Importance of Gene Duplication in the Evolution of Genomic Imprinting Revealed by Molecular Evolutionary Analysis of the Type I MADS-Box Gene Family in *Arabidopsis* Species

Takanori Yoshida, Akira Kawabe[*]

Faculty of Life Science, Kyoto Sangyo University, Kyoto, Kyoto, Japan

Abstract

The pattern of molecular evolution of imprinted genes is controversial and the entire picture is still to be unveiled. Recently, a relationship between the formation of imprinted genes and gene duplication was reported in genome-wide survey of imprinted genes in *Arabidopsis thaliana*. Because gene duplications influence the molecular evolution of the duplicated gene family, it is necessary to investigate both the pattern of molecular evolution and the possible relationship between gene duplication and genomic imprinting for a better understanding of evolutionary aspects of imprinted genes. In this study, we investigated the evolutionary changes of type I MADS-box genes that include imprinted genes by using relative species of *Arabidopsis thaliana* (two subspecies of *A. lyrata* and three subspecies of *A. halleri*). A duplicated gene family enables us to compare DNA sequences between imprinted genes and its homologs. We found an increased number of gene duplications within species in clades containing the imprinted genes, further supporting the hypothesis that local gene duplication is one of the driving forces for the formation of imprinted genes. Moreover, data obtained by phylogenetic analysis suggested "rapid evolution" of not only imprinted genes but also its closely related orthologous genes, which implies the effect of gene duplication on molecular evolution of imprinted genes.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: akiraka@cc.kyoto-su.ac.jp

Introduction

Genomic imprinting is a phenomenon that causes complete or partial uniparental gene expression of particular genes (called "imprinted genes"). In plants, genomic imprinting mainly occurs in the endosperm of developing seeds that nourish the embryo during and after seed development. Recent studies using genome-wide surveys have identified more than one hundred candidate imprinted genes in *Arabidopsis thaliana*, *Oryza sativa*, and *Zea mays* [1–6].

The basis of regulation of genomic imprinting in developing seeds has been intensively investigated, and two epigenetic mechanisms have been identified as factors that regulate genomic imprinting in plants: (1) DNA methylation and (2) histone methylation [7–12]. In plants, the methylation status of the differently methylated regions in the proximal region of the imprinted genes between maternal and paternal alleles affects the expression of some imprinted genes [7,8]. In addition to the

methylation status of the DNA, trimethylation of H3K27 catalyzed by Polycomb repressive complex 2 (PRC2) also affects the expression of imprinted genes by silencing either the paternal or maternal allele [12]. It is reported that transposable elements (TEs) are methylated in the embryo but extensively demethylated in the endosperm, affecting the imprinting status of the nearby genes [1]. These results may support the defense theory that the status of imprinting arises as a byproduct of silencing of the invading DNA fragments such as TEs [13]. Recent studies carried out using genome-wide analysis of imprinted genes have focused on understanding the relationship between gene duplication and the gene imprinting status [4]. Together with TEs, gene duplication might also affect the imprinting status and evolution of imprinted genes.

Another question concerning genomic imprinting is the pattern of molecular evolution of the imprinted genes. The parental conflict theory predicts that conflict between maternal
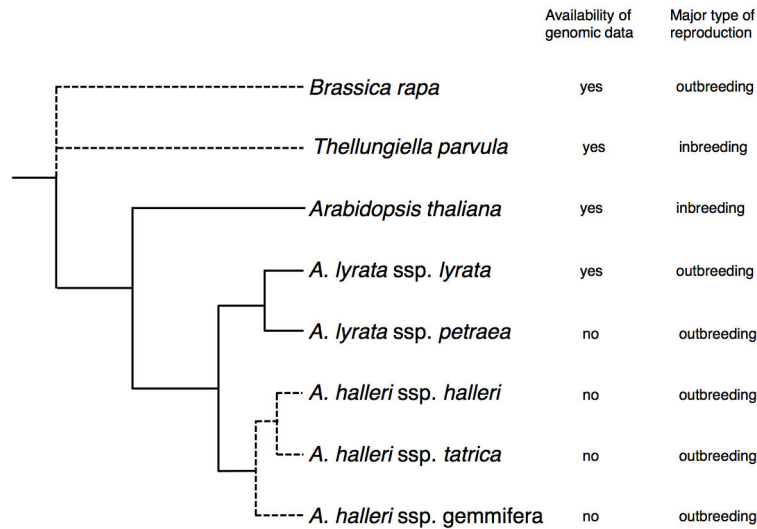
Figure 1. Schematic diagram showing the phylogenetic relationship among species used in this study. The reproduction type and availability of genomic data are also indicated. Dashed lines are used to indicate partly ambiguous phylogenetic relations.

and paternal genomes causes evolutionary "arms races" and that the imprinted genes show accelerated molecular evolution [14]. For example, Spillane et al. suggested that an imprinted locus *MEDEA* in *Arabidopsis* is under the influence of positive Darwinian selection with neo-functionalization after its generation by whole-genome duplication [15]. However, in the analysis of molecular evolution of some imprinted genes of both mammals and plants, no evidence for antagonistic coevolution was detected [16,17]. Thus, a comprehensive understanding of the effect of genomic imprinting on the pattern of molecular evolution of imprinted genes remains elusive.

In this study, we investigated the evolutionary changes of type I MADS-box genes, which include imprinted genes, by using relative species of *A. thaliana*. MADS-box genes encode transcription factors that contain the conserved MADS-box domain. Members of type I MADS-box gene family in plants have been reported to be involved in reproductive development and expressed in developing seeds (Gene networks in seed development website; available from: http://seedgenenetwork.net/) [18,19]. Some genes in the family (*PHERES1*, *At1g59930*, *AGL36*, and *AGL92*) show imprinted gene expression in *A. thaliana* [3,20,21]. *PHERES1* and *AGL36* were identified as imprinted genes by the observation of expression patterns in the endosperms [20–22]. *At1g59930* and *AGL92* were identified by deep sequencing of the developing seeds [3]. *PHERES1* and *AGL92* are paternally expressed, whereas *AGL36* and *At1g59930* are maternally expressed. The available data suggest that in both *O. sativa* and *A. thaliana* type I MADS-box, gene duplications occurred independently [23]. One of the duplicated genes, *OsMADS87*, is also reported as a maternally expressed imprinted gene [24]. Imprinted genes of these two species emerged independently among duplicated gene clusters. Unlike single-copy imprinted genes, duplicated genes enable us to compare DNA sequences of imprinted genes and their non-imprinted

homologs. Thus, type I MADS-box genes are a useful resource to study molecular evolution of imprinted genes and the relationship between gene duplication and genomic imprinting. First, we focused on identifying the relationship between gene duplication and the genomic imprinting status. The incidence of gene duplication in each clade of type I MADS-box genes was estimated to assess its relationship to genomic imprinting. Next, we investigated the effect of genomic imprinting on molecular evolution. Our results suggest a positive relationship between gene duplication and the incidence of imprinted genes and "rapid evolution" in clades containing imprinted genes. We discuss possible driving forces that trigger these correlations, thereby bringing about the evolution of imprinted genes.

Materials and Methods

Plant materials

Closely related species of *A. thaliana* were used in this study (Figure 1). Publicly available genomic data of *A. thaliana*, *A. lyrata* ssp. *lyrata*, *Thellungiella parvula* (salt cress), and *Brassica rapa* were used in this study for comparisons [25–28]. The origin of other plant materials was as follows: four *A. lyrata* ssp. *petraea* strains from Plech Germany (kindly provided by J. de Meaux); *A. halleri* ssp. *gemmifera* strain 144-1 isolated in Mino-shi, Osaka, Japan; *A. halleri* ssp. *tatrica* strain T-PLDH1 isolated in Vysoke Tatry, Poland; *A. halleri* ssp. *halleri* strain H–RB isolated in Bistrita, Romania; *Turritis glabra* strain OM isolated in Ohmi-Shirahama, Shiga, Japan; and *Crucihimalaya wallichii* strain SJS00500 obtained from RIKEN BioResource Center.

Sequence analysis of orthologous type I *MADS*-box genes from a publicly available database

Sequences of type I MADS-box genes of *A. thaliana*, *A. lyrata* ssp. *lyrata*, *T. parvula*, and *B. rapa* were collected from a publicly available NCBI Genome database (http://www.ncbi.nlm.nih.gov/genome). For *A. lyrata* ssp. *lyrata*, *T. parvula*, and *B. rapa*, we performed a BLAST search optimized for somewhat similar sequences (blastn) to search database of each genome with sequences of *At5g26575* (*AGL34*), *At5g26630* (*AGL35*), *At5g26650* (*AGL36*), *At1g65330* (*PHE1*), *At1g65300* (*PHE2*), *At3g05860* (*AGL45*), *At2g28700* (*AGL46*), *At5g48670* (*AGL80*), *At1g31630* (*AGL86*), *At5g27960* (*AGL90*), and *At1g31640* (*AGL92*) as queries. We also used MADS-box family genes *At5g35120*, *At1g59920*, and *At1g59930* that lack the MADS-box motif.

Cloning and sequencing of type I MADS-box genes

In this study, we analyzed type I MADS-box genes in *A. thaliana* because these genes contain several imprinted genes and are phylogenetically related to each other. To obtain orthologous genes of each type I MADS-box gene of *A. thaliana* from relative species, eight primer pairs were used for PCR (Table S1 in File S1). Genomic DNAs were isolated from fresh leaves by using DNeasy Plant Mini Kit (QIAGEN, USA). DNA fragments orthologous to type I MADS-box genes were amplified by PCR and cloned using T-Vector pMD20 (TaKaRa, Japan) and Ligation Mighty Mix (TaKaRa, Japan). Multiple DNA sequences from *A. lyrata* ssp. *petraea* and three subspecies of *A. halleri* (ssp. *halleri*, ssp. *tatrica*, and ssp. *gemmifera*) were cloned. For each primer pair, we sequenced up to 15 clones from each relative species. Only sequences that were verified at least twice from independent clones were used for the following analyses. When diverged sequence(s) was obtained from the initial 15 clones, additional clones were sequenced. The sequence data was aligned by manual inspection. Newly determined sequences were deposited in DDBJ under the accession numbers AB830633-AB830706.

Computational analyses of molecular evolution

Phylogenetic relationship was estimated by the neighbor-joining method with Jukes and Cantor distance [29] by using the MEGA5 program [30]. Pairwise ω (d_N/d_S) ratios of type I MADS-box genes and other MADS transcription factor family genes between *A. thaliana* and *A. lyrata* ssp. *lyrata* were calculated using DnaSP ver.5 [31]. BLAST searches were performed to find pairwise genes between *A. thaliana* and *A. lyrata* ssp. *lyrata*. Annotated *A. thaliana*'s 46 type I MADS-box genes and 63 other MADS transcription factor family genes (TAIR *Arabidopsis* Gene Family Information; available from: http://www.arabidopsis.org) were used as queries. For each clade of type I MADS-box gene, we estimated the ratio of non-synonymous (d_N) to synonymous (d_S) substitution rate (ω) by the maximum likelihood method implemented in program codeml in program package PAML ver. 4.4b [32]. A sequence of *A. lyrata* that showed similarity to the *At1g59930* gene was excluded from the analysis because of the large deletion. Sequences from *T. glabra* were used as out-groups. We could not obtain the *T. glabra* orthologous sequence of *AGL46*;

therefore, the sequence obtained from *C. wallichii* was used as an out-group. Tree topologies of each clades constructed by the neighbor-joining method were used in simulations of codeml. We applied a free-ratio model in which it is assumed that n parameters of the ω-ratio are equal to the total number of branches in the phylogeny.

Our data suggested that branches in the clade containing imprinted genes have evolved faster (higher ω-ratio) than those in the clade without imprinted genes. To verify this tendency, we estimated ω-ratio using the following models: Model 0, one-ratio model with single ω-ratio for all branches and Model 1, two-ratio model with two ω-ratios (ω_1 and ω_2). A likelihood ratio test between Model 0 and Model 1 was conducted.

Results

Phylogenetic analyses of orthologous type I *MADS*-box genes

The phylogenetic relationship of the relative species of *A. thaliana* used in this study is shown in Figure 1. We used gene sequences from the whole-genome sequences of *A. thaliana*, *A. lyrata* ssp. *lyrata*, *T. parvula*, and *B. rapa* to estimate phylogenetic clustering and the number of duplications within and among species. Phylogenetic relationships of type I MADS-box genes close to known imprinted genes were estimated by the neighbor-joining method (Figure 2). We identified eight distinguishable clades (designated as clades I–VIII in Figure 2) when *A. thaliana* and *A. lyrata* genes were considered. Each clade from I to IV contained one imprinted gene, whereas there were no imprinted genes in clades V–VIII. Thus, we compared clades I–IV with clades V–VIII to assess the effect of imprinting on gene duplication and molecular evolution. Meanwhile, *B. rapa* and *T. parvula* genes showed a completely different phylogenetic relation between these two clusters. A clade composed of six orthologous sequences from *B. rapa* and two sequences from *T. parvula* was clustered with clades III and IV (designated as clade B). The phylogenetic position of this clade is close to the base of clades that contain the imprinted genes, implying that *B. rapa* and *T. parvula* would have experienced independent duplication events during their evolutionary histories. In contrast to sequences in clade B, other *B. rapa* and *T. parvula* sequences were clustered with each of clade V–VIII. For each numbered clade, a detailed phylogenetic relation was separately investigated using only *A. thaliana* and its close relatives that showed high similarity of sequences (Figure S1 in File S1). *B. rapa* and *T. parvula* were not included, because these two species were highly diverged from *A. thaliana* causing inaccurate phylogenetic estimations.

Estimated number of gene duplication in each clade

The result of phylogenetic analyses obtained using a publicly available database indicated that the incidence of gene duplication and genomic imprinting is variable among clades and species. To investigate the variation in gene duplication among more close relatives, orthologous sequences of *A. lyrata* ssp. *petraea* and three subspecies of *A. halleri* were amplified. The number of sequences obtained for each clade is shown in Table 1. Allelic variants might be included in the count
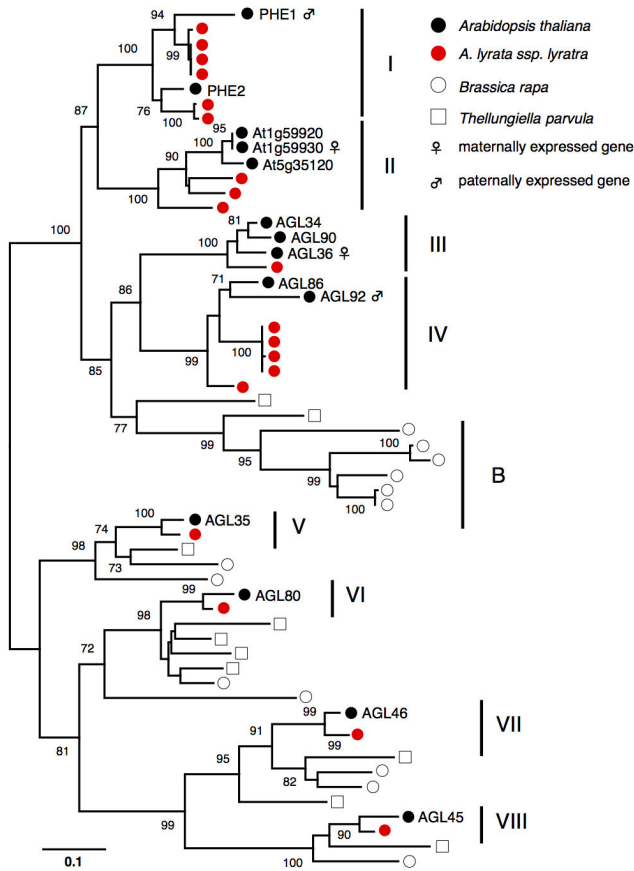
Figure 2. Neighbor-joining tree of type I MADS-box genes. Black circles; *A. thaliana*, red circles; *A. lyrata* ssp. *lyrata*, empty circles; *B. rapa*, and empty squares; *T. parvula*. Bootstrap values (%) were estimated by 500 replications for each clade and shown at corresponding nodes. Nine distinguishable clades shown as thick vertical lines are designated as clade I–VIII and clade B. The scale bar is shown below the tree.

doi: 10.1371/journal.pone.0073588.g002

because these species are outbred. Interestingly, there were several duplicated sequences in clades I–IV that contain imprinted genes. In some cases, duplicated sequences clustered to each other and seldom had segregating sites among them (For example, five sequences of *A. lyrata* ssp. *lyrata* in clade IV; Figure S2 in File S1) implying recent tandem gene duplication within species. However, the gene duplication does not always occur in orthologous sequences of imprinted genes but is also found in the paralogous sequences of the clade containing known imprinted genes (clade IV in Figure S1 in File S1). In contrast to abundant gene duplications in clades I–IV, there were fewer gene duplications in clades V–VIII. No duplicated gene was found in *A. lyrata* ssp. *petraea*. The apparent difference in the gene duplication patterns between clade I-IV and clade V-VIII indicates that different evolutionary mechanisms might contribute to the status of genomic imprinting in these clades.

Table 1. Estimated number of duplication in each clade.

	A. thaliana	A. lyrata		A. halleri			B. rapa	T. parvula
		ssp. lyrata	ssp. petraea	ssp. gemmifera	ssp. tatrica	ssp. halleri		
I	PHE1♂, PHE2	6 (4, 2)	5 (4, 1)	2 (1, 1)	1	3 (2, 1)	-	-
II	At5g35120, At1g59920, At1g59930♀	3 (1, 1, 1)	3 (1, 1, 1)	3 (1, 1, 1)	3 (1, 1)	3 (1, 1, 1)	-	-
III	AGL34, AGL36♀, AGL90	1	2 (2)	3 (3)	3 (3)	3 (3)	-	-
IV	AGL86, AGL92♂	5 (4, 1)	4 (3, 1)	1	7 (6, 1)	6 (5, 1)	-	-
V	AGL35	1	1	1	1	1	2 (1, 1)	1
VI	AGL80	1	1	no data	2 (2)	2 (2)	1	4 (1, 1, 1, 1)
VII	AGL46	1	1	1	1	1	2 (1, 1)	2 (1, 1)
VIII	AGL45	1	1	1	1	1	1	1
B	-	-	-	-	-	-	6 (1, 2, 1, 2)	2 (1, 1)

Numbers in parentheses represent a set of sequences with divergence less than 0.1.
♂. paternally expressed gene.
♀. maternally expressed gene.

Molecular evolution of type I MADS-box genes

Pairwise ω-ratio between each type I MADS-box gene of *A. thaliana* and *A. lyrata* was compared with other MADS transcription factor family genes (Figure 3). Apparent orthologous pairs between the two species were used to estimate the ω-ratio: six pairs of type I MADS-box genes in clades I–IV, 29 pairs of other type I MADS-box genes, and 36 pairs of other MADS transcription factor family genes (type II MADS-box genes and others). A median ω-ratio of orthologous gene pairs in clades I–IV was 0.73, whereas those of other type I MADS-box genes and other MADS transcription factor family genes were 0.41 and 0.23, respectively. A median ω-ratio of gene pairs between *A. thaliana* and *A. lyrata* in clades I–IV was significantly higher than those of other type I MADS-box gene pairs and other MADS transcription factor family gene pairs, respectively (p = 0.0047 and p = 0.0003, respectively; Wilcoxon Rank Sum). This difference suggested that the type I MADS-box gene clades, including imprinted genes, were under positive selection or under relaxed selective constraint.
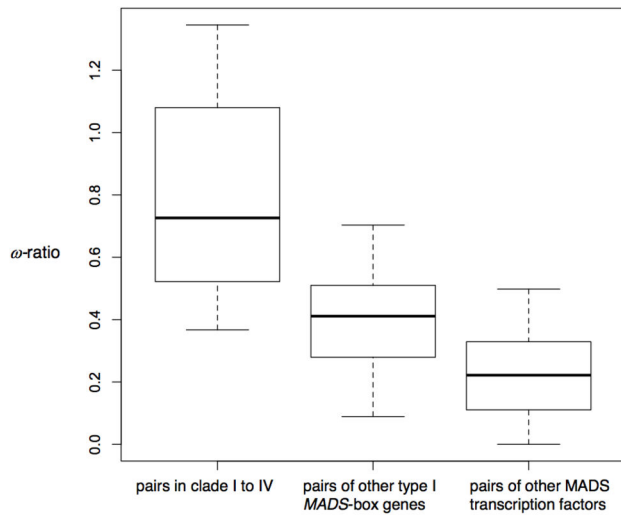
Figure 3. Boxplot representation of pairwise ω-ratio of MADS-box genes between *A. thaliana* and *A. lyrata*. Left; gene pairs in clade I–IV, center; gene pairs in clade V–VIII, and right; gene pairs of other MADS transcription factor family genes. Bold lines represent medians and thin lines of edges of the box represent the first and third quartiles. Lines from whiskers represent minimum and maximum values.

doi: 10.1371/journal.pone.0073588.g003

To investigate the rate of evolutionary changes on the branches of each clade, computational simulations were conducted using the program package PAML. The estimated ω-ratio and tree topology is shown in Figure 4 with synonymous and non-synonymous distances (d_S and d_N). Branches in clades containing imprinted genes (I, II, III, and IV) showed relatively high ω-ratio values compared to those in other clades (V, VI, VII, and VIII). Interestingly, some external branches of non-imprinted locus, sister to imprinted genes showed ω > 1 indicative of adaptive evolution (for example clade I and IV). The result suggested that branches in the clade containing imprinted genes evolved faster (higher ω-ratio) than those in the clade without imprinted genes regardless of the imprinting status of the locus. To verify this tendency, a likelihood ratio test was performed using the test statistics 2ΔL = $2(l_{\text{two-ratio}} - l_{\text{one-ratio}})$, where $l_{\text{one-ratio}}$ and $l_{\text{two-ratio}}$ are log likelihood values of each model. A one-ratio model assuming a single ω-ratio (ω_0) for all branches was compared with a two-ratio model assuming two ω-ratio: ω_1 for branches in clades containing imprinted genes and branches leading to these clades, and ω_2 for other branches (Table 2). In a two-ratio model, the estimated value of ω_1 was higher than the estimated value of ω_0 (ω_1 = 0.53 and ω_2 = 0.27). The two-ratio model presented a significantly better fit to the data than the one-ratio model with a single ω-ratio for the whole phylogeny (2ΔL = 17.9799; *p* = 0.0000). The significant difference in ω-ratio between the two clades point to the role of different evolutionary forces acting on these clades and the observed differences could be due to the imprinting status of the genes.

Discussion

Relationship between gene duplications and imprinted genes

In this study, we observed a positive relationship between number of gene duplication events and presence of imprinted gene in the clades of type I MADS-box gene family. The effect of gene duplications on the evolution of imprinted genes has been firstly discussed in studies of placental mammals [33,34] and later in a genome-wide survey of imprinted genes in plants [4]. In mammals, Walter et al. suggested that imprinted genes have many paralogs that are imprinted or are close to imprinted genes [33]. In a genome-wide survey of *A. thaliana*, significantly higher number of imprinted genes was found in clustered genes than expected by chance [4]. Interestingly, homologous genes of most gene clusters are not always imprinted but include non-imprinted genes. These imprinted genes have a significantly increased numbers of close homologs in comparison to the genome-wide average. From these results, Wolff et al. suggested local gene duplication as a driving force for formation of genomic imprinting [4].

Our finding in this study might, in part, support this hypothesis. The tendency for increased gene duplication in clades containing imprinted genes was detected not only in *A. thaliana*, but also in its close relatives. The question is why and how gene duplication could promote the formation of imprinted genes. There are two ways to understand the relationship between gene duplication and genomic imprinting: (1) gene duplication followed by genomic imprinting, or (2) genomic imprinting leads to gene duplication. In both cases, the first event (gene duplication or genomic imprinting) can change the gene dosage but the direction is opposite. In the former case, gene duplication increases the amount of transcripts and genomic imprinting may control the expression by complete/ partial mono-allelic expression. In contrast, in the latter case the newly generated genomic imprinting reduces the expression level. If the reduction is deleterious, a gene duplication that compensates the gene dosage may be selectively advantageous and the frequency of the duplicated mutants may increase. It should be noted that gene duplications could be considered as driving force for the formation of imprinted genes only in the former case.

Previous studies showed that the formation of imprinted genes is tightly associated with the presence of TEs [1,3,4]. The gene duplications and the presence of TEs might not be entirely distinct, but rather, these factors could be linked with each other. Wolff et al. showed a significant enrichment of TEs in the vicinity of imprinted genes [4]. TEs can cause TE-mediated gene duplication that affect evolution of imprinted genes. The possible interplay between gene duplication and presence of TEs during imprinted gene evolution should be analyzed in the future study.

Molecular evolution of imprinted genes and its homologs

The pattern of molecular evolution of imprinted genes had been analyzed in both plants and placental mammals [4,15–17,35–37]. However, the entire picture of the
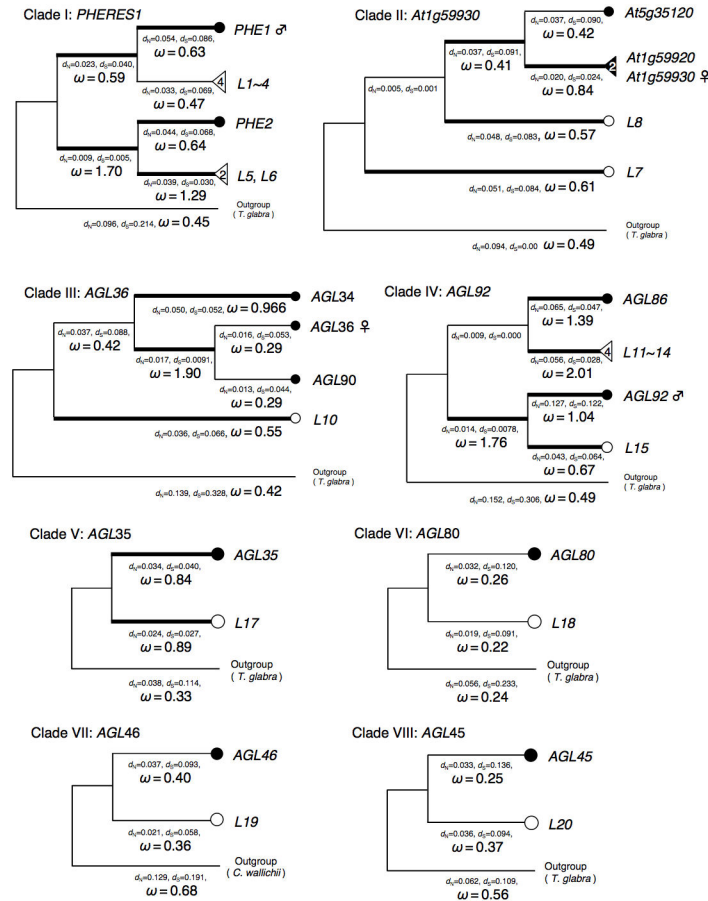
Figure 4. Phylogenetic trees of the type I MADS-BOX genes. For each clade, phylogenetic tree was estimated by Neighbor-joining method with Jukes and Canter's distance using MEGA5. For convenience, sequences of *A. lyrata* were numbered from *L1* to *L20*. Orthologous sequences from *T. glabra* and *C. wallichii* were used as out-groups for each clade. Imprinting status of *A. thaliana* are shown after gene name as follows: ♂; paternally expressed gene, ♀; maternally expressed gene. ω-ratio, d_N, and d_S values estimated by PAML are shown below the branches. Origin of the species is indicated as follows: Black circles and triangles; *A. thaliana*, empty circles and triangles; *A. lyrata* ssp. *lyrata*. Triangles represent cluster of recently duplicated genes. Copy number of the cluster is shown in the triangle. Branches with high ω-ratio (ω > 0.5) are shown as thick lines.

doi: 10.1371/journal.pone.0073588.g004

evolutionary pattern of imprinted genes is still controversial. In plant, some studies suggested natural selection acting on the imprinted gene *MEDEA* (*MEA*) [15,35,36]. Spillane et al. indicated a positive Darwinian selection on an external branch, leading to *A. lyrata MEA* [15], while Kawabe et al. found high diversity in the *MEA* promoter region and suggested balancing selection acting on the promoter [35]. In contrast, Haun et al. concluded that the Enhancer-of-zeste, orthologous gene of Arabidopsis *MEA*, is under pressure of purifying selection [17]. Recently, molecular evolution of other imprinted genes has been analyzed. The genome-wide survey of imprinted gene suggests rapid evolution of candidate genes relative to other backgrounds [4].

In the previous studies, it has been assumed that the formation of imprinting is the main cause of observed pattern of molecular evolution such as high ω-ratio. Our result might bring a new insight into the pattern of molecular evolution of

Table 2. Maximum likelihood parameter estimation for the two models.

	ω-ratios			Likelihood ratio test	
	ω_0				
model	ω_1	ω_2	lnL	2(lnL$_{two-ratio}$ -lnL$_{one-ratio}$)	p
one-ratio	0.39221		-5211.025458		
two-ratio	0.53408	0.26901	-5202.035533	17.97985	0.0000

imprinted genes. The results in this study imply a rapid evolution of not only imprinted genes but also its paralogous genes. For example, ω-ratio of the branch leading to *L5* and *L6* in clade I is higher than 1, while the ω-ratio of branches of *AtPHE1* and *L1 ~ 4* is approximately 0.5. In the case of clade

IV, ω-ratio of the branch leading to *L11~14* is higher than 2 (Figure 4) while its sister locus *AGL86* was not imprinted. In addition, model selection test between one-ratio and two-ratio models by PAML suggested the higher ω-ratio of clades I–IV than that of clades V–VIII. These high ω-ratios might not be directly due to genomic imprinting but due to gene duplication followed by neo-functionalization or relaxation of selective constraint. In theoretical studies of gene duplication, the trajectory of duplicated genes and effects of duplications on its molecular evolution have been investigated [38–45]. Gene duplication is a main source of new genes [46] and causes higher evolutionary rate in one or both duplicates [45]. Innan et al. classified gene duplications and its evolutionary trajectory into four categories (categories I to IV) [47]. Most models predict elevations of ω-ratio before fixation of newly derived mutations in duplicates by relaxed selective pressure, positive selection on duplications or pre-duplicational variations, and pseudogenization of duplicates [47]. Although the result in this study did not specify the most suitable model, the homologous genes of imprinted genes in clades I–IV might evolve faster mainly by the effect of gene duplication.

Possibility of rapid turnover of imprinted status

Another possibility of discordance between imprinted status and rapid molecular evolution is a rapid turnover of imprinted status among orthologous type I MADS-box genes if the imprinted status is a cause of rapid molecular evolution. A comparison of identified imprinted genes of rice, maize and *A. thaliana* suggests that the conservation of imprinted status among these plant species is limited [12]. For example, Luo et al. suggested 165 candidate imprinted loci of rice but only 27 loci have significant sequence homology with the candidate imprinted loci of *A. thaliana* [5]. These limited conservations across species may represent the result of a rapid formation and degradation of imprinted status. The high values of ω-ratio in internal branches observed in this study might imply vestiges of past rapid evolution caused by genomic imprinting and succeeding rapid molecular evolution. It is important to note that the detail of turnover of imprinting status is still unknown because species mentioned above are phylogenetically divergent and the limited conservation of imprinted loci might reflect independent origin of imprinting for each locus. The survey of expression patterns of candidate imprinted genes in closely related species will provide the estimation of turnover rate of imprinting status to test this hypothesis.

Conclusion

Our results support the view that there exists a relationship between gene duplication and the generation of genomic imprinting. In addition, the clades including imprinted genes tend to evolve faster. However, genomic imprinting is not always the cause of acceleration of molecular evolution. Instead, our results support the view that gene duplication before or after the generation of new genomic imprinting could cause relaxation of selective constraint or non/sub-functionalization, thus leading to increased ω-ratios that were observed in this study. Gene duplication could be one of the driving forces causing evolution of imprinted genes. In the future, analysis of other imprinted genes using *A. thaliana* and closely related species is necessary to test the generality of this hypothesis.

Supporting Information

File S1. Table S1, Figures S1-S2. Table S1. Primer pairs used in this study. Figure S1. Neighbor-joining trees of clades I–VIII. Phylogenetic relationship was estimated with the Jukes and Cantor distance. Bootstrap values (%) were estimated by 500 replications for each clade and shown at corresponding nodes. All trees are shown in a same scale. A distance bar is shown at the bottom. Sequence of each gene excluding alignment gaps or indels was used for estimations. Black circle; *A. thaliana*, red circle; *A. lyrata* ssp. *lyrata*, red triangle; *A. lyrata* ssp. *petraea*, green circle; *A. halleri* ssp. *gemmifera*, open green circle; *A. halleri* ssp. *halleri*, green triangle circle; *A. halleri* ssp. *tatrica*, empty diamond; *C. wallichii*, black diamond; *T. glabra*. Figure S2. Location of the homologs in the genome of *A. lyrata* ssp. *lyrata*. Each line represents large scaffolds covering the majority of each of the 8 chromosomes of *A. lyrata*. The numbered boxes from L1 to L20 representing homologous sequences are identical with sequences in Figure 4. Tandem duplicated sequences are shown in scaffold 1 and 2.
(PPTX)

Acknowledgements

We thank Dr. Tetsu Kinoshita for comments and suggestions for the early version of this manuscript. We also thank anonymous reviewers for helpful suggestions.

Author Contributions

Conceived and designed the experiments: TY AK. Performed the experiments: TY AK. Analyzed the data: TY AK. Contributed reagents/materials/analysis tools: TY AK. Wrote the manuscript: TY AK.

References

1. Gehring M, Bubb KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324: 1447-1451. doi:10.1126/science.1171609. PubMed: 19520961.

2. Gehring M, Missirian V, Henikoff S (2011) Genomic Analysis of Parent-of-Origin Allelic Expression in *Arabidopsis thaliana* Seeds. PLOS ONE 6: e23687. Available: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0023687. Accessed 1 May 2013. doi: 10.1371/journal.pone.0023687. PubMed: 21858209.

3. Hsieh TF, Shin J, Uzawa R, Silva P, Cohen S et al. (2011) Regulation of imprinted gene expression in *Arabidopsis* endosperm. Proc Natl Acad Sci U S A 108: 1755-1762. doi:10.1073/pnas.1019273108. PubMed: 21257907.

4. Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C et al. (2011) High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm. PLOS Genet 7: e1002126. Available: http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1002126. Accessed 1 May 2013.

5. Luo M, Taylor JM, Spriggs A, Zhang H, Wu X et al. (2011) A Genome-Wide Survey of Imprinted Genes in Rice Seeds Reveals Imprinting Primarily Occurs in the Endosperm. PLOS Genet 7: e1002125. Available: http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1002125#close. Accessed 1 May 2013.

6. Zhang M, Zhao H, Xie S, Chen J, Xu Y et al. (2011) Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. Proc Natl Acad Sci U S A 108: 20042-20047. doi:10.1073/pnas.1112186108. PubMed: 22114195.

7. Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X et al. (2004) One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. Science 303: 521-523. doi:10.1126/science.1089835. PubMed: 14631047.

8. Jullien PE, Kinoshita T, Ohad N, Berger F (2006) Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting. Plant Cell 18: 1360-1372. doi:10.1105/tpc.106.041178. PubMed: 16648367.

9. Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y et al. (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell 124: 495-506. doi: 10.1016/j.cell.2005.12.034. PubMed: 16469697.

10. Baroux C, Gagliardini V, Page DR, Grossniklaus U (2006) Dynamic regulatory interactions of *Polycomb* group genes: *MEDEA* autoregulation is required for imprinted gene expression in *Arabidopsis*. Genes Dev 20: 1081-1086. doi:10.1101/gad.378106. PubMed: 16651654.

11. Weinhofer I, Hehenberger E, Roszak P, Hennig L, Köhler C (2010) H3K27me3 profiling of the endosperm inplies exclusion of polycomb group protein targeting by DNA methylation. PLOS Genet 6: e1001152. Available: http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1001152. Accessed 1 May 2013.

12. Ikeda Y (2012) Plant imprinted genes identified by genome-wide approaches and their regulatory mechanisms. Plant Cell Physiol 53: 809-816. doi:10.1093/pcp/pcs049. PubMed: 22492232.

13. Barlow DP (1993) Methylation and imprinting: from host defense to gene regulation? Science 260: 309-310. doi:10.1126/science.8469984. PubMed: 8469984.

14. Haig D, Westoby M (1989) Parent-specific gene expression and the triploid endosperm. Am Nat 134: 147-155. doi:10.1086/284971.

15. Spillane C, Schmid KJ, Laoueillé-Duprat S, Pien S, Escobar-Restrepo JM et al. (2007) Positive Darwinian selection at the imprinted *MEDEA* locus in plants. Nature 448: 349-352. doi:10.1038/nature05984. PubMed: 17637669.

16. McVean GT, Hurst LD (1997) Molecular evolution of imprinted genes: no evidence for antagonistic coevolution. Proc R Soc Lond B Biol Sci 264: 739-746. doi:10.1098/rspb.1997.0105. PubMed: 9178545.

17. Haun WJ, Laoueillé-Duprat S, O'connell MJ, Spillane C, Grossniklaus U et al. (2007) Genomic imprinting, methylation and molecular evolution of maize Enhancer of zeste (Mez) homologs. Plant J 49: 325-337. doi:10.1111/j.1365-313X.2006.02965.x. PubMed: 17181776.

18. Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF et al. (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. Proc Natl Acad Sci U S A 107: 8063-8070. doi:10.1073/pnas.1003530107. PubMed: 20385809.

19. Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ et al. (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. Proc Natl Acad Sci U S A 110: E435-E444. doi:10.1073/pnas.1222061110. PubMed: 23319655.

20. Köhler C, Page DR, Gagliardini V, Grossniklaus U (2005) The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nat Genet 37: 28-30. PubMed: 15619622.

21. Shirzadi R, Andersen ED, Bjerkan KN, Gloeckle BM, Heese M et al. (2011) Genome-wide transcript profiling of endosperm without parental contribution identifies parent-of-origin-dependent regulation of *AGAMOUS-LIKE36*. PLOS Genet 7: e1001303. Available: http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1001303. Accessed 1 May 2013.

22. Köhler C, Hennig L, Spillane C, Pien S, Gruissem W (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. Genes Dev 17: 1540-1553. doi:10.1101/gad.257403. PubMed: 12815071.

23. Arora R, Agarwal P, Ray S, Singh AK, Singh VP et al. (2007) MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. BMC Genomics 8: 242. Available: http://www.biomedcentral.com/1471-2164/8/242. Accessed 1 May 2013. doi: 10.1186/1471-2164-8-242. PubMed: 17640358.

24. Ishikawa R, Ohnishi T, Kinoshita Y, Eiguchi M, Kurata N et al. (2011) Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. Plant J 65: 798-806. doi:10.1111/j.1365-313X.2010.04466.x. PubMed: 21251103.

25. Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M et al. (2008) The Arabidopsis information resource (TAIR): gene structure and function annotation. Nucleic Acids Res 36: D1009-D1014. Available: http://nar.oxfordjournals.org/content/36/suppl_1/D1009.long. Accessed 1 May 2013. PubMed: 17986450.

26. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF et al. (2011) The Arabidopsos lyrata genome sequence and the basis of rapid genome size change. Nat Genet 43: 476-481. doi:10.1038/ng.807. PubMed: 21478890.

27. Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H et al. (2011) The genome of the extremophile crucifer *Thellungiella parvula*. Nat Genet 43: 913-918. doi:10.1038/ng.889. PubMed: 21822265.

28. Wang X, Wang H, Wang J, Sun R, Wu J et al. (2011) The genome of the mesopolyploid crop species Brassica rapa. Nat Genet 43: 1035-1039. doi:10.1038/ng.919. PubMed: 21873998.

29. Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: HN Munro. Mammalian protein metabolism. New York: Academic Press. pp. 21-132.

30. Tamura K, Peterson D, Peterson N, Stecher G, Nei M et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28: 2731-2739. doi:10.1093/molbev/msr121. PubMed: 21546353.

31. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452. doi: 10.1093/bioinformatics/btp187. PubMed: 19346325.

32. Yang Z (2007) PAML4: aprogram package for phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-1591. doi:10.1093/molbev/msm088. PubMed: 17483113.

33. Walter J, Paulsen M (2003) The potential role of gene duplications in the evolution of imprinting mechanisms. Hum Mol Genet 12: 215-220. doi:10.1093/hmg/ddg296. PubMed: 12944422.

34. O'Connell MJ, Loughran NB, Walsh TA, Donoghue MTA, Schmid KJ et al. (2010) A phylogenetic approach to test for evidence of parental conflict or gene duplications associated with protein-encoding imprinted orthologous genes in placental mammals. Mamm Genome 21: 486-498. doi:10.1007/s00335-010-9283-5. PubMed: 20931201.

35. Kawabe A, Fujimoto R, Charlesworth D (2007) High diversity due to balancing selection in the promoter region of the Medea gene in Arabidopsis lyrata. Curr Biol 17: 1885-1889. doi:10.1016/j.cub.2007.09.051. PubMed: 17949979.

36. Miyake T, Takebayashi N, Wolf DE (2009) Possible Diversifying Selection in the Imprinted Gene, MEDEA, in Arabidopsis. Mol Biol Evol 26: 843-857. doi:10.1093/molbev/msp001. PubMed: 19126870.

37. Smith NG, Hurst LD (1998) Molecular Evolution of an Imprinted Gene: Repeatability of Patterns of Evolution Within the Mammalian Insulin-Like Growth Factor Type II Receptor. Genetics 150: 823-833. PubMed: 9755212.

38. Ohno S (1970) Evolution by gene duplication. New York: Springer Verlag.

39. Nei M (1969) Gene duplication and nucleotide substitution in evolution. Nature 221: 40-42. doi:10.1038/221040a0. PubMed: 5782607.

40. Kimura M, King JL (1979) Fixation of a deleterious allele at one of two 'duplicate' loci by mutation pressure and random drift. Proc Natl Acad

Sci U S A 76: 2858-2861. doi:10.1073/pnas.76.6.2858. PubMed: 288072.

41. Li WH (1980) Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. Genetics 95: 237-258. PubMed: 7429144.

42. Stoltzfus A (1999) On the possibility of constructive neutral evolution. J Mol Evol 49: 169-181. doi:10.1007/PL00006540. PubMed: 10441669.

43. Force A, Lynch M, Pickett FB, Amores A, Yan YL et al. (1999) Preservation of duplicated genes by complementary, degenerative mutations. Genetics 151: 1531-1545. PubMed: 10101175.

44. Tanaka KM, Takahasi KR, Takano-Shimizu T (2009) Enhanced fixation and preservation of a newly arisen duplicate gene by masking

deleterious loss-of-function mutations. Genet Res 91: 267-280. doi: 10.1017/S0016672309000196.

45. Gossmann TI, Schmid KJ (2011) Selection-driven divergence after gene duplication in Arabidopsis thaliana. J Mol Evol 73: 153-165. doi: 10.1007/s00239-011-9463-2. PubMed: 21965041.

46. He X, Zhang J (2005) Rapid Subfunctionalization Accompanied by Prolonged and Substantial Neofunctionalization in Duplicate Gene Evolution. Genetics 169: 1157-1164. doi:10.1534/genetics.104.037051. PubMed: 15654095.

47. Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11: 97-108. doi:10.1038/ni0110-97d. PubMed: 20051986.