JSE Journal of Systematics and Evolution

Journal of Systematics and Evolution 51 (6): 681–692 (2013) doi: 10.1111/jse.12045

Research Article

Evolutionary divergence of the PISTILLATA-like proteins in *Hedyosmum orientale* (Chloranthaceae) after gene duplication

[1,2]Shu-Jun LIU [1]Xiao-Qiu DU [1]Feng WU [1,2]Xue-Lei LIN [1,3]Qi-Jiang XU* [1]Zheng MENG

[1](*Key Laboratory of Plant Molecular Physiology*, *Institute of Botany, Chinese Academy of Sciences*, Beijing 100093, China)
[2](*Graduate University of Chinese Academy of Sciences*, Beijing 100049, China)
[3](*College of Life Sciences*, *Northeast Forestry University*, Harbin 150040, China)

Abstract Gene duplication and diversification is a significant aspect of gene and genome evolution, even though the evolutionary mechanism following duplication almost certainly varies by genes or gene families. We have identified three *PISTILLATA* (*PI*)-like genes, *HoPI_1*, *HoPI_2*, and *HoPI_3*, from basal angiosperm *Hedyosmum orientale* (Chloranthaceae) and found a functional divergence of these duplicate paralogs in the B-function process. However, the mechanism underlying the divergence of the three *HoPI* copies remains unclear. In this study, we carefully compared the coding sequence of the three duplicate *HoPI* paralogs, and investigated their evolutionary process after duplication. Molecular evolutionary analyses suggested that the coding sequence of the relatively recent duplicate paralogs *HoPI_1* and *HoPI_3* may evolve under purifying selection, but the functional constraint that acted on the coding sequence of *HoPI_2* may have been relaxed. Domain swapping and site-directed mutagenesis experiments further showed that *HoPI_2* has lost class B-function activity through a gradual process of mutational degradation in MADS and I domains. Although we currently have no direct *in vivo* evidence that *HoPI_2* has lost function in floral development, the relaxed selection, in combination with its inability to rescue *pi-1* mutant phenotype, indicate that this distant paralog *HoPI_2* might have been converted to be functionless in the B-function process. This study provides a detailed molecular characterization of duplicate *PI* lineage gene divergence in a single individual level, and our results suggest that mutations in the MI region are more likely to have negative effects on the biochemical activity of PI-like proteins.

Key words B-function, coding sequence divergence, gene duplication, *PISTILLATA*-like genes, relaxed selection.

Duplication has long been thought to be a primary driving force in the evolution of genes and genomes (Ohno, 1970; Gu et al., 2003). It can provide the raw genetic material for mutation, drift, and selection to act upon, resulting in increased diversity and functional innovation (Lynch & Conery, 2000; Wagner, 2001; Zhang, 2003; Irish & Litt, 2005). Gene duplications are prevalent in eukaryotic genomes in general (Lynch & Conery, 2000) and in flowering plants in particular (Blanc & Wolfe, 2004; Moore & Purugganan, 2005; Shiu et al., 2005; Cui et al., 2006). The genomes of most extant angiosperm species are the result of a series of short segmental/whole genome and individual gene duplication events (Soltis et al., 2009; Jiao et al., 2011; Proost et al., 2011). These processes have created many of the large gene families, such as the MIKC-type MADS-box gene family whose members have been shown to be the key regulators throughout plant development, especially in reproductive development (Sommer et al., 1990; Coen & Meyerowitz, 1991; Theissen et al., 2000; Theißen & Saedler, 2001; Becker & Theißen, 2003).

The sequence of MIKC-type MADS-box gene products consists of four distinct regions with specific functions: the MADS domain (M) that is essential for binding DNA to CarG promoter sequences; the intervening region (I) that is involved in determining the specificity of protein dimerization; the keratin-like domain (K) that folds into three amphipathic α-helices (K1, K2, and K3) with different roles in protein–protein interactions (dimerization); and the most variable C-terminal region (C) that is involved in multimeric complex formation and in transcriptional activation (Riechmann et al., 1996a, 1996b; Egea-Cortines et al., 1999; Honma & Goto, 2001; Lamb & Irish, 2003; Jack, 2004; Yang & Jack, 2004; Kaufmann et al., 2005).

Received: 23 April 2013 Accepted: 3 June 2013
* Author for correspondence. E-mail: qijiangxu@ibcas.ac.cn. Tel.: 86-10-62836556. Fax: 86-10-62836691.

Within the MIKC-type MADS-box gene family, floral homeotic class B genes constitute one of the best-studied subfamilies (Kramer et al., 1998; Stellari et al., 2004; Zahn et al., 2005). It comprises *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) sublineages whose products form obligate AP3–PI heterodimers to promote petal and stamen identities in eudicot flower development (Jack et al., 1992, 1994; Goto & Meyerowitz, 1994; Riechmann et al., 1996a; Theissen et al., 2000). The evolutionary history of class B genes involves multiple gene duplication events (whole genome or localized duplication) and subsequent divergence at different taxonomic levels. Especially in basal groups, including basal eudicots, monocots, magnoliids, and the ANITA grade, more than one *PI* and/or *AP3* members have been isolated, and most of these duplicate paralogs show differential expression and interaction patterns (Kramer et al., 2003, 2007; Tsai et al., 2004; Kim et al., 2005; Mondragón-Palomino & Theißen, 2011; Sharma et al., 2011). Class B genes of angiosperms provide an excellent model system for investigating the evolutionary mechanism following duplication events.

After a duplication event, duplicated genes accumulate changes in coding regions and/or regulatory regions, and they can generally lead to one of four possible fates: retaining the original function (functional redundancy); becoming silenced by degenerative mutations (non-functionalization/pseudogenization); acquiring a novel function (neofunctionalization); or retaining different subfunctions of the original gene (subfunctionalization) (Force et al., 1999; Lynch & Conery, 2000). Considerable theoretical works have been carried out concerning the mechanisms behind evolution by gene duplication and diversification (Force et al., 1999; Lynch & Conery, 2000; Lynch & Force, 2000; Lynch et al., 2001; Wagner, 2001; Gu et al., 2002, 2005; Lynch, 2002; Moore & Purugganan, 2003; Zhang, 2003; Blanc & Wolfe, 2004; Lynch & Katju, 2004; He & Zhang, 2005; Li et al., 2005b; Innan & Kondrashov, 2010; Liu & Adams, 2010; Qian et al., 2010), but it is notable that each gene or gene family has its own evolutionary history and could be affected by duplication in specific ways that cannot be described by globally observed patterns. Therefore, exploration of an appropriate model for specific genes is necessary for direct consideration of a specific gene or gene family. As more precise data accumulate from various plant model systems, we will better understand the spectrum of genic effects of duplication events and gain additional insight into the divergence patterns following duplication at a gene-specific or family-specific level.

We have isolated three *PI*-like genes *HoPI_1*, *HoPI_2*, and *HoPI_3* from basal angiosperm *Hedyosmum orientale* (Chloranthaceae) and found a coding sequence divergence between these duplicate paralogs: overexpression of *HoPI_1* or *HoPI_3* in wild-type *Arabidopsis* promotes petaloid organs in the first floral whorl and they can restore petal and stamen development in *Arabidopsis pi-1* mutants, whereas *HoPI_2* cannot induce any phenotype in transgenic *Arabidopsis*. However, it remains unknown which differences in the coding region are the major contributors for the observed divergence and how the differences were accumulated. Here, we carefully compared the coding sequence of the three duplicate *HoPI* paralogs, and investigated their evolutionary process since gene duplication. We used several different models of codon evolution to assess what evolutionary forces are associated with the evolution of coding regions of the three duplicate *HoPI* paralogs following gene duplications. Our results suggested that *HoPI_1* and *HoPI_3* may be subject to purifying selection, whereas *HoPI_2* may have been evolving under relaxed functional constraint. Domain-swapping and site-directed mutagenesis experiments further indicated that mutational accumulation that occurred in the MADS and I domains has contributed to the functional divergence of these *HoPI* genes in the class B-function process. The study improves our understanding of functional divergence of duplicate genes belonging to the floral homeotic *PI* lineage in basal angiosperm *H. orientale*, and the foundations may be laid for future evaluation of other MADS-box gene duplications in other angiosperms.

