Duplication and Functional Divergence in the Floral Organ Identity Genes

Kelsey Diane Galimba

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Verónica Di Stilio, Chair

Billie Swalla

Takato Imaizumi

Program Authorized to Offer Degree:

Biology

University of Washington

Abstract

Duplication and Functional Divergence in the Floral Organ Identity Genes

Kelsey Diane Galimba

Chair of the Supervisory Committee:

Dr. Verónica Di Stilio

Department of Biology

Among all extant land plant lineages, the flowering plants, or angiosperms, are by far the most abundant and diverse. A leading force in their speciation has been attributed to morphological changes in flowers, a key reproductive innovation. The floral organ identity genes are thus a logical focus to gain insight into the angiosperm radiation, since alterations to both the structure and regulation of developmental genes are known to underlie morphological diversity. Genetic studies on flower development in the model plants *Arabidopsis* and *Antirrhinum* have yielded the 'ABC model', which predicts that the four types of floral organs (sepals, petals, stamens and carpels) are specified by the products of four gene classes, A, B, C, and the subsequently added E class, that act in combination. All but one of these transcription factors belong to the MADS-box family, their function analogous to the metazoan body patterning of HOX genes.

Plant genomes are prone to duplication and paralogs are particularly widespread in the floral organ identity genes. My aim was to elucidate the function of representatives of an early clade of

duplicated floral organ identity genes, with a focus on determining the degree of redundancy vs. divergence amongst them and of conservation in relation to the later-diverging models. To that end, I use the species *Thalictrum thalictroides,* from the clade sister to all other eudicots, with two C class gene orthologs and three B class gene orthologs. In the C class genes, one paralog showed deep conservation of the dual C class function, affecting both reproductive organ identity and meristem determinacy, while the second revealed specialization to ovule identity. This sub-functionalization is unlike the fate of the B class genes following duplication, which retained partial redundancy, showed deep conservation in stamen identity, and gained a novel role in ectopic petaloidy. Taken together, these studies illustrate the different fates that floral organ identity gene paralogs can assume following duplication that can ultimately be relevant to the generation of biodiversity. Moreover, they underscore the need to survey the functional relevance of all branches of a gene's evolutionary history.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I have received an enormous amount of support and encouragement from a great number of individuals throughout the realization of this dissertation. I would particularly like to thank Dr. Verónica Di Stilio for the enormous amount of time and energy she has dedicated to mentoring me, and for the lasting positive effects she will always have on me a scientist. I would also like to thank my committee members, Dr. Billie Swalla, Dr. Takato Imaizumi and Dr. Willie Swanson, for their insight and inspiration, which influenced both my research and my writing.

DEDICATION

I would like to dedicate this dissertation to my grandfather and grandmother, Roy and Loretta Miller. Grandpa shared with me his love for the natural world, and his seemingly boundless knowledge regarding the plants and animals of the California foothills helped inspire my pursuit of biology. Grandma always encouraged my love of flowers, and some of my fondest memories are being allowed to gather bouquets and play in her gardens to my heart's content.

Loss of deeply conserved C class floral homeotic gene function and C and E class protein interaction in a double-flowered ranunculid mutant

Kelsey D. Galimba, Theadora R. Tolkin, Alessandra M. Sullivan, Rainer Melzer, Günter Theißen, and Verónica S. Di Stilio

Abstract

In the model plant *Arabidopsis thaliana*, a core eudicot, the floral homeotic C class gene *AGAMOUS (AG)* has a dual role specifying reproductive organ identity and floral meristem determinacy. We conduct a functional analysis of the putative *AG* ortholog *ThtAG1* from the ranunculid *Thalictrum thalictroides*, a representative of the sister lineage to all other eudicots. Down-regulation of *ThtAG1* by virus-induced gene silencing resulted in homeotic conversion of stamens and carpels into sepaloid organs and loss of flower determinacy. Moreover, flowers exhibiting strong silencing of *ThtAG1* phenocopied the double-flowered ornamental cultivar *T. thalictroides* 'Double White.' Molecular analysis of 'Double White' *ThtAG1* alleles revealed the insertion of a retrotransposon causing either nonsense- mediated decay of transcripts or alternative splicing that results in mutant proteins with K-domain deletions. Biochemical analysis demonstrated that the mutation abolishes protein–protein interactions with the putative E class protein ThtSEP3. C and E class protein heterodimerization is predicted by the floral quartet model, but evidence for the functional importance of this interaction is scarce outside the core eudicots. Our findings therefore corroborate the importance and conservation of the interactions between C and E class proteins. This study provides a functional description of a full C class mutant in a noncore ("basal") eudicot, an ornamental double flower, affecting both organ identity and meristem determinacy. Using complementary forward and reverse genetic approaches, this study

demonstrates deep conservation of the dual C class gene function and of the interactions between C and E class proteins predicted by the floral quartet model.

Introduction

Current understanding of floral patterning has emerged primarily from studies in the core eudicot model plants *Arabidopsis thaliana* and *Antirrhinum majus*. In these species, the genetic ABCE model predicts how combinatorial expression of four classes of transcription factors specifies organ identity in the floral meristem (1–4). According to the latest *Arabidopsis* model, which incorporates the role of the E class proteins, once flowering has initiated, A and E class proteins specify sepals; A, B, and E class proteins specify petals; B, C, and E class proteins specify stamens; and C and E class proteins specify carpels and terminate floral meristem development (2, 5, 6). The underlying biochemical mechanism for specifying organ identity has been described by the floral quartet model, which predicts that correct transcription of organ-specific genetic programs requires the formation of hetero-multimeric complexes between these four interacting classes of transcription factors (5, 7–9). Mutations affecting the class A, B, C, and E functions are homeotic, resulting in the replacement of one organ type by another. Loss of expression of the *Arabidopsis* C class gene *AGAMOUS (AG)* results in conversion of stamens and carpels to petals and sepals and indeterminacy of the floral meristem (10), leading to showy "double flowers" with excess petals.

Because all but one of the genes in the ABCE model belong to the same gene family, it has been hypothesized that the evolutionary success of the angiosperm clade depends, to a great extent, on the proliferation and coevolution of this MIKC type MADS-box family of transcription factors (11, 12). However, insufficient taxonomic breadth of functional studies involving organ-identity genes currently

hinders the generation of new hypotheses on how genetic-level changes may be responsible for angiosperm floral diversity.

There have been multiple duplication events within the C class lineage (13) followed by partial redundancy (14–16), potential or proven sub-functionalization (17–24), or switching of functional equivalence among duplicates in different species (25). Even when a single gene is present, alternative transcripts are a source of genetic variation (14). Nonetheless, and presumably related to its crucial role in reproductive organ development, mounting evidence points to high overall functional conservation of C class gene function across angiosperms (14, 26, 27), including basal eudicots (14, 15) and monocots (16), and possibly more broadly across seed plants (28–30).

In the Ranunculaceae, a duplication event preceding the diversification of the family gave rise to two C class genes with distinct expression patterns; this duplication is independent of the core eudicot duplication that gave rise to *Arabidopsis AG* and its *Antirrhinum* functional equivalent, *PLENA (PLE)* (13). In *Thalictrum*, *ThtAG1* is expressed in stamens and carpels throughout development, whereas *ThtAG2* has ovule-specific expression (20). The fact that recruitment to ovule function has occurred independently multiple times throughout the angiosperm phylogeny (26), combined with the divergent expression data (20), led us to suspect that *ThtAG1* is the C class floral homeotic gene. Here, we focus on the functional evolution of this gene and its interaction with the putative E class gene *ThtSEP3*.

Flowers with aberrant phenotypes, also known as "teratomorphs," have been described for at least the last 2,000 y, double flowers being the first documented (ref. 31 and references therein). Double-flowered cultivars have become popular garden plants because of the attractiveness imparted by the extra petals (e.g., roses, peonies, carnations, and camellias). Moreover, scientists have put similar natural deviations from normal development to good use to help elucidate the genetic basis of normal flower development (32, 33). Although morphological, developmental, and/or genetic aspects of double-

flower cultivars have been investigated (22, 34–44), no functional evidence for the underlying molecular mechanism of this familiar phenotype is available to date in a noncore eudicot.

The species *Thalictrum thalictroides* includes cultivars that exhibit homeotic floral phenotypes suggestive of defects in the canonical organ-identity genes of the ABCE model. Among these cultivars, we identified the *Thalictrum thalictroides* cultivar 'Double White' (also known as 'Snowball') as a candidate for loss of C class function, based on its double-flower phenotype. This cultivar is a sterile homeotic mutant with flowers consisting entirely of multiple white petaloid sepals (the genus *Thalictrum* is apetalous). It presumably occurred spontaneously in natural populations, where it was collected and clonally propagated by plant enthusiasts because of the attractive nature of its double flowers.

In this study, we set out to characterize functionally *ThtAG1*, a putative *AGAMOUS* ortholog from the ranunculid *Thalictrum thalictroides*, and to test whether a mutation at this locus is responsible for a double-flower variety. In a forward genetic approach, we gathered strong evidence for the 'Double White' cultivar being affected in the *AGAMOUS*-like gene *ThtAG1*. In the complementary reverse genetics approach, targeted silencing of *ThtAG1* caused a double-flowered phenocopy of the cultivar. Importantly, this constitutes a description of a full C class mutant, affecting both organ identity and determinacy, in a noncore ("basal") eudicot.

Our results provide strong evidence for high conservation of C class gene function and of the interaction of C and E class proteins between core eudicot model plants and a ranunculid, including comparable roles in reproductive organ identity and flower meristem determinacy. In addition, we identify the genetic and biochemical basis of an ornamental double-flower variety, suggesting that mutations in C class genes likely underlie other widespread double-flower varieties.

Materials and Methods

Plant Materials

T. thalictroides wild-type, 'Double White,' and 'Snowball' plants were purchased from nurseries and grown in the University of Washington greenhouses. Voucher specimens (WTU 376535, WTU 376544, and WTU-387674) were deposited in the University of Washington Herbaria.

Cloning and Sequencing of the *ThtAG1* Genomic Locus

The complete wild-type and 'Double White' *ThtAG1* locus was PCR amplified and sequenced from genomic DNA. Intron 1 was amplified using primers ThtAG1int-cloneF7 and ThtAG1int-cloneR7, cloned, and sequenced using M13 and internal primers. The remainder of the wild-type gene was amplified using primers ThtAG1int-cloneF6 and ThtAG1int-cloneR4, cloned, and sequenced using M13F and internal primers. The remainder of the 'Double White' gene was PCR amplified and direct sequenced using the primer combinations ThtAG1 Exon 2 forward and ThtAG1 Exon 3 reverse; ThtDWAG1 Intron 2 forward and ThtDWAG1 Intron 4 reverse; and ThtAG1 Exon 4 forward and ThtAG1 coding stop reverse (Table S1). Sequence fragments were assembled into a contig in Sequencher (v. 4.9), and consensus sequences by plurality were deposited in GenBank (accession nos. JQ002519 and JQ002520).

Cloning of 'Double White' *ThtAG1* Transcripts from cDNA

The coding sequence for the *ThtAG1* locus (GenBank accession no. AY867878) was completed on the 5′ end by PCR on two accessions of wild-type cDNA using a forward primer on the 5′UTR of the related species *Thalictrum dioicum*: TdAG-1–5′ UTR-F2: 5′-TGATCATTCCCCCCAAAGA 3′ and the reverse primer TdAG-1-RWhole Gene: 5′TAACAAAGTCCAGTTTGAAGGCA 3′. The same primers

were used on two accessions of 'Double White' to obtain complete coding regions. PCR products were cloned into pCRII using the TA cloning kit (Invitrogen), following the manufacturer's instructions. Three to six clones from each of the two accessions were sequenced at the University of Washington High-throughput Genomics Unit (Seattle). Sequences were aligned in Sequencher (v. 4.9), and consensus sequences by plurality were deposited in GenBank (accession nos. JN887118– JN887121). The expressed allelic forms of *ThtAG1* were amplified from cDNA of two wild-type and three 'Double White' individuals using primers specific to each of two wild-type and two mutant transcripts (Table S2).

VIGS of *ThtAG1*

A TRV2-*ThtAG1* construct was prepared using a 421-bp fragment of the cDNA (AY867878) (20), half of which comprised the end of the coding region and the other half most of the 3′UTR. The fragment was amplified by PCR from a representative clone using primers with added XbaI and BamHI restriction sites, was ligated to TRV2, and was transformed into *Agrobacterium tumefaciens* strain GV3101.

VIGS was performed on 60 *T. thalictroides* tubers (bare-root plants) that had been kept at 4°C for 8 wk, as described previously (45). Tubers were wounded lightly near the bud using a clean razor blade, were placed in infiltration medium so that all parts were submerged, and were infiltrated under full vacuum (−100 kPa) for 10 min. Five untreated plants were potted and grown alongside treated plants. Ten mock-treated control plants were infiltrated identically, with empty TRV2 vector. Plants were placed in a Conviron growth chamber at 21 °C and 16 h light and were monitored daily for the following 2 mo. Homeotic floral phenotypes were first observed 2.5 wk after infiltration. When flowers

appeared to show an abnormal phenotype (shortly after anthesis), they were photographed, collected, and frozen for molecular validation or were fixed in formaldehyde/acetic acid/ alcohol (FAA) for SEM.

Morphological Characterization of Floral Phenotypes

Flowers of *T. thalictroides* 'Double White,' silenced plants, and controls were photographed using a Nikon SMZ800 dissecting microscope equipped with a QImaging MicroPublisher 3.3 RTV digital camera.

In preparation for SEM, floral tissue was fixed overnight or longer in FAA, dehydrated for 30 min through an alcohol series, critical-point dried, mounted, and sputter coated. Tissue preparation and observations were conducted in a JEOLJSM-840A scanning electron microscope at the University of Washington microscopy facility. Images were processed minimally in Adobe Photoshop CS5 v.11.0 and assembled using Adobe Illustrator CS5 v14.0.0.

Flower Organ Counts

Floral organs were counted once flowers had reached maturity, defined as having fully developed anthers before dehiscence in wildtype flowers or expanded outer sepals in fully silenced flowers that lacked stamens. Organs were identified as sepals, stamens, carpels, or homeotic organs. A flower was considered to possess a secondary (nested) flower if a pedicel was observed separating the innermost whorls; organs within these secondary flowers were not included in counts. One-way ANOVA was performed to test the statistical significance of differences in the total number of flower organs among silenced TRV2-*ThtAG1* (n = 14), wild-type (n = 21), and 'Double White' (n = 21) flowers using JMP v. 7 statistical software (SAS Institute, Inc).

Molecular Validation of VIGS.

Total RNA was prepared from 50–100 mg of frozen floral tissue using TRIzol (Invitrogen),

following the manufacturer's instructions. One microgram of the resulting total RNA was treated with

amplification-grade DNase I (Invitrogen) to eliminate potential genomic contamination and was reverse-

transcribed to cDNA using the SuperScript III first-strand synthesis kit (Invitrogen) with Oligo(dT)$_{20}$ or

specific primers to pTRV1, OYL 198 (5′-GTAAAATCATTGATAACAACACAGACAAAC 3′) (70), or

pTRV2, PYL156R (5′-GGACCGTAGTTTAATGTCTTCGGG-3′) (71).

To test for the presence of the TRV1 and TRV2 viral RNA in treated plants, RT-PCR was

carried out on 1 µL of cDNA using the TRV1-specific primers OYL195 (5′-

CTTGAAGAAGAAGACTTTCGAAGTCTC 3′) (70) and OYL198 and the TRV2-specific primers

pYL156F (5′-TTACTCAAGGAAGCACGATGAGC 3′) (71) and pYL156R for 30 cycles at 51°C

annealing temperature.

Quantification of *ThtAG1* expression in the different treatments was performed using real-time

PCR, as previously described (47). Briefly, each 30-µL reaction contained 15 µL of SYBR Green PCR

Master Mix (Bio-Rad), 0.9 µL (10 µM) of the gene-specific primers, 1 µL of template cDNA, and 12.2

µL of water. Locus-specific primers used for qPCR were ThtAG1 forward 5′-

AGTCTCTCAGCAATCTCAATATCAGG-3′ and ThtAG1 reverse 5′-

GCCCTGAGATACTTGTTATCAGTCTGC 3′.

Samples were amplified by 40 cycles in triplicate, including a no-template control, under the

following conditions: 94°C for 10 min, followed by 45 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C

for 30 s on the MJ Research Chromo4 PCR at the University of Washington Comparative Genome

Center. Melting curve analysis of the primers (50–95°C in 0.5°C increments of 1 s each) yielded a single

peak. Reactions were normalized to two housekeeping genes, *ACTIN* and *ELONGATION FACTOR 1*

(EEF1), using the ΔΔCT relative quantification method (72). Average values and SEs were graphed and compared statistically by one-way ANOVA with Tukey contrasts using JMP v. 7.

