JIPB Journal of Integrative Plant Biology

Journal of Integrative Plant Biology 2011, **53** (8): 682–693

Research Article

Evolutionary History and Complementary Selective Relaxation of the Duplicated *PI* Genes in Grasses

Ruo-Xun Wei[1,2] and Song Ge[1*]

[1]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, the Chinese Academy of Sciences, Beijing 100093, China

[2]Graduate University of the Chinese Academy of Sciences, Beijing 100049, China

[*]Corresponding author

Tel: +86 10 6283 6097; Fax: +86 10 6259 0843; E-mail: gesong@ibcas.ac.cn

Available online on 26 May 2011 at www.jipb.net and www.wileyonlinelibrary.com/journal/jipb

doi: 10.1111/j.1744-7909.2011.01058.x

Abstract

Gene duplication plays an important role in the evolution of organisms by allowing functional innovation and the divergence of duplicate genes. Previous studies found two *PI*-like genes in grass species, suggesting functional divergence between the paralogous copies. Here, we reconstructed the evolutionary history of two *PI* genes from major lineages of grasses and other monocot species, and demonstrated that two *PI* genes (*PI1* and *PI2*) arose from a whole genome duplication that occurred in a common ancestor of extant grasses. Molecular evolutionary analyses at the family and tribal levels found strong purifying selection acting on two genes in grasses, consistent with the conserved class B function of the *PI* genes. Importantly, we detected different patterns of selective relaxation between the duplicated *PI* genes although no signature of positive selection was found. Likelihood ratio tests revealed that the ω ratio for M domain is significantly higher in *PI1* than in *PI2* but that for K domain is significantly higher in *PI2* than in *PI1*. These findings imply that complementary selective relaxation occurs in two *PI* genes after duplication, and provide additional molecular evidence for the subfunctionalization of the duplicated *PI* genes in grasses.

Keywords: evolution; selective relaxation; duplication; *PI genes*; grasses.

Wei RX, Ge S (2011) Evolutionary history and complementary selective relaxation of the duplicated *PI* genes in grasses. *J. Integr. Plant Biol*. **53**(8), 682–693.

Introduction

Gene duplication plays an important role in the evolution of organisms by allowing functional innovation and the divergence of duplicate genes (Ohno 1970; Hughes 1999). Classic models of gene duplication predict that one of the duplicate genes might either acquire a new function because of positive Darwinian selection (neofunctionalization) (Ohno 1970; Hughes 1994) or becomes a pseudogene by accumulation of deleterious mutations (nonfunctionalization or pseudogenization) (Nei and Roychoudhury 1973; Li 1980; Lynch and Conery 2000). Additional possible fates of duplicate genes have been proposed, including maintenance of the ancestral function by maintaining both copies (redundancy) and by subdivision of the ancestral function between copies (subfunctionalization and subneofunctionalization) (Hughes 1994; Force et al. 1999; Lynch et al. 2001; Zhang 2003; Innan and Kondrashov 2010). Given the important effect of gene duplication on evolutionary novelties, studies aiming at understanding the fates of duplicate genes and the evolution of gene families may shed important light on the association between molecular evolution and morphological novelty.

In angiosperms, the ABC model of floral development is a complex regulatory network that shows multiple gene duplications and different protein-protein interactions. According to the classic ABC model of floral development, three classes of homeotic genes (A, B, and C) determine the identity of the four whorls of floral organs in a combinatorial manner (Bowman et al. 1989; Coen and Meyerowitz 1991), i.e. class A genes alone specify sepal formation, classes A and B genes

together determine petal identity, classes B and C genes together regulate stamen development, and class C genes alone control carpel formation. Additional two classes of genes (D and E) were found to specify ovule identity and to be involved in the formation of petal, stamen, and carpel, respectively (Colombo et al. 1995; Pelaz et al. 2000; Theissen and Saedler 2001). All the genes involved in the flower identity except for *AP2* and its orthologs, belong to the MADS-box gene family (Theissen 2001). Of them, class B genes include two members, *APETALA3*-like (*DEFICIENS*-like) (*AP3*-like) and *PISTILLATA*-like (*GLOBOSA*-like) (*PI*-like) genes based on their functional characterization in *Arabidopsis* or *Antirrhinum* (Schwarz-Sommer et al. 1992; Davies et al. 1996; Riechmann et al. 1996). Previous studies showed that *AP3*/*PI*-like genes played conserved roles in identities of petal and stamen during floral development (Jack et al. 1992; Ambrose et al. 2000; Whipple et al. 2004, 2007) and that gene duplication and functional variation of *AP3*/*PI*-like genes have occurred frequently in the evolutionary history of angiosperms (Kramer et al. 1998; Lamb and Irish 2003; Kim et al. 2004, Stellari et al. 2004, Kanno et al. 2007; Kramer et al. 2007; Mondragón-Palomino and Theissen 2008). A recent study showed genome synteny at *AP3* and *PI* containing loci from various species in dicots, revealing the evolutionary pattern of B-function genes (Causier et al. 2010).

The grass family (Poaceae) has morphologically unique floral structure termed the spikelet, with the specialized lodicule occurring after the divergence of the subfamily Anomochlooideae and other grasses (GPWG 2001; Kellogg 2001, 2009). Evidence showed that morphological transition in the spikelet happened recurrently in the evolutionary history of grasses, including variation of the number and morphology of lodicules and stamen (GPWG 2001; Kellogg 2009). Previous molecular studies suggested that two class B MADS-box genes (*AP3* and *PI*) involving the identity of petal and stamen had conserved B-function on lodicule and stamen identity in Poaceae (Ambrose et al. 2000; Nagasawa et al. 2003; Whippel et al. 2004). It is noted, however, that two *PI*-like genes (*PI1* and *PI2*) were found in grass species relative to a single copy in *Arabidopsis* (Kang et al. 1998; Münster et al. 2001; Yadav et al. 2007). In particular, recent studies demonstrated that functional divergence occurred between the paralogous *PI* genes in grasses (Whipple et al. 2007; Yadav et al. 2007; Yao et al. 2008). These observations raise a few interesting questions regarding the evolution and functional divergence of the *PI*-like genes. First, when did the *PI* duplication happen in history and what mode of duplication (tandem, segmental or whole-genome duplication) is involved? Reconstructing the phylogenetic relationships between the *PI* genes will help elucidate their duplication history and generation mode. Second, we were interested in the relative importance of positive selection and relaxation of purifying selection in evolution of the *PI* genes. Previous studies have revealed positive Darwinian selection on the genes after their duplication (e.g. Zhang et al. 1998; Bielawski and Yang 2003; Shiu et al. 2006) but empirical study is scarce in plants (see reviews by Hughes 1999; Zhang 2003; Yang 2006). Finally, we ask what model (neofunctionalization, nonfunctionalization, and subfunctionalization) fits the evolutionary fate of the duplicated *PI* genes. Although studies on grass *PI* genes revealed that two *PI* genes might have gained either divergent functions (Münster et al. 2001; Lee et al. 2003; Yadav et al. 2007) or unequal redundancy of class B function (Yao et al. 2008), implicative of subfunctionalization, evidence of molecular evolution of the two genes is still lacking. Analyzing molecular evolution of duplicate genes will contribute to a better understanding of their evolutionary fates and functional divergence after duplication.

