 O. rufipogon lines, the Δ18 and Δ29 deletions were observed only three times each, accounting for 0.5% of the collection. Among all other species (with 188 lines altogether), neither allele was found. These deletions are thus very rare in nature. Importantly, the Phr1 deletion correlates perfectly with the PHR-negative phenotype for the accesses where both genotype and phenotype data are available.

The Recent Origin of the Loss-of-Function Mutations in Phr1

The three apparently independent origins of Phr1 loss-of-function mutations raise several interesting questions about the evolution of the PHR-negative phenotype. Did the independent mutations all emerge during domestication or did they preexist in the wild rice? If the former, were the mutations selected strongly, and why were they favored in one subspecies only? If the latter, how old are these mutations and how are these polymorphisms maintained in the wild species?

To answer these questions, we sequenced the Phr1 alleles from 14 japonica, seven indica, 27 O. rufipogon, and seven other wild species lines (Table 1). Lines bearing deletions were preferentially included for sequencing. The haplotypes that bear the Δ18 and Δ29 deletions were referred to as H18s and H29s, respectively (both in plural form denoting multiple different haplotypes sharing the same deletion). The genealogy of all haplotypes is given in Figure 5. An obvious feature of the genealogy is that H18s and H29s are both clustered, indicating recent origins of H18s and H29s, respectively. The origins are sufficiently recent that most of the H18s, for example, have not had time to recombine with other haplotypes.

For the H18s cluster (Figure 5A, in red), four features are noteworthy. First, the H18s in O. rufipogon are all similar (ranging from 0 to 3 differences between lines); by contrast, the japonica lines are more diverse (differing by 3 to 10 bp among lines; Figure 5B). If Δ18 had existed in O. rufipogon prior to domestication, one would have expected this mutation to be more common and more diverse in wild rice. Since H18s in O. rufipogon are very rare (0.5%; Table 1) and are a subset of H18s in japonica (as shown in the genealogy of Figure 5), it seems likely that Δ18 emerged during domestication and that the presence of the deletion in O. rufipogon reflects introgression from the crop into wild populations.

Second, the genealogical cluster of H18s (labeled red in Figure 5A) harbors a few alleles that do not bear the Δ18 deficiency (i.e., O. rufipogon w0154f, indica dl, and indica K2406f). Similarly, the H29 cluster (labeled blue) contains a non-H29 sequence, O. rufipogon w0634f. One explanation might be that Δ18 first occurred on a common H18-like haplotype and some of these ancestral haplotypes that did not bear Δ18 have persisted until now. However, a closer inspection of the genealogical pattern of Figure 5A and the zoom-in picture of the H18s cluster in the
haplotype tree (Figure 5B) suggest a different explanation. In this haplotype tree, three of the four non Δ18-bearing haplotypes are embedded among H18 types, and each is most closely related to another H18 haplotype. A more parsimonious explanation may be that these non-Δ18-bearing alleles were originally H18s themselves but have subsequently lost the Δ18 deficiency by recombination with a rice strain carrying a functional Phr1 allele. Because Δ18 is at the very 3' end of the Phr1 sequence (see Supplemental Figure 1 online), single crossover is sufficient to remove this deletion without introducing many other changes to the haplotype. In other words, an H18 haplotype could lose Δ18 (and, by definition, would no longer be considered H18) but retains its genealogical relationship with other H18 haplotypes. The samples not labeled red in Figure 5B (one O. rufipogon and two indica lines) may be such examples. We also note that all japonica lines retain Δ18 (except for the 1-bp insertion line; see below), and the loss of Δ18 was observed only in the other two taxa.