1 Material and methods

1.1 Sequence alignment and phylogenetic analysis

To investigate the evolutionary processes of three *HoPI* genes, we first carried out phylogenetic analysis using sequences from the present study and publicly available databases (Table S1). Three *AP3*-like gene sequences of Chloranthaceae were used as the representative of outgroups. Amino acid sequence alignment was carried out with MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) and then adjusted manually using GeneDoc (Nicholas et al., 1997). The poorly aligned positions were eliminated by the program Gblocks in combination with manual edition. The corresponding DNA matrix was generated on the basis of the well-aligned protein matrices and then used to construct the maximum likelihood (ML) tree by the program PHYML (Guindon & Gascuel, 2003). The $GTR + I + \Gamma$ model was chosen and ML parameter

values were estimated to optimize. One thousand bootstrap replicates were carried out.

1.2 Selective pressure analysis

A total of 17 *PI*-like sequences of basal angiosperms Chloranthaceae, Nymphaeaceae, Schisandraceae, Cabombaceae, and Amborellaceae retrieved from the NCBI database were used for selective pressure analyses. The *PI*-like gene *AmbtPI* from *Amborella trichopoda* (Amborellaceae) was used as the outgroup representative. Accession numbers as well as species names and plant families for each gene are presented in Table S1. The alignment result of full length coding regions was used to calculate the pairwise synonymous (d*S*) and non-synonymous (d*N*) nucleotide substitution rates and their ratios d*N*/d*S* (ω) by the CODEML program from PAML version 4 (Yang, 2007). We used several different models of codon evolution to infer hypotheses about the evolution forces. A likelihood ratio test (LRT) calculated using $2\Delta\ell$ was used to detect significant differences in fit between nested models (Yang, 1998). The branch-specific models allow for variable ω ratios among branches but invariable ω ratios in sites in the tree and can be implemented for the study of changes in selective pressures in specific lineages (Yang & Nielsen, 2002). It involved the "free-ratio", "one-ratio", and "two-ratio" models to represent the different presumed hypothesis. The free-ratio model specifies an independent ω value for each branch in the phylogenetic tree, the one-ratio model assumes the same ω value for all branches, and the two-ratio model assigns different ω values for the foreground (ω_1) and background branches (ω_0), respectively.

To identify positively selected amino acid sites in MADS, I, K, and C domains of PI proteins, three pairs of comparisons were carried out in site-specific models: M1a (a nearly neutral model) versus M2a (a positive selection model) was used to detect whether a positive selection is present among the different amino acid sites; M7 (beta) versus M8 (beta and ω) was used to identify specific amino acid sites that are potentially under positive selection; and M0 (the one-ratio model) versus M3 (discrete) was used to test for selection heterogeneity among amino acid sites. Two and three site classes were used in M3. Positive selection is indicated when a freely estimated ω is greater than 1 and the LRT reaches a statistically significant level.

The branch-site models (models A and B) allow the ω value to vary both among sites and among lineages, and were used to detect the occurrence of positive selection on individual codons within specific branches (Yang & Nielsen, 2002). In model A, ω_0 was assigned $0 < \omega < 1$, and ω_1 was fixed at 1; hence, positive selection was permitted only in the foreground branch (Zhang et al., 2005). In model B, ω_0 and ω_1 are free and, thus, some sites may evolve by positive selection across the entire phylogeny, whereas other sites may evolve by positive selection in just the foreground branch. Model A is compared with M1a (nearly neutral) or model A null and model B is compared with M3 (discrete).

1.3 Domain swapping and site-directed mutagenesis plasmid construction

Site-directed mutagenesis was obtained using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA). Domain swapping between *HoPI_1/3* and *HoPI_2* was obtained by amplifying separately of two cDNA fragments with primers that have a 30–40 nucleotide overlap. Subsequently, the products from the two polymerase chain reactions (PCRs) were purified and mixed, then used as templates for the second amplification. The gene-specific forward and reverse primers were added to amplify the full fragment of cDNAs with 30 PCR cycles. The PCR fragments were cloned into binary vector SN1301 (possessing CaMV *35S* promoter) to drive nearly ubiquitous expression. In complementation experiments, the *Arabidopsis PI* promoter (*pAtPI*, from position −1573 to −3 relative to the translation start codon ATG of the *Arabidopsis PI* gene) was used to replace the *35S* promoter of SN1301 to drive expression of constructs in *Arabidopsis pi-1* mutants. The final constructs were verified by sequencing and restriction analyses. Primers used are presented in Table S2.

1.4 *Arabidopsis thaliana* transformation and genotyping

The plasmid constructs described above were transformed into wild-type *A. thaliana* (Landsberg erecta), and heterozygous *pi-1* mutants, respectively, by the floral dip method (Clough & Bent, 1998). The seeds of transgenic plants were selected on solid 0.6× MS medium containing 25 mg/L hygromycin B and 1 g/L carbenicillin disodium salt and were genotyped by PCR with transgene-specific primers. Homozygous *pi-1* transformants were identified using the dCAPS finder program as described previously (Lamb & Irish, 2003). Then the homozygous mutant plants containing transgenic cDNAs were analyzed.

1.5 Yeast two-hybrid analysis

Yeast two-hybrid assays were carried out using the GAL4-based MATCHMAKER Two-Hybrid System (Clontech, Mountain View, CA, USA). *Saccharomyces*

cerevisiae strain AH109, GAL4 activation domain (AD) expression vector pGADT7 and GAL4 DNA-binding domain (DNA-BD) expression vector pGBKT7 were used. The coding sequences of *A. thaliana AtAP1*, *AtAP3*, *AtPI*, and *Hedyosmum orientale HoPIs*, site-directed mutagenesis *HoPI_2-*Q_{86} and *HoPI_3* + Q_{86} were amplified and fused into the pGADT7 and pGBKT7 vectors, respectively. All constructs were verified by restriction enzyme analyses and sequencing. Primers used are listed in Table S2. Yeast transformation and two-hybrid experiments were carried out as described elsewhere (Shan et al., 2006). The transformants cotransformed with plasmids encoding AtAP3 and AtPI from *A. thaliana* without MADS domains were used as a positive control, and the transformants containing plasmids pGADT7 and pGBKT7 were used as a negative control.

2 Results

2.1 Sequence and phylogenic analysis of *PI*-like genes

Phylogenetic analysis showed that the three *HoPI* genes were most closely related to *CsPI*, the *PI* ortholog of *Chloranthus spicatus* (Chloranthaceae) (Li et al., 2005a), and all these *PI*-like genes of the Chloranthaceae family grouped together with *PI* homologs from monocots (Fig. 1: A).

Sequence alignment of the three deduced HoPI amino acid sequences showed that these HoPI proteins display the typical MIKC domain structure of Type II MADS domain proteins (Fig. 1: B). They share 70%–80% identity over the whole length, including a relatively high level of similarity in the MIK region (>80%) and a low level of similarity in the C-terminal region (40%–70%; Fig. 1: C).

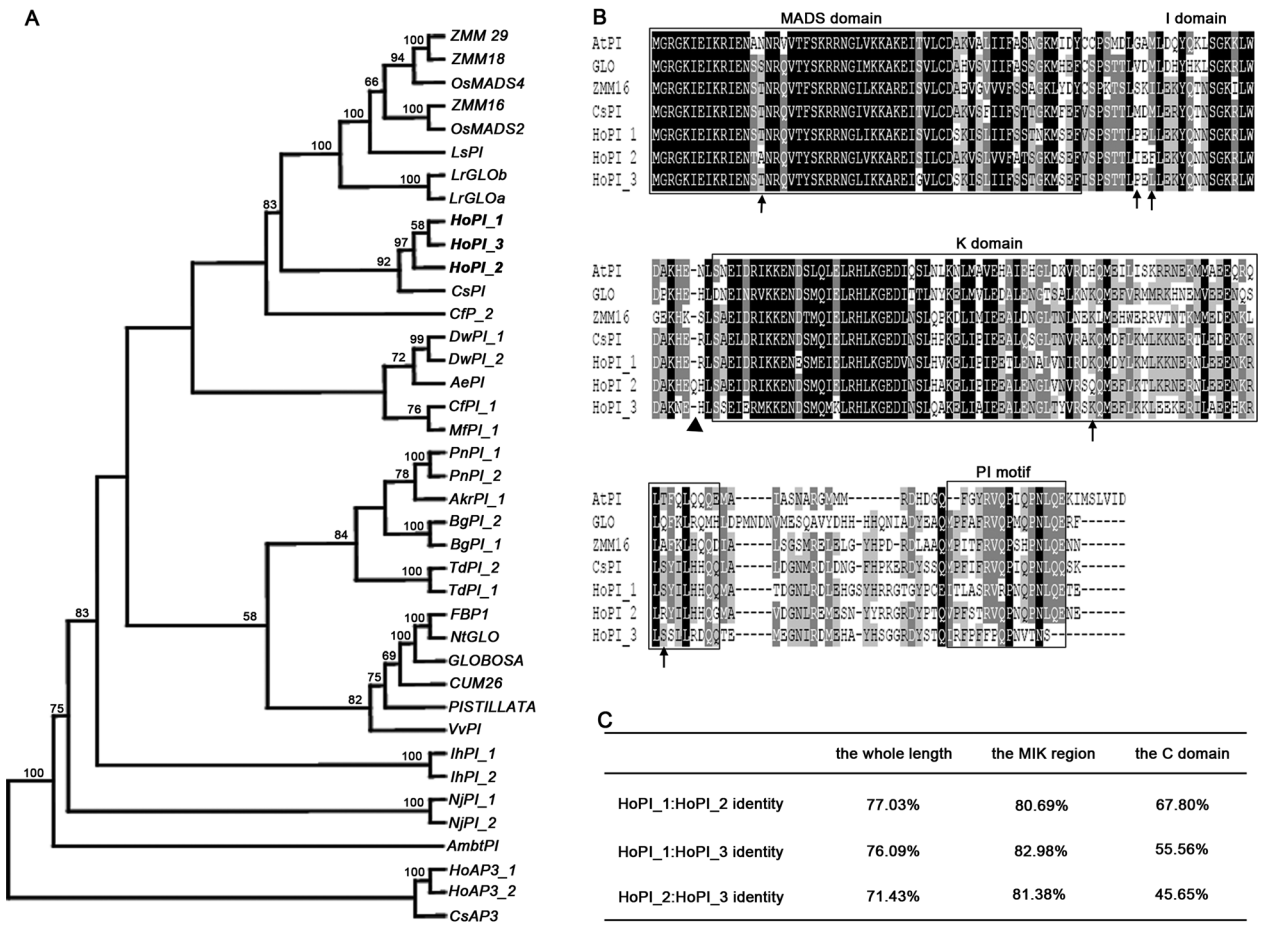

Fig. 1. Sequence alignment and phylogenic analyses of *PISTILLATA* (*PI*)-like genes. **A,** Maximum likelihood phylogeny of *PI*-like genes in angiosperms. Bootstrap values greater than 50 are noted on the nodes. The *APETALA3* (*AP3*)-like genes *CsAP3*, *HoAP3_1*, and *HoAP3_2* are used as representatives of the outgroup. **B,** Alignment of the predicted PI-like protein sequences. Codon changes in the MIK region among HoPI_2 and other proteins are indicated with triangle and arrow. **C,** Level of protein sequence similarity among three duplicate *HoPI* paralogs.

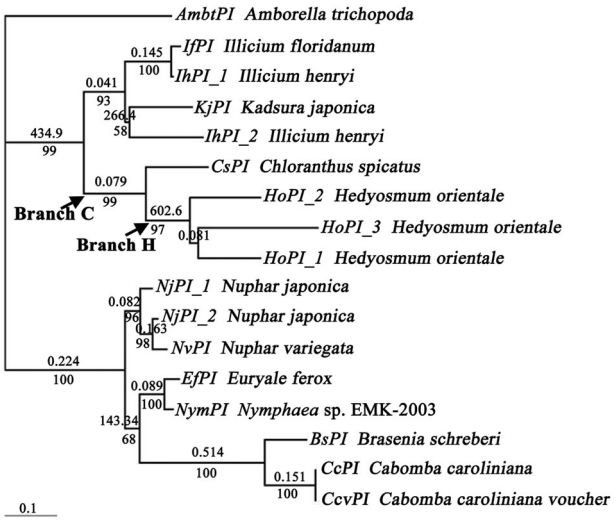

Fig. 2. Maximum likelihood tree used in selection pressure analysis. The number shown above each branch is the estimated ratio (ω) of pairwise synonymous (d*S*) and non-synonymous (d*N*) nucleotide substitution rates. Numbers below the branches are estimated bootstrap values. Bootstrap values greater than 50 are noted.

2.2 Selective pressure analysis acting upon *HoPI* genes

We treated branch H (including *HoPI_1*, *HoPI_2*, and *HoPI_3*, Fig. 2) as the foreground branch and all other branches in the phylogeny as background branches (Fig. 2). Likelihood values and parameters, as well as LRT statistics, are shown in Tables 1 and 2. The free-ratio model was found to be significantly better than the one-ratio model ($2\Delta \ln L = 45.05337$; degree of freedom ($d.f.$) = 30; $P = 0.038 < 0.05$, Table 2), supporting the hypothesis of variable selective constraints across different branches in the phylogeny. Branch H showed a ω value higher than 1 ($\omega_1 = 11.99823$, Table 1), but the M2-M0 and M2-null M2 comparisons indicated that M2 for branch H was not significantly better than M0 ($2\Delta \ln L = 0.447228$; $d.f. = 1$; $P = 0.504$, Table 2) and null M2 ($2\Delta \ln L = 0.013378$; $d.f. = 1$; $P = 0.908$, Table 2), suggesting that branch H might be acted upon by relaxation of functional constraint rather than positive selection.

In site-specific models, the discrete model (M3) with $K = 2$ or 3 site classes was used as an alternative hypothesis against M0, and the LRT statistics were significant (Table 2), suggesting a selection

Table 1 Likelihood values and parameter estimates for *HoPI* genes

Model	Parameters	ts/tv	n_p	ln *L*
Branch model				
M0: one-ratio	$\omega = 0.16171$	1.65825	33	−4238.698849
M1: free-ratio		1.66721	63	−4216.172164
M2: two-ratio branch H	$\omega_0 = 0.16009$, $\omega_1 = 11.99823$	1.65606	34	−4238.475235
M2: two-ratio null branch H	$\omega_0 = 0.16033$, $\omega_1 = 1.00000$	1.65642	33	−4238.481924
M2: two-ratio branch C	$\omega_0 = 0.17060$, $\omega_1 = 0.05407$	1.66750	34	−4235.443331
M2: two-ratio null branch C	$\omega_0 = 0.15750$, $\omega_1 = 1.00000$	1.66532	33	−4247.774873
Branch-site branch H				
Model A	$p_0 = 0.00013$, $p_1 = 0.00002$ $p_2 = 0.88976$, $p_3 = 0.11009$ $\omega_0 = 0.15222$, $\omega_2 = 1.00000$	1.81410	36	−4219.897944
Model A null	$p_2 = 0.88989$, $p_3 = 0.11011$	1.81410	35	−4219.897939
Model B	$p_0 = 0.39190$, $p_1 = 0.60810$ $p_2 + p_3 = 0$, $\omega_0 = 0.04297$ $\omega_1 = 0.27788$, $\omega_2 = 0.00000$	1.73252	37	−4193.160657
M3 ($K = 2$)	$p_0 = 0.40715$, $p_1 = 0.59285$ $\omega_0 = 0.04427$, $\omega_1 = 0.26968$	1.68317	35	−4193.160657
Site models				
M1a: nearly neutral	$p_0 = 0.88832$, $p_1 = 0.11168$ $\omega_0 = 0.15350$, $\omega_1 = 1.00000$	1.81981	34	−4220.157576
M2a: positive selection	$p_0 = 0.88832$, $p_1 = 0.10926$ $p_2 = 0.00242$, $\omega_0 = 0.15350$ $\omega_1 = 1.00000$, $\omega_2 = 1.00000$	1.81981	36	−4220.157576
M0: one ratio	$\omega = 0.16171$	1.65825	33	−4238.698849
M3: discrete ($K = 3$)	$p_0 = 0.24295$, $p_1 = 0.57267$ $p_2 = 0.18438$, $\omega_0 = 0.01929$ $\omega_1 = 0.16812$, $\omega_2 = 0.46679$	1.68315	37	−4186.975626
M7: beta	$p = 1.03493$, $q = 4.43917$	1.68412	34	−4188.208895
M8: beta and ω	$p_0 = 0.99999$, $p = 1.03493$ $q = 4.43917$, $\omega = 22.94430$	1.68413	36	−4188.210435

ln *L*, log likelihood; n_p, number of parameter estimates for the ω ratios; p_0, p_1, p_2, proportions for site classes ω_0, ω_1, ω_2, respectively; ts/tv, transition/transversion ratio; ω, substitution rate ratio calculated under each model. In M7 (beta) and M8 (beta and ω), the beta distribution for ω is determined by the parameters p and q.

Table 2 Likelihood ratio tests between different models of codon evolution

Compared models	2Δ ln *L*	*d.f.*	*P*-value
Branch-specific models			
M1-M0	45.05337	30	0.038*
M2-M0 branch H	0.447228	1	0.504
M2-M2 null branch H	0.013378	1	0.908
M2-M0 branch C	6.511036	1	0.011*
M2-M2 null branch C	24.663084	1	7E-07*
Site-specific models			
M3 ($K=3$)-M0	103.446446	4	2E-21*
M3 ($K=2$)-M0	91.076384	2	2E-20*
M2a-M1a	0	2	1
M8-M7	−0.00308	2	N.A.
Branch-site branch H			
Model A-Model A null	−1E-05	1	N.A.
Model A-M1a	0.519264	2	0.771
Model B-M3 ($K=2$)	0	2	1

d.f., degree of freedom; N.A., not applicable; *Significant at $P \leq 0.05$.

heterogeneity among codon sites. However, positive selection was not detected at any of the sites. The M2a and M8 models were not significantly better than the M1a or M7 models, respectively. As the more complex models did not fit the data better for either M2a-M1a or M8-M7 comparisons, M1a and M7 were accepted, respectively.

Branch-site models that allow positive selection only in the foreground branches (Model A) did not result in a better fit than the null Model A or M1a. Moreover, the LRT statistic of comparison of Model B versus M3 ($K=2$) was not significant (Table 2). These comparisons further suggested that no positive selection pressure acted on the *HoPI* coding sequences.

The branch model tests without *HoPI_2* were carried out and the result showed that when *HoPI_2* was eliminated from the matrix, a mean ω value of 0.336 was detected for branch H' (including *HoPI_1* and *HoPI_3*, Fig. S1), but the LRT statistic comparisons did not show significant statistical support (Tables S3 and S4). One possible explanation is that the species of basal magnoliids and ANITA grades we sampled in molecular evolution analyses are relatively distant from *Chloranthus* and *Hedyosmum* of Chloranthaceae, and there is a low sampling number of species included in the current alignment file. Successive evolutionary analyses have shown that both *AP3* and *PI* loci are subjected to strong purifying selection in the evolution of the class B MADS-box gene subfamily, even though relaxation of constraint or even positive selection were also inferred on particular internal branches or in certain lineages within the phylogeny across major angiosperms (Hernández-Hernández et al., 2007; Mondragón-Palomino et al., 2009; Shan et al., 2009). It is possible that the observed $\omega < 1$ of the branch leading to *HoPI_1* and *HoPI_3* likely reflects purifying selection dominating the evolution of these genes. In both site-specific and branch-site models, no positive selection sites were detected (Fig. S1; Tables S3 and S4).

2.3 Point mutations accumulated in MADS and I domains influence the B-function activity of HoPI_2 protein

In transgenic analyses, *HoPI_1*, *HoPI_2*, and *HoPI_3* show different abilities to promote petal and stamen development when expressed under the same constitutive *35S* or *Arabidopsis PI* (*pAtPI*) promoters, clearly suggesting that their differences in biochemical activity are driven by changes in their protein sequences (Tables 3 and 4). Protein sequence alignment shows that the most distinct difference between HoPI_2 and HoPI_1 (or HoPI_3) is at position 86 (Fig. 1: B, triangle). There is a single glutamine amino acid residue (Q_{86}) present in putative HoPI_2 protein but absent in both HoPI_1 and HoPI_3 as well as all other PI-like proteins. It is reasonable to propose that the failure of HoPI_2 to promote petals and stamens in transgenic *Arabidopsis* may be due to the single Q_{86} insertion. We tested this hypothesis by adding a Q at the appropriate position in HoPI_1 or HoPI_3 (HoPI_1/3+Q_{86} in Fig. 3: A) and, conversely, removing Q_{86} from HoPI_2 (HoPI_2-Q_{86} in Fig. 3: A), and observed the consequences of their expression in wild-type *Arabidopsis* and *pi-1* mutants. As shown in Tables 3 and 4, Q_{86} insertion resulted in the specific loss-ability of HoPI_1 and HoPI_3 to induce petaloid organs in the first whorl of wild-type *Arabidopsis* flowers and to restore petal and stamen identities in *Arabidopsis pi-1* mutants. However, removing Q_{86} from HoPI_2 did not switch its inability to activate petal and stamen identities in transgenic *Arabidopsis*. These experiments indicate that the presence of Q_{86} causes a defect in biochemical activity of HoPI proteins, and besides the Q_{86} insertion, there must be other disablements accumulated in the

Table 3 Summary of the overexpression phenotypes in domain swapping and site-directed mutagenesis analyses

Plant name	No. of plants with petaloid sepals (%)	No. of plants
35S::HoPI_2	0 (0)	51
35S::HoPI_2-Q_{86}	0 (0)	47
35S::HoPI_1	29 (74)	39
35S::HoPI_1+Q_{86}	0 (0)	12
35S::HoPI_1(K_{138}-Q)	18 (58)	31
35S::HoPI_1(S_{161}-R)	44 (65)	68
35S::HoPI_1(2MI)	0 (0)	29
35S::HoPI_1(2C)	23 (87)	26
35S::HoPI_3	78 (66)	118
35S::HoPI_3 + Q_{86}	0 (0)	26
35S::HoPI_3(K_{138}-Q)	9 (35)	26
35S::HoPI_3(S_{161}-R)	15 (38)	39

Table 4 Summary of rescue phenotypes in domain swapping and site-directed mutagenesis analyses

Plant name	Weak/no rescue (%)	Medium rescue (%)	Strong rescue (%)	No. of plants
pAtPI::HoPI_2;pi-1	38 (100)	0 (0)	0 (0)	38
pAtPI::HoPI_2-Q_{86};pi-1	9 (100)	0 (0)	0 (0)	9
pAtPI::HoPI_1;pi-1	4 (21)	4 (21)	11 (58)	19
pAtPI::HoPI_1+Q_{86};pi-1	8 (100)	0 (0)	0 (0)	8
pAtPI::HoPI_1(K_{138}-Q);pi-1	3 (20)	2 (13)	10 (67)	15
pAtPI::HoPI_1(S_{161}-R);pi-1	3 (20)	4 (27)	8 (53)	15
pAtPI::HoPI_1(2MI);pi-1	16 (100)	0 (0)	0 (0)	16
pAtPI::HoPI_1(2C);pi-1	3 (25)	3 (25)	6 (50)	12
pAtPI::HoPI_3;pi-1	12 (21)	36 (62)	10 (17)	58
pAtPI::HoPI_3 + Q_{86};pi-1	15 (88)	2 (12)	0 (0)	17
pAtPI::HoPI_3(K_{138}-Q);pi-1	2 (22)	5 (56)	2 (22)	9
pAtPI::HoPI_3(S_{161}-R);pi-1	2 (17)	6 (50)	4 (33)	12

HoPI_2 coding region that compromise its B-function activity.

To further explore the regions of HoPI_2 protein responsible for loss of B-function activity, we created a series of domain-swapping and site-directed mutagenesis constructs among HoPI_1, HoPI_2, and HoPI_3 (as shown in Fig. 3: A), and tested the consequences of their expression in wild-type *Arabidopsis* and *pi-1* mutant backgrounds, respectively. A number of independent transgenic lines were isolated for each of the constructs. The results showed that HoPI_1 (2C), HoPI_1 (K_{138}-Q), and HoPI_1 (S_{161}-R) were able to induce petaloid organs in the first whorl (Table 3; Fig. 3: C), and almost fully rescued the petal and stamen

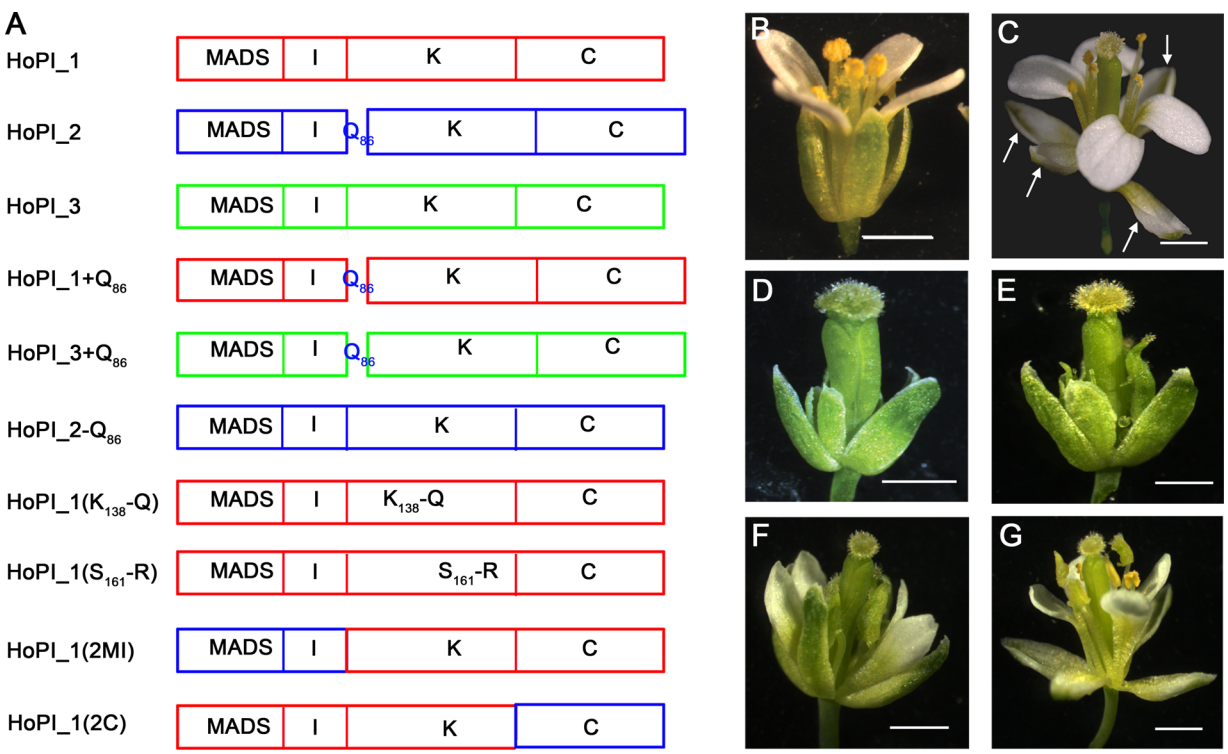

Fig. 3. PISTILLATA (PI) constructs used in transgenic experiments and phenotypes of transgenic *Arabidopsis* flowers.**Fig. 3. A,** PI constructs. Red boxed sections represent the region of HoPI_1 protein, blue represents the region of HoPI_2 protein, and green represents the region of HoPI_3 protein. HoPI_1 + Q_{86} and HoPI_3 + Q_{86} represent the variants of HoPI_1 and HoPI_3 where a single glutamine amino acid residue Q is inserted at position 86. HoPI_2-Q_{86} represents the variant of HoPI_2 where a single glutamine amino acid residue Q is removed from position 86; HoPI_1(K_{138}-Q) and HoPI_1 (S_{161}-R) represents the variants of HoPI_1 where a single lysine amino acid residue K is replaced by Q, and a single serine amino acid residue S is replaced by arginine R at position 138 and 161, respectively. HoPI_1(2MI) and HoPI_1(2C) represent constructs that replace the MI region, or the C-terminal region of HoPI_1 with those from HoPI_2, respectively. **B,** Wild-type flower. **C,** Flower of strong overexpression plants with petaloid sepals in the first whorl (white arrows). **D,** *pi-1* homozygous flower. **E–G,** *pi-1* homozygous mutant flowers of transgenic plants with different degrees of rescue. **E,** Flower of weak rescue lines: no rescue of petals, and carpeloid stamens in the third whorl. **F,** Flower of medium rescue lines: white but often short petals, and mosaic stamens. **G,** Flower of strong rescue lines: petals indistinguishable from wild-type petals, mosaic stamens or pollen produced stamens. Bar = 1 mm.

development in *pi-1* mutants (Table 4; Fig. 3: G), similar to the HoPI_1 transgenic plants. Likewise, most of the HoPI_3 (K_{138}-Q) and HoPI_3 (S_{161}-R) complementation plants showed a medium rescue in *Arabidopsis pi-1* mutants, which was similar to that observed in *HoPI_3* complementation plants (Table 4; Fig. 3: F). However, HoPI_1 (2MI) was not able to induce any homeotic transformation in transgenic *Arabidopsis* (Tables 3 and 4). These findings clearly indicated that, in addition to the Q_{86} insertion, decisive changes in the MADS and I domains (Fig. 1: B, arrows) also limit the B-function activity of HoPI_2 protein. Changes in the K and C domains (Fig. 1: B, arrows) are not deleterious to its B-function activity.