Plasmid Construction for Yeast Two-Hybrid Assays.

cDNA sequences encoding *T. thalictroides* MADS-domain proteins were cloned into plasmids pGADT7 and pGBKT7 for yeast two-hybrid analyses using EcoRI and BamHI recognition sites. Exceptions are ThtAG1, ThtAG1$_{\Delta 3 \Delta 13}$, and ThtSEP3, which were cloned into pGBKT7 using NcoI and SalI recognition sites. ThtSEP3 was cloned into pGADT7 using EcoRI and SacI recognition sites. Primers used for the PCR cloning procedure are listed in Table S3. Except for ThtSEP3ΔC, which lacked the last 195 bp of the coding sequence of *ThtSEP3*, all sequences cloned spanned the region from the MADS domain to at least the stop codon of the respective cDNAs. The region encoding the N-terminal extension present in some *AG*-like genes was not included.

Yeast Two-Hybrid Assays.

Yeast two-hybrid assays were carried out as described previously (73) with the following modifications: Cells were dissolved in 100 µL water and were serially diluted 10-fold to up to 1:10,000 in water before 2 µL of the undiluted suspension and of each dilution was plated on Leu/Trp/His-free medium supplemented with 3mM 3-Amino-1,2,4-triazole (3-AT) to test for an interaction and on Leu/Trp-free plates to control for yeast growth. Interactions were scored after 6–9 d of incubation at 28°C.

Results

'Double White' Flower Morphology and Development Are Consistent with a Loss-of-Function Mutation in a C Class Gene.

To characterize *T. thalictroides* 'Double White,' we compared its morphology and development with wild type. *T. thalictroides* is apetalous, with typically 5–12 white or pink petaloid sepals enclosing 45–76 filamentous stamens and 3–11 free, uniovulate carpels (n = 21), the last two spirally arranged (Fig. 1 A, E, and I). *T. thalictroides* 'Double White' flowers are sterile, with 56–105 petaloid sepals (n = 12)—no stamens or carpels are formed—resulting in a double-flower phenotype (Fig. 1 B, F, and J). The outer sepals of 'Double White' are comparable to wild-type sepals, whereas internal sepals have a slender base of variable length (Fig. 1 E and F).

Scanning electron microscopy (SEM) of young floral buds shows marked differences between wild type and 'Double White' early in development. Wild-type floral meristems develop flattened sepal primordia on the outside, followed by multiple, cylindrical, spirally arranged stamen primordia, then carpel primordia (distinguished by a central depression) (Fig. 1 I). In 'Double White' all organ primordia are flattened and sepal-like (Fig. 1 J).

The development of perianth organs in lieu of stamens and carpels is reminiscent of *Arabidopsis* C class *ag* mutants (10), leading us to hypothesize that 'Double White' is mutated in a C class gene. To test this hypothesis, we (i) silenced the *AG* ortholog *ThtAG1* in wild-type *T. thalictroides* plants in an attempt to phenocopy the cultivar; (ii) sequenced the 'Double White' *ThtAG1* genomic locus and transcripts in search of mutations; and (iii) tested the biochemical interactions of mutant proteins as compared with interactions in wild type.

Virus-Induced Gene Silencing of ThtAG1 Phenocopies 'Double White' Flowers.

T. thalictroides wild-type plants were subjected to virus induced gene silencing (VIGS) to test whether the 'Double White' floral phenotype results from loss of function of the putative *AG* ortholog *ThtAG1* and, more generally, to test the degree of functional conservation of C class genes in a ranunculid.

Survival after agroinfiltration with TRV1 and TRV2 silencing vectors was 92% for plants treated with TRV2-ThtAG1, 90% for TRV2-empty controls, and 100% for untreated control plants. Flowering started 6 d after infiltration of the bare root plants. Phenotypes observed in TRV2-empty plants included asymmetric reduction in sepal size and occasional brown necrotic spots, both previously reported as viral effects for this species (45). Initially, 43% of TRV2-*ThtAG1*-treated plants showed overall stunted growth habit and lack of stamen filament elongation as compared with untreated controls. The early stunted growth phenotype also was observed in 86% of TRV2-empty-treated plants. This early phenotype did not affect subsequent growth; plants initially may have been responding to the vacuum infiltration treatment. The lack of stamen elongation, which was not observed in TRV2-empty plants, was interpreted as a late-stage function of *AG* consistent with that reported in *Arabidopsis* (46). This phenotype occurred shortly after infiltration when, in preformed buds, *ThtAG1* presumably would have been silenced late in stamen development.

Homeotic floral phenotypes indicative of gene silencing were seen in 53% of treated plants. Flowers showing complete homeotic conversion of stamens and carpels to sterile organs (Fig. 1 C and G) were first observed 4 wk after infiltration and continued for at least 12 wk post infiltration, resulting in 35 documented silenced flowers across 16 TRV2-*ThtAG1*–treated plants. In these flowers, outer organs resembled wild-type sepals, middle organs (in the position of stamens) had slender bases as in 'Double White,' and central organs (in the position of carpels) resembled small sepals (Fig. 1 G). Early development in strongly silenced flowers shows that all primordia initiate with the flattened shape

characteristic of perianth organs rather than the cylindrical shape characteristic of reproductive organs (Fig. 1 K). Prior to the appearance of the fully homeotic phenotypes described above, approximately 2 wk after infiltration, we began to observe partial homeotic phenotypes consisting of loss of carpel and stamen identity in a chimeric pattern within flowers (Fig. S1). Carpels remained open and lacked their characteristic single apical ovule, no differentiated stigmas with papillae were visible, and nearby carpels were closed and ended in the normal stigmas with white papillae (Fig. S1 A and B). Stamen filaments became flattened and lanceolate, resembling sepals, and anthers were missing or highly reduced (Fig. S1 C and D). This chimeric phenotype was observed in approximately half of TRV2-*ThtAG1*-treated plants. After 3 wk, 13% of experimental plants had flowers showing intermediate phenotypes, consisting of partial conversion of reproductive organs to sterile organs throughout the whole flower (Fig. 1 D and H). Sepaloid organs replacing stamens were white to light pink, with a long narrow stalk and a distal oval blade (Fig. 1 H, two middle organs). Sepaloid organs in the center of the flower, where carpels normally are found, were comparable in shape to outermost sepals, except that they were much smaller, green, and curved inward (Fig. 1 H, rightmost organ). SEM detail of the center of a strongly silenced mature flower reveals perianth-like organs through the center and a floral meristem still differentiating perianth primordia (Fig. 1 L). None of the phenotypes described above were observed on either TRV2-empty or untreated controls.

Close examination of *T. thalictroides* tubers by dissecting the bud located at the "crown" (at the stage used for agroinfiltration) revealed several floral primordia already present at various stages of development before treatment (Fig. S2). This observation presented a likely explanation for phenotypes becoming stronger over time after infiltration, because flowers emerging after treatment would be affected by the lack of *ThtAG1* expression from inception, whereas those present before infiltration would be affected at a later developmental stage.

Wild-type adaxial sepal epidermal cell morphology of *T. thalictroides* was variable, consisting of regular to interlocking cells (Fig. 1 M, Upper) that, in some cases, became asymmetrically papillate toward the base (Fig. 1 M, Lower) (20, 47). Stamen filaments had slender, elongated cells (Fig. 1 N), whereas carpels had more rounded cells and stomata (Fig. 1 O). 'Double White' sepals had an epidermal phenotype comparable to that of the more irregular types found in wild type (Fig. 1 P). Sepal-like interlocking cells also were found in the blade of organs occupying the position of stamens (Fig. 1 Q), whereas slender cells resembling those of stamen filaments were found in their stalks (Fig. 1 R). Sepal-like cells also were found in the organs occupying the position of carpels (Fig. 1 S). Epidermal morphology of outer sepals in silenced flowers (Fig. 1 T) was comparable to the more regular morphology found in wild-type sepals (Fig. 1 M, Lower). Homeotic organs that replaced stamens consisted of sepal-like cells in the blade (Fig. 1 U) and more elongated cells at the stalk (Fig. 1 V), resembling those found in comparable organs of 'Double White' (Fig. 1 R) and in wild-type stamen filaments (Fig. 1 N); organs replacing carpels had regular cells similar to those found in the sepaloid organs (Fig. 1 W). Flowers with weak phenotype (Fig. 1 D) had partially homeotic organs (Fig. 1 H, third from left) exhibiting patchy epidermis with mixed stamen (elongate cells) and sepal cell types (Fig. 1 X) and mixed types of elongated cells in their stalks (Fig. 1 Y). Partially fused central organs with intermediate carpel–sepal identity also were observed (Fig. 1 Z). In summary, the SEM results show that, in both 'Double White' and full double flowers resulting from *ThtAG1* silencing, cell morphology of homeotic organs matched that of the outer sepals (compare Fig. 1 P, Q, and S with Fig. 1 T, U, and W), except for the stalked bases of organs in place of stamens, which more closely resembled cells in stamen filaments of wild-type plants (compare Fig. 1 N, R, and V).

Most 'Double White' flowers had homeotic perianth-like organs through the center (Fig. 2 A), and approximately one third of the flowers dissected (n = 14) had a secondary pedicel within the center

of the floral receptacle, giving rise to a nested bud that reiterates the pattern of homeotic organs (Fig. 2 B) and eventually develops into a mature 'Double White' flower (Fig. 2 C). Similarly, although most strongly silenced flowers seemed to continue to produce homeotic organs toward the center of the flower (Fig. 2 D), beginning at ~6 wk post infiltration we observed a floral bud growing on a pedicel from the center of the flower meristem (Fig. 2 E). The organs of the secondary flowers also were entirely homeotic, with a whorl of larger outer petaloid sepals surrounding smaller inner perianth-like organs (Fig. 2 F). This phenotype was observed in 42% of flowers (n = 19) and 67% of plants surveyed (n = 9). In wild-type flowers, on the other hand, the floral meristem is consumed in the production of carpels, which are surrounded by stamens and sepals, never forming a nested flower (Fig. 2 G). Consistent with the indeterminacy that results in a secondary bud within a flower, developmental SEM of VIGS treated flowers shows an extreme case of over proliferation of the floral meristem, which likely would have led to more than one nested flower (Fig. 2 H). Organ counts (nested-flower organs excluded) followed by one-way ANOVA found no significant difference in the total number of organs (P = 0.69, F = 0.36) in wild-type, silenced, and 'Double White' flowers (Fig. 2 I), consistent with a one-to-one homeotic conversion of reproductive organs into perianth-like organs.

ThtAG1 Is Down-Regulated in Double Flowers Resulting from VIGS.

To test whether the floral phenotypes described above were, in fact, caused by VIGS, we proceeded first to detect TRV transcripts and then to test for down-regulation of *ThtAG1*. TRV1 and TRV2 viral transcripts were detected in a subset of treated silenced plants (n = 18) and mock-treated plants (empty TRV2, n = 4) but were absent from untreated controls (n = 4) (Fig. S3). To test whether flowers exhibiting homeotic phenotypes expressed lower levels of *ThtAG1* than controls and levels similar to 'Double White,' we used quantitative PCR (qPCR) to quantify the expression of *ThtAG1* in

the different treatments, untreated and empty TRV controls, and the cultivar, relative to housekeeping genes (Fig. 3).

As expected from a VIGS experiment, a range of phenotypes (described above) was observed. Flowers with very weak or patchy phenotypes (Fig. S1) were not included in the expression analyses. The remaining silenced flowers collected (n = 35) were divided into the following categories by phenotype level: category 1, partial homeotic conversions of stamens and carpels (e.g., Fig. 1 D); category 2, complete homeotic conversion of stamens and carpels to sepaloid organs (e.g., Fig. 1 C); and category 3, complete homeotic conversions and containing a secondary nested flower within (e.g., Fig. 2 E and F). To test whether the increased intensity of the phenotypes resulted from increased silencing (lower levels of target gene expression), we compared the relative expression of *ThtAG1* in phenotype category levels 1–3 by one-way ANOVA. The result was not significant (P = 0.78, F = 0.25; n = 35), suggesting the phenotype likely is affected by timing rather than by intensity of silencing (Fig. S4).

One-way ANOVA among the four experimental groups - untreated (n = 12), mock-treated (n = 9), TRV2-*ThtAG1*-treated (n = 35), and 'Double White' cultivar (n = 4) - yielded a highly significant effect of treatment (P < 0.0001, F = 36.55). Tukey contrasts resulted in a significant difference between silenced TRV2-*ThtAG1*-treated plants and controls (α = 0.05). The *ThtAG1* expression level in the 'Double White' cultivar did not differ significantly from that of silenced flowers and also was significantly lower than controls under Tukey contrasts (α = 0.05) (Fig. 3).

In conclusion, the relative expression level of *ThtAG1* was decreased by almost fivefold in double-flower phenotypes resulting from VIGS as compared with the average in controls. These results confirm that the observed phenotypes are caused by the down-regulation of the endogenous copy of *ThtAG1* as a result of VIGS (Fig. 3). 'Double White' expressed 2.5-fold less *ThtAG1* than controls, providing additional evidence supporting the hypothesis that its double-flower phenotype is caused by

reduced or lacking C class function. Although 'Double White' appeared to express more *ThtAG1* than TRV2-*ThtAG1*-treated double flowers, the difference was not statistically significant.

Retrotransposon Insertion Underlies the 'Double White' Mutation.

To investigate the molecular basis of the 'Double White' mutant, we compared sequences of wild-type and 'Double White' *ThtAG1* genomic loci. The wild-type gene structure consists of seven exons and six introns, with a total length of 7,394 bp. The 'Double White' gene is 9,608 bp, the extra length resulting primarily from an insertion in exon 4 (Fig. 4 A). Closer inspection of the 2,209 bp insert revealed that it is the long terminal repeat (LTR) of a retrotransposon, a "solo" LTR, flanked by a 5-bp target site duplication (CTCTC) and exhibiting the signature TG...CA terminal dinucleotides (48). The insertion provides an early stop codon 42 bp into exon 4, resulting in a nonsense mutation that likely causes nonsense-mediated decay of the resulting transcript, since we did not recover it from cDNA (49). Two cryptic splice acceptor sites (asterisks in Fig. 4 A) are present 7 and 22 bp after the LTR insertion.

Other differences between wild-type and mutant genomic loci include intronic indels: three in intron 1 (15-bp, 59-bp, and 7-bp long), one in intron 2 (17-bp long), and one in intron 5 (53-bp long). An additional 91 small indels (≤3 bp) and SNPs occur throughout the introns. When all genetic differences were considered, the major insertion in exon 4 was suspected to be the most likely cause of the double-flower phenotype.

mRNA Analysis Reveals Putative Mutant ThtAG1 Proteins in 'Double White.'

Because the 'Double White' cultivar does express a low level of *ThtAG1* (Fig. 3), we wanted to investigate whether the transcripts resulted from the use of the two cryptic splice acceptor sites identified at the genomic level (Fig. 4 A). Sequence comparison of cDNA of wild-type *T. thalictroides*,

two accessions of 'Double White,' and one of 'Snowball' revealed two types of shortened *ThtAG1* transcripts that translate into putative proteins lacking either 8 or 13 amino acids within the K domain (ThtAG1Δ_8 and ThtAG1$\Delta_3\Delta_{13}$) (Fig. 4 B). These transcripts are consistent with the use of the two cryptic splice-acceptor sites (asterisks in Fig. 4 A). The resulting deletion at the protein level encompasses the end of the K1 subdomain and the intervening region between K1 and K2 subdomains (Fig. 4 B) (50). In addition, an upstream FYQ sequence at the boundary of the intervening region (the I domain) and the K domain also found in *AG* from other plants such as *Aquilegia, Papaver, Eschscholzia*, and *Vitis* (AY464110, GU123602, DQ088996, and HM192806, respectively), is missing in one of the wild-type alleles (ThtAG1Δ_3), and occurs in combination with the missing 13-amino acids in one of the 'Double White' accessions (ThtAG1$\Delta_3\Delta_{13}$) (Fig. 4B). Five additional single-nucleotide changes in the coding region, representing synonymous substitutions, occur between wild type and 'Double White'/'Snowball.' Because the K domain has been implicated in protein–protein interactions (50, 51), the eight- and 13 amino acid deletion alleles in the K domain were considered likely candidates for the loss of *AG* function in 'Double White' (and 'Snowball') and were tested further by yeast two-hybrid analysis.

'Double White' Mutations Disable Interactions Between C and E Proteins That Are Necessary for Normal Function.

AGAMOUS proteins form homodimers and interact with SEPALLATA proteins to function in floral organ specification (5, 7, 52). To establish whether the lesion in the K domain of ThtAG1 proteins in the double flowered mutants has a consequence at the level of protein–protein interactions, a GAL4-based yeast two-hybrid system was used. Interactions among the two wild-type alleles ThtAG1Δ_3 and

ThtAG1, the mutant alleles ThtAG1$\Delta_3\Delta_{13}$ and ThtAG1Δ_8, and the SEPALLATA protein ThtSEP3 were assessed (Fig. 5).

ThtSEP3 exhibited autoactivation when fused to the DNA binding domain of GAL4. Thus, ThtSEP3ΔC, a C terminal–deleted version in which autoactivation was abolished, also was included in the analyses (Fig. S5). The wild-type proteins ThtAG1Δ_3 and ThtAG1 were equally capable of interacting with ThtSEP3 and with ThtSEP3ΔC, albeit only in one direction (i.e., when ThtSEP3ΔC was fused to the activation domain of GAL4) (Fig. S5), indicating that the loss of FYQ does not have a detrimental effect on this interaction. In contrast, ThtSEP3 or ThtSEP3ΔC did not interact or interacted only very weakly with either of the 'Double White' proteins with a mutated K domain (Fig. 5 and Fig. S5). This result suggests that the missing eight or 13 amino acids in the K domain of 'Double White' ThtAG1 are essential for mediating interaction with ThtSEP3. Importantly, it also provides much needed biochemical evidence in a noncore eudicot for the C–E interaction predicted by the floral quartet model, in line with findings of high conservation of protein–protein interactions in other basal eudicots (53).

Discussion

We present evidence for strong conservation of C class function across distantly related angiosperms, i.e., the model core eudicots *Arabidopsis thaliana* and *Antirrhinum majus* and the ranunculid *Thalictrum thalictroides*. Moreover, we present a molecular dissection of a double-flower cultivar, the oldest class of homeotic floral phenotype on record (31).

In a reverse genetics approach, we used VIGS to reduce endogenous levels of *ThtAG1* mRNA significantly in wild-type *T. thalictroides* (Fig. 3). Flowers experiencing silencing of *ThtAG1* phenocopied the 'Double White' cultivar, with one-to-one conversion of reproductive to sepal-like organs and reinitiation of the floral meristem (Figs. 1 and 2). In the complementary forward genetic

approach, we uncovered a solo LTR retrotransposon insertion underlying the double-flower cultivar. A parsimonious explanation of the mutant phenotype comprises a reduction of the amount of functional mRNA caused by nonsense-mediated decay resulting from the insertion. However, even if considerable amounts of a protein were made (i.e., through the use of the cryptic splice-acceptor sites), the results of our yeast two-hybrid assays indicate that the mutant protein would be compromised in its interaction capacities and hence its function (Fig. 5). Therefore, C class function in the double-flower variety is twice handicapped.

LTR retrotransposons have been implicated in inducing alternative splicing when inserted in an intron, disrupting the original splice-acceptor site (54). Although in 'Double White' the insertion is in an exon, and the original splice acceptor site is maintained (Fig. 4 A), the recovery of almost full-length transcripts from cDNA of 'Double White' (Fig. 4 B) implies that two cryptic splice-acceptor sites downstream of the inserted transposon are active. Although LTR retrotransposons are abundant in plant genomes (48), they rarely have been shown to affect phenotype significantly, a notable exception being the loss of red skin color during the evolution of white grape varieties (55). Therefore, our demonstration that an LTR retrotransposon is the cause of a flower homeotic mutation provides further evidence supporting the potential role of these elements in plant domestication and evolution.

Protein–protein interactions are essential for the function of MADS-domain transcription factors (5, 8, 56, 57). Importantly, the K domain is involved in conferring protein–protein interactions among MADS-domain proteins (12, 50, 58). Evidence for the functional importance of the K domain in AG comes from *Arabidopsis thaliana*, where the partial loss-of-function allele ag-4 possesses a lesion in the K domain (59). Moreover, the presence of a single additional amino acid in the K domain of the C class-related protein FARINELLI (FAR) from *Antirrhinum*, also resulting from a change in splicing, alters its interaction affinities and influences its ability to specify floral organ identity (60). Thus, a deletion in the

K domain of ThtAG1 may well explain the mutant phenotype observed in 'Double White' cultivars. This possibility is supported by the notion that the interaction between AG and SEP orthologs, considered essential for determination of reproductive organ identity (5, 8, 60), is impaired in ThtAG1 mutant proteins that lack part of the K domain. In addition, *AG* expression levels were significantly lower in 'Double White' than in wild-type plants (Fig. 3). In *Arabidopsis*, AG forms protein complexes with other transcription factors to act in a positive autoregulatory loop that maintains and amplifies *AG* expression (61). If the K-domain mutation affects the ability of the AG protein to form such complexes by inhibiting protein–protein interactions, this effect also would explain the low expression rates in 'Double White' flowers. However, an alternative explanation (which is not mutually exclusive with the former one) for the reduced *ThtAG1* levels in 'Double White' flowers is that the nonsense mutation generated by the insertion of a solo LTR (Fig. 4A) renders the truncated *ThtAG1* transcripts unstable so that they are targeted for degradation.

Homeotic mutants have been invaluable in increasing the understanding of the role of floral organ-identity genes, including the elucidation of the central ABC model of flower development (1, 3). Partial C class mutants have been described previously in the monocot grasses rice (24) and maize (21), and recently, the rice double-mutant *mads3 mads58* was shown to have the full C class loss-of-function phenotype, including loss of organ identity and loss of floral determinacy (16). Such mutations typically have been ignored as potential mechanisms of evolution because of the assumption that such a radical transformation of floral organs invariably would cause a decrease in fitness (62). For this reason, homeotic mutants occurring in the wild typically have been considered evolutionary dead ends, and such a fate would seem particularly likely for any C class mutant lacking reproductive organs that might arise spontaneously within a given population. 'Double White' originally was isolated by growers from the wild as a spontaneous mutant; the clonal growth habit of the species makes it possible for the mutant to

persist in a population despite being sterile. The existence of similar mutants has been documented in natural populations of other plant species; in one notable instance, a double-flowered *Vinca* mutant has persisted for more than 160 y (22).

Functional studies in the Papaveraceae have found that lack of C function in the fourth whorl results in petals (14, 15), as in the snapdragon *ple far* double mutant (23), rather than in sepals, as in *Arabidopsis* (10). In *Thalictrum* the perianth is unipartite; only one type of organ is present, historically described as sepals, based on evolutionary and developmental evidence (63). The sepals of *T. thalictroides*, as is typical of Ranunculaceae, are the attractive part of the flower: They are large, white or pink, and sometimes have conical-papillate cells (47). In our VIGS experiments, stamens and carpels were replaced by organs that resembled additional sepals, except that those in the position of stamens were stalked, which appears to be a staminoid characteristic based on epidermal morphology (compare Fig. 1 N and U). Because petals have been lost before the diversification of *Thalictrum* (64), the gene and protein networks required for petal identity presumably would no longer be functional. Two other double flowers of Ranunculaceae have been described in which carpels were interpreted as suppressed and stamens as replaced by petals (44). In 'Double White,' we find no evidence of carpel suppression or petals being formed; instead, the meristem produces sepaloid organs in lieu of stamens and carpels.

The pre-formation of floral meristems in the tubers of our perennial-plant study system (Fig. S2) provides an opportunity to dissect the timing of C class gene functions in *Thalictrum*. Two distinct roles have been attributed to *AGAMOUS* in core eudicots: reproductive organ identity and flower determinacy (10, 23, 65–67). Evidence from *Arabidopsis* indicates that *AG* has different timing and threshold levels for its identity vs. determinacy roles, with determinacy occurring earlier (stage 3) (68) and organ identity later (stage 6) (46). Although our experiments did not test directly for *AG* thresholds, they do provide insight into the timing of these two fundamental functions. The fact that flowers displaying phenotypes

of different severity had comparable levels of *ThtAG1* mRNA (Fig. S4) does suggest a timing effect, as opposed to a delay in the triggering of silencing. In *Thalictrum*, loss of reproductive organ differentiation was the first evident phenotype (Fig. S1), indicating a developmentally late function of *AG*, because the first flowers to emerge presumably were pre-formed at treatment and therefore were affected by silencing late in their development. In fact, lack of filament elongation was among the earliest phenotypes observed, consistent with the late role of *AG* in *Arabidopsis* stamen maturation (46). Floral meristem reinitiation, on the other hand, occurred later after the day of infiltration; these flowers would have emerged after silencing had initiated and therefore show the early effects of *ThtAG1*. This timing for the determinacy and identity functions therefore is similar to that reported for *Arabidopsis* and further confirms the conservation not only of overall function but also of the timing C class roles, despite the distinct floral traits present in these early- and late-diverging lineages of eudicot angiosperms.

Variable numbers of spirally arranged floral organs and free uniovulate carpels with ascidiate development are ancestral floral traits present in *Thalictrum* and are distinct from the fixed number of whorled organs and the syncarpous gynoecia with multiple ovules of core eudicots (69). These marked differences in flower morphology and development in the two lineages make the similarity in underlying genetic mechanisms of reproductive organ identity and flower meristem determinacy demonstrated in this study especially remarkable.

None of the myriad of double-flowered horticultural varieties has been dissected at a molecular level comparable to *agamous* in *Arabidopsis* (3) or *plena* in *Antirrhinum* (67). Here we provide a likely explanation for a case of double flower formation in the ranunculid *Thalictrum thalictroides*. More broadly, functional analysis of a C class gene showed strong conservation of the reproductive organ identity and floral-determinacy roles described for core eudicots (10, 23, 67) and provided vital

experimental evidence for the importance of the interaction between C and E class proteins in a noncore eudicot.

References

1. Coen ES, Meyerowitz EM (1991) The war of the whorls: Genetic interactions controlling flower development. Nature 353:31–37.
2. Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require SEPALLATA MADS-box genes. Nature 405:200–203.
3. Bowman JL, Smyth DR, Meyerowitz EM (1991) Genetic interactions among floral homeotic genes of Arabidopsis. Development 112(1):1–20.

4. Schwarz-SommerZ, Huijser P, NackenW, Saedler H,Sommer H (1990)Genetic control of flower development by homeotic genes in Antirrhinum majus. Science 250:931–936.

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

6. Ditta G, Pinyopich A, Robles P, Pelaz S,Yanofsky M F(2004)The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. Curr Biol 14:1935–1940.

7. Theissen G (2001) Development of floral organ identity: Stories from the MADS house. Curr Opin Plant Biol 4(1):75–85.

8. Theissen G, Saedler H (2001) Plant biology. Floral quartets. Nature 409:469–471.

9. Melzer R, Theissen G (2009) Reconstitution of 'floral quartets' in vitro involving class B and class E floral homeotic proteins. Nucleic Acids Res 37:2723–2736.

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

11. Kramer EM, Hall JC (2005) Evolutionary dynamics of genes controlling floral development. Curr Opin Plant Biol 8:13–18.

12. Kaufmann K, Melzer R, Theissen G (2005) MIKC type MADS-domain proteins: Structural modularity, protein interactions and network evolution in land plants.Gene347:183–198.

13. Kramer EM, Jaramillo MA, Di Stilio VS (2004) Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. Genetics 166:1011–1023.

14. Hands P, Vosnakis N, Betts D, Irish VF, Drea S (2011) Alternate transcripts of a floral developmental regulator have both distinct and redundant functions in opium poppy. Ann Bot (Lond) 107:1557–1566.

15. Yellina AL, et al. (2010) Floral homeotic C function genes repress specific B function genes in the carpel whorl of the basal eudicot California poppy (Eschscholzia californica). Evodevo 1:13.

16. Dreni L, et al. (2011) Functional analysis of all AGAMOUS subfamily members in rice reveals their roles in reproductive organ identity determination and meristem determinacy. The Plant Cell 23(8):2850–2863.

17. Lü S, Du X, Lu W, Chong K, Meng Z (2007) Two AGAMOUS-like MADS-box genes from Taihangia rupestris (Rosaceae) reveal independent trajectories in the evolution of class C and class D floral homeotic functions. Evol Dev 9:92–104.

18. Pan IL, McQuinn R, Giovannoni JJ, Irish VF (2010) Functional diversification of AGAMOUS lineage genes in regulating tomato flower and fruit development. J Exp Bot 61:1795–1806.

19. Wang SY, et al. (2011) Duplicated C class MADS-box genes reveal distinct roles in gynostemium development in Cymbidium ensifolium (Orchidaceae). Plant Cell Physiol 52:563–577.

20. Di Stilio VS, Kramer EM, Baum DA (2005) Floral MADS box genes and homeotic gender dimorphism in Thalictrum dioicum (Ranunculaceae) - a new model for the study of dioecy. Plant J 41:755–766.

21. Mena M, et al. (1996) Diversification of C function activity in maize flower development. Science 274:1537–1540.

22. Wang YQ, Melzer R, Theissen G (2011) A double-flowered variety of lesser periwinkle (Vinca minor fl. pl.) that has persisted in the wild for more than 160 years. Ann Bot (Lond) 107:1445–1452.

23. Davies B, et al. (1999) PLENA and FARINELLI: Redundancy and regulatory interactions between two Antirrhinum MADS-box factors controlling flower development. EMBO J 18:4023–4034.

24. Yamaguchi T, et al. (2006) Functional diversification of the two C class MADS box genes OSMADS3 and OSMADS58 in Oryza sativa. Plant Cell 18:15–28.

25. Causier B, et al. (2005) Evolution in action: Following function in duplicated floral homeotic genes. Curr Biol 15:1508–1512.

26. Zahn LM, et al. (2006) Conservation and divergence in the AGAMOUS subfamily of MADS-box genes: Evidence of independent sub- and neo-functionalization events. Evol Dev 8:30–45.

27. Ferrario S, Immink RG, Angenent GC (2004) Conservation and diversity in flower land. Curr Opin Plant Biol 7:84–91.

28. Zhang PY, Tan HTW, Pwee KH, Kumar PP (2004) Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. Plant J 37:566–577.

29. Tandre K, Svenson M, Svensson ME, Engström P (1998) Conservation of gene structure and activity in the regulation of reproductive organ development of conifers and angiosperms. Plant J 15:615–623.

30. Rutledge R, et al. (1998) Characterization of an AGAMOUS homologue from the conifer black spruce (Picea mariana) that produces floral homeotic conversions when expressed in Arabidopsis. Plant J 15:625–634.

31. Meyerowitz EM, Smyth DR, Bowman JL (1989) Abnormal flowers and pattern formation in floral development. Development 106(2):209–217.

32. Hintz M, et al. (2006) Catching a 'hopeful monster': Shepherd's purse (Capsella bursapastoris) as a model system to study the evolution of flower development. J Exp Bot 57:3531–3542.

33. Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401:157–161.

34. Akita Y, Horikawa Y, Kanno A (2008) Comparative analysis of floral MADS-box genes between wild-type and a putative homeotic mutant in lily. J Hortic Sci Biotechnol 83(4):453–461.

35. Innes RL, Remphrey WR, Lenz LM (1989) An analysis of the development of single and double flowers in Potentilla fruticosa. Can J Bot 67(4):1071–1079.

36. MacIntyre JP, Lacroix CR (1996) Comparative development of perianth and androecial primordia of the single flower and the homeotic double-flowered mutant in Hibiscus rosa-sinensis (Malvaceae). Can J Bot. 74(12):1871–1882.

37. Lehmann NL, Sattler R (1993) Homeosis in floral development of Sanguinaria canadensis and S. canadensis 'Multiplex' (Papaveraceae). Am J Bot 80(11):1323–1335.

38. Ma GY, et al. (2011) Analysis of the Petunia hybrida double flower transcriptome using suppression subtractive hybridization. Sci Hortic (Amsterdam) 127(3):398–404.

39. Akita Y, Nakada M, Kanno A (2011) Effect of the expression level of an AGAMOUS like gene on the petaloidy of stamens in the double-flowered lily, 'Elodie'. Sci Hortic (Amsterdam) 128(1):48–53.

40. Dubois A, et al. (2010) Tinkering with the C function: A molecular frame for the selection of double flowers in cultivated roses. PLoS ONE 5:e9288.

41. Lehmann NL, Sattler R (1994) Floral Development and homeosis in Actaea rubra (Ranunculaceae). Int J Plant Sci 155(6):658–671.

42. Scutt CP, Oliveira M, Gilmartin PM, Negrutiu I (1999) Morphological and molecular analysis of a double-flowered mutant of the dioecious plant white campion showing both meristic and homeotic effects. Dev Genet 25:267–279.

43. Scovel G, Ben-Meir H, Ovadis M, Itzhaki H, Vainstein A (1998) RAPD and RFLP markers tightly linked to the locus controlling carnation (Dianthus caryophyllus) flower type. Theor Appl Genet 96(1):117–122.

44. Smith GH (1928) Vascular Anatomy of Ranalian Flowers. II. Ranunculaceae (Continued), Menispermaceae, Calycanthaceae, Annonaceae. Bot Gaz 85(2):152–177.

45. Di Stilio V, et al. (2010) Virus-induced gene silencing as a tool for comparative functional studies in Thalictrum. PloS ONE 5(8):e12064.

46. Ito T, Ng KH, Lim TS, Yu H, Meyerowitz EM (2007) The homeotic protein AGAMOUS controls late stamen development by regulating a jasmonate biosynthetic gene in Arabidopsis. Plant Cell 19:3516–3529.

47. Di Stilio VS, Martin C, Schulfer AF, Connelly CF (2009) An ortholog of MIXTA like2 controls epidermal cell shape in flowers of Thalictrum. New Phytol 183:718–728.

48. Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 33:479–532.