Molecular evolutionary analyses, in combination with phylogenetic reconstruction, have been successfully used to uncover the duplication history of gene family and to test the alternative explanations for retention and evolution of duplicate genes (e.g. Kramer et al. 1998; Bielawski and Yang 2004; Kim et al. 2004; Stellari et al. 2004; Hernandez-Hernandez et al. 2007; Sun and Ge 2010). In the present study, we investigate the molecular evolution of the *PI* genes in grasses at the family and tribal levels. We confirmed that the duplication of an ancestral *PI* into *PI1* and *PI2* happened in the common ancestor of extant grasses through whole genome duplication. In particular, we found different patterns of selective relaxation across four domains of two *PI* genes and demonstrated that complementary selective relaxation happened during the evolution of the *PI* genes after duplication. These results provide additional molecular evidence for the subfunctionalization of the duplicated *PI* genes in grasses.

Results

Characteristics of sequences

We cloned and sequenced 31 *PI*-like sequences from 12 species in the subfamily Ehrhatoideae, and one species in the subfamily Bambusoideae (Table S1). These sequences consist of all introns and most of the coding region including partial MADS-box, complete I, K, and C domains, and partial 3′ intergene region, ranging from 1 725 to 3 353 bp in length, with the exception that four sequences (*Phyllostachys propinqua PpPI1B Luziola leiocarpa LlPI1B*, *Luziola leiocarpa LlPI2B* and *Zizania latifolia ZlPI2B*) had shorter sequences due to the failure of PCR amplification. In some cases, we found multiple copies within species for a given gene (Table S1), but phylogenetic analyses demonstrated that they were either recent duplicates occurring within species (e.g. *Luziola leiocarpa*) or polyploidy species (e.g. *Zizania latifolia* and

Phyllostachys propinqua). In these cases, one sequence with the average branch length across multiple copies was used in our subsequent analyses. In addition, based on BlastN search in GenBank, we obtained 33 *PI*-like sequences from 18 grass species, 26 sequences from 17 monocot species and four sequences from three dicots. In total, 94 *PI*-like sequences were used in the present study and their names and GenBank accession numbers are listed in Table S1.

Phylogeny and duplication patterns of *PI*-like genes

We aligned all the *PI*-like sequences and generated a phylogeny of the *PI*-like sequences of monocots using sequences of dicots as the outgroups. As expected, all monocot *PI*-like sequences formed one monophyletic clade, and all grass *PI*-like sequences, as one monophyletic group, divided into two subclades (Figure S1). To better resolve the phylogenetic relationship of grass species, we further reconstructed the phylogeny of the *PI*-like sequences of the Poaceae species using *Joinvillea* and *Elegia* species as the outgroups. As shown in **Figure 1**, all grass sequences formed two independent clades (*PI1* and *PI2*), with each clade consisting of a majority of the Poaceae species sampled, although some branches were not fully resolvable within the *PI1* and *PI2* clades. BI generated a tree with similar topologies with some modification of minor branches (Figure S2). The phylogenetic analysis indicated clearly that a duplication event giving rise to two *PI* genes occurred in an ancestor of grasses.

To determine the pattern of the *PI* duplication in grasses, we aligned flanking sequences of the *PI* genes (*PI2* on chromosome 1 and *PI1* on chromosome 5 of rice) using 70% cumulative identity percentage (CIP) and 70% cumulative alignment length percentage (CALP) criterions and found high homology between the two chromosome regions (**Figure 2**). By screening the rice genome, we found a conserved colinearity at the *PI1* and *PI2* region in rice, and confirmed that the duplicated segments belong to the duplicated regions between chromosomes 1 and 5, which arose from the whole genome duplication identified in previous studies (Wang et al. 2005; Yu et al. 2005; Salse et al. 2008), i.e. the two *PI* genes arose from a whole genome duplication. Therefore, in conjunction with phylogenetic analysis, we concluded that the *PI*-like paralogs in grasses originated from a whole genome duplication prior to the common ancestor of extant grasses and posterior to divergence of Poaceae and Joinvilleaceae.

Selective relaxation and positive selection of *PI* genes after duplication

We used a site-specific model and branch-site model (Yang et al. 2000; Yang and Nielsen 2002; Bielawski and Yang 2004; Yang et al. 2005) to detect the signatures of selective relaxation and positive selection following the *PI* duplication in Poaceae. In these analyses, we used a pruned phylogenetic tree in which 16 species from eight subfamilies were selected, with no more than five species from one subfamily to keep a balanced sample from the subfamilies (**Figure 1** and Table S1).

We first used the site-specific models to evaluate the positive selection and variation of selective pressure among sites (**Table 1**). Both one ratio model (M0) and the discrete models (M3a and M3b) show that purifying selection dominates the evolution of two *PI* genes, with $\omega = 0.008$–0.392. The selection model (M2a) does not have significantly higher likelihood scores than the neutral model (M1a), indicating no signature of positive selection across sites. However, model M1a fits the data significantly better than model M0 ($P < 0.001$), suggesting relaxation in some sites across the genes. Similarly, likelihood ratio tests (LRTs) of M0 against M3a ($k = 2$) and M3b ($k = 3$) also indicate significant variation in selective pressure among sites for both genes (**Table 1**). These models provide strong evidence of variable selective pressure among sites across the genes without positive selection involved.

To detect the signatures of positive selection following the duplication, we compared selective pressures of the two branches (B1 and B2), in which *PI1* or *PI2* evolved from ancestral species of single-copy *PI* to common ancestors of extant grasses, with remaining branches based on branch model and branch-site models (**Table 1**). LRTs of branch model against M0 ($P = 0.68$ for B1; $P = 0.63$ for B2) and model A against M1a ($P = 0.14$ for B1; $P = 0.32$ for B2) show no significant difference of ω ratios between either branch B1 or B2 and background branches, indicating no evidence of positive selection in the two branches.

We then used the branch-site models A and D to detect the heterogeneity of selective constraint between two *PI* clades. As shown in **Table 1**, model A is a significantly better model than M1a ($P < 0.01$) for both *PI1* and *PI2* clades, implicative of potential positive selection. However, estimated ω values ($\omega_{3b} = 1.000$) and no significant difference between models A and A0 ($P = 1$) suggest that selective relaxation rather than positive selection is more likely to result in such an increase of ω values. For *PI2* clade, model D is significantly better than M3a ($P = 0.05$), indicating significant relaxation on partial codons (14.3%) in *PI2* clade ($\omega_{3b} = 0.509$) relative to background clade ($\omega_{3a} = 0.294$). For *PI1* clade, however, there is no significant relaxation on codons of the *PI1* gene compared to the background clade ($P = 0.27$). These results indicate different evolutionary patterns acting on the two *PI* genes, with the *PI2* but not *PI1* invoking significantly selective relaxation.