Third, among japonica lines surveyed, all have either a large in-frame deletion (Δ18) or frame-shift (a 1-bp insertion or a 29-bp deletion) indel in Phr1, and they have all lost PHR activity. The contrast with indica and O. rufipogon, in which such mutations are rather uncommon, demands an explanation (see below). The 1-bp insertion in the coding region, observed only once, is particularly noteworthy (Table 1; see Supplemental Figure 1 online). If Δ18 and this 1-bp insertion arose independently, one would have to assume that, by coincidence, both occurred on a H18-like haplotype (Figures 5A and 5B). However, this haplotype is relatively uncommon in O. rufipogon. The haplotype on which the 1-bp insertion occurred is deep with the H18s genealogical cluster (Figure 5B). A more interesting scenario is as follows: an H18 haplotype underwent two changes: the gain of the 1-bp insertion and then the loss of Δ18. If this is indeed the case, this japonica haplotype would have remained nonfunctional throughout its evolution.

Fourth, the H29s cluster, characterized by another frame-shift mutation, Δ29 (Figure 5A, labeled in blue) is very young, with <1 bp of nucleotide changes per kilobase since the emergence of Δ29. There has not been sufficient time for H29s of O. rufipogon, which is not a strictly selfing species, to experience recombination with other haplotypes (see Supplemental Results online for details). Furthermore, Δ29 is only one-tenth as common in O. rufipogon as it is in japonica. While three H29s from O. rufipogon are shown in Figure 5A, the frequency of H29s in that species is only 2.5/517, vis-à-vis the frequency of 2/32 in japonica (Table 1). All these observations suggest that Δ29 is a relatively recent mutation that probably originated in japonica.

Positive Selection on the Δ18 Mutation: Signature of Differentiation

Given that all japonica lines suffer a lesion in Phr1, we ask if the spread of such lesions was driven by positive selection (i.e., was adaptive specifically to the environment of japonica). We first consider the differentiation of the Phr1 gene across populations in this section. Sites of unusually strong differentiation between populations are often suggested to be under taxon-specific selection (Hara et al., 1984; Akey et al., 2004; Beaumont, 2005).

We ask whether the Δ18 region is unusually differentiated between japonica and either indica or O. rufipogon. We first calculated the extent of population differentiation for each variant site in the Phr1 gene among the three taxa. In Figure 6, we show the distribution of the Fst statistic (Weir and Cockerham, 1984) between japonica and the other two populations. High Fst values close to 1 indicate strong genetic differentiation among populations, and low Fst near 0 indicate homogeneity.

For comparison, we used two different data sets. The first set contains 10 reference genes for which we have collected polymorphism data from the three taxa ourselves (Tang et al., 2006) (see Methods for the bootstrapping procedure). In the second set, we retrieved randomly chosen DNA fragments from the data of Caicedo et al. (2007). The two data sets differ in the lines used and in genes chosen. In Tang et al. (2006), the lines are mostly so-called elite rice, whereas Caicedo et al. (2007) included many land races. The sequences from Tang et al. (2006) are all well-annotated genes, whereas those from Caicedo et al.’s collection are random DNA fragments.

![Figure 4. Discoloration of Transgenic Grains during Storage.](image-url)
Figure 5. The Genealogy of Phr1 Alleles in Cultivars and Their Wild Relatives.

(A) The main panel (large tree, right) shows the optimal tree obtained using the neighbor-joining method with 1000 bootstraps. The bootstrap values are...
We shall compare *japonica* and *O. rufipogon* first. *Phr1* is shown to have more high Fst sites than other genes, and the high peak near Fst = 1 in Figure 6A can be attributed mainly to Δ18. This peak is absent in the two reference data sets (Figures 6C and 6E). Thus, the strong differentiation observed around Δ18 is unusual for *japonica* and *O. rufipogon*. Differentiation between *japonica* and *indica* is expected to be more complex because both taxa have been influenced by domestication (population bottlenecks, selection, and hitchhiking; Lu et al., 2006). In all three sets of data, the *indica-japonica* differentiation is larger than the *japonica-O. rufipogon* comparison (Figures 6A versus 6B, 6C versus 6D, and 6E versus 6F; all with P < 0.001 by the one-sided Kolmogorov-Smirnov test). The Fst profile of *Phr1* is even more distinctive from the two reference sets (Figures 6A versus 6C or 6E, and 6B versus 6D or 6F; all with P < 10^{-6} by the one-sided Kolmogorov-Smirnov test). The results suggest that *Phr1* is unusually differentiated between *indica* and the other two taxa when compared with other parts of the genome.