49. Chang Y-F, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76:51–74.

50. Yang YZ, Jack T (2004) Defining subdomains of the K domain important for protein protein interactions of plant MADS proteins. Plant Mol Biol 55:45–59.

51. Fan HY, Hu Y, Tudor M, Ma H (1997) Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. Plant J 12:999–1010.

52. Melzer R, Verelst W, Theissen G (2009) The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro. Nucleic Acids Res 37:144–157.

53. Liu C, et al. (2010) Interactions among proteins of floral MADS-box genes in basal eudicots: Implications for evolution of the regulatory network for flower development. Mol Biol Evol 27:1598–1611.

54. Varagona MJ, Purugganan M, Wessler SR (1992) Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. Plant Cell 4:811–820.

55. Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. Science 304:982–982.

56. de Folter S, et al. (2005) Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. Plant Cell 17:1424–1433.

57. Immink RGH, Kaufmann K, Angenent GC (2010) The 'ABC' of MADS domain protein behaviour and interactions. Semin Cell Dev Biol 21:87–93.

58. Yang YZ, Fanning L, Jack T (2003) The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. Plant J 33: 47–59.

59. Sieburth LE, Running MP, Meyerowitz EM (1995) Genetic separation of third and fourth whorl functions of AGAMOUS. Plant Cell 7:1249–1258.

60. Airoldi CA, Bergonzi S, Davies B (2010) Single amino acid change alters the ability to specify male or female organ identity. Proc Natl Acad Sci USA 107:18898–18902.

61. Gómez-Mena C, de Folter S, Costa MMR, Angenent GC, Sablowski R (2005) Transcriptional program controlled by the floral homeotic gene AGAMOUS during early organogenesis. Development 132:429–438.

62. Theissen G (2006) The proper place of hopeful monsters in evolutionary biology. Theory Biosci 124:349–369.

63. Tamura M (1965) Morphology, ecology, and phylogeny of the Ranunculaceae IV. Scientific Reports of Osaka University 14(1):53–71.

64. Wang W, Chen Z-D (2007) Generic level phylogeny of Thalictroideae (Ranunculaceae) implications for the taxonomic status of Paropyrum and petal evolution. Taxon 56(3): 811–821.

65. Mizukami Y, Ma H (1997) Determination of Arabidopsis floral meristem identity by AGAMOUS. Plant Cell 9:393–408.

66. Mizukami Y, Ma H (1995) Separation of AG function in floral meristem determinacy from that in reproductive organ identity by expressing antisense AG RNA. Plant Mol Biol 28:767–784.

67. Bradley D, Carpenter R, Sommer H, Hartley N, Coen E (1993) Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the plena locus of Antirrhinum. Cell 72:85–95.

68. Sun B, Xu YF, Ng KH, Ito T (2009) A timing mechanism for stem cell maintenance and differentiation in the Arabidopsis floral meristem. Genes Dev 23:1791–1804.

69. Endress PK, Doyle JA (2009) Reconstructing the ancestral angiosperm flower and its initial specializations. Am J Bot 96:22–66.

70. Hileman LC, Drea S, Martino G, Litt A, Irish VF (2005) Virus-induced gene silencing is an effective tool for assaying gene function in the basal eudicot species Papaver somniferum (opium poppy). Plant J 44:334–341.

71. Gould B, Kramer EM (2007) Virus-induced gene silencing as a tool for functional analyses in the emerging model plant Aquilegia (columbine, Ranunculaceae). Plant Methods 3:6.

72. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using realtime quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402–408.

73. Wang YQ, Melzer R, Theissen G (2010) Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm Gnetum gnemon provide a clue to the evolutionary origin of 'floral quartets'. Plant J 64:177–190.

Figure Captions

Figure 1. Silencing of the *T. thalictroides AGAMOUS* ortholog *ThtAG1* results in homeotic conversions of floral organs that phenocopy the 'Double White' cultivar. Morphology and development of *T. thalictroides* wild-type, 'Double White' and TRV2-*ThtAG1* flowers. (A) Wild-type flower with petaloid sepals (se), stamens (st), and carpels (ca). (B) 'Double White' flower with petaloid sepals throughout. (C) Fully silenced TRV2-*ThtAG1* flower showing complete homeotic conversion of reproductive organs to sterile organs. (D) Partially silenced TRV2-*ThtAG1* flower showing incomplete homeotic conversion of reproductive organs. (E–H) Dissected organs, outer to inner (Left to Right). (E) Wild-type sepal, stamen, and carpel. (F) 'Double White' outer, middle, and inner sepaloid organs. (G) Fully silenced TRV2-*ThtAG1* outer, middle, and inner sepaloid organs. (H) Partially silenced TRV2-*ThtAG1* outer sepal, middle sepaloid stamens, and inner sepaloid carpel. (I–L) SEM of developing flowers. (I) Wild-type bud with sepal, stamen, and carpel primordia. (J) 'Double White' bud with sepal primordia. (K) TRV2-*ThtAG1* bud with sepal primordia. (L) TRV2-*ThtAG1* open flower with homeotic organs. (M–Y) Adaxial epidermis cellular types by SEM. (M– O) Wild-type sepal (M), stamen filament (N), and carpel (O). (P–S) 'Double White' outer sepal (P), blade of middle homeotic organ (Q), stalk of middle homeotic organ (R), and ventral organ (S). (T–W) TRV2-*ThtAG1* outer sepal (T), blade of middle homeotic organ (U), stalk of middle homeotic organ (V), and central homeotic organ (W). (X) Mixed cell types in partially homeotic middle organ depicted in H. (Y) Mixed cell types in stalk of middle organ in H. (Z) Central partially homeotic organ. (Scale bars: 1 cm in A–D; 1 mm in E–H; 100 μm in I–L and Z; and 10 μm in M–Y.)

Figure 2. Floral meristem reinitiation and one-to-one homeotic conversion of organs in flowers subjected to VIGS of *ThtAG1* compared with that of 'Double White' flowers. Longitudinal sections of 'Double White' (A–C) and silenced (D–F) flowers of *T. thalictroides* reveal floral meristem determination defects as compared with wild type (G). (A) Multiple sepaloid organs through the flower center in 'Double White.' (B) Reinitiation of the floral meristem. (C) Fully developed nested 'Double White' flower. (D) Silenced flower with sterile homeotic organs in lieu of stamens and carpels. (E) Reinitiation of the floral meristem in silenced flower leading to another floral bud that reiterates the pattern of homeotic organs. (F) Fully developed silenced nested flower. (G) Wild-type flower with sepals, stamens, and central carpels. (H) SEM detail of extreme over proliferation of the floral meristem in a TRV2-*ThtAG1*-treated flower. (I) Total number of floral organs (average ± SE) in wild-type (n = 21), 'Double White' (n = 21), and silenced pTRV2-*ThtAG1* (n = 14) flowers were not significantly different in a one-way ANOVA ($P = 0.69$, $F = 0.36$). (Scale bars: 1 mm in A–G; 100 μm in H.)

Figure 3. Molecular validation of VIGS shows down-regulation of *ThtAG1* in homeotic flowers of *T. thalictroides* compared with controls. Relative levels of *ThtAG1* mRNA by qPCR in untreated (n = 12), TRV2-empty (n = 9), TRV2-*ThtAG1* silenced (n = 35), and 'Double White' flowers (n = 4), normalized with the housekeeping genes *ThtACTIN* and *ThtEEF1*. Mean and SE are shown; different letters indicate a significant difference ($P < 0.0001$, $F = 36.55$) in a one-way ANOVA with Tukey contrasts ($\alpha = 0.05$). The expression in TRV2-*ThtAG1* silenced and 'Double White' samples (B) is significantly lower than the expression in untreated and TRV2-empty samples (A). Silenced and 'Double White' expression levels were not significantly different from each other, nor were the two controls (untreated and TRV2-empty) ($P > 0.05$).

Figure 4. Comparative genomic structure of *ThtAG1* reveals a retrotransposon insertion in the 'Double White' mutant that disrupts mRNA accumulation and protein function. (A) Genomic structure of the *ThtAG1* locus in wild type (Upper) and 'Double White' (Lower). Lines represent introns, black boxes represent exons, arrow represents donor site, and the white box represents an insertion. The K domain is roughly encoded by exons 3–5 shown here. A detail of the insertion shows a solo LTR of a retrotransposon coding for an early stop codon (TGA) and two cryptic splice-acceptor sites (asterisks). (B) Translated amino acid alignment of the K-domain region of *T. thalictroides* wild-type and 'Double White' *ThtAG1* recovered from cDNA. Note that two 'Double White' transcripts that result from the alternate acceptor sites in (A) translate into shorter proteins, lacking either eight or 13 amino acids. The three K subdomains are indicated (50), showing that the missing amino acids in 'Double White' affect the area near the end of the K1 subdomain and the region between K1 and K2 subdomains.

Figure 5. Mutant C class proteins of *T. thalictroides* are unable to interact with their putative E class partner proteins, whereas wild-type proteins interact normally. Interactions between different ThtAG1 proteins and ThtSEP3 were determined with the yeast two-hybrid system. Colony growth on selective Leu/Trp/His-free + 3 mM 3-AT medium is shown. Yeast cells were spotted in 10-fold serial dilution (from left to right) for each interaction tested. ThtSEP3 was expressed as fusion with the GAL4 activation domain (AD); ThtAG1 proteins were expressed as fusions with the GAL4 DNA binding domain (BD). For the complete set of interactions tested, see Fig. S5.

Supplemental Figure Captions

Figure S1. Incomplete silencing effects on TRV2-*ThtAG1*-treated flowers of *Thalictrum thalictroides*. (A) Weak carpel phenotype consisting of open carpels without stigmas in lower half of flower center. (B) Detail of flower center in A. (C) Partial homeotic conversion of stamens in right half of the flower; (D) Homeotic organs among stamens. (Scale bars: 1 mm.)

Fig. S2. Tubers with floral primordia. (A) Large cluster of radially arranged tubers (T) supporting one bud (B) on the proximal end, containing primordia protected by enclosing bracts. (B) Floral (FP) and leaf (LP) primordia in various stages of development with bracts removed. (C) Floral primordium showing differentiated carpel (CP), stamen (StP), and sepal (SeP) primordia, with leaf primordia (LP) enveloping the flower.

Fig. S3. Detection of TRV transcripts in virus-induced gene silencing (VIGS)-treated flowers showing homeotic phenotypes. RT-PCR with TRV1- and TRV2-specific primers on untreated (n = 4), mock-treated (n = 4), and VIGS-treated silenced plants (n = 18). The smaller band in TRV2-empty plants corresponds to the distance between primers spanning the MCS region in the absence of a target-gene insert; the difference between this band and the bigger band in treated plants corresponds to the length of the target-gene insert (421 bp). A reverse-transcription control confirms the lack of genomic DNA contamination. Size of PCR products, in base pairs (bp), is indicated on the right side.

Fig. S4. Relative levels of *ThtAG1* transcript by quantitative PCR in VIGS flowers with increasing intensity of homeotic phenotypes. TRV2-*ThtAG1*-treated flowers in Fig. 3 have been divided

according to the intensity of their homeotic phenotype. Mean and SE of phenotype level 1 (n = 9), level 2 (n = 20), and level 3 (n = 6) are shown, normalized with the housekeeping genes *ThtACTIN* and *ThtEEF1*. Average control line represents the mean expression of untreated and empty TRV2 plants (in Fig. 3). The different VIGS phenotypes express comparable amounts of *ThtAG1*. Differences were not significant in a one-way ANOVA; P = 0.78, F = 0.25. See main text for description of phenotype categories.

Fig. S5. Interactions among *Thalictrum* MADS-domain proteins as determined with the yeast two-hybrid system. Colony growth on selective -Leu/-Trp/-His + 3 mM 3-AT medium is shown. As in Fig. 5, yeast cells were spotted in 10-fold serial dilution (from left to right) for each interaction tested. Constructs that expressed proteins as fusions with the GAL4 activation domain (AD) are shown horizontally; constructs that express GAL4 binding-domain (BD) fusions are shown vertically. Δ, negative controls in which empty vectors that did not contain a MADS-box gene cDNA insert were used.

Table S1. Primers for amplification and sequencing of *ThtAG1* genomic locus from *T. thalictroides* wild type and 'Double White'

Table S2. Primers for detection of expressed alleles of *ThtAG1* in individual plants *of Thalictrum thalictroides* wild type, 'Double White,' and 'Snowball'

Table S3. Primers for cloning of *T. thalictroides* MADS-box gene cDNAs into yeast two-hybrid vectors
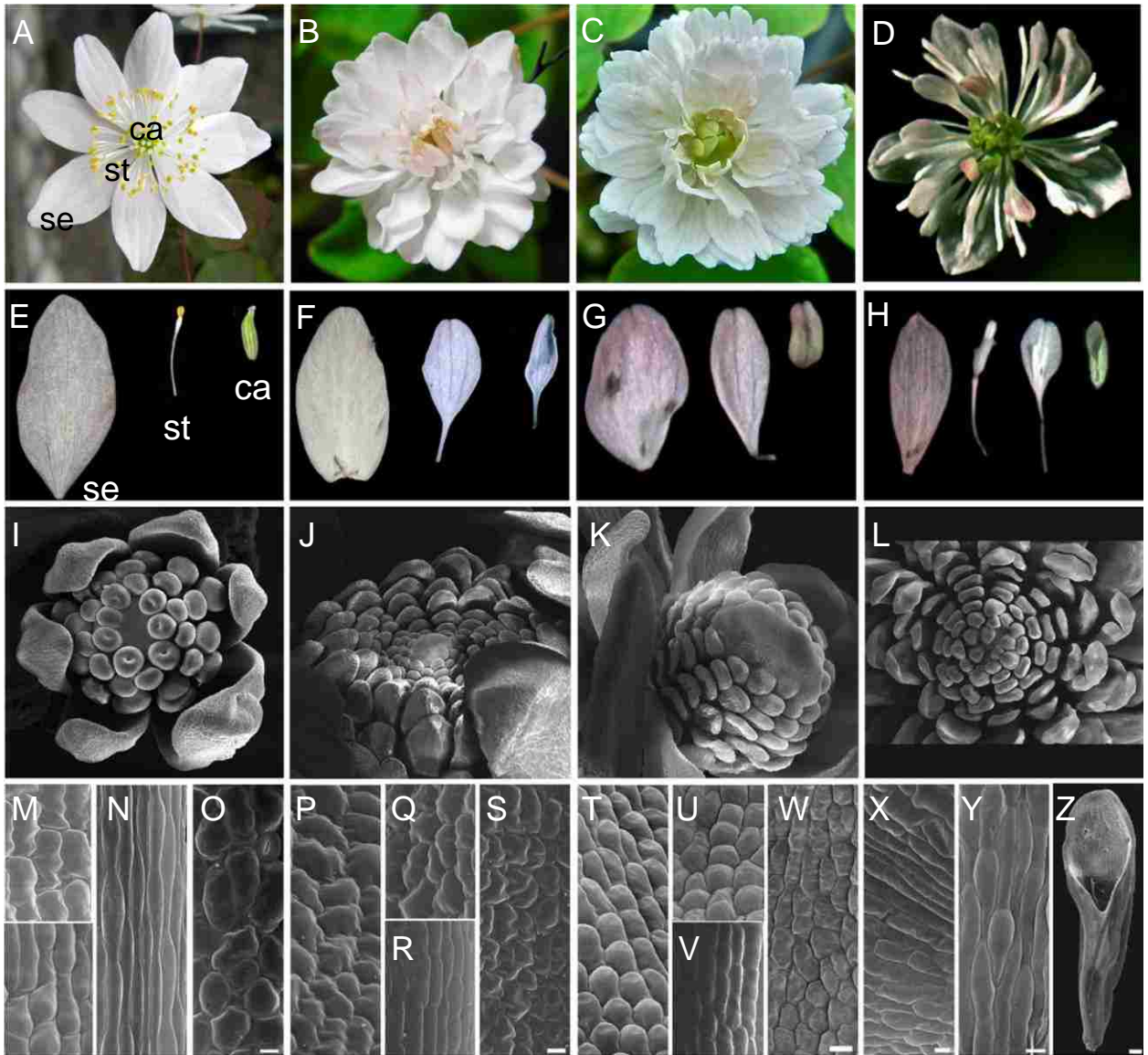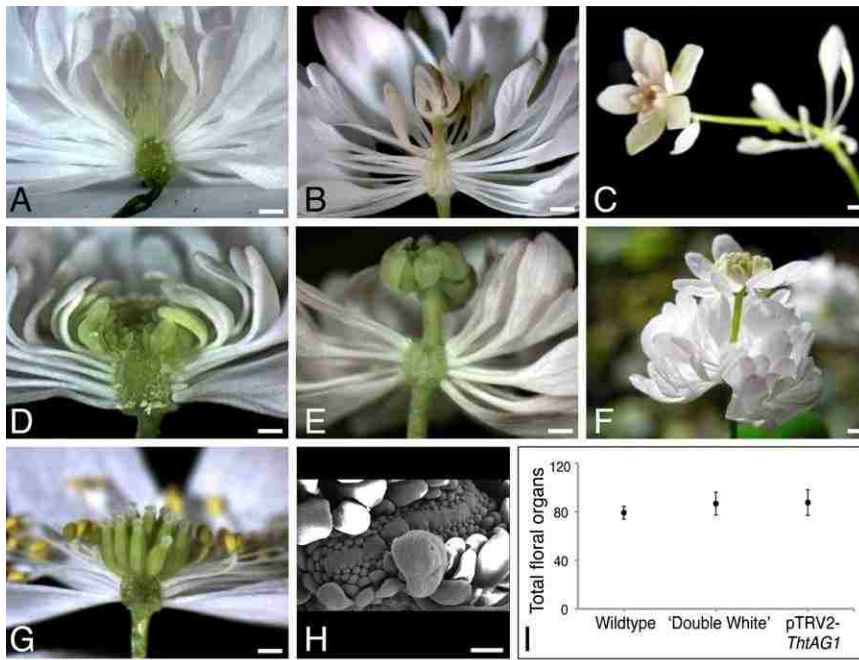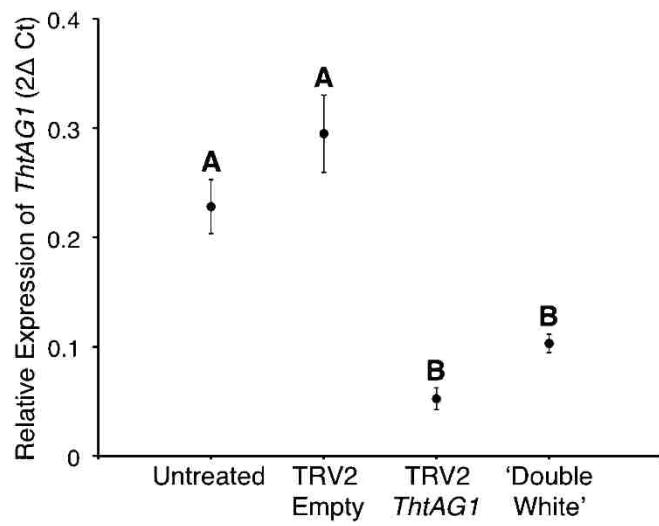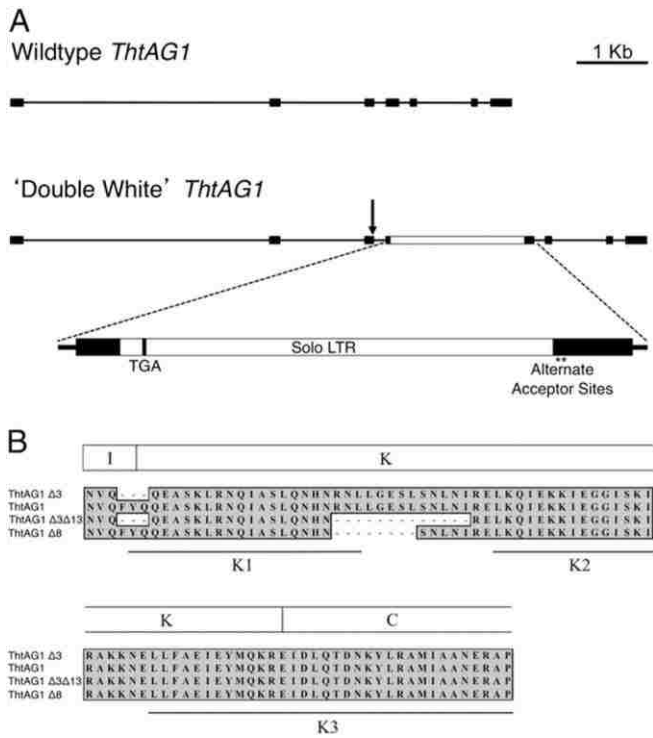
Figure 1