Comparisons of selective pressure among domains and between genes

To investigate the heterogeneity of selective pressure across four domains for two *PI* genes, we conducted likelihood ratio

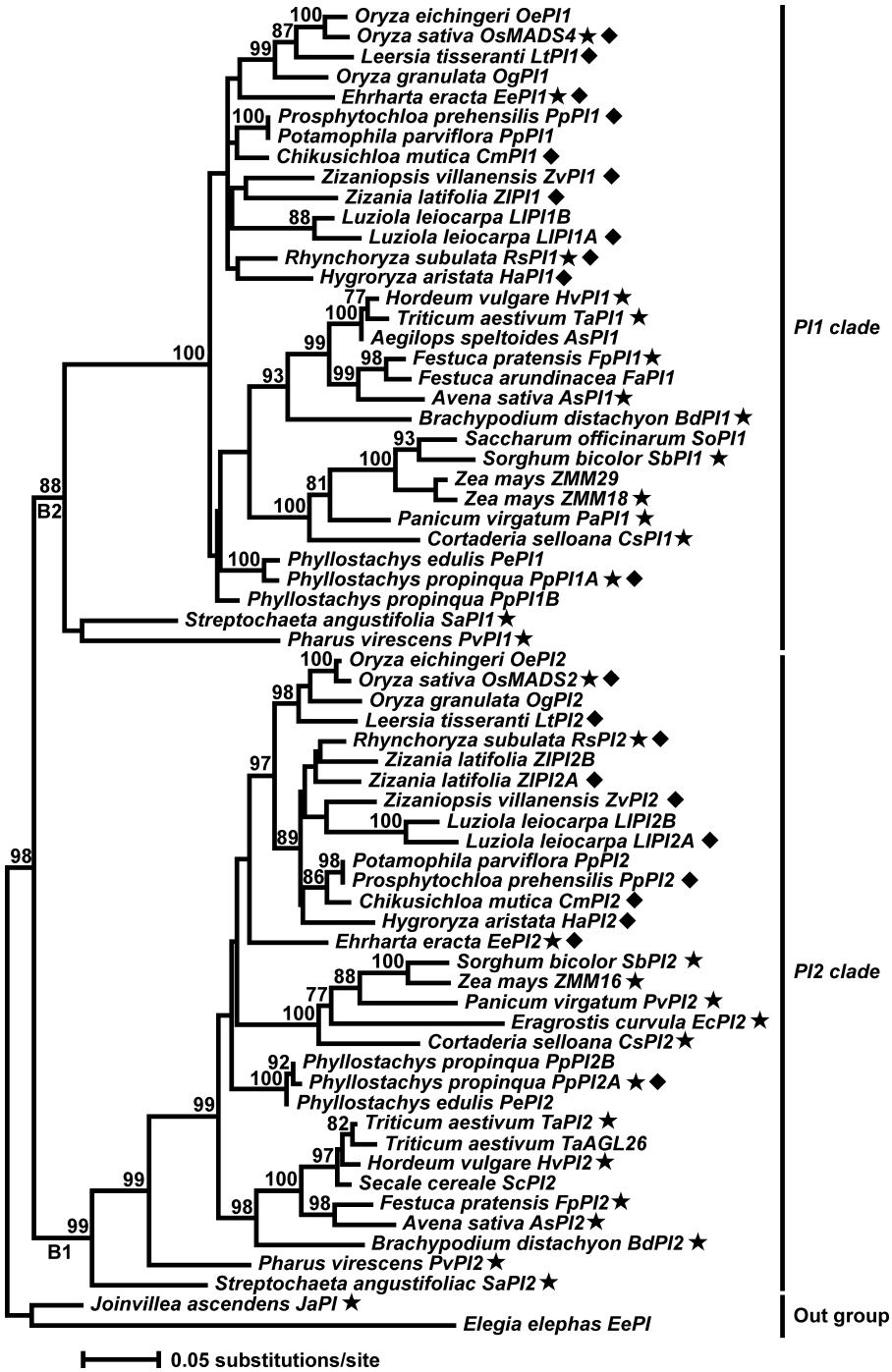

Figure 1. Phylogenetic tree of the Poaceae *PI*-like sequences inferred by maximum likelihood (ML) under GTR+I+G model.

Bootstrap values (≥75%) are shown above the branches. "★" and "◆" indicated that these sequences were selected to evaluate the molecular evolution of duplicated *PI* genes at the family and tribal level, respectively.

tests using fixed-site models and found that similar evolutionary patterns are found at either the family or the tribal level although no significant difference is found among four domains for *PI1* at the tribal level (**Figure 3**). Of the four domains, M domain possesses the lowest ω ratio, with the *PI2* M domain suffering stronger purifying selection than that of *PI1*. It is noted that different evolutionary patterns among domains were found for two *PI* genes. For *PI1*, the ω ratio is significantly lower in

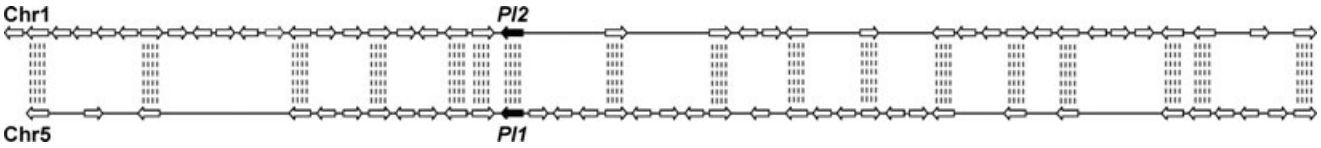

Figure 2. Duplicated genes identified in the flanking regions of *PI1* and *PI2*.

Dotted lines indicate duplicated gene pairs. Black arrows indicate the chromosomal locations of *PI1* (*OsMADS4*) and *PI2* (*OsMADS2*), respectively.

Table 1. Parameters and likelihood ratio test (LRT) for the *PI* genes

Model	*P*	ℓ	Estimate of parameters	LRT		Divergent or positively selective sites
				Comparison	*P*	
Site-specific models						
M0	1	−5 835.546	$\omega = 0.097$	-	-	-
M1a	2	−5 794.295	$\omega_1 = 0.086$, $p_1 = 0.927$; $\omega_2 = 1$, $p_2 = 0.073$	M0	***<0.01****	-
M2a	4	−5 794.295	$\omega_1 = 0.086$, $p_1 = 0.927$; $\omega_2 = 1$, $p_2 = 0.073$; $\omega_3 = 17.165$; $p_3 = 0.000$	M1a	1.00	No
M3a ($k = 2$)	3	−5 732.397	$\omega_1 = 0.029$, $p_1 = 0.647$; $\omega_2 = 0.252$, $p_2 = 0.353$	M0	***<0.01****	-
M3b ($k = 3$)	5	−5 721.096	$\omega_1 = 0.008$, $p_1 = 0.417$; $\omega_2 = 0.121$, $p_2 = 0.449$; $\omega_3 = 0.392$, $p_3 = 0.134$	M3a	***<0.01****	-
Branch models and Branch-site models						
Foreground branches: *PI1* clade (all *PI1* lineages)						
Branch model	2	−5 835.127	$\omega_1 = 0.104$, $\omega_{PI1} = 0.091$	M0	0.36	
Model A	4	−5 785.877	$\omega_1 = 0.081$, $p_1 = 0.896$; $\omega_2 = 1$, $p_2 = 0.060$; $\omega_{3b} = 1.000$, $p_3 = 0.044$	M1a Model A0	***<0.01**** 1	**161A, 185D**
Model A0	4	−5 785.877	$\omega_1 = 0.081$, $p_1 = 0.896$; $\omega_2 = 1$, $p_2 = 0.060$; $\omega_{3b} = 1$, $p_3 = 0.044$	-	-	-
Model D	6	−5 719.787	$\omega_1 = 0.007$,$p_1 = 0.406$; $\omega_2 = 0.115$,$p_2 = 0.451$; $\omega_{3a} = 0.461$, $\omega_{3b} = 0.317$, $p_3 = 0.144$	M3b	0.27	**129Q, 146L, 147R, 148R, 150G, 169V, 184P, 185D, 189A, 190P**
Foreground branch: B1 (*PI1* branch in which *PI1* evolved from ancestral species of single-copy *PI* to common ancestor of extant grass species)						
Branch model	2	−5 835.459	$\omega_1 = 0.097$, $\omega_2 = 0.140$	M0	0.68	
Model A	4	−5 792.309	$\omega_1 = 0.085$, $p_1 = 0.912$; $\omega_2 = 1$, $p_2 = 0.073$; $\omega_{3b} = 13.240$, $p_3 = 0.015$	M1a	0.14	**208N**
Foreground branches: *PI2* clade (all *PI2* lineages)						
Branch model	2	−5 834.814	$\omega_1 = 0.090$, $\omega_{PI2} = 0.106$	M0	0.23	
Model A	4	−5 769.986	$\omega_1 = 0.076$, $p_1 = 0.877$; $\omega_2 = 1$, $p_2 = 0.034$; $\omega_{3b} = 1.000$, $p_3 = 0.088$	M1a Model A0	***<0.01**** 1	**136R, 147R, 167Q, 177D, 209K**
Model A0	4	−5 769.986	$\omega_1 = 0.076$, $p_1 = 0.877$; $\omega_2 = 1$, $p_2 = 0.034$; $\omega_{3b} = 1.000$, $p_3 = 0.088$			
Model D	6 6	−5 718.116	$\omega_1 = 0.007$, $p_1 = 0.407$; $\omega_2 = 0.115$, $p_2 = 0.450$; $\omega_{3a} = 0.294$, $\omega_{3b} = 0.509$, $p_3 = 0.143$	M3b	***0.05****	**47I, 129Q, 146L, 147R, 148R, 150G, 169V, 184P, 189A, 190P**
Foreground branch: B2 (*PI2* branch in which *PI2* evolved from ancestral species of single-copy *PI* to common ancestor of extant grass species)						
Branch model	2	−5 835.432	$\omega_1 = 0.098$, $\omega_2 = 0.075$	M0	0.63	
Model A	4	−5 793.144	$\omega_1 = 0.086$, $p_1 = 0.920$; $\omega_2 = 1$, $p_2 = 0.073$; $\omega_{3b} = 11.261$, $p_3 = 0.008$	M1a	0.32	**123S**

Note: Bold letters indicate the sites that are involved in selective relaxation or positive selection.

* $0.01 < P \leq 0.05$; ** $P \leq 0.01$.

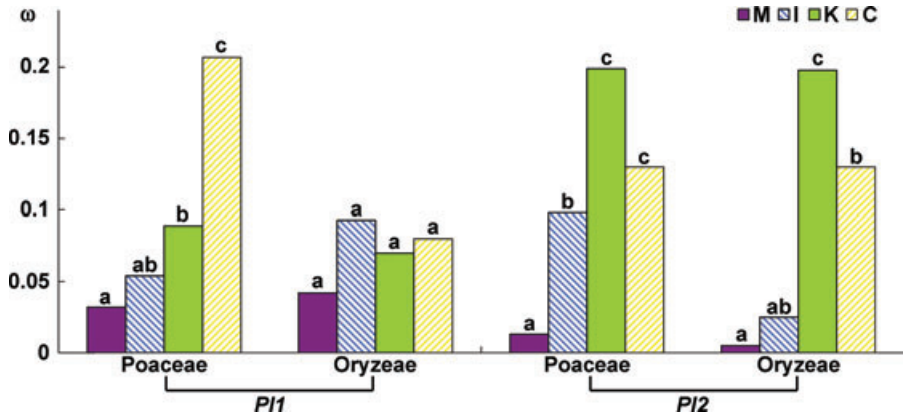

Figure 3. The difference of *dN*/*dS* ratios (ω) among the four domains of *PI* genes in Poaceae and Oryzeae.

The letters above the columns stand for the significance of difference in ω between the domains, with the same letters indicating no significant difference and different letters, significant difference.

K domain than in C domain, suggesting stronger purifying selection on K domain. For *PI2*, on the contrary, higher ω ratio was found for K domain relative to C domain, implying stronger purifying selection on C domain. These observations provide new evidence of domain divergence due to different selective pressures on different domains of two *PI* genes.

We further evaluated the difference of selective pressure on each of four domains between two *PI* genes (**Table 2**). Similar to the model tests above, strong purifying selection acts on two genes for M domain ($\omega < 0.05$), consistent with its functional importance (Riechmann et al. 1996; Kim et al. 2004). It is evident that difference of the ω ratio is not significant between two genes for the I and C domains. However, significantly different ω ratios were detected between two genes for M and K domains at both taxonomic levels ($P < 0.05$). Interestingly, the ω ratio for M domain is significantly higher in *PI1* than in *PI2* ($P < 0.05$), suggestive of selective relaxation of *PI1* relative to *PI2* in M domain. On the contrary, the ratio for K domain is significantly higher in *PI2* than in *PI1* ($P < 0.01$), implicative of selective relaxation of *PI2* relative to *PI1* in K domain (**Table 2**). These observations indicate that complementary selective relaxation occurs in the M and K domains of two *PI* genes.

Table 2. Difference of selective constraint on each of four domains and the entire region between genes using fixed-site models

Domain	Level	Model	*P*	ℓ	ω	$2\Delta\ell$	*P*
Total	Poaceae	Model C	50	−6 444.860	$\omega = 0.088$	4.712	0.09
		Model E	52	−6 442.504	$\omega_{PI1} = 0.078$, $\omega_{PI2} = 0.099$		
	Oryzeae	Model C	40	−3 908.358	$\omega = 0.079$	6.614	**0.04***
		Model E	42	−3 905.051	$\omega_{PI1} = 0.066$, $\omega_{PI2} = 0.096$		
M	Poaceae	Model C	48	−1 353.645	$\omega = 0.020$	6.338	**0.04***
		Model E	50	−1 350.476	$\omega_{PI1} = 0.029$, $\omega_{PI2} = 0.010$		
	Oryzeae	Model C	40	−767.564	$\omega = 0.014$	13.142	**<0.01****
		Model E	42	−760.993	$\omega_{PI1} = 0.034$, $\omega_{PI2} = 0.004$		
I	Poaceae	Model C	50	−539.392	$\omega = 0.064$	3.540	0.17
		Model E	52	−537.622	$\omega_{PI1} = 0.047$, $\omega_{PI2} = 0.111$		
	Oryzeae	Model C	40	−327.432	$\omega = 0.025$	2.948	0.23
		Model E	42	−325.958	$\omega_{PI1} = 0.026$, $\omega_{PI2} = 0.027$		
K	Poaceae	Model C	50	−3 001.465	$\omega = 0.109$	11.036	**<0.01****
		Model E	52	−2 995.947	$\omega_{PI1} = 0.078$, $\omega_{PI2} = 0.155$		
	Oryzeae	Model C	40	−1 914.919	$\omega = 0.106$	14.114	**<0.01****
		Model E	42	−1 907.862	$\omega_{PI1} = 0.067$, $\omega_{PI2} = 0.168$		
C	Poaceae	Model C	50	−1 412.664	$\omega = 0.143$	2.530	0.28
		Model E	52	−1 411.399	$\omega_{PI1} = 0.164$, $\omega_{PI2} = 0.117$		
	Oryzeae	Model C	40	−796.186	$\omega = 0.089$	0.776	0.68
		Model E	52	−795.798	$\omega_{PI1} = 0.079$, $\omega_{PI2} = 0.115$		

Note: $2\Delta\ell$ shows twice the difference of likelilood ratios between Model C and Model E. *$0.01 < P \leq 0.05$; **$P \leq 0$.

Discussion

This study identified two copies of the *PI* genes for all grass species, which formed two monophyletic groups corresponding to the rice *PI1* and *PI2* genes; whereas a single copy was detected for the majority of remaining species outside the grass clade in monocots. Phylogenetic reconstruction of all *PI*-like sequences demonstrated that the duplication event giving rise to *PI1* and *PI2* occurred in the common ancestor of extant grasses and after the divergence of the families Poaceae and Joinvilleaceae (**Figure 1**). In addition, our colinearity analysis based on the comparison of rice genome sequences further confirmed that the *PI* duplication originated from a whole genome duplication, which occurred in a common ancestor of extant grasses (Wang et al. 2005; Yu et al. 2005; Salse et al. 2008).

Molecular evolution analyses indicated that strong purifying selection acted on two *PI* genes in grasses, consistent with previous conclusions that the *PI*-like genes performed the conserved class B function through the protein heterodimer between *PI* and *AP3* in plants (Moon et al. 1999; Nagasawa et al. 2003; Whipple et al. 2004; Yadav et al. 2007; Yao et al. 2008). Theoretical and empirical studies showed that selective relaxation or positive selection might occur after gene duplication (Lynch and Conery 2000; Yang 2006; Innan and Kondrashov 2010). Although we were unable to detect signature of positive selection for both genes, different model tests detected significant heterogeneity in selective pressure among sites for both genes. Interestingly, model tests based on domains showed that the ω ratio was significantly different across four domains for both genes, with the lower ω value in K domain than in C domain for *PI1* gene but the higher value in K domain than in C domain for *PI2* gene (**Figure 3**). More importantly, model tests comparing two genes revealed that the ω ratio for M domain is significantly higher in *PI1* than in *PI2* but that for K domain is significantly higher in *PI2* than in *PI1*. These findings imply that complementary selective relaxation occurs in two *PI* genes after duplication.

It has been well established that duplicate genes possess three different fates: non-, neo- and sub-functionalization (see reviews by Zhang 2003; Lynch 2007; Innan & Kondrashov 2010). Previous studies showed that rice *PI1* (*OsMADS4*) and *PI2* (*OsMADS2*) were expressed in lodicule (the second whorl floral organ) and stamen (the third whorl) although two genes showed slight divergence at the tissue level (Yadav et al. 2007; Yao et al. 2008). In addition, several lines of evidence showed that rice *PI2* could alone perform the function of *PI*-like gene based on the phenotypes of *OsMADS4* RNAi plants (Yoshida *et al.* 2007; Yao et al. 2008), and rice *PI1* could bind with OsMADS16 (rice *AP3*) by protein–protein interaction (Moon et al. 1999; Lee et al. 2003; Yoshida et al. 2007; Yao et al. 2008) and was sufficient for normal stamen development (Prasad and Vijayraghavan 2003; Yoshida et al. 2007). It is worth noting that stamen and lodicule were transformed into carpel and palea-like organ in OsMADS2 RNAi and OsMADS4 RNAi double knockdown lines (Yao et al. 2008). Therefore, an ancestral B function that is conserved in angiosperms, is shared by two duplicated *PI* genes (Yoshida et al. 2007; Yao et al. 2008). These observations indicated that two duplicated *PI* genes possessed conserved B-class function, and thus nonfunctionalization seems an unlikely explanation for the evolution of the *PI* genes in grasses. The hypotheses of neofunctionalization is not supported either by our molecular evolution analyses in which no signature of positive selection was detected by various models (**Table 1**). Consequently, subfunctionalization might be involved in the *PI* evolution in grasses. The case of grass *PI* genes agrees well with previous observations that divergent expressions between two *PI* genes occurred at specific cells and tissues, and different phenotypes were invoked when *PI1* and *PI2* were knocked out, respectively (Münster et al. 2001; Lee et al. 2003; Prasad and Vijayraghavan 2003; Yadav et al. 2007).

Based on the duplication-degeneration-complementation (DDC) model, preservation of duplicate genes by subfunctionalization is driven by complementary degenerative mutation or selective relaxation (Force et al. 1999; Zhang 2003). Selective relaxation or degenerative mutation might result from alternative splicing or mutation in regulatory and coding regions of proteins (Zhang 2003; Lynch 2007). Although alternative splicing and mutation in regulatory regions have been reported to explain subfunctionalization (Force et al. 1999 and see review by Lynch 2007), subfunctionalization arising from the mutation in coding regions has rarely been observed (Korswagen et al. 2000; van Hoof 2005). Previous studies revealed that M and K domains of MADS-box gene were highly conserved in angiosperms (Riechmann et al. 1996; Kim et al. 2004) and very strong negative selection acted on M domain and K domain in dicots (Jaramillo and Kramer 2007). The present study detected significantly higher selective pressure on M domain of *PI1* than *PI2* at two taxonomic levels; on the contrary, K domain of *PI2* possessed significantly higher d_N/Ds ratios than *PI1* (**Table 2**). These observations indicated selective relaxation that occurs simultaneously on M domain of *PI1* and K domain of *PI2* in grasses in addition to the mutations in regulatory regions. Consistent with the DDC model, this finding provides clear evidence that complementary relaxation on different domains or sites of proteins might lead to subfunctionalization of duplicate genes. For MADS-box genes, M domain is responsible for DNA binding, while K domain participates in the interaction with *AP3* gene (Riechmann et al. 1996; Theissen 2001). Therefore, it is of great interest to ask how the duplicated *PI* genes retain protein-protein interaction with both downstream genes

and *AP3* gene in grass. The answer will require substantial investigation through multiple functional and evolutionary approaches.

Materials and Methods

Species samples

We sampled 31 species that represent the major lineages of Poaceae, including eight subfamilies: Ehrhatoideae (13 species), Pooideae (eight species), Panicoideae (four species), Bambusoideae (two species), Anomochlooideae (one species), Pharoideae (one species), Arundinoideae (one species), and Chloridoideae (one species). To facilitate the molecular analyses at the tribal level, we densely collected 12 species from the rice tribe (Oryzeae, Ehrhatoideae) representing 10 genera. To infer the evolutionary history of the *PI* genes, we selected additional 17 monocots and three dicots to generate the phylogenetic tree. In total, 31 sequences from 13 species in Ehrhatoideae (12 species) and Bambusoideae (one species) were isolated and sequenced here and the remaining 63 sequences were retrieved from GenBank by BLAST searches (Altschul et al. 1990). All information concerning the species sampled, the *PI*-like sequences and their GenBank accession numbers is listed in Table S1.

Isolation and sequencing of the *PI*-like genes

Total genomic DNA was extracted from silica gel-dried leaves, using the hexadecyltrimethylammonium bromide method as described in Ge et al. (1999). Based on the *PI*-like sequences of the Poaceae species, we designed three pairs of polymerase chain reaction (PCR) primers to amplify the *PI*-like genes. For grass species, M and C domains were highly conserved between the two genes but K domain was very divergent between them. Therefore, we first designed a pair of universal primers M4F4 (M domain) and M4R5 (C domain) for obtaining all *PI*-like sequences. Then, an additional two pairs of specific primers, M4F2 (K domain) and M4R1 (K domain) for *PI1*, and M2F2 (K domain) and M2R1 (K domain) for *PI2*, were designed for amplifying the two genes separately (**Table 3**). To obtain the entire C domain, we designed another universal reverse primer HouR1 on the conserved coding-regions on 3′-downstream of the *PI* genes and two specific forward primers M4F11 of *PI1* (K domain) and M2F11 of *PI2* (K domain). Based on the combinations of these primers, we generated the sequences that spanned all seven exons and introns as well as the downstream of two *PI* genes. The regions amplified and locations of all the amplifying and sequencing primers are shown in Figure S3.

All PCR amplifications were performed in a total volume of 25 uL on a Tpersonal thermocycler (Biometra, Germany). The reaction mixture contains 50 mM KCl, 1.5 mM $MgCl_2$,

Table 3. Summary of the genes sequenced and the primers used for polymerase chain reaction (PCR) and sequencing

Gene	Chromosome	Primer name	Region		Length[a] (bp)	Primer sequence
			Exon	Domain		
PI1	5	M4F4	exon1	M	1 125	gCggSAAgATCgAgATCAAg
		M4R1	exon6	K		gYCARYATCTTgTgYTCgTCCT
		M4F2	exon4	K	1 083	CAgRCCAAYCTgCgSgASAAgA
		M4R5	exon7	C		ASCCKgAAggTgAACggCAT
		M4F11	exon6	K	1 417	gCTSSAggAYgARCAYAARAT
		HouR1	3′	downstream		HYggMgAYATgCTCMgRTgA
		PIR1	exon3	K	–	gCTSRATCTgCATRTTRYCRTTCT
		M4R10	exon2	I	–	KCTCgATCTgCATgTTgTCRT
PI2	1	M4F4	exon1	M	1 174	gCggSAAgATCgAgATCAAg
		M2R1	exon6	K		TCTCRTCTTCCAgCATCTTAYC
		M2F2	exon4	K	1 401	AgYgCGgAgATTgATCgAA
		M4R5	exon7	C		ASCCKgAAggTgAACggCAT
		M2F11	exon6	K	1 601	RTAAgATgCTggARgAYgAgA
		HouR1	3′	downstream		HYggMgAYATgCTCMgRTgA
		PIR1	exon3	K	–	gCTSRATCTgCATRTTRYCRTTCT
		M2R6	exon2	I	–	CRgARTTSgTCTggTACTTCT

Note: all parameters are identified based on the information of rice genome database. [a] indicates the length of polymerase chain reaction (PCR) products in rice.

10 mM Tris-Cl (pH = 8.3), 0.2 μM of each primer, 200 μM of each deoxyribonucleotide triphosphate, 10–25 ng genomic DNA, and 0.75 U exTaq polymerase (TaKaRa, Otsu, Shiga, Japan)). A hot-start step, 2 min incubation at 94 °C was followed by 35 cycles at 94 °C (35 s), touchdown from 57 °C to 50 °C (35 s) for promoting PCR efficiency, 72 °C (2 min) and a final extension step of 10 min at 72 °C. The PCR products were cleaned with DNA purification kit (Tiangen, Beijing, China) and directly sequenced. In cases where direct sequencing failed or multiple copies occurred, the products were ligated into pGEM-T easy vector following manufacturer's protocols (Promega, Madison, WI, USA) and at least six independent clones were sequenced.

Sequencing reactions were performed by a MegaBACE 1000 automated sequencer (Amersham Pharmacia Biotech, Piscataway, NJ, USA) or an ABI3730 automated sequencer (Applied Biosystems, Foster City, CA, USA). All sequences were aligned using Muscle 3.6 (Edgar, 2004) and then manually adjusted. Exons, introns and intergenic regions were identified based on the cDNA and genome sequences of rice, maize and sorghum.

Phylogenetic reconstruction and colinearity analysis

Phylogenetic trees were reconstructed using maximum likelihood (ML) method, implemented in PhyML 2.4.4 (Guindon and Gascuel 2003), and Bayesian inference (BI) with MrBayes v.3.12 (Huelsenbeck and Ronquist 2001). Heuristic searches were run for ML analyses, with random taxon addition, tree bisection reconnection swap for 100 replications. Reliability of branches was evaluated by 1 000 bootstrap replications. For BI, Markov chain Monte Carlo (MCMC) analysis was run for 1 000 000 generations, sampled very 10 generations with the first 50 000 generations set as burn-in. For both ML and BI analyses, GTR+I+G model of nucleotide evolution was determined using jModelTest 0.1 (Posada 2008). Three angiosperm species (Table S1) were used as outgroups when phylogenetic trees of all *PI*-like sequences were generated to explore the duplication history of the *PI* genes.

Duplicate genes can be generated by different modes such as crossing over, retroposition and chromosomal or genome duplication (Zhang 2003). Thus, we conducted a colinearity analysis to clarify the generation of the duplicated *PI* genes. Based on the Rice Annotation Project (http://rice.plantbiology.msu.edu/), we first identified the chromosomal position and the flanking genes of *OsMADS2* (*PI2*) and *OsMADS4* (*PI1*) in rice. Then we compared these genes with all other genes in the rice genome by BLASTN. Three parameters, including aligned length (AL), cumulative identity percentage (CIP) and cumulative alignment length percentage (CALP), were used to increase the stringency and significance of BLASTN (Salse et al. 2008). We finally used the stringent values (70% CIP and 70% CALP) in identification of the duplicate genes and the homology of flanking genes in the rice genome.

Tests for selection

To characterize variation of selective pressure and positive selection, we estimated the ratios of nonsynonymous to synonymous substitution rates (*dN*/*dS* or ω) using the codeml program of PAML version 4 (Yang 2007). If the ω ratio is significantly less than 1, purifying selection is evoked, whereas neutral evolution is likely to occur if the ratio is approximately equal to 1. Positive selection is inferred if nonsynonymous substitution rate is significantly larger than synonymous substitution rates ($\omega > 1$).

To evaluate variation of selective pressure among sites, we first used the site-specific models, where ω ratios are assumed to be the same for all branches but an independent ω ratio is estimated for each site (Yang and Swanson 2002; Yang 2006). The nested models include M0, one-ratio model assuming equal κ (ratio of transition to transversion substitutions) and a single ω ratio for all codons and all branches; M1a, a neutral model assuming two classes of sites that are either constrained ($0 < \omega < 1$) or neutral ($\omega = 1$); M2a, a selection model that adds a third class of sites with positive selection ($\omega > 1$) comparing with M1a; M3, a discrete model that assigns all sites into several categories, each with a different ω ratio. Positive selection is indicated if M2a is significantly better than M1a and the estimated ω ratios for individual sites were > 1. A comparison between M3 and M0 indicates whether the ω ratio is homogeneous across different parts of the genes.

Because selective pressure is commonly variable among sites and positive selection may act only on a few sites of a specific branch, it is of little power to detect positive selection using an average ω ratio for all sites or an identical ω value for all branches (Yang 2006). Therefore, we further used modified branch-site models to detect positive selection on sites in specific branches (Bielawski and Yang 2004). The modified branch-site model A0 and model A add a third class sites with $\omega = 1$ and $\omega > 1$, respectively, in special branches, and are compared with M1a and A0, respectively. To compare the difference of selective constraint between branches we used modified clade model D (Bielawski and Yang 2004) that extends M3 by adding one class of sites that assume different ω ratios between branches.

For evaluating heterogeneity of selective pressure among domains, we used the fixed-site models of Yang and Swanson (2002). We first estimated the ω values of each domain using model C that assumes the same ω ratio but different substitution rates (*rs*) and variant codon frequencies (πs) among domains, and model E allowing different parameters (ω, *rs* and πs) among domains. Thus, a significant degree of the heterogeneity in selective pressure among domains can be inferred through likelihood ratio test (LRT) between models

C and E. Similarly, we used the same models to evaluate the difference of selective pressure on each of four domains between genes.

Acknowledgements

We thank Qihui Zhu, Ziqiang Wu, Ning Zhang, and other members of Ge's group for their assistance during the experiment and data analyses. We are also grateful for Zhe Li, Fumin Zhang, and Hong-Zhi Kong for their valuable comments on the manuscript. We thank the International Rice Research Institute (Los Banos, Philippines) for providing seed samples. This work was supported by the National Natural Science Foundation of China (30990242 and 30121003) and the State Key Basic Research and Development Plant of China (973 2007CB815704).

Received 3 May 2011 Accepted 13 May 2011

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Ambrose BA, Lerner DR, Ciceri P, Padilla CM, Yanofsky MF, Schmidt RJ (2000) Molecular and genetic analyses of the *Silky1* gene reveal conservation in floral organ specification between eudicots and monocots. *Mol. Cell* **5**, 569–579.

Bielawski JP, Yang Z (2003) Maximum likelihood methods for detecting adaptive evolution after gene duplication. *J. Struct. Funct. Genomics* **3**, 201–212.

Bielawski JP, Yang Z (2004) A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.* **59**, 121–132.

Bowman JL, Smyth DR, Meyerowitz EM (1989) Genes directing flower development in *Arabidopsis*. *Plant Cell* **1**, 37–52.

Causier B, Castillo R, Xue Y, Schwarz-Sommer Z, Davies B (2010) Tracing the evolution of the floral homeotic B- and C-function genes through genome synteny. *Mol. Biol. Evol.* **27**, 2651–2664.

Coen ES, Meyerowitz M (1991) The war of the whorls: Genetic interactions controlling flower development. *Nature* **353**, 31–37.

Colombo L, Franken J, Doyle S, Koetje E, van Went J, Dons H, Angenent G, van Tunen A (1995) The prtunia MADS box gene *FBP11* determines identity. *Plant Cell* **7**, 1859–1868.

Davies B, Egea-Cortines M, de Andrade Silva E, Saedler H, Sommer H (1996) Multiple interactions amongst floral homeotic MADS box proteins. *EMBO J*. **15**, 4330–4343.

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative replacements. *Genetics* **151**, 1531–1545.

Ge S, Oliveira GCX, Schaal BA, Gao LZ, Hong DY (1999) RAPD variation within and between natural populations of wild rice (*Oryza rufipogon*) from China and Brazil. *Heredity* **82**, 638–644.

Grass Phylogeny Working Group (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann. Mo. Bot. Gard.* **88**, 373–457.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol*. **52**, 696–704.

Hernandez-Hernandez T, Martinez-Castilla LP, Alvarez-Buylla ER (2007) Functional diversification of B MADS-box homeotic regulators of flower development, Adaptive evolution in protein-protein interaction domains after major gene duplication events. *Mol. Biol. Evol.* **24**, 465–481.

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.

Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. B* **256**, 119–124.

Hughes AL (1999) *Adaptive Evolution of Genes and Genomes*. Oxford University Press, NY.

Innan H, Kondrashov F (2010) The evolution of gene duplications: Classifying and distinguishing between models. *Nat. Rev. Genet.* **11**, 97–108.

Jack T, Brockman LL, Meyerowitz EM (1992) The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* **68**, 683–697.

Jaramillo MA, Kramer EM (2007) Molecular evolution of the petal and stamen identity genes, *APETALA3* and *PISTILLATA*, after petal loss in the Piperales. *Mol. Phylogenet. Evol*. **44**, 598–609.

Kang HG, Jeon JS, Lee S, An G (1998) Identification of class B and class C floral organ identity genes from rice plants. *Plant Mol. Biol.* **38**, 1021–1029.

Kanno A, Nakada M, Akita Y, Hirai M (2007) Class B gene expression and the modified ABC model in nongrass monocots. *Sci. World J.* **19**, 268–279.

Kellogg EA (2001) Evolutionary history of grasses. *Plant Physiol*. **125**, 1198–1205.

Kellogg EA (2009) The evolutionary history of Ehrhartoideae, Oryzeae, and *Oryza*. *Rice* **2**, 1–14.

Kim S, Yoo M-J, Albert VA, Farris JS, Soltis PS, Soltis DE (2004) Phylogeny and diversification of B-function MADS-box genes in angiosperms: evolutionary and functional implications of a 260-million-year-old duplication. *Am. J. Bot.* **91**, 2102–2118.

Korswagen HC, Herman MA, Clevers HC (2000) Distinct beta-catenins mediate adhesion and signalling functions in *C. elegans*. *Nature* **406**, 527–532.

Kramer EM, Dorit RL, Irish VF (1998) Molecular evolution of genes controlling petal and stamen development: Duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. *Genetics* **149**, 765–783.

Kramer EM, Holappa L, Gould B, Jaramillo MA, Setnikov D, Santiago PM (2007) Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot Aquilegia. *Plant Cell* **19**, 750–766.

Lamb RS, Irish VF (2003) Functional divergence within the *APETALA3*/*PISTILLATA* floral homeotic gene lineages. *Proc. Natl. Acad. Sci. USA* **100**, 6558–6563.

Lee S, Jeon JS, An K, Moon YH, Lee S, Chung YY, An G (2003) Alteration of floral organ identity in rice through ectopic expression of *OsMADS16*. *Planta* **217**, 904–911.

Li WH (1980) Rate of gene silencing at duplicate loci: A theoretical study and interpretation of data from tetraploid fishes. *Genetics* **95**, 237–258.

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155.

Lynch M, O'Hely M, Walsh B, Force A (2001) The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**, 1789–1804.

Lynch M (2007) *The Origins of Genome Architecture.* Sinauer Associates, Inc., Sunderland, MA.

Mondragón-Palomino M, Theissen G (2008) MADS about the evolution of orchid flowers. *Trends Plant Sci.* **13**, 51–59.

Moon YH, Jung JY, Kang HG, An G (1999) Identification of a rice APETALA3 homologue by yeast two-hybrid screening. *Plant Mol. Biol.* **40**, 167–177.

Münster T, Wingen LU, Faigl W, Werth S, Saedler H, Theissen G (2001) Characterization of three *GLOBOSA*-like MADS-box genes from maize: evidence for ancient paralogy in one class of floral homeotic B-function genes of grasses. *Gene* **262**, 1–13.

Nagasawa N, Miyoshi M, Sano Y, Satoh H, Hirano H, Sakai H, Nagato Y (2003) *SUPERWOMAN1* and *DROOPING LEAF* genes control floral organ identity in rice. *Development* **130**, 705–718.

Nei M, Roychoudhury AK (1973) Probability of fixation of nonfunctional genes at duplicate loci. *Am. Nat.* **107**, 362–372.

Ohno S (1970) *Evolution by Gene Duplication*. Springer-Verlag, NY.

Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* **405**, 200–203.

Posada D (2008) jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256.

Prasad K, Vijayraghavan U (2003) Double-stranded RNA interference of a rice *PI*/*GLO* paralog, *OsMADS2,* uncovers its second-whorl-specific function in floral organ patterning. *Genetics* **165**, 2301–2305.

Riechmann JL, Krizek BA, Meyerowitz EM (1996) Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins *APETALA1, APETALA3, PISTILLATA*, and *AGAMOUS*. *Proc. Natl. Acad. Sci. USA* **93**, 4793–4798.

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *Plant Cell* **20**, 11–24.

Schwarz-Sommez Z, Hue I, Huijser P, Flor PJ, Hansen R, Tetens F, Loggin WE, Saedler H, Sommer H (1992) Characterization of the *Antirrbinum* floral homeotic MADS-box gene *deficiens*: Evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *EMBO J.* **11**, 251–263.

Shiu SH, Byrnes JK, Pan R, Zhang P, Li WH (2006) Role of positive selection in the retention of duplicate genes in mammalian genomes. *Proc. Natl. Acad. Sci. USA* **103**, 2232–2236.

Stellari GM, Jaramillo MA, Kramer EM (2004) Evolution of the *APETALA3* and *PISTILLATA* lineages of MADS-box-containing genes in the basal angiosperms. *Mol. Biol. Evol.* **21**, 506–519.

Sun HZ, Ge S (2010) Molecular evolution of the duplicated *TFIIA*γ genes in Oryzeae and its relatives. *BMC Evol. Biol.* **10**, 128.

Theissen G (2001) Development of floral organ identity: Stories from the MADS house. *Curr. Opin. Plant Biol.* **4**, 75–85.

Theissen G, Saedler H (2001) Floral quartets. *Nature* **409**, 469–471.

van Hoof A (2005) Conserved functions of yeast genes support the duplication, degeneration and complementation model for gene duplication. *Genetics* **171**, 1455–1461.

Wang X, Shi X, Hao B, Ge S, Luo J (2005) Evidence that rice and other cereals have a common ancient tetraploid progenitor. *New Phytol.* **165**, 937–946.

Whipple CJ, Ciceri P, Padilla CM, Ambrose BA, Bandong SL, Schmidt RJ (2004) Conservation of B-class floral homeotic gene function between maize and *Arabidopsis*. *Development* **131**, 6083–6091.

Whipple CJ, Zanis MJ, Kellogg EA, Schmidt RJ (2007) Conservation of B class gene expression in the second whorl of a basal grass and outgroups links the origin of lodicules and petals. *Proc. Natl. Acad. Sci. USA* **104**, 1081–1086.

Yadav SR, Prasad K, Vijayraghavan U (2007) Divergent regulatory *OsMADS2* functions control size, shape and differentiation of the highly derived rice floret second-whorl organ. *Genetics* **176**, 283–294.

Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449.

Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**, 908–917.

Yang Z, Swanson WJ (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol. Biol. Evol.* **19**, 49–57.

Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118.

Yang Z (2006) *Computational Molecular Evolution.* Oxford University Press, Oxford.

Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.

Yao SG, Ohmori S, Kimizu M, Yoshida H (2008) Unequal genetic redundancy of rice *PISTILLATA* orthologs, *OsMADS2* and *OsMADS4,* in lodicule and stamen development. *Plant Cell Physiol.* **49**, 853–857.

Yoshida H, Itoh J, Ohmori S, Miyoshi K, Horigome A, Uchida E, Kimizu M, Matsumura Y, Kusaba M, Satoh H, Nagato Y (2007) superwoman1-cleistogamy, a hopeful allele for gene containment in GM rice. *Plant Biotechnol. J.* **5**, 835–846.

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, Zhang J, Zhang Y, Li R, Xu Z, Li S, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Wang J, Deng Y, Ran L, Shi X, Wang X, Wu Q, Li C, Ren X, Wang J, Wang X, Li D, Liu D, Zhang X, Ji Z, Zhao W, Sun Y, Zhang Z, Bao J, Han Y, Dong L, Ji J, Chen P, Wu S, Liu J, Xiao Y, Bu D, Tan J, Yang L, Ye C, Zhang J, Xu J, Zhou Y, Yu Y, Zhang B, Zhuang S, Wei H, Liu B, Lei M, Yu H, Li Y, Xu H, Wei S, He X, Fang L, Zhang Z, Zhang Y, Huang X, Su Z, Tong W, Li J, Tong Z, Li S, Ye J, Wang L, Fang L, Lei T, Chen C, Chen H, Xu Z, Li H, Huang H, Zhang F, Xu H, Li N, Zhao C, Li S, Dong L, Huang Y, Li L, Xi Y, Qi Q, Li W, Zhang B, Hu W, Zhang Y, Tian X, Jiao Y, Liang X, Jin J, Gao L, Zheng W, Hao B, Liu S, Wang W, Yuan L, Cao M, McDermott J, Samudrala R, Wang J, Wong GK, Yang H (2005) The genomes of *Oryza sativa*: A history of duplications. *PLoS Biol.* **3**, 266–281.

Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95**, 3708–3713.

Zhang J (2003) Evolution by gene duplication: An update. *Trends Ecol. Evol.* **18**, 292–298.

Supporting Information

Additional Supporting Information may be found in the online version of this article

Figure S1. Maximum likelihood (ML) phylogeny of the monocot *PI* sequences under the GTR+I+G model. Bootstrap values (>75%) are shown above the branches.

Figure S2. Phylogenetic tree of the monocot *PI*-like sequences reconstructed by Bayesian inference under GTR+I+G model. Bayesian posterior probabilities (≥95%) are shown above the branches.

Figure S3. The structure of the *PI* genes and the locations of the primers for amplifying and sequencing *PI1* and *PI2*. The black, white and grey blocks indicate exons, introns and intergenic regions, respectively.

Table S1. *PI*-like sequences included in this study

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

(Co-Editor: Hongya Gu)