2.4 Point mutations accumulated on coding region of HoPI_2 do not affect its dimerization

Q_{86} is located in the 3′-terminus end of I domain that spans the junction of I domain and the amphipathic α-helix K1 of the K domain. As the I and K domains of MADS-box proteins have been shown to be important for dimerization (Riechmann et al., 1996a; Yang & Jack, 2004), we speculated that the insertion of Q_{86} might alter the specificity and affinity of dimerization. We tested the influence of the presence or absence of Q_{86} on dimerization between HoPI and *A. thaliana* AP1 (AtAP1) and AP3 (AtAP3) proteins by yeast two-hybrid experiments. As the MADS domain has been shown to interfere with detection of protein–protein interaction of class B gene products (Lamb & Irish, 2003), we used class B gene constructs lacking this domain. As shown in Fig. 4, MADS-deleted HoPI_2 and HoPI_3 showed an almost identical strong dimerization capability with AtAP1 and AtAP3 proteins in yeast two-hybrid analysis. In addition, the dimerization of chimeric HoPI_2+Q_{86} and HoPI_3-Q_{86} was not altered by removing or adding Q_{86}. These results suggest that the deleterious mutations accumulated on coding region of HoPI_2 do not affect its dimerization capability.

3 Discussion

We have isolated three *PI*-like duplicate genes from basal angiosperm *Hedyosmum orientale*. The phylogenetic analysis showed that the three *HoPI* paralogs likely resulted from two duplications (Fig. 1: A). The first duplication gave rise to *HoPI_2* and the progenitor gene of *HoPI_1* and *HoPI_3*, and the second, more recent, duplication gave rise to the duplicate gene pair *HoPI_1* and *HoPI_3*. As there is no molecular information of *PI*-like genes in other species of genus *Hedyosmum*, besides in *H. orientale*, we are unable to determine whether the duplications are genus-specific or species-specific. In addition, we also cannot estimate the exact time the gene duplication events occurred. However, the relatively high level of sequence similarity (70%–80% identity with each other at the protein level, Fig. 1: C) suggests that the two duplications may be relatively recent events.

After duplication events, duplicate genes could be expressed at equal levels, or there could be unequal expression or silencing of one copy (Kashkush

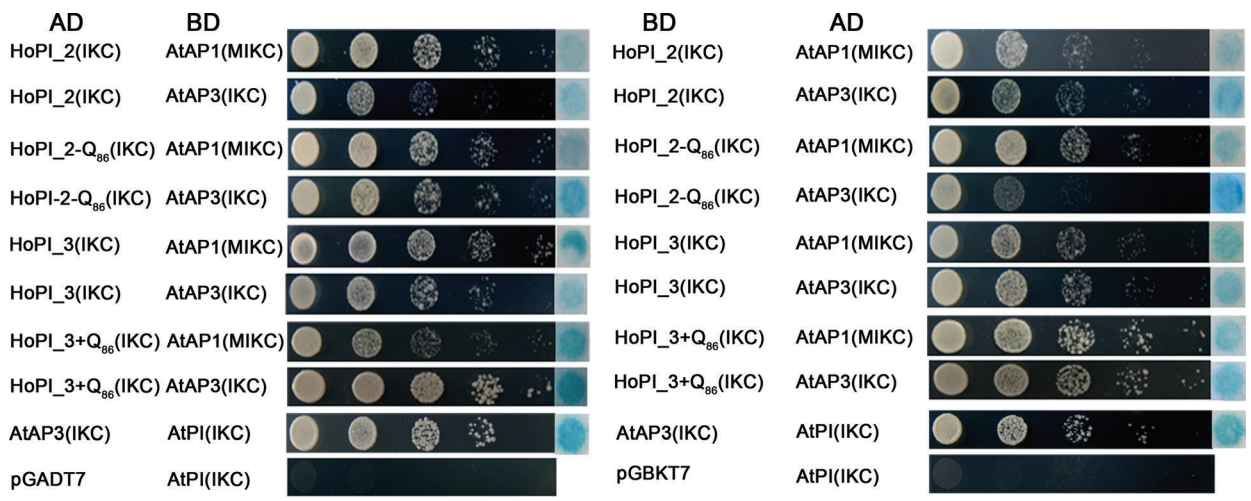

Fig. 4. Influence of the presence or absence of Q_{86} on dimerization between PISTILLATA-like protein from *Hedyosmum orientale* (HoPI) and *Arabidopsis thaliana* AP1 (AtAP1), AP3 (AtAP3) by yeast two-hybrid experiments. All of the class B proteins are MADS-deleted. HoPI_2 and HoPI_3 show an identical strong dimerization affinity with *Arabidopsis* class A and B proteins. The dimerization of HoPI proteins is not altered by removing or adding Q_{86}.

et al., 2002; Adams et al., 2003, 2004; Wang et al., 2004). In *H. orientale*, *HoPI_1* and *HoPI_3* are expressed at almost equally high levels, but the expression of *HoPI_2* is very low (Liu et al., 2013). In addition, the coding sequence of *HoPI_1* and *HoPI_3* could function as well as the *Arabidopsis PI*, but *HoPI_2* could not induce any phynotype in transgenic analysis (Tables 3 and 4). These findings suggest that *in vivo* the recent duplicate paralogs *HoPI_1* and *HoPI_3* may be preserved and functionally redundant in specifying stamens by interacting with *HoAP3* genes, whereas the relatively distant paralog *HoPI_2* has been converted to be weak or functionless in this process because we detected no B-function activity of the *HoPI_2* coding sequence in transgenic analysis. However, this duplicate copy has not been completely silenced or even lost because we can still detect its weak expression in *H. orientale* and have obtained its DNA sequence (Doc. S1) by amplification from the genome of *H. orientale*. Alternatively, it is also possible that *HoPI_2* has differentiated a novel function in floral development by activating unique targets or interacting with novel protein partners in *H. orientale* (neofunctionalization), and these neofunctions would be impossible to assess in the context of *Arabidopsis*. This question can only be answered by further detailed experimental characterization of this gene copy *in vivo*.

It has been postulated that initially after gene duplication, one copy is redundant (not essential), and as such, this copy will experience a phase of relaxed selection (Lynch & Conery, 2000) or even be relieved from functional constraint (neutral genetic drift) (Ohno, 1973). In the present study, selective pressure analysis showed that the *HoPI* branch has undergone selective pressure that is significantly different from that of other branches, and these *HoPI* paralogs have accumulated more non-synonymous than synonymous mutations, but there is no evidence of positive selection acting upon codon sites of *HoPI* genes (Fig. 2; Tables 1 and 2). The coding sequences of duplicate *HoPI_1* and *HoPI_3* may evolve under purifying selection, but the functional constraint on *HoPI_2* may have been relaxed (Fig. S1; Tables S3 and S4). Then the question arises as to whether the relaxation of selection revealed in the coding sequence of *HoPI_2* reflects simply a long-term accumulation under a relaxed selection pressure, or an abrupt increase in an episodic period for functional divergence following the duplication event. Relaxed purifying selection following gene duplication has been reported in many studies (dePamphilis et al., 1997; Yokoyama & Blow, 2001; Leebens-Mack & dePamphilis, 2002) and it seems to be a generalized feature in the early stage of duplication evolution (reviewed in Zhang, 2003), so the relaxed selection we observe in *HoPI_2* is to some extent due to the occurrence of the first duplication event. It is notable, however, that the relatively recent duplicate gene pair *HoPI_1* and *HoPI_3* does not show evidence of relaxed selection, as that of *HoPI_2*. Therefore, the relaxation of functional constraints observed in this relatively distant copy *HoPI_2* may actually be a long-term accumulation under a relaxed selection. Nevertheless, we cannot rule out the possibility that after the first duplication event, *HoPI_2* had been maintained in the genome for a long time for functional redundancy, then maybe after the *HoPI_1/3* duplication event, becoming non-functional in the B-function developmental process because of the relaxation of functional constraints.

In principle, unless there is selection for their products, one copy might eventually be led to complete non-functionalization by accumulating a sufficient number of deleterious mutations, and be further set on the course of pseudogenization (Lynch & Conery, 2000; Lynch & Katju, 2004). Based on our domain swapping and site-directed mutagenesis experiments, we found that inactivating point mutations in the MADS and I domains of the *HoPI_2* coding sequence, including the Q_{86} insertion mutation in the 3′-terminus end of I domain, render it non-functional in terms of B-function activity (Tables 3 and 4). Previous studies have reported that in MADS and I domains, as well as in K1 and K2 subdomains involved in DNA binding and dimerization, there may be functional constraints greater than those acting on the K3 subdomain and the C-terminal region (Riechmann & Meyerowitz, 1997; Yang & Jack, 2004). Therefore, mutations in the MI region are more likely to have negative effects on the biochemical activity of PI-like proteins, making this area the "hotspot region" that is susceptible to deleterious mutations. However, the yeast two-hybrid results showed that the dimerization affinities to AP1- and AP3-like proteins of HoPI proteins are almost identical, and the dimerization is not affected by the presence/absence mutation of Q_{86} (Fig. 4). These findings suggest that mutations accumulated on the coding region of HoPI_2 compromise the B-function activity of this protein, but do not affect its dimerization. Perhaps these naturally occurring variations limit its ability in the B-function process by influencing DNA binding, or protein complex formation between AP3-PI heterodimers and other MADS-box proteins such as SEPALLATA (de Folter et al., 2005; Immink et al., 2009).

In this study, we revealed the evolutionary mechanism underlying the coding sequence divergence of three duplicate *HoPI* copies: the recent duplicate

paralogs *HoPI_1* and *HoPI_3* under purifying selection might have been fixed because of their beneficial property of functional redundancy, whereas the relatively distant copy *HoPI_2* might have accumulated degenerative mutations in MADS and I domains as a result of relaxed functional constraints. Although we currently have no direct *in vivo* evidence that *HoPI_2* has lost function in floral development, it is clear that this copy has become non-functional in the B-function developmental process, and deleterious mutations accumulated in the MI region are responsible for the loss-of-function of *HoPI_2* in this process. However, the three *HoPI* gene copies not only diverge in biochemical activity, but also in expression levels (Liu et al., 2013), which implies that the regulatory regions may also be involved in their divergence. Therefore, further detailed molecular study on their regulatory regions is needed for a comprehensive understanding of the mechanism underlying the divergence of the three *HoPI* gene copies.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 31270280 and 31100867) and the Ministry of Science and Technology of China (Grant No. 2011CB100405).

References

Adams KL, Cronn R, Percifield R, Wendel JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proceedings of the National Academy of Sciences USA 100: 4649–4654.

Adams KL, Percifield R, Wendel JF. 2004. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics 168: 2217–2226.

Becker A, Theißen G. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Molecular Phylogenetics and Evolution 29: 464–489.

Blanc G, Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. The Plant Cell 16: 1679–1691.

Clough SJ, Bent AF. 1998. Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. The Plant Journal 16: 735–743.

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

Cui L, Wall K, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW. 2006. Widespread genome duplications throughout the history of flowering plants. Genome Research 16: 738–749.

de Folter S, Immink RGH, Kieffer M, Parenicova L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM, Davies B, Angenenta GC. 2005. Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. The Plant Cell 17: 1424–1433.

dePamphilis CW, Young ND, Wolfe AD. 1997. Evolution of plastid gene *rps2* in a lineage of hemiparasitic and holoparasitic plants: Many losses of photosynthesis and complex patterns of rate variation. Proceedings of the National Academy of Sciences USA 94: 7367–7372.

Egea-Cortines M, Saedler H, Sommer H. 1999. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. The EMBO Journal 18: 5370–5379.

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

Goto K, Meyerowitz EM. 1994. Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. Genes and Development 8: 1548–1560.

Gu X, Zhang Z, Huang W. 2005. Rapid evolution of expression and regulatory divergences after yeast gene duplication. Proceedings of the National Academy of Sciences USA 102: 707–712.

Gu Z, Nicolae D, Lu HH-S, Li W-H. 2002. Rapid divergence in expression between duplicate genes inferred from micro-array data. Trends in Genetics 18: 609–613.

Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li W-H. 2003. Role of duplicate genes in genetic robustness against null mutations. Nature 421: 63–66.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52: 696–704.

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169: 1157–1164.

Hernández-Hernández T, Martínez-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. Molecular Biology and Evolution 24: 465–481.

Honma T, Goto K. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature 409: 525–529.

Immink RGH, Tonaco IAN, de Folter S, Shchennikova A, van Dijk ADJ, Busscher-Lange J, Borst JW, Angenent GC. 2009. SEPALLATA3: The 'glue' for MADS box transcription factor complex formation. Genome Biology 10: R24.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: Classifying and distinguishing between models. Nature Reviews Genetics 11: 97–108.

Irish VF, Litt A. 2005. Flower development and evolution: Gene duplication, diversification and redeployment. Current Opinion in Genetics and Development 15: 454–460.

Jack T. 2004. Molecular and genetic mechanisms of floral control. The Plant Cell 16: 1–17.

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.

Jack T, Fox GL, Meyerowitz EM. 1994. *Arabidopsis* homeotic gene *APETALA3* ectopic expression: Transcriptional and posttranscriptional regulation determine floral organ identity. Cell 76: 703–716.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473: 97–100.

Kashkush K, Feldman M, Levy A. 2002. Gene loss, silencing, and activation in a newly synthesized wheat allotetraploid. Genetics 160: 1651–1659.

Kaufmann K, Melzer R, Theissen G. 2005. MIKC-type MADS-domain proteins: Structural modularity, protein interactions and network evolution in land plants. Gene 347: 183–198.

Kim S, Koh J, Yoo M-J, Kong H, Hu Y, Ma H, Soltis PS, Soltis DE. 2005. Expression of floral MADS-box genes in basal angiosperms: Implications for the evolution of floral regulators. The Plant Journal 43: 724–744.

Kramer EM, Di Stilio VS, Schluter PM. 2003. Complex patterns of gene duplication in the *APETALA3* and *PISTILLATA* lineages of the Ranunculaceae. International Journal of Plant Sciences 164: 1–11.

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*. The Plant Cell 19: 750–766.

Lamb RS, Irish VF. 2003. Functional divergence within the *APETALA3/PISTILLATA* floral homeotic gene lineages. Proceedings of the National Academy of Sciences USA 100: 6558–6563.

Leebens-Mack J, dePamphilis C. 2002. Power analysis of tests for loss of selective constraint in cave crayfish and nonphotosynthetic plant lineages. Molecular Biology and Evolution 19: 1292–1302.

Li G-S, Meng Z, Kong H-Z, Chen Z-D, Theissen G, Lu A-M. 2005a. Characterization of candidate class A, B and E floral homeotic genes from the perianthless basal angiosperm *Chloranthus spicatus* (Chloranthaceae). Development Genes and Evolution 215: 437–449.

Li W-H, Yang J, Gu X. 2005b. Expression divergence between duplicate genes. Trends in Genetics 21: 602–607.

Liu SL, Adams KL. 2010. Dramatic change in function and expression pattern of a gene duplicated by polyploidy created a paternal effect gene in the Brassicaceae. Molecular Biology and Evolution 27: 2817–2828.

Liu S, Sun Y, Du X, Xu Q, Wu F, Meng Z. 2013. Analysis of the *APETALA3*- and *PISTILLATA*-like genes in *Hedyosmum orientale* (Chloranthaceae) provides insight into the evolution of the floral homeotic B-function in angiosperms. Annals of Botany. doi:10.1093/aob/mct182.

Lynch M. 2002. Gene duplication and evolution. Science 297: 945–947.

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

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics 154: 459–473.

Lynch M, Katju V. 2004. The altered evolutionary trajectories of gene duplicates. Trends in Genetics 20: 544–549.

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

Mondragón-Palomino M, Hiese L, Härter A, Koch MA, Theißen G. 2009. Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses. BMC Evolutionary Biology 9: 81.

Mondragón-Palomino M, Theißen G. 2011. Conserved differential expression of paralogous *DEFICIENS*- and *GLOBOSA*-like MADS-box genes in the flowers of Orchidaceae: Refining the 'orchid code.' The Plant Journal 66: 1008–1019.

Moore RC, Purugganan MD. 2003. The early stages of duplicate gene evolution. Proceedings of the National Academy of Sciences USA 100: 15682–15687.

Moore RC, Purugganan MD. 2005. The evolutionary dynamics of plant duplicate genes. Current Opinion in Plant Biology 8: 122–128.

Nicholas KB, Nicholas HB Jr, Deerfield DW II. 1997. GeneDoc: Analysis and visualization of genetic variation. EMBnet. NEWS 4: 1–4.

Ohno S. 1970. Evolution by gene duplication. New York: Springer.

Ohno S. 1973. Ancient linkage groups and frozen accidents. Nature 244: 259–262.

Proost S, Pattyn P, Gerats T, Van de Peer Y. 2011. Journey through the past: 150 million years of plant genome evolution. The Plant Journal 66: 58–65.

Qian WF, Liao B-Y, Chang AYF, Zhang JZ. 2010. Maintenance of duplicate genes and their functional redundancy by reduced expression. Trends in Genetics 26: 425–430.

Riechmann JL, Krizek BA, Meyerowitz EM. 1996a. Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. Proceedings of the National Academy of Sciences USA 93: 4793–4798.

Riechmann JL, Meyerowitz EM. 1997. MADS domain proteins in plant development. Biological Chemistry 378: 1079–1101.

Riechmann JL, Wang M, Meyerowitz EM. 1996b. DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Research 24: 3134–3141.

Shan H, Su K, Lu W, Kong H, Chen Z, Meng Z. 2006. Conservation and divergence of candidate class B genes in *Akebia trifoliata* (Lardizabalaceae). Development Genes and Evolution 216: 785–795.

Shan H, Zahn L, Guindon S, Wall PK, Kong H, Ma H, dePamphilis CW, Leebens-Mack J. 2009. Evolution of plant MADS box transcription factors: Evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. Molecular Biology and Evolution 26: 2229–2244.

Sharma B, Guo C, Kong H, Kramer EM. 2011. Petal-specific subfunctionalization of an *APETALA3* paralog in the Ranunculales and its implications for petal evolution. New Phytologist 191: 870–883.
Shiu S-H, Shih M-C, Li W-H. 2005. Transcription factor families have much higher expansion rates in plants than in animals. Plant Physiology 139: 18–26.
Soltis D, Albert V, Leebens-Mack J, Bell C, Paterson A, Zheng C, Sankoff D, dePamphilis C, Wall P, Soltis P. 2009. Polyploidy and angiosperm diversification. American Journal of Botany 96: 336–348.
Sommer H, Beltran J-P, Huijser P, Pape H, Lonnig W-E, Saedler H, Schwarz-Sommer Z. 1990. *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: The protein shows homology to transcription factors. The EMBO Journal 9: 605–613.
Stellari GM, Alejandra Jaramillo M, Kramer EM. 2004. Evolution of the *APETALA3* and *PISTILLATA* lineages of MADS-box-containing genes in the basal angiosperms. Molecular Biology and Evolution 21: 506–519.
Theißen G, Saedler H. 2001. Floral quartets. Nature 409: 469–471.
Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter K-U, Saedler H. 2000. A short history of MADS-box genes in plants. Plant Molecular Biology 42: 115–149.
Tsai W-C, Kuoh C-S, Chuang M-H, Chen W-H, Chen H-H. 2004. Four *DEF*-Like MADS box genes displayed distinct floral morphogenetic roles in *Phalaenopsis* orchid. Plant and Cell Physiology 45: 831–844.
Wagner A. 2001. Birth and death of duplicated genes in completely sequenced eukaryotes. Trends in Genetics 17: 237–239.
Wang J, Tian L, Madlung A, Lee H, Chen M, Lee J, Watson B, Kagochi T, Comai L, Chen Z. 2004. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. Genetics 167: 1961–1973.
Yang Y, Jack T. 2004. Defining subdomains of the K domain important for protein–protein interactions of plant MADS proteins. Plant Molecular Biology 55: 45–59.
Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular Biology and Evolution 15: 568–573.
Yang Z. 2007. PAML4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24: 1586–1591.
Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Molecular Biology and Evolution 19: 908–917.
Yokoyama S, Blow NS. 2001. Molecular evolution of the cone visual pigments in the pure rod-retina of the nocturnal gecko, *Gekko gekko*. Gene 276: 117–125.
Zahn LM, Leebens-Mack J, dePamphilis CW, Ma H, Theissen G. 2005. To B or not to B a flower: The role of *DEFICIENS* and *GLOBOSA* orthologs in the evolution of the angiosperms. Journal of Heredity 96: 225–240.
Zhang J. 2003. Evolution by gene duplication: An update. Trends in Ecology and Evolution 18: 292–298.
Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular Biology and Evolution 22: 2472–2479.

Supplementary Material

The following supplementary material is available for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12045/suppinfo:

Fig. S1. Maximum likelihood tree after eliminating gene *HoPI_2* from the matrix. The number shown above each branch is the estimated ratio (ω) of pairwise synonymous (d*S*) and non-synonymous (d*N*) nucleotide substitution rates. Numbers below the branches are estimated bootstrap values. Bootstrap values greater than 50 are noted.

Doc. S1. DNA sequence of *HoPI_2* (exon3–exon7). The section in red represents the exon sequence.

Table S1. Accession information for all loci used in the phylogenetic and evolutionary analysis.

Table S2. Primers used in this study.

Table S3. Likelihood values and parameter estimates for *HoPI* genes after eliminating *HoPI_2* from the matrix.

Table S4. Likelihood ratio tests between different models after eliminating *HoPI_2* from the matrix.