Figure 2

Figure 3

Relative Expression of *ThtAG1* (2Δ Ct)

Untreated TRV2 Empty TRV2 *ThtAG1* 'Double White'

Figure 4

Figure 5

BD \ AD	ThtSEP3
ThtAG1$_{\Delta 3}$	
ThtAG1	
ThtAG1$_{\Delta 3 \Delta 13}$	
ThtAG1$_{\Delta 8}$	

Figure S1

Figure S2

Figure S3

Figure S4

Figure S5

BD \ AD	ThtAG1$_{\Delta3}$	ThtAG1	ThtAG1$_{\Delta3\Delta13}$	ThtAG1$_{\Delta8}$	ThtSEP3	ThtSEP3ΔC	Δ
ThtAG1$_{\Delta3}$					●	●	
ThtAG1					●	●	
ThtAG1$_{\Delta3\Delta13}$							
ThtAG1$_{\Delta8}$							
ThtSEP3ΔC						●	
Δ							

BD-ThtSEP3/ AD-Δ ● ● ●

Table S1

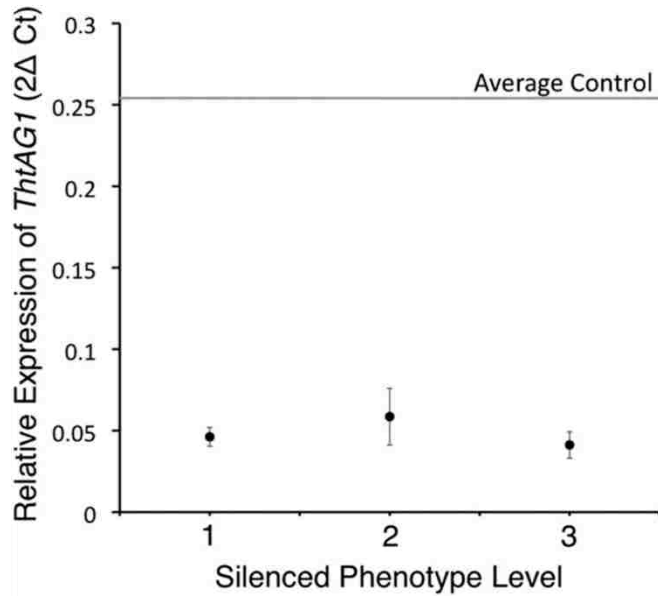| Primer Name	Primer Sequence
ThtAG1int-cloneF4	5′-GCAAACGAGAGAGCCCCAGAACAT-3′
ThtAG1int-cloneR4	5′-TATAAGATAGGGAGAGATTAAGACACGGCAC 3′
ThtAG1int-cloneF5	5′-CAGCAAGAAGCGTCAAAACTG-3′
ThtAG1int-cloneR5	5′- GCATATACTCTATCTCAGCAAACAGC 3′
ThtAG1int-cloneF6	5′-AACTCTGGATCTGTTTCTGAAGCTAATG-3′
ThtAG1int-cloneF7	5′-GTCTTCTCTACGCGTGGTCG-3′
ThtAG1int-cloneR7	5′-CATTAGCTTCAGAAACAGATCCAGAGTT-3′
ThtAG1int-cloneF8	5′-ATTTCTCTCTCTCTCTTGCTTTC 3′
ThtAG1int-cloneR8	5′-GCCATCTCTTGGTCTATTACTCC 3′
ThtAG1int-cloneF9	5′-ATGCTTCACATAGAACCTCG-3′
ThtAG1int-cloneR9	5′-TTGGTTGCTTCCTGATAGTC 3′
ThtAG1int-cloneF10	5′-CCAAAGGGTAAGAATCAAGACC 3′
ThtAG1int-cloneR10	5′-GCACATTATGCTCTTGGG-3′

Table S2

Primer Sequence	Description
5′-GCTAATGTGCAGTTTTACCAGC 3′	AG1_allele_F1: binding in the Δ3 region (FYQ), designed to amplify only when the FYQ sequence is present
5′-AAAGGGTGCAGAAGACATGAC 3′	AG1_allele_R1_all: binding downstream of both deletions, designed to amplify all alleles
5′-ATTCGGCTTCAAAGAAAAATG-3′	AG1_allelE F2_all: binding upstream of both deletions, designed to amplify all alleles
5′-GCTCCCTGATATTGAGATTGC 3′	AG1_allele_R2: binding in Δ13 region. Designed to amplify only when Δ13 deletion is absent
5′-AGACTCCCCCAACTCTGGAT-3′	AG1 allele test forward:: binding upstream of both deletions
5′-TCTTTTTGGCTCGGATTTTG-3′	AG1 allele test reverse: binding downstream of both deletions

Table S3

Primer Sequence	Description
5'-TAGCGCGAATTCATGGGAAGGGGAAAGATTGAAA 3'	Binding at the beginning of the MADS box of ThtAG1; used for cloning of ThtAG1; ThtAG1$_{\Delta3}$; ThtAG1$_{\Delta8}$; ThtAG1$_{\Delta3\Delta13}$.
5'-AATCGCGGATCCCAAAAGCTTTCACAGAATATCACC 3'	Binding at the end of the coding sequence and in the 3′ UTR of ThtAG1; used for cloning of ThtAG1; ThtAG1$_{\Delta3}$; ThtAG1$_{\Delta8}$; ThtAG1$_{\Delta3}\Delta_{13}$.
5'-AGTGTAGAATTCAGCCATGGGAAGGGGAAAGATTGAA 3'	Binding at the beginning of the MADS box of ThtAG1; used for cloning of ThtAG1; ThtAG1$_{\Delta3}\Delta_{13}$.
5'-ACTGCAGTCGACCACAGAATATCACCCAAGTTG-3'	Binding at the end of the coding sequence and in the 3′ UTR of ThtAG1; used for cloning of ThtAG1; ThtAG1$_{\Delta3}\Delta_{13}$.
5'-CCTAGTGAATTCATGGGAAGAGGAAGAGTTGAGT-3'	Binding at the beginning of the MADS box of ThtSEP3; used for cloning of ThtSEP3 and ThtSEP3ΔC.
5'-AGTCGAGAATTCAGCCATGGGAAGAGGAAGAGTTGAG-3'	Binding at the beginning of the MADS box of ThtSEP3; used for cloning of ThtSEP3.
5'-TAGCTTGAGCTCTTAAGGTACGAGAGACACCACT-3'	Binding at the 3′ UTR of ThtSEP3; used for cloning of ThtSEP3.
5'-CTATCAGTCGACCTTGCGGATCAGCTTCAGC 3'	Binding at the 3′ UTR of ThtSEP3; used for cloning of ThtSEP3.
5'-AAGCTTGGATCCGTCGACTTAATTTGGTTGGCTACCTTCCTC 3'	Used for cloning of ThtSEP3ΔC

Sub-functionalization to ovule development following duplication of a floral organ identity gene

Kelsey D. Galimba and Verónica S. Di Stilio

Abstract

Gene duplications result in paralogs that may be maintained due to the gain of novel functions (neo-functionalization) or the partitioning of ancestral function (sub-functionalization). Plant genomes are especially prone to duplication; paralogs are particularly widespread in the floral MADS box transcription factors that control organ identity through the ABC model of flower development. C class genes establish stamen and carpel identity and control floral meristem determinacy, and are largely conserved across the angiosperm phylogeny. An additional D class had been identified as controlling ovule identity; yet subsequent studies indicated that both C and D lineage genes more commonly control ovule development redundantly. The ranunculid *Thalictrum thalictroides* has two orthologs of the *Arabidopsis thaliana* C class gene *AGAMOUS (AG)*, *ThtAG1* and *ThtAG2 (Thalictrum thalictroides AGAMOUS1/2)*. We previously showed that *ThtAG1* exhibits typical C class function; here we examine the role of its paralog, *ThtAG2*. Our phylogenetic analysis shows that *ThtAG2* falls within the C lineage, together with *ThtAG1*, and is consistent with previous findings of a Ranunculales-specific duplication in this clade. However, *ThtAG2* is not expressed in stamens, but rather solely in carpels and ovules. This female specific expression pattern is consistent with D lineage genes, and with other C lineage genes known to be involved in ovule identity. Given the divergent expression of *ThtAG2*, we tested the hypothesis that it has acquired ovule identity function. Molecular evolution analysis showed evidence of positive selection on *ThtAG2* - a pattern that supports divergence of function by sub-functionalization. Down-regulation of *ThtAG2* by virus-induced gene silencing resulted in homeotic conversions of ovules

into carpel-like structures. Taken together, our results suggest that, although *ThtAG2* falls within the C

lineage, it has diverged to acquire "D function" as an ovule identity gene, and does not appear to require

a direct interaction with the ThtAG1 protein. We therefore present a functional example of ovule

identity being specified by either a single gene or a gene pair within the C lineage, with no D lineage

contribution. Following a Ranunculales-wide duplication in the *AG* lineage, functional divergence has

led to the evolution of ovule identity-specificity in a *T. thalictroides* C lineage gene.

Introduction

Gene duplication has long been recognized as an important contributor to the evolution of

biological complexity. While redundant duplicates are more likely to become pseudogenes, they can

sometimes partition the role of the original gene (sub-functionalization), or gain novel roles (neo-

functionalization) (1,2). In angiosperms, gene duplications are responsible for the expansion of the

MADS-box floral organ identity gene family, which contains instances of all three post-duplication fates

(3–5). Moreover, duplications in the MADS-box gene family appear to correlate closely with

angiosperm diversification (6–8). Evo-devo studies focused on the functional fate of duplicated MADS-

box genes in key taxa should therefore contribute to a better understanding of the evolution of flowering

plants (3).

The first model describing the role of organ identity genes in the patterning of the floral axis

emerged from studies in *Arabidopsis thaliana* and *Antirrhinum majus*. The original ABC model

predicted how the expression of specific combinations of three classes of transcription factors

determines floral organ identity. A class proteins were characterized as specifying sepals; A and B class,

petals; B and C class, stamens; and C class, carpels and floral meristem determinacy (9–11). Since then,

modifications of the ABC model include the addition of a D class for ovule identity (12), and an E class,

working in association with the ABC proteins to form the DNA binding tetramers necessary for functionality (13,14).

Whether genes belonging to the D clade (from here on referred to as D lineage) should be recognized as a separate class in the ABC(D)E model has been controversial, mostly due to taxon-specific differences. Both the clade comprised of *AG* and its orthologs (C lineage) and the clade containing *SEEDSTICK* (*STK)* and its orthologs (D lineage) belong to the larger *AGAMOUS* subfamily (15). The D class was originally described based on loss of function mutants of the *Petunia FLORAL BINDING PROTEIN* (*FBP*) genes, *FBP7* and *FBP11*. These mutants consisted of homeotic conversion of ovules into carpels, clearly suggesting a role in ovule identity (12). However, a more recent study revealed that *Petunia* ovule identity is actually redundantly shared among four genes within the *AGAMOUS* subfamily: the C lineage genes *PMADS3* and *FBP6*, in addition to the originally described D lineage genes *FBP7* and *FBP11* (16). A similar redundancy is observed in *A. thaliana,* where ovule identity is shared among C lineage *SHATTERPROOF (SHP) SHP1* and *SHP2,* and D lineage *STK* (17). In rice, where ovules develop directly from the floral meristem rather than from carpels, ovule identity is also redundantly controlled by the C lineage genes *OsMADS3* and *OsMADS58* in combination with the D lineage gene *OsMADS13* (18,19).

A duplication event in the C lineage gave rise to two sub clades that predate the diversification of the order Ranunculales (15), the earliest diverging group within the eudicots (20,21). This duplication is independent, and much more recent, than the one that gave rise to the separate C and D lineages. In the ranunculid species *T. thalictroides*, *ThtAG1* is expressed in stamens, carpels[1], and ovules throughout development, controlling stamen and carpel identity and floral meristem determinacy (23,24). In the related species *T. dioicum, ThdAG2* has ovule-specific expression (24), leading us to hypothesize that

---

[1] For the purpose of this article, carpels comprise the ovary wall, style and stigma, and are distinct from ovules (22)

this paralog might be responsible for ovule identity throughout the genus. Interestingly, no D lineage gene has been recovered in Ranunculaceae (15), although other families in the order do have genes that fall within this clade (8,25).

The goal of this study was to test whether a C class gene duplicate has sub- or neo-functionalized to become an ovule-identity gene in the ranunculid species *T. thalictroides*. We addressed this question by defining the nature of the duplication phylogenetically, investigating patterns of selection by molecular evolution, testing protein-protein interactions and characterizing gene expression and function. Our results illustrate a case of decoupling of genetic orthology from functional equivalence and represent a novel example of a C lineage gene taking up ovule identity function with no D lineage contribution.

Materials and Methods

Plant materials

T. thalictroides bare root plants were purchased from nurseries and grown in the University of Washington Greenhouses or in Conviron chambers at 21°C under long day conditions (100 μmol/m^2 for 16 hrs/day). A voucher specimen for this species had been previously deposited at the UW Herbarium (WTU 376542).

Cloning of *ThtAG2*

In order to isolate the complete *ThtAG2* coding sequence, we designed primers to the 5' and 3' untranslated regions (UTRs) in the related species *T. dioicum, ThdAG2* (KP006675). That full transcript was in turn generated by 5' Rapid Amplification of cDNA Ends (RACE) and alignment to a previously generated partial sequence (AY464094.1) (15) (Table S2).

Phylogenetic analysis

In order to understand the phylogenetic position and duplication history of the *Thalictrum* C

class genes, *ThtAG1* (JN887118.1) (23), *ThtAG2* (KM656465) and 21 additional sequences from non-

core eudicots (GenBank and 1KP Project) were aligned by hand to previously published nucleotide

alignments (8,15) using MacClade version 4.08 (26). Orthologs were obtained using BLAST search with

ThtAG1, ThtAG2 and *STK* (NM_001203767.1) (27) as the query sequences (Table S1).

Bayesian analyses on the nucleotide alignment were conducted in MrBayes using the

TIM2+I+G model of evolution as determined by Modeltest version 3.7 (28) in CIPRES Science

Gateway version 3.1 (29) with four chains of Markov Chain Monte Carlo and sampling one tree every

1000 generations for 100,000,000 generations. The first 25% of trees were discarded before convergence

(standard deviation = 0.01). In order to identify the degree of conservation of previously identified AG

motifs I and II (15), the C terminal regions from selected taxa were translated and hand-aligned using

MacClade version 4.08. The degree of amino acid residue conservation was highlighted using

BOXSHADE, version 3.21, setting the fraction of sequences that must agree for shading to 0.5.

Molecular evolution analyses

In order to detect the type of selection operating on the C class genes and to test the hypothesis

that *ThtAG2* is under positive selection, we used the codeml branch-site model in PAML-X (the

graphical user interface for PAML 4.7, (30). A subset of thirty-one C lineage genes was hand aligned in

MacClade (twenty-nine Ranunculales, one Sabiaceae and one Proteaceae), and a Bayesian analysis was

performed to produce the user-tree input for codeml. The TRN + G model of evolution was determined

using Modeltest version 3.7 (28). Bayesian analyses were conducted as explained above for 10,000,000

generations. The first 10% of trees were discarded before divergence (standard deviation = 0.01).

Positive selection was tested on the branch leading to *ThtAG2*, and additionally on the branches leading to the Ran*AG2* clade, *ThtAG1*, *ThdAG2* and *ThtAG2-ThdAG2*. Default settings with fixed values of κ=2, α=0 and ω=1, model =2 (2 or more dN/dS) and Nsites=2 were used for the null model, and identical settings, except an unfixed ω for the alternative model (30). To determine if the difference in likelihood scores among the two models was statistically significant, Likelihood Ratio Tests were performed and p values corresponding to the resulting ratios were corrected using the Holm-Bonferroni method (31). A Bayes Empirical Bayes (BEB) analysis was used to determine which residues of *ThtAG2* were under positive selection (ω>1). To determine which lineages in the phylogeny were evolving under positive selection, we used the Branch-Site REL method (Adaptive Evolution Server at Datamonkey.org) on the Ranunculales alignment in addition to the PAML analyses. This method identifies lineages and nodes at which a proportion of sites evolve with dN/dS > 1, without any assumptions (32).

Since the two species compared are phylogenetically distant (33) and their *AG2* orthologs are also rather divergent, we added the *AG2* ortholog from a third species from an intermediate phylogenetic position, *T. hernandezii (ThhAG2).* The sequence was obtained from a *T. hernandezii* transcriptome and added to the Ranunculales alignment, then analyzed using PAML as described above. Positive selection was tested on the branch leading to *ThtAG2-ThhAG2*, as well as on branches leading to *ThdAG2, ThtAG2-ThhAG2-ThdAG2* and the whole Ran*AG2* clade.

Expression analysis of C lineage genes in wild-type flowers

***In Situ* hybridization**

Young buds were dissected from the eyes of dormant *T. thalictroides* tubers. Buds were fixed in FAA (50% ethanol, 10% formalin, 5% acetic acid) with vacuum infiltration for 15 minutes and 4 hours of oscillation. Samples were taken through a dehydration series of 30 minutes each: 50%, 70%, 85%,

90% and 100% ethanol. After the 100% ethanol step, samples were placed in fresh 100% ethanol and left overnight at 4°C. Solution was exchanged for fresh 100% ethanol and incubated at room temperature for 1 hour, followed by subsequent room temperature incubations in 50% ethanol/50% CitriSolv for 2 hours, then 100% CitriSolv for 2 hours. They were then embedded in Paraplast Plus (VWR Scientific, CA) at 60°C, changing Paraplast three times each day for two days.

Gene-specific probes were designed to exclude regions of high similarity between *ThtAG1* and *ThtAG2*. A 439-bp region from *ThtAG1* and a 432-bp region from *ThtAG2* comprising approximately equal amounts of the end of the coding region (C terminus) and the 3'UTR were amplified from plasmids, using primers designed to contain either T3 (sense) or T7 (antisense) RNA polymerase promoters (Table S2). The transcription reaction was performed using DIG RNA Labeling Mix (Roche 11277073910), following manufacturer's protocol. The probe was hydrolyzed to 150bp and re-suspended in 50% deionized formamide.

In Situ hybridization was performed as described in (34) with the following modifications: 125 µg/ml of protease was used for enzyme digestion, 0.5 ng/µl/kb of probe was used for hybridization reactions and slides were hybridized for 13 hours at 42° C. Stringency washes were done at 52° C. Color reaction was stopped at 48 hours.

Quantitative PCR

Young open flowers of *T. thalictroides* were dissected into sepals, stamens, carpels and ovules. Three samples were processed for each organ type, each consisted of pooled organs from two flowers from one plant. Total RNA was extracted using TRIzol (Invitrogen, Life Technologies, CA), following manufacturer instructions. Total RNA was treated with amplification grade DNase (Invitrogen, Life

Technologies, CA). First strand cDNA was synthesized from 1 µg of total mRNA using Superscript III (Invitrogen, Life Technologies, CA) with Oligo-dT, following manufacturer instructions.

Locus specific primers were used to test expression levels of *ThtAG1* and *ThtAG2* by quantitative PCR (qPCR) (Table S2). Each reaction contained 15 µl iQ SYBR Green Supermix (Bio-Rad170-8882), 12.2 µl H_2O, 0.9 µl each of forward and reverse primer and 1 µl of cDNA. Samples were amplified on an MJ Research Chromo4 Detector for 45 cycles, in triplicate, including a no-template control. Cycling conditions were: 94ºC for 10 minutes and 45 cycles of 94ºC for 30 seconds, 54ºC for 30 seconds and 72ºC for 30 seconds. Relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method (35), and normalized against *ThtACTIN* and *ThteEF1α (Elongation Factor 1α)*.

Ontogeny of wild-type ovules

Wild-type carpels were dissected from flowers at different developmental stages, ranging from unopened buds to fully opened flowers with dehisced anthers. Ovules were photographed using auto fluorescence under confocal microscopy (see details below) to determine a baseline ontogeny for the characterization of phenotypes emerging from the targeted gene silencing experiments.

Virus-induced gene silencing (VIGS)

A TRV2-*TtPDS*-*ThtAG2* construct was prepared using a 429 bp fragment comprising the C terminal region (176 bp) and 3'UTR (253 bp) of *ThtAG2*. A 425 bp fragment of *T. thalictroides PHYTOENE DESATURASE* (*TtPDS*) (HM488111.1) was added to the construct as a reporter, as it causes photobleaching of photosynthetic tissue, facilitating screening of plants undergoing gene silencing (36). The *ThtAG2* fragment was amplified by PCR from a representative clone using primers with added XbaI and BamHI restriction sites (Table S2), ligated to an existing TRV2-*TtPDS* construct,

and transformed into *Agrobacterium tumefaciens* strain GV3101. Thirty-three *T. thalictroides* tubers (bare root plants) that had been kept in the dark at 4°C for 8 weeks, were treated with TRV2-*TtPDS-ThtAG2* as described previously (36). Briefly, tubers were nicked near the bud using a clean razor blade, submerged in infiltration medium containing appropriate *Agrobacterium* cultures, and infiltrated in a chamber under full vacuum (−100 kPa) for 10 min. Twenty five plants were infiltrated identically with empty TRV2 vector (mock-treated control, to detect background viral effects) and these, along with 20 untreated plants, were grown in identical conditions to treated plants. Plants were placed in a Conviron growth chamber at 21°C under long day conditions (100 μmol/m^2 for 16hrs/day). Flowers from all treatments were observed under a Nikon SMZ800 dissecting scope and photographed using a Q-Imaging MicroPublisher 3.3 digital camera. Eighteen flowers (from seven plants) were chosen for analysis, based on their strongly photobleached floral tissue. Nine flowers (from six plants) were frozen and processed for RNA, and nine flowers (from four plants, three of the plants also used for RNA) were processed for confocal microscopy. Eight flowers (from five mock-treated plants) and five flowers (from five untreated plants) were collected and processed as controls.

Molecular validation of VIGS lines

Total RNA was extracted from whole frozen flowers using TRIzol (Invitrogen, Life Technologies, CA), following manufacturer instructions. First strand cDNA was synthesized as explained above using Superscript III (Invitrogen, Life Technologies, CA) with primers OYL198 for TRV1, pYL156R for TRV2 (37) or Oligo-dT. Locus specific primers were used to test expression levels of *ThtAG1* and *ThtAG2* by Reverse-transcription PCR (RT-PCR) on 2.5 μl of cDNA, taking care to place one of the primers outside of the gene fragment present in the TRV2 construct (Table S2). RT-PCR was performed on 2.5 μl of cDNA using Platinum Taq DNA Polymerase (Invitrogen, Life

Technologies, CA), following manufacturer instructions. Cycling conditions were: 95ºC for 2 minutes and 30 cycles of 94ºC for 30 seconds, 52ºC for 1 minute and 72ºC for 2 minutes and a final extension at 72ºC for 5 minutes. Five microliters of PCR product was run on a 1.2% agarose gel and stained with ethidium bromide for visualization, then photographed using Photodyne Photography System, model 702. Gel band intensities were quantified using ImageJ 1.48 (NIH), normalizing against *ThtACTIN* controls. For treated samples, flowers with similarly high levels of photobleaching as those used for confocal microscopy (below) were used.

Scanning Electron Microscopy (SEM)

Floral tissue was fixed overnight in FAA, dehydrated for 30 min through an alcohol series, critical point dried, mounted, and sputter coated. Observations were conducted in a JEOLJSM-840A scanning electron microscope at the University of Washington microscopy facility.

Confocal microscopy

The single ovules contained in each carpel were dissected, fixed in FAA for at least 24 hrs, dehydrated through an ethanol series and cleared with methyl salicylate. Ovules were mounted on slides using Cytoseal Mounting Medium (Richard-Allan Scientific, MI) and scanned for auto fluorescence using a Leica TCS SP5 II Laser Scanning Confocal Microscope (LSCM) equipped with a DPSS561 Laser and a DD/488/561 filter. Images were captured at 0.2 to 0.4 μm optical sections using a 20×0.7 objective in LAS AF version 2.6 (Leica Microsystems Inc., IL). Analysis and processing of the images were done using LAS AF software and ImageJ 1.48 (38).

Histological Sectioning

In order to compare infertile carpels derived from gene silencing to wild-type carpels after hand pollinations, seed morphology was analyzed in paraffin sections. Mature fruits were collected and fixed in FAA, followed by dehydration through an ethanol series, then 50/50% ethanol CitriSolv, ending in CitriSolv (Fisher Scientific, PA). Fruits were embedded in Paraplast Plus at 60ºC, replacing the wax three times daily for two days. On the third day they were placed into Peel-A Way molds (Fisher Scientific, PA) and allowed to harden to room temperature. Fruits were sectioned using an American Optical Company microtome (8 µm) and mounted on glass slides (Fisher Scientific, PA). Wax was removed with CitriSolv and sections were rehydrated through an ethanol series. Slides were then washed and mounted in Phosphate buffered saline (PBS). Sections were observed under bright-field settings with a Nikon Microphot-FX microscope and photographed using a Q-Imaging MicroPublisher 3.3 digital camera.

Yeast two-hybrid assays

Plasmids pGADT7 and pGBKT7 containing cDNA sequences encoding *T. thalictroides* MADS-domain proteins had been previously generated and kindly provided for this study by Rainer Melzer and Günter Theißen (23).

cDNA sequences encoding proteins for *ThtAG1*, *ThtAG2* and *ThtSEP3ΔC* were cloned into pGADT7 using EcoRI and BamHI recognition sites; *ThtAG2* and *ThtSEP3ΔC* were cloned into pGBKT7 using EcoRI and BamHI recognition sites and *ThtAG1* was cloned into pGBKT7 using NcoI and SalI recognition sites (primers listed in Table S2). *ThtSEP3ΔC* is a C terminal-deleted version of *ThtSEP3* lacking the last 195 bp of the coding sequence, used previously to avoid auto activation (23,39). All other sequences cloned spanned the region from the beginning of the MADS domain to at least the stop codon of the respective cDNAs.

Yeast two-hybrid assays were carried out using Matchmaker Gold Yeast Two-Hybrid System (Clontech, CA). Cells were co-transformed and plated on Leu/Trp-free media to select for diploid colonies. Single colonies were serially diluted 10-fold to up to 1:10,000 in water before 5 μl of each dilution were plated on Leu/Trp/His-free media and Leu/Trp/His-free media supplemented with Aurebasidin A (AbA) to test for protein interaction, and on Leu/Trp-free media to control for yeast growth. Interactions were scored and photographed after 4 days of incubation at 28°C.

Results

***ThtAG2* belongs to the C lineage of flower organ identity genes**

To determine the evolutionary relationship between the two *T. thalictroides* C class genes and other members of the *AG* subfamily, we generated a phylogenetic analysis of *ThtAG1* and *ThtAG2* (Table S1) (Fig. 1). The Bayesian majority rule consensus tree showed that *ThtAG1* and *ThtAG2* fall within the C lineage, in two separate clades representing a Ranunculales-wide duplication event (Fig. 1, red arrows). No D lineage *T. thalictroides* genes were recovered from the 1KP transcriptome nor a draft genome (BGI). Our analysis confirmed that the duplication that led to the Ranunculaceae *AG1* (Ran*AG1)* and *AG2* (Ran*AG2)* orthologs occurred prior to the diversification of the order Ranunculales (15), which includes the families Ranunculaceae, Berberidaceae, Menispermaceae, Lardizabalaceae, Circaeasteraceae, Papaveraceae and Eupteleaceae (15,21). With an increased representation of non-core eudicot taxa, and based on currently available data, our analysis revealed a number of apparent gene losses following the duplication event in several families within the order. Orthologs corresponding to the Ran*AG2* clade are absent in taxa from Berberidaceae, Lardizabalaceae and Eupteleaceae, while orthologs corresponding to the Ran*AG1* clade are absent from Papaveraceae. Based on available data (no sequences are available for Menispermaceae and Circaeasteraceae), Ranunculaceae is the only

family to have retained both *AG1* and *AG2* in the Ranunculales. This hypothetical pattern of gene losses may change in the future, once more complete genomic coverage of additional genera in each family becomes available.

In our analysis, the rice gene *OsMADS13*, one of the few D lineage genes that has been tested functionally, falls within the C lineage whether or not we include the C terminus in the nucleotide alignment. This and other grass genes have historically been labile in placement between C and D lineages, and the C terminus region has been omitted from alignments completely or in part in different studies due to its divergent nature (8,15). A more recent examination shows that these genes are most likely orthologous to other D lineage genes based on the presence of a number of K-domain features (18). There are multiple factors that could be affecting the placement of *OsMADS13* in our alignment: the more prominent one being higher sampling coverage in the Ranunculales clade (the focus of this study) as compared to the grass clade.

To further characterize *ThtAG1* and *ThtAG2* we analyzed the C terminal region of their protein products, which is involved in protein-protein interactions (40). Compared to core eudicots, ThtAG1 exhibits conservation of the aspartic acid residue in the second amino acid of AG motif I (Fig. 2, dot), while ThtAG2 has a serine at this position; in core eudicot D lineage proteins this residue is highly variable (15). ThtAG2 lacks the highly conserved lysine residues in positions five and six of AG motif II (Fig. 2, dots) that have also been identified as a D lineage feature in core eudicots (15). Taken together, ThtAG2 exhibits features associated with both C and D lineages at the protein level; the biological relevance of these features will require a more in-depth functional assessment.

The enrichment in Ranunculales taxa allowed us to identify a single amino acid change specific to this order. A valine residue is highly conserved in the last position of the AG motif II in the core

eudicot C lineage; non-core eudicot sequences exhibit a conserved glycine and all members of the Ran*AG2* clade have a cysteine at this position (Fig. 2, dot).

ThtAG2 is under positive selection.

In order to determine patterns of selection in the Ranunculales *AG* phylogeny, we conducted molecular evolution analyses on two different trees, containing *AG*2 sequences from two or three species of *Thalictrum*. The Likelihood Ratio (LR) for positive vs. neutral models on the *ThtAG2* branch for the first gene phylogeny was statistically significant (8.401, p=0.015), which indicated that *ThtAG2* experienced positive selection. Five sites were identified under positive selection: one is located in the K-domain and four are in the C terminus. A lysine at position 211 is highly statistically significant (Table 1). There was no evidence of positive selection for the branches leading to *ThtAG1*, *ThdAG2* nor *ThtAG2-ThdAG2* (LR=0). Additionally, the Ran*AG2* clade as a whole is not under positive selection (LR= 0.202, p=1).

To complement the above approach, we used the Branch-Site REL method to determine whether any Ranunculales C lineage genes are under positive selection. The resulting phylogeny indicated two branches under significant positive selection: *ThtAG2* (p=0.02) and the *AG* ortholog from *Ranunculus scleratus*, *RascAG* (p=0.01) (Fig. S1). Taken together, these results suggested that out of all Ranunculales C lineage genes, positive selection occurs only in two genes within the RanAG2 clade: *ThtAG2* and *RascAG*.

The presence of a large deletion in the AG Motif I of the available ThdAG2 protein (Fig. 2) caused us to suspect this gene may be skewing our analysis (causing positive selection to falsely appear species-specific in the *ThAG2* clade). In order to have better representation for the genus, we added an ortholog from a third species, *T. hernandezii*. *ThhAG2* was extracted from a transcriptome (BGI), added

to the Ranunculales alignment and analyzed in PAML. The three species, *T. thalictroides*, *T. dioicum* and *T. hernandezii* are representatives of distinct clades in the species phylogeny (33,41). The coding sequence of *ThhAG2* (KR150759) is identical to *ThtAG2,* though differences were noted in the UTRs. The Likelihood Ratio (LR) for positive vs. neutral models on the *ThtAG2-ThhAG2* branch of the gene phylogeny was statistically significant (8.457, p=0.014), indicating that both *ThtAG2* and *ThhAG2* are under positive selection. The same five amino acid sites were identified under positive selection (Table 1). There was no evidence of positive selection for the branches leading to *ThdAG2* or *ThtAG2-ThhAG2-ThdAG2* (LR=0). Again, the Ran*AG2* clade as a whole is not under positive selection (LR= 0.238 p=1). The results from the molecular analysis on this second phylogeny suggest that *AG2* experienced positive selection in at least two distantly related species of *Thalictrum*, *T. thalictroides* and *T. hernandezii*, whose common ancestor dates back 18.4 my to the origin of the whole genus (41).

ThtAG2 is expressed in carpels and ovules.

To test the spatial and temporal expression patterns of *ThtAG1* and *ThtAG2*, we performed *in situ* hybridization on buds and qPCR on dissected organs of young open flowers (Fig. 3). In buds (Fig. 3 A), neither gene was detected in sepals or stamens (Fig. 3 B, E); both genes were expressed in the ovule and the placenta – the section of carpel wall where the ovule is attached (Fig. 3 C, F). In ovules that are developing integuments, expression appears excluded from the nucellus and concentrated in the integuments (Fig. 3 D, G).

In the qPCR analysis of dissected organs from young open flowers (Fig 3 H), neither gene could be detected in sepals (Fig. 3 I). *ThtAG1* was present in stamens while *ThtAG2* was not and both genes were expressed in carpels and ovules (Fig. 3 I). Ovules had higher expression of both genes than carpels: approximately eleven times as much *ThtAG1* and three times as much *ThtAG2*. Differences in

expression observed by the two methods are likely related to flower developmental stage and different sensitivity of each method to detecting low levels of mRNA. Because these plants are perennial spring ephemerals that form their flowers underground, the earliest stage at which flowers can be harvested is already passed the organ primordia stage when MADS box genes are typically expressed the most. This would explain why we did not detect expression of *ThtAG1* in stamens and carpels by *in situs*, as expected from the ABC model, instead detecting its later role in ovule development. It could also explain why we only detected *ThtAG2* expression in the integuments of the ovule. The fact that *ThtAG1* is detected by qPCR at a later developmental stage in stamens and carpels suggests that this assay was more sensitive to low levels of mRNA.

Ovule ontogeny in *T. thalictroides*

An ovule ontogeny was generated as a baseline comparison against ovules in flowers for which *ThtAG2* had been silenced. The *T. thalictroides* gynoecium is apocarpous (free carpels), consisting of 4-15 uniovulate carpels that develop into achenes (indehiscent, dry fruits) after fertilization (42). In young buds, small ovules (approximately 20 µm long) had two visible integuments surrounding the nucellus and embryo sac (the female gametophyte). A hypostase appeared as an auto fluorescing region below the embryo sac (Fig. 4 A, F) that became brighter as development progressed (Fig. 4 H-J). Hypostases consist of thick-walled cells and are likely involved in nutrient transfer to the developing seed (43). Two integuments continued to completely envelope the nucellus and embryo sac at all stages (Fig. 4 F-J). Just prior to anthesis (Fig. 4 A), the embryo sac was visibly cellularized (Fig. 4 F). In older flowers with partially elongated stamens (Fig. 4 C, H) and at the beginning of stamen dehiscence (Fig. 4 D, I), the embryo sac became enlarged and the central cell transitioned from two polar nuclei (Fig. 4 H) to a secondary nucleus (derived from the fusion of these two nuclei, Fig. 4 I). In mature flowers that were

shedding pollen (Fig. 4 E, J) embryo sacs seemed fully developed, having increased approximately nine

times in size (180 μm) and showing a Polygonum-type morphology (seven cells, eight nuclei) (44).

Individual cells could be identified at the micropylar end (interpreted as one synergid and one egg), in

the center (central cell with secondary nucleus) and at the chalazal end (interpreted as two antipodal

cells). Other members of Ranunculaceae have been reported to degenerate synergid and antipodal cells

as a normal part of development (45), so it is possible that synergid and/or antipodal cells that are not

visible in our preparations have actually degraded.

Down-regulation of *ThtAG2* results in homeotic conversion of ovules into carpel-like organs.

VIGS was used to down-regulate expression of *ThtAG2* in *T. thalictroides* plants, using *TtPDS* as a

reporter gene. VIGS-treated flowers were undistinguishable from untreated or mock-treated flowers to

the naked eye, except for photobleaching of carpels and young stamens due to down-regulation of

TtPDS (compare Figs. 5 A, B). Of 33 plants infiltrated with TRV2-*TtPDS-ThtAG2*, 30 survived to

flowering. Photobleaching of leaves was first observed 21 days after infiltration in 20 plants (67%) and

nine of these (30% of total) had photobleached flowers.

No homeotic phenotypes were apparent in flowers from any treatment, so we used confocal

microscopy to investigate the presence of an ovule phenotype. As explained above, untreated ovules of

wild-type plants are bitegmic, with both integuments enveloping the nucellus and a distinct, cellularized

embryo sac at maturity (Fig. 5 C). Of the nine TRV2-*TtPDS-ThtAG2* treated flowers fixed for confocal

microscopy, all exhibited abnormalities in ovule development (different ovules from the same flower

often exhibited different phenotypes). Eight flowers contained ovules with an elongated structure

protruding slightly from the micropyle, an amorphous micropylar region and an unorganized, under-

developed embryo sac (Fig. 5 E   F). In two of these flowers, no nuclei were visible in any of the ovules

(Fig. 5 E), whereas in the other six at least some of the ovules contained embryo sacs where one to three nuclei were visible (Fig. 5 F). Six flowers (five of which also had weak phenotypes) exhibited a stronger phenotype, in which the ovules had been replaced by a carpel-like structure with no embryo sac (Fig. 5 G, H). We interpreted this structure as a homeotic conversion of an ovule into a carpel-like chimeric organ, based on the morphological similarity to a wild-type carpel, with the elongated terminal region resembling a style and a stigma (Fig. 5 D). Some ovule features are still present on these converted organs (i.e. integuments and hypostase), indicating that they are chimeric, with mixed carpel and ovule identity.

In an attempt to confirm carpel identity, we looked for morphological markers of carpel development and carpel-specific cell-types using three-dimensional (3D) confocal reconstructions of ovule morphology. We first compared surface cells of the style-like protrusion to those of wild-type styles and ovule (outer) integuments. Since surfaces of all three organs consisted of mostly isodiametric cells with slightly irregular borders (Fig. S3 A C), surface cell morphology could not be used to determine the identity of the protrusions. We then examined five style-like protrusions by 3D imaging and near-360° views for three of them (Fig. 5 I-L). These reconstructions revealed a distinct and irregular cleft (Fig. 5 J, L), which we interpreted as the furrow present in normal developing wild-type ascidiate carpels of *Thalictrum* and other non-core eudicots (46) (Fig. 5 C).

Hand pollination with wild-type pollen resulted in mature achenes in untreated flowers (Fig. 5 M) and in green carpels (presumably not experiencing silencing) of chimeric flowers in plants treated by VIGS (Fig. 5 N). Fully photobleached carpels, in which *ThtAG2* was presumably down-regulated, did not develop achenes and appeared atrophied (Fig. 5 N). Seeds of mature untreated fruits were enlarged and plump, containing endosperm and a small embryo upon sectioning (Fig. 5 O, P). The underdeveloped embryo is consistent with the morphological dormancy of seeds described for other

Ranunculaceae, in which the embryo develops fully only after seed dispersal (45,47), and was therefore considered normal. Within the atrophied carpels of VIGS-treated plants, ovules appeared aborted, elongated and withered, with no visible embryo (Fig. 5 Q, R).

To confirm the down-regulation of *ThtAG2* by VIGS, we analyzed its expression and tested for the presence of TRV1 and TRV2 transcripts in flowers from treated, untreated and mock-treated (empty TRV2) plants (Fig. 6, S5). All treated flowers (n=9) had *ThtAG2* expression that was visibly lower than untreated (n=5) and mock-treated controls (n=3) (Fig. 6 A). Gel band brightness quantified using ImageJ, showed that the average mean *ThtAG2* expression in treated flowers was statistically lower than controls (12.5-fold decrease) (Fig 6 B). We also analyzed *ThtAG1* expression levels, which were lower than controls in some of the treated flowers (Fig. 6 A, flowers 1a, 2a, 4a and 1b). Average *ThtAG1* expression in treated flowers was not statistically different from controls (Fig. 6 B), yet individual flower expression levels ranged greatly, from a 9-fold decrease to a 2-fold increase compared to average untreated (Fig. S5). Treated plants were chimeric, which is not uncommon in VIGS experiments (36), having a mix of flowers with mild and strong phenotypes.

TRV1 and TRV2 transcripts were present in all TRV2-*TtPDS-ThtAG2* treated flowers. TRV2 product sizes were approximately 1244 bp, yet varied slightly in size from whole (429 bp) to truncated (314 bp) versions of the *ThtAG2* insert (Fig. S6). Truncation of inserts has been reported in *Agrobacterium*-mediated transformation of other species (48,49) and it is not expected to affect silencing in the observed range. Of the eight mock-treated flowers screened, three had TRV1 and empty TRV2 transcripts (390 bp); no viral transcripts were detected in the five untreated flowers screened (Fig. S6).

ThtAG2 interacts with the E class protein ThtSEP3, but not with ThtAG1

Because in *A. thaliana*, SEP3 is required for ovule development (50) and is known to interact directly with AG (51,14,52), we tested their interactions in *T. thalictroides*. We used a GAL4-based yeast two-hybrid system to characterize interactions among ThtAG1, ThtAG2 and ThtSEP3ΔC (Fig. 7). ThtAG2 did not interact with itself or with ThtAG1 when fused to either the activation or binding domain of GAL4. ThtAG2 did interact with ThtSEP3ΔC, although only when fused to the activation domain – a phenomenon reported previously in ThtAG1 (23).

Discussion

We present evidence that, following gene duplication in the C lineage of Ranunculales, the *Thalictrum thalictroides AG* ortholog *ThtAG2* was constrained to ovule identity (a function traditionally described for D class genes). This change in gene function is supported by: 1) positive selection on the branch of the gene tree leading to *ThtAG2*, 2) restriction of *ThtAG2* expression to female reproductive organs (carpels and ovules) and 3) conversion of ovules into carpel-like structures as a result of targeted gene silencing of *ThtAG2*.

In the absence of D lineage genes, *ThtAG2* assumed ovule identity function

Given the angiosperm-wide nature of the duplication leading to the C and D lineages (15), it is likely that a *Thalictrum* ancestor once possessed a D lineage gene (Fig. 1, 8 B). Functional studies of other D lineage genes (16,17,19) suggest that this putative D gene would have had an ovule identity role that was redundantly shared with the C lineage gene. Based on the widespread absence of D lineage genes from Ranunculaceae, Berberidaceae and Eupteleaceae and their presence in Papaveraceae and Lardizabalaceae, the most parsimonious explanation (based on sequences available at the time of this study) is that the D lineage was lost at least twice in the order Ranunculales (Fig. 8 A). Based on current

understanding of the Ranunculales phylogeny (21), we propose that these gene losses would have occurred after the *AG1/AG2* duplication at the base of the order.

Since the original ABC model was proposed, more extensive sampling of orthologs from divergent taxa has resulted in gene phylogenies with specific clades that are generally identified by the same class-based organization. However, it is apparent from the literature that the actual function of such genes does not always correspond to the phylogenetic lineage in which they fall (53,54). This issue of genetic versus functional orthology is also clearly illustrated by *ThtAG2*, which falls within the C lineage in the *AG* subfamily, but has a D class function. In other taxa, such as *Arabidopsis, Petunia* and rice, ovule identity is specified redundantly by genes in the C and D lineages (17,16,19).

Examples of D lineage genes with expression in both ovules and carpels are also common, and include hyacinth (*HoMADS1*), lisianthus (*EGMADS1*) and grape (*VvMADS5*) (55–57). For the reasons stated above, the distinction of the D from the C class has been controversial. This study provides an additional argument against the designation of the D lineage as a separate class, as it illustrates another example of D function, i.e. ovule identity, taken up by a gene that falls within the C lineage.

Patterns of molecular evolution support sub-functionalization of *ThtAG2*

As mutations accumulate in duplicated genes, a variety of fates are possible for each copy, each of which corresponds to a specific pattern of selection. For this reason, we investigated the molecular signatures of selection on *ThtAG2*. Among the *Thalictrum* genes tested, both *ThtAG2* and *ThhAG2* showed evidence of positive selection, suggesting divergence of function that occurred before the diversification of these two species. The species phylogeny of *Thalictrum* has one major split: *T. thalictroides* belongs to one of the major clades that has traits considered ancestral for the genus (such as being diploid, hermaphroditic and insect pollinated, while *T. dioicum* and *T. hernandezii* belong to the

other major clade of derived, mostly polyploid, dioecious or andromonoecious (and wind-pollinated) species (33,41). The fact that our analysis detected positive selection in the common ancestor of both *ThhAG2* and *ThtAG2*, but not in *ThdAG2,* suggests that positive selection (and sub-functionalization) occurred before the diversification of the genus and was subsequently relaxed in the *T. dioicum* lineage.

Our analyses indicated that five specific amino acids are under positive selection in *ThtAG2* and *ThhAG2*, one of which is a threonine that falls within the second helix of the K domain and has been implicated in tetramerization (58). We propose that this residue could be altering the ability of ThtAG2 to form higher-order protein complexes, a mechanism by which MADS-box proteins auto-regulate their own expression and cross-regulate each other's expression (59,60). The four remaining amino acids under positive selection are located in the C terminus, which has been recognized as having multiple functions, including protein-protein interactions (40). Interestingly, two of these (serine and phenylanine) are located in AG Motif I. Although the precise role of the AG motif is unknown, its high conservation suggests that positive selection on these sites might be relevant to the functional novelty revealed by our VIGS experiments.

Positive selection can be a signature for both neo- or sub-functionalization (61). Under neo-functionalization, we would expect *ThtAG2* to be under positive selection (dN/dS>1) and its paralog *ThtAG1*, to be under neutral selection (dN/dS=1). While this scenario would potentially fit our molecular evolution results (Table 1, Fig. S2), so would sub-functionalization, which we argue is more plausible (Fig. 8 B). In a sub-functionalization scenario, the ancestral C lineage alone controlled stamen and carpel identity, while the C and D lineage genes redundantly controlled ovule identity. This hypothetical multiplicity of function of the ancestral C lineage (control of stamen, carpel and ovule identity) may cause "adaptive conflict," as mutations optimizing one function can be detrimental to other functions (62). When the ancestral C lineage gene duplicated into *AG1* and *AG2* (Fig 8, step 1), carpel and stamen

identity would have been under the control of both paralogs, while ovule identity would have been redundantly shared among at least three genes. The next step in this model would be the loss of the D lineage (Fig. 8, step 2) and the reduction to two ovule identity genes, *AG1* and *AG2*. Subsequently, the two paralogs would have sub-functionalized to escape adaptive conflict, resulting in *AG1* retaining stamen and carpel (and possibly ovule) identity functions, while *AG2* became specialized for ovule identity (Fig 8, step 3).

This evolutionary path fits our molecular evolution evidence of positive selection in *ThtAG2*. It is also possible that the ancestral D lineage played a unique role in ovule development, which the C lineage took over, in a neo-functionalization event that preceded the *AG1/AG2* split in Ranunculales. This scenario could explain the frequent loss of either the D lineage or one of the C lineage paralogs that occurs among the Ranunculid families (Fig. 8). Taken together, given current data, we conclude that sub-functionalization is an accurate description of the functional divergence of the *ThtAG2* paralog, either on its own or following a neo-functionalization event.

Diverse forms of sub-functionalization are present among Ranunculid C lineage duplicates

Although there are no additional functional studies of *AG* orthologs in Ranunculaceae, we can draw comparisons from evidence in the closely related Papaveraceae (within the order Ranunculales). In *Eschscholzia californica, EScaAG1* and *EScaAG2* represent a genus-specific duplication that occurred an estimated 51 mya. Down-regulation of these paralogs by VIGS demonstrated that they are redundantly necessary for stamen and carpel identity, and meristem determinacy (63). In *Papaver somniferum,* two *AG* homologs that are the product of alternative splicing are highly expressed in stamens and carpels and appear to be partially sub-functionalized. One of the copies is responsible for stamen and carpel identity and floral meristem determinacy; the other is partially redundant in these

functions, while also having a distinct role in carpel development (64). Therefore, although the above examples are similar to that of *ThtAG1/ThtAG2* in having sub-functionalized, they differ both in the pattern of gene duplication and in the function of their protein products. While it appears likely that Ran*AG1* and Ran*AG2* are present among most genera within Ranunculaceae, additional functional studies will be needed to infer the ancestral condition for the family and to ascertain whether sub-functionalization (and a shift to ovule identity) has occurred more than once. The *Ranunculus RascAG* gene would be a particularly interesting candidate for functional studies, given that our molecular evolution analysis suggests that it is the only other Ranunculales C lineage gene under positive selection (Fig. S1).

Down-regulation of *ThtAG2* suggests a function in ovule-identity

We show that down-regulation of *ThtAG2* results in homeotic conversion of ovules into chimeric carpel-like structures (Fig. 5). While a carpel marker gene would have allowed us to ascertain carpel identity of the homeotic organ at the molecular level, the marker that has been used in other species, *CRABSCLAW* (*CRC*) (18,65), is not carpel-specific in Ranunculids that are more closely related to our study system (66,67). Nevertheless, we find that several features of morphological resemblance of the homeotic organs to carpels (compare Fig. 5 G-L to C, D), in combination with verification that *ThtAG2* was down-regulated (Fig. 6), is highly suggestive that this is an ovule-identity gene. Without properly forming ovules, we predicted that seeds would not mature and carpels would atrophy. Indeed, photobleached carpels failed to mature into fruits (Fig. 5 N).

An unexpected side effect of the down-regulation of *ThtAG2* by VIGS was the apparent concurrent down-regulation of *ThtAG1* in a subset of the samples (4/9, Fig. 6). It is unlikely that *ThtAG1* was silenced by the *ThtAG2* construct due to low levels of homology between the two genes, especially

within the fragment chosen for the VIGS construct. The sequence used for VIGS shares 44.5% identity between *ThtAG1* and *ThtAG2*, with no stretches of contiguous homologous nucleotides greater than sixteen throughout its length. The lower size limit for post transcriptional gene silencing in plants is twenty-three nucleotides (68). The alternative explanation for lowered levels of *ThtAG1* is that *ThtAG2* may promote expression of *ThtAG1*. For two of the flower samples (2a and 1b), expression levels of *ThtAG1* were actually lower than the average down-regulation seen in *ThtAG1* VIGS-treated flowers in a previous study, where floral phentoypes were comparable to a total loss of *AG* function (double flowers without stamens nor carpels) (Fig. S5, flowers 2a and 2b) (23). One of those samples, had fewer organs overall and an abnormal floral meristem that seemed to have terminated early, producing only five carpels and leaving undifferentiated tissue in the flower center (Fig S4, flower 1b). We hypothesize that silencing of *ThtAG2* would have indirectly caused a reduction in *ThtAG1* expression late in development and that we are therefore seeing the late effects of *ThtAG1* (termination of floral meristem) and not the early effects (establishment of stamen and carpel identity) that would have resulted in double flowers. The next lowest expression of *ThtAG1* was a 5.5-fold decrease, and the morphology of this flower looked comparable to untreated controls (Fig S4, flower 2a). Ultimately, we regarded this indirect down-regulation of *ThtAG1* as non-significant, since it did not affect carpel and stamen identity.

Given that ovules show abnormal phenotypes in *ThtAG2* knock-down experiments, redundancy between *ThtAG1* and *ThtAG2* in ovule development can be ruled out (Fig. 4 B). Nevertheless, *ThtAG1* is expressed in ovules (Fig. 3), suggesting it may have a distinct function from *ThtAG2* in ovule development or, alternatively, that ovule expression is vestigial. While we have shown that *ThtAG2* is necessary for ovule identity, it is not possible to disprove, with current data, that *ThtAG2* could have a redundant early role with *ThtAG1* in carpel and stamen identity. *ThtAG2* is not expressed in fully developed young stamens, prior to anther dehiscence (Fig. 3 I), nor at earlier stages of development by

in situ hybridization (Fig.3 B, E), yet we cannot rule out expression when stamen and carpel primordia are beginning to develop. Early expression of *ThdAG2* is not present in *T. dioicum* stamen and carpel primordia (24), yet we have shown that this ortholog is highly divergent within the genus and therefore direct comparisons of these two genes may not be fully informative.

That down-regulation of *ThtAG2* does not cause changes in morphology to the stamens or carpels may be interpreted as redundancy with *ThtAG1* in conferring the identity of these organs. Down-regulation of *ThtAG1* causes a full conversion of carpels (and stamens) into sepals (23), and *ThtAG2* expression is absent from these flowers (data not shown). This may be simply due to the physical absence of carpels within which ovules would normally develop, however, it could also signify that *ThtAG1* is an upstream initiator of *ThtAG2*. Analysis of protein-protein interactions suggests that ThtAG1 and ThtAG2 do not interact directly, since they do not dimerize (Fig.7). It should be noted, however, that in *A. thaliana* C and D lineage genes that do not form dimers in yeast-two hybrid assays are still able to form multimeric protein complexes when SEP proteins are added (50). Since ThtAG2 does interact with ThtSEP3 in our yeast two hybrid experiments, we cannot rule out a redundant role of ThtAG2 early in carpel and stamen identity with current evidence. Under this alternative scenario, *ThtAG1* would be required to initiate *ThtAG2* in carpel (and possibly stamen) primordia, after which both genes would redundantly promote C class function. This scenario is consistent with our data and with the fact that MADS-box genes are known to have complex auto and cross-regulatory interactions (60,59).

Expression patterns vary among paralogs and between species of *Thalictrum*

T. thalictroides ThtAG1 is expressed in all reproductive organs (stamens, carpels and ovules) (Fig. 3)*,* and exhibits the canonical C class function; its down-regulation results in homeotic conversion

of carpels and stamens into perianth organs and floral indeterminacy (23). In a previous analysis, the expression of *T. dioicum ThdAG2* was shown to occur late in flower development and to be spatially limited to ovules (24). In *T. thalictroides,* we find that even though *ThtAG2* is expressed in carpels and ovules (Fig. 2), its down regulation does not affect carpel development (Fig. 5 A).

It is possible that the difference in expression patterns between *AG2* orthologs in *T. dioicum* and *T. thalictroides* is related to their divergent sequences. *ThdAG2* is lacking most of the AG Motif I (Fig. 2) and this could potentially cause a change in protein-protein interactions required for self-regulation (40). Inclusion of *AG2* orthologs from three additional species and phylogenetic considerations (33) suggest that *ThtAG2* is representative of the ancestral ortholog and *ThdAG2* is a divergent ortholog, present in either just this species or a small clade containing it. Additional studies on *AG2* orthologs and their protein-binding partners in other *Thalictrum* species would be needed to determine the potential role in the expression and function of *AG2* orthologs across the genus. The integument-specific expression of both *ThtAG1* and *ThtAG2* in buds (Fig. 3 D, G) is intriguing, given that down-regulation of *ThtAG2* seems to affect the nucellus and embryo sac and that no abnormalities were found in the integuments. This result could be an effect of lack of co-localization of the mRNA and its protein product or the result of analyzing flowers after ovule development has initiated.

Conclusions

In spite of ovules being undoubtedly crucial organs in plants that reproduce through seeds, functional studies determining the genetic control of ovule identity are rare. Our study presents a novel example of a C lineage gene that has acquired an exclusive function in ovule identity. Analyzing the ways in which floral transcription factors, such as these *AG* orthologs, have evolved following duplication can ultimately help to determine their potential role in angiosperm diversification.

Acknowledgements

References

1. Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. Genetics. 1999 Apr 1;151(4):1531–45.

2. Ohno S. Evolution by gene duplication. Springer-Verlag; 1970. 184 p.

3. Airoldi CA, Davies B. Gene Duplication and the Evolution of Plant MADS-box Transcription Factors. J Genet Genomics. 2012 Apr 20;39(4):157–65.

4. Parenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, et al. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell. 2003 Jul;15(7):1538–51.

5. Theissen G, Kim JT, Saedler H. Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. J Mol Evol. 1996 Nov;43(5):484–516.

6. Kim S, Koh J, Yoo M-J, Kong H, Hu Y, Ma H, et al. Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. Plant J. 2005 Sep 1;43(5):724–44.

7. Zahn LM, Kong H, Leebens-Mack JH, Kim S, Soltis PS, Landherr LL, et al. The Evolution of the SEPALLATA Subfamily of MADS-Box Genes A Preangiosperm Origin With Multiple Duplications Throughout Angiosperm History. Genetics. 2005 Apr 1;169(4):2209–23.

8. Zahn LM, Leebens-Mack JH, Arrington JM, Hu Y, Landherr LL, DePamphilis CW, et al. Conservation and divergence in the AGAMOUS subfamily of MADS-box genes: evidence of independent sub- and neofunctionalization events. Evol Dev. 2006 Jan 1;8(1):30–45.

9. Bowman JL, Smyth DR, Meyerowitz EM. Genetic interactions among floral homeotic genes of Arabidopsis. Development. 1991 May 1;112(1):1–20.

10. Coen ES, Meyerowitz EM. The war of the whorls: genetic interactions controlling flower development. Nature. 1991;353(6339):31–7.

11. Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H. Genetic Control of Flower Development by Homeotic Genes in Antirrhinum majus. Science. 1990 Nov 16;250(4983):931–6.

12. Angenent GC, Franken J, Busscher M, Dijken A van, Went JL van, Dons HJ, et al. A novel class of MADS box genes is involved in ovule development in petunia. Plant Cell Online. 1995 Oct 1;7(10):1569–82.

13. Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. The SEP4 Gene of Arabidopsis thaliana Functions in Floral Organ and Meristem Identity. Curr Biol. 2004 Nov 9;14(21):1935–40.

14. Theißen G, Saedler H. Plant biology: Floral quartets. Nature. 2001 Jan 25;409(6819):469–71.

15. Kramer EM, Jaramillo MA, Stilio VSD. Patterns of Gene Duplication and Functional Evolution During the Diversification of the AGAMOUS Subfamily of MADS Box Genes in Angiosperms. Genetics. 2004 Feb 1;166(2):1011–23.

16. Heijmans K, Ament K, Rijpkema AS, Zethof J, Wolters-Arts M, Gerats T, et al. Redefining C and D in the Petunia ABC. Plant Cell Online. 2012 Jun 1;24(6):2305–17.

17. Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, et al. Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature. 2003 Jul 3;424(6944):85–8.

18. Dreni L, Jacchia S, Fornara F, Fornari M, Ouwerkerk PBF, An G, et al. The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice. Plant J Cell Mol Biol. 2007 Nov;52(4):690–9.

19. Dreni L, Pilatone A, Yun D, Erreni S, Pajoro A, Caporali E, et al. Functional Analysis of All AGAMOUS Subfamily Members in Rice Reveals Their Roles in Reproductive Organ Identity Determination and Meristem Determinacy. Plant Cell Online. 2011 Aug 1;23(8):2850–63.

20. Maia VH, Gitzendanner MA, Soltis PS, Wong Gane Ka-Shu, Soltis DE. Angiosperm Phylogeny Based on 18S/26S rDNA Sequence Data: Constructing a Large Data Set Using Next-Generation Sequence Data. Int J Plant Sci. 2014 Jul 1;175(6):613–50.

21. Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, et al. Angiosperm phylogeny: 17 genes, 640 taxa. Am J Bot. 2011 Apr 1;98(4):704–30.

22. Sattler R, Perlin L. Floral development of Bougainvillea spectabilis Willd., Boerhaavia diffusa L. and Mirabilis jalapa L. (Nyctaginaceae). Bot J Linn Soc. 1982 Apr 1;84(3):161–82.

23. Galimba KD, Tolkin TR, Sullivan AM, Melzer R, Theißen G, Di Stilio VS. Loss of deeply conserved C-class floral homeotic gene function and C- and E-class protein interaction in a double-flowered ranunculid mutant. Proc Natl Acad Sci U S A [Internet]. 2012 Aug 1 [cited 2012 Aug 8]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/22853954

24. Di Stilio VS, Kramer EM, Baum DA. Floral MADS box genes and homeotic gender dimorphism in Thalictrum dioicum (Ranunculaceae) - a new model for the study of dioecy. Plant J Cell Mol Biol. 2005 Mar;41(5):755–66.

25. Pabón-Mora N, Wong GK-S, Ambrose BA. Evolution of fruit development genes in flowering plants. Plant Evol Dev. 2014;5:300.

26. Maddison DR, Maddison WP. MacClade v. 4.08. Sinauer Assoc. 2005;

27. Mayer K, Schüller C, Wambutt R, Murphy G, Volckaert G, Pohl T, et al. Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana. Nature. 1999 Dec 16;402(6763):769–77.

28. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinforma Oxf Engl. 1998;14(9):817–8.

29. Miller MA, Pfeiffer W, Schwartz T. The CIPRES Science Gateway: A Community Resource for Phylogenetic Analyses. Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery [Internet]. New York, NY, USA: ACM; 2011 [cited 2014 Aug 5]. p. 41:1–41:8. Available from: http://doi.acm.org/10.1145/2016741.2016785

30. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007 Aug;24(8):1586–91.

31. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scand J Stat. 1979 Jan 1;6(2):65–70.

32. Poon A, Frost S, Kosakovsky Pond S. Detecting Signatures of Selection from DNA Sequences Using Datamonkey - Springer. In: Posada D, editor. Humana Press; 2009 [cited 2015 May 28]. Available from: http://link.springer.com/protocol/10.1007%2F978-1-59745-251-9_8

33. Soza VL, Brunet J, Liston A, Salles Smith P, Di Stilio VS. Phylogenetic insights into the correlates of dioecy in meadow-rues (Thalictrum, Ranunculaceae). Mol Phylogenet Evol. 2012 Apr;63(1):180–92.

34. Kramer E. Methods for studying the evolution of plant reproductive structures: Comparative gene expression techniques. Methods in Enzymology. Elsevier Inc; 2005. p. 617–36.

35. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2−\Delta\Delta CT$ Method. Methods. 2001 Dec;25(4):402–8.

36. Di Stilio VS, Kumar RA, Oddone AM, Tolkin TR, Salles P, McCarty K. Virus-induced gene silencing as a tool for comparative functional studies in Thalictrum. PloS One. 2010;5(8):e12064.

37. Hileman LC, Drea S, de Martino G, Litt A, Irish VF. Virus-induced gene silencing is an effective tool for assaying gene function in the basal eudicot species Papaver somniferum (opium poppy). Plant J. 2005 Oct 1;44(2):334–41.

38. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012 Jul;9(7):671–5.

39. Causier B, Davies B. Analysing protein-protein interactions with the yeast two-hybrid system. Plant Mol Biol. 2002 Dec 1;50(6):855–70.

40. Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM. The Emerging Importance of Type I MADS Box Transcription Factors for Plant Reproduction. Plant Cell Online. 2011 Mar 1;23(3):865–72.

41. Soza VL, Haworth KL, Stilio VSD. Timing and Consequences of Recurrent Polyploidy in Meadow-Rues (Thalictrum, Ranunculaceae). Mol Biol Evol. 2013 Aug 1;30(8):1940–54.

42. Lubbers AE, Christensen NL. Intraseasonal Variation in Seed Production Among Flowers and Plants of Thalictrum thalictroides (Ranunculaceae). Am J Bot. 1986 Feb 1;73(2):190–203.

43. Haig D, Westoby M. Inclusive Fitness, Seed Resources, and Maternal Care. Plant Reproductive Ecology: Patterns and Strategies. New York, New York: Oxford University Press; 1988. p. 60–79.

44. Strasburger E. Die Angiospermen und die Gymnospermen. Jena: Fischer; 1879.

45. Unal M, Vardar F. Embryological analysis of Consolida regalis L. (Ranunculaceae). ACTA Biol CRACOVIENSIA. 2006;48(1):27–32.

46. Endress PK, Igersheim A. Gynoecium