**Positive Selection on the Δ18 Mutation: Level and Pattern of Polymorphism**

If *Phr1* is driven by positive selection, there should be a signature of a selective sweep in *japonica*, but not in *indica* or *O. rufipogon*. We thus compared the within-population polymorphism for the three taxa. The signature can be observed in three forms, shown below.

First, the level of polymorphism near *Phr1* should be reduced relative to other parts of the genome in *japonica*, and this reduction should be seen in *japonica* only. In Figure 7A, it can be seen that the level of polymorphism in *Phr1* in *japonica* is indeed much lower than that in either *indica* or *O. rufipogon*. Measured against the average polymorphism of the 10 reference genes from the same taxa (Tang et al., 2006), the level of *Phr1* polymorphism in *japonica* remains significantly lower than in the other two taxa (P < 0.001 by bootstrapping; see Methods). Note that *Phr1* in non-*japonica* taxa is more polymorphic than the 10 reference genes. As reported in Tang et al. (2006), the rice genome is a mosaic of regions of very low to very high polymorphism, and the 10 reference genes were chosen from regions of normal polymorphism. The high-polymorphism genes can be 6 to 10 times as variable as the ten reference genes (Tang et al., 2006).

Second, in the gene suspected to have undergone a recent selective sweep, not only is the level of polymorphism expected to be reduced but the reduction should exhibit an excess of very low frequency and very high frequency mutant sites (Fay and Wu, 2000; Przeworski et al., 2001; Zeng et al., 2006). This trend is often summarized by two different statistics: Tajima’s D (Tajima, 1989) and Fay and Wu’s H (Fay and Wu, 2000; Zeng et al., 2006). Significantly negative D and H indicate an excess of very low and very high frequency variants, respectively. In *japonica*, both statistics for *Phr1* are indeed significantly negative by the bootstrapping procedure described in Methods (Figures 7B and 7C). Furthermore, both D and H are much higher in *indica* and *O. rufipogon*. In neither of the two is D or H significantly different from 0.

The result that both D and H tests show the same trend is important for inferring positive selection (Figures 7B and 7C). Between the two tests, Fay and Wu’s H is considered a more specific test for positive selection (Zeng et al., 2006, 2007). However, deep population subdivision with migration may sometimes lead to false positive by the H test (Przeworski et al., 2001). Indeed, Caicedo et al. (2007) found many loci in domesticated rice to be associated with a significantly negative H value. Although these authors suggested pervasive positive selection during domestication, they did not reject gene flow between the two subspecies as a major contributing factor.

Recently, Zeng et al. (2006, 2007) pointed out that the joint test of D and H, referred to as the DH compound test, is robust against most demographic factors (see Supplemental Results online for a brief explanation). The expected D or H value at the neutral equilibrium is 0 (or very close to 0) but, after a recent selective sweep, both values would often be significantly less than 0 (Tajima, 1989; Fay and Wu, 2000). When both are significantly negative, the compound statistic, DH (Zeng et al., 2006), will also be highly significant. DH has the added advantage of being insensitive to demographic influences. In *japonica*, but not in *indica* or *O. rufipogon*, DH is significantly negative (see Supplemental Results online). According to the simulations of Zeng et al. (2006, 2007), the most likely explanation for the simultaneous significance of D and H in *japonica* is positive selection.

**DISCUSSION**

The negative PHR phenotype in *japonica* cultivars is associated with functional loss of *Phr1*, a consequence of indels in the coding regions. While PPOs exist as a large family of functionally redundant genes in tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*) (Thygesen et al., 1995; Thipyapong et al., 1997), the redundancy appears absent in rice. In Figure 2, it can be seen that indels in the *Phr1* gene alone can nullify PPO activity.

Since rice cultivars are selfers and their wild progenitors often outcross, the impact of this breeding structure on various aspects of our observations, including coalescence time and genetic hitchhiking, is further considered in the supplemental
material online. The overall population genetic patterns of \textit{Phr1} alleles suggest that the loss-of-function mutations arose three times in \textit{japonica} in the recent past. Based on four different population genetic tests, we conclude that at least one of them was driven to high frequency by positive selection, presumably associated with human activities.

In our study, samples of \textit{indica}, \textit{japonica}, and \textit{O. rufipogon} lines came predominantly from within China. Hence, the interpretation of positive selection, for example, is not confounded by possible geographical differentiation. Nevertheless, the conclusion may apply to China (or at most eastern Asian) populations only. Whether it is applicable to domesticated rice in general will have to await further studies.

An interesting parallel with the Phr1 deletions in our collection of \textit{japonica} lines has been reported for the six-rowed spike phenotype in barley (\textit{Hordeum vulgare}) (Komatsuda et al., 2007). Each of the three independent loss-of-function mutations in the \textit{Vrs1} gene, a homeodomain-leucine zipper gene, leads to the six-rowed phenotype, which presumably increases grain production. The difference between the rice and barley systems is that the loss of Phr1 function is restricted to only one subspecies of rice. This dichotomous distribution offers a

---

**Figure 6.** Distributions of the \textit{Fst} Statistic in \textit{Phr1} vis-á-vis 10 Reference Genes and Randomly Chosen Gene Fragments.

(A) \textit{Fst} of \textit{Phr1} between \textit{japonica} and \textit{O. rufipogon}.

(B) \textit{Fst} of \textit{Phr1} between \textit{japonica} and \textit{indica}.

(C) \textit{Fst} of the 10 reference genes between \textit{japonica} and \textit{O. rufipogon}.

(D) \textit{Fst} of the 10 reference genes between \textit{japonica} and \textit{indica}.

(E) \textit{Fst} of randomly chosen gene fragments referred to as STS from Caicedo et al. (2007) between \textit{japonica} and \textit{O. rufipogon}.

(F) \textit{Fst} of the same STS between \textit{japonica} and \textit{indica}. \textit{Phr1} sequences from each population were resampled following a multinomial distribution with 100 permutations (see Methods).

The \textit{Fst} profiles of \textit{Phr1} (A) and (B) show a peak near 1, which is distinctive from the two reference data sets (C) and (D) for reference genes; (E) and (F) for the STS loci.)
contrast and permits some insight into subspecies differentiation during domestication. For Phr1, at least one of the mutations has spread very rapidly in japonica, likely aided by positive selection associated with human activities. Some of these mutations may have been introduced into indica and O. rufipogon. Most interesting in our finding is that the introduced alleles tend to re-acquire the functional site by recombination (Figure 5).

Why, then, is this loss of function in Phr1 associated only with japonica, but not with indica, or their immediate progenitor, O. rufipogon? We tentatively propose a hypothesis based on our observations. Traits selected by humans during domestication may sometimes be those of aesthetic appeal (Sweeney et al., 2006). Grains of the cultivars of japonica subspecies are refractory to discoloration during storage, making the Phr1 gene a plausible target of artificial selection. In this sense, Phr1 may be analogous to Rc, a domestication-related gene required for red pericarp in rice (Sweeney et al., 2006).

The retention of PHR activity in indica, in contrast with its ubiquitous loss in japonica, demands an explanation. One might argue that the history of domestication may not be long enough for every desired mutation to emerge. However, three independent losses have occurred in japonica in a rather brief period of domestication and an indica line does carry the deletion (Table 1), potentially allowing it to spread in this subspecies. Thus, the Phr1 function is retained in indica not because of the lack of time for it to be lost, but because of other reasons. Plant PPOs have been reported to be associated with disease resistance (Thipyapong et al., 1995; Li and Steffens, 2002). It may be that agriculture in the tropical and subtropical zones still put a premium on disease resistance, much like the case in the wild. Another intriguing possibility is that Phr1 activities are needed in warmer climates to maintain seed dormancy (Gu et al., 2004, 2005). Hence, the appeal of white grains might not compensate for the cost of premature seed germination in storing indica grains under some environmental conditions.

Figure 7. Population Genetic Tests on the Phr1 Sequences for All Three Populations.

(A) Nucleotide diversity (\(\pi\)) at silent sites by the box-and-whisker diagrams for Phr1 (white) and reference genes (gray). The variance is estimated by the bootstrapping procedure. For comparison, the corresponding estimates for 10 reference genes are also given.

(B) and (C) Tajima’s D (B) and Fay and Wu’s H (C) tests for the Phr1 sequences. The dashed lines indicated the expected values of 0 for Tajima’s D and Fay and Wu’s H under the standard neutral model. Both D and H statistics for Phr1 are significantly negative in japonica (\(P < 0.05\)) but much higher in indica and O. rufipogon by the bootstrapping procedure described in Methods. Significantly negative D and H indicate an excess of very low and very high frequency variants, respectively.
Finally, high frequency major mutations like the Phr1 indels in *japonica* might be a common source of phenotypic divergence among domesticated breeds or cultivars. Indeed, a widely discussed view is the less-is-more hypothesis (Olson, 1999), which posits that domestication is accompanied by extensive gene losses. A recent survey of the molecular basis of trait differentiation between cultivars of crops (rice, wheat, etc.) has suggested that loss-of-function mutations could be associated with several known traits (Yano et al., 2000, 2004; Sasaki et al., 2002; Ueguchi-Tanaka et al., 2005). In a separate study based on the two fully sequenced rice genomes, we have also found that large in-frame indels or frame-shift indels in coding regions are unusually frequent among rice cultivars (Huang et al., 2008). In light of these observations, the result on Phr1 mutations could be seen as part of a common trend in the studies of domestication.

**METHODS**

**Material**

All seeds or DNA used in this study were collected by our own lab or provided by the International Rice Research Institute. In total, 35 *indica* lines, 20 * indica* lines, 523 *Oryza rufipogon* lines, and 188 others were used. All of * indica* lines were collected in China. For *japonica* lines, three are from Japan and the rest from China. For *rufipogon*, the lines were from Bangladesh (17 lines), Burma (21 lines), Cambodia (1 line), China (146 lines), India (159 lines), Indonesia (five lines), Khmer Republic (one line), Malay (15 lines), Nepal (two lines), New Guinea (three lines), Philippines (one line), Sri Lanka (nine lines), and Thailand (115 lines). In addition, 28 * rufipogon* lines were from uncertain locales. Hence, the geographical distribution of *O. rufipogon* lines is properly extensive.

For * indica* and *japonica* lines, we give the identifier for each line used in Supplemental Table 2 online. For * rufipogon* and other groups, the inference made in this study depends mainly on those lines chosen for DNA sequencing. Each sequenced line is also individually listed. The number of lines chosen for genotyping is much larger, but these lines provide little information, other than the frequencies of the two deletions; hence, these lines are not individually listed. The more interesting but rare * rufipogon* lines with deletions are identified in Supplemental Table 2 online.

**Cloning of Phr1**

To clone Phr1, the mapping population of 5589 F2 plants was derived from the cross between an * indica* variety Minghui63 (MH63) and a * japonica* variety Chunjiang06 (CJ06). Ten F2 seeds per line were soaked in 2% (v/v) phenol solution and observed after 5 d according to the manufacturer's instructions. RT primers used were synthesized from immature rice kernels as previous reported (Li et al., 2003b). The RT reaction was performed using 2 μg of total RNA with oligo(dT) and SuperScript III RNaseH (−) reverse transcriptase (Invitrogen) according to the manufacturer’s instructions. RT primers used were synthesized based on the cDNA sequences (see Supplemental Table 1 and Supplemental Figure 1 online).

**Sequence Manipulations**

The nonredundant peptide sequences were searched using BLASTX (Schaffer et al., 2001). The Phr1 molecular mass was calculated using the method of Kyte and Doolittle (1982) and the Genetics Computer Group software of the University of Wisconsin (Devereux et al., 1984). The signal peptide and cleavage site were predicted according to the method described previously (Nielsen et al., 1997; Nair and Rost, 2003). The comparison of Phr1 sequences was performed with the Multalin program (Corpet, 1988).

The deletions of Δ18 or Δ29 in Phr1 sequences were surveyed by PCR amplification in a total of 55 Asian cultivars, one Africa cultivar, and 710 wild rice individuals with primers pSTS18 and pSTS29 (see Supplemental Table 1 online). We further sequenced the Δ18- or Δ29-containing Phr1 genes with the primers F217, Cxp3, Cxp5, and c5311 (see Supplemental Table 1 online) from 21 *Oryza sativa* cultivars (seven * indica* and 14 * japonica*), the 27 *O. rufipogon* lines, and randomly selected lines from one African cultivated rice *O. glaberrima*, two *O. nivara*, two *O. barthii*, and two more distant wild relatives with a CC or CCDD genome (one *O. officinalis* and one *O. alta*) (Table 1). Two of the 27 *O. rufipogon* accessions and one of the seven * indica* accessions were also heterozygotes (Table 1); each of the two alleles was included in the sequence analysis. Contigs were assembled using SeqMan (DNASTAR), and multiple sequences were aligned using the ClustalX program (Thompson et al., 1997).

**Phylogenetic Analyses of DNA Sequences**

A phylogenetic tree of the sequenced lines was reconstructed by the neighbor-joining (NJ) method (Saitou and Nei, 1987) based on Kimura’s two-parameter distances (Kimura, 1980). MEGA version 4.0 (Tamura et al., 2007) was used to perform the phylogenetic reconstruction. Bootstrap values were estimated (with 1000 replicates) to assess the relative support for each branch (Felsenstein, 1985). All positions containing gaps were eliminated in pairwise sequence comparisons in NJ analyses. The NJ tree was shown rooted by the midpoint to improve clarity.

In the cluster of lines containing Δ18 (H18 lines), a finer resolution is needed for visualizing the origin of the non-H18 lines embedded in that cluster. To achieve the resolution, we used the cladistic approach on the haplotypes by means of statistical parsimony (Templeton et al., 1992) with the aid of TCS v1.21 (Clement et al., 2000). Indels were treated as a single mutation. The root of this parsimonious haplotype tree, which is a cluster of the larger NJ tree, is set at the midpoint of the longest branch. This root is the same as the one given by the larger tree that encompasses all sequences. The branch length represents the number of observed mutations on the branch. This cluster is shown separately in Figure 5B.
Sequence Resampling by Bootstrapping

For phylogenetic analysis, lines with indels were chosen preferentially for sequencing. These DNA sequences are hence biased for population genetic analysis. We hence corrected this bias by resampling according to the known frequencies of these indels. (Because indels were genotyped from larger samples, their frequencies were determined with greater accuracy.) All population genetics statistics are hence presented as distributions from 100 such resamplings. The detail of this resampling scheme is as follows:

Each population contains \( k = 3 \) mutually exclusive and exhaustive classes of genotypes, for example, wild type (\( G^1 \)), \( G^{18} \), and \( G^{29} \), with the probability distribution of \( P = (p_{\text{wild}}, p_{G^{18}}, p_{G^{29}}) \), where \( p_{\text{wild}} + p_{G^{18}} + p_{G^{29}} = 1 \). Although \( G^{29} \) individuals were not observed in the \( \text{indica} \) sample, this genotype could have been missed due to the small sample size. Let the sampling event, \( X \), be random for \( N_{\text{indica}} = 14 \), \( N_{\text{indica}} = 7 \), and \( N_{\text{rufipogon}} = 27 \) in \( \text{japonica} \), \( \text{indica} \), and \( \text{O.rufipogon} \), respectively. Then, \( X = (X_{\text{wild}}, X_{G^{18}}, X_{G^{29}}) \) is multinomially distributed with index \( k = 3 \) and parameter \( P = (p_{\text{wild}}, p_{G^{18}}, p_{G^{29}}) \), i.e., \( X \sim \text{Mult} (k, P) \), where \( X_{\text{wild}} = \text{number} \) of trials in which \( G^1 \) occurs, \( X_{G^{18}} = \text{number} \) of trials in which \( G^{18} \) occurs, \( X_{G^{29}} = \text{number} \) of trials in which \( G^{29} \) occurs, and \( X_{\text{wild}} + X_{G^{18}} + X_{G^{29}} = N \). To infer the distribution of each genotype in the wild populations, we simulated the multinomial sampling process for each population with 100 permutations. According to the results of genotyping survey (Table 1), \( p_{\text{wild}} = (1/35, 32/35, 2/35) \), \( p_{G^{18}} = (518/523, 2.5/523, 2.5/523) \), respectively. The missing data of \( P_{29} \) in \( \text{indica} \) population was arbitrarily assigned within the range from 0 to 0.5/20. Increasing the value to a maximum of 0.5/20 did not change the pattern reported in this study. We first classified the available sequences data for each population into the three genotypes of \( G^1 \), \( G^{18} \), and \( G^{29} \) and then generated the 100 multinomially distributed random number vectors for each genotype in each population. These numbers were used as the specified sizes for sampling our available sequence data with replacement. R scripts were written to perform the simulation in R 2.6.1 environment.

Population Genetic Tests

Each set of the resampled sequences was sequentially submitted to population genetic analyses. \( Fst \) statistic (Weir and Cockerman, 1984) was calculated for each site to detect genetic differentiation among \( \text{japonica} \) with \( \text{indica} \) and \( \text{O.rufipogon} \) populations. To examine the genomic distribution of \( Fst \), randomly chosen fragments referred to as STSs were retrieved from GenBank (EF000002 to EF010509) (Caicedo et al., 2007). Alignments of \( O. \text{sativa} \) ssp \( \text{indica} \), \( O. \text{sativa} \) ssp \( \text{japonica} \), and the wild progenitor \( O. \text{rufipogon} \) per locus were prepared with ClustalW2 (Larkin et al., 2007) and then used to calculate the \( Fst \) statistic for each site. The alignments of STS loci range from 391 to 667 bp, with an average of 493 bp. Nucleotide diversity (Tajima, 1983) was estimated using synonymous sites and noncoding regions for each population. Tajima’s \( D \) statistic (Tajima, 1989) and Fay and Wu’s \( H \) statistic (Fay and Wu, 2000) were computed using the \( H \)-test program (http://www.genetics.wustl.edu/flab/h/test.html). DH test (Zeng et al., 2006, 2007) was computed using Java scripts kindly provided by Kai Zeng. To calculate the \( H \) statistic, \( O. \text{barthii} \) was used as outgroup to infer ancestral character state of the \( O. \text{sativa}/O. \text{rufipogon} \) complex. The character states of segregating sites within \( O. \text{barthii} \) were resolved by checking their counterparts in more distant \( O. \text{alta} \) and \( O. \text{officialis} \). All in the population genetics analyses above, alignment gaps were excluded and indels were scored as binary characters. Scripts used in this study are available upon request.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under the following accession numbers: MH63 (DQ532375), Nipponbare (DQ532376), CJ06 (DQ532377), J1506 (DQ532378), Jnxx (DQ532379), Qfr (DQ532380), R4 (DQ532381), Saon (DQ532382), Sun (DQ532383), Tx3 (DQ532384), Tx36 (DQ532385), Tzn (DQ532386), w11 (DQ532387), Zh11(DQ532388), Zhen5125 (DQ532389), dl(DQ532390), 93-11 (DQ532391), GLA (DQ532392), k2406d (DQ532393), k2406f (DQ532394), Tx7 (DQ532395), Tx9 (DQ532396), A3 (DQ532397), A8 (DQ532398), G01037 (DQ532399), G01049 (DQ532400), G01054 (DQ532401), G01060 (DQ532402), G01067 (DQ532403), G01084 (DQ532404), G02016 (DQ532405), G02068 (DQ532406), G02104 (DQ532407), G02115 (DQ532408), G02177 (DQ532409), G12 (DQ532410), G25 (DQ532411), G30 (DQ532412), G40 (DQ532413), G52030 (DQ532414), G7113 (DQ532415), G7134 (DQ532418), G7232 (DQ532417), G7251 (DQ532418), G9014 (DQ532419), w0009 (DQ532420), w0154d (DQ532421), w0154f (DQ532422), w0509 (DQ532423), w0634d (DQ532424), w0634f (DQ532425), w1 (DQ532426), w1125 (DQ532427), w1727 (DQ532428), w1862 (DQ532429), w6 (DQ532430), yaoyong (DQ532431), and G7 (DQ532432).

Supplemental Data

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

Supplemental Figure 1. Alignment of Phr1 cDNA Sequences of \( \text{Indica-Type Rice MH63} \) and GLA with \( \text{Japonica-Type Rice Nipponbare} \).

Supplemental Figure 2. Alignment of PHR1 with Its Homologous Proteins Identified in the Rice Genome.

Supplemental Figure 3. The Frequency Spectrum of Phr1 Mutations.

Supplemental Table 1. PCR-Based Molecular Markers Developed in This Study.

Supplemental Table 2. Summary of Samples for Phr1 Study.

Supplemental Data Set 1. Text File Corresponding to Alignment in Figure 3.

Supplemental Data Set 1. Text File Corresponding to Alignment in Supplemental Figure 1.

Supplemental Data Set 1. Text File Corresponding to Alignment in Supplemental Figure 2.

Supplemental Results.

ACKNOWLEDGMENTS

We thank Zhukuan Cheng (Institute of Genetics and Developmental Biology), Chuanqing Wang and Yuzo Minobe (Plant Gene Center, Tsukuba, Japan) for providing some seeds or DNA samples of wild rice and cultivar lines. This work was supported by grants from the National Natural Science Foundation of China (30330040, 30500049, and 30730008/C0102) and the Ministry of Science and Technology of China (2005CB1208 and 2007CB815701).

Received April 30, 2008; revised October 9, 2008; accepted November 7, 2008; published November 25, 2008.

REFERENCES


Correction


The last line of the Acknowledgments section is corrected to read:

This work was supported by grants from the National Natural Science Foundation of China (30330040, 30500049, and 30730008/C0102) and the Ministry of Science and Technology of China (2005CB1208 and 2007CB815701).
Independent Losses of Function in a Polyphenol Oxidase in Rice: Differentiation in Grain Discoloration between Subspecies and the Role of Positive Selection under Domestication
Yanchun Yu, Tian Tang, Qian Qian, Yonghong Wang, Meixian Yan, Dali Zeng, Bin Han, Chung-I Wu, Suhua Shi and Jiayang Li

Plant Cell 2008;20;2946-2959; originally published online November 25, 2008;
DOI 10.1105/tpc.108.060426

This information is current as of July 11, 2017

Supplemental Data
/content/suppl/2008/11/21/tpc.108.060426.DC1.html

References
This article cites 77 articles, 34 of which can be accessed free at:
/content/20/11/2946.full.html#ref-list-1

Permissions

eTOCs
Sign up for eTOCs at:
http://www.plantcell.org/cgi/alerts/ctmain

CiteTrack Alerts
Sign up for CiteTrack Alerts at:
http://www.plantcell.org/cgi/alerts/ctmain

Subscription Information
Subscription Information for The Plant Cell and Plant Physiology is available at:
http://www.aspb.org/publications/subscriptions.cfm

Errata
An erratum has been published regarding this article. It is appended to this PDF and can also be accessed at:
/content/21/3/1020.full.pdf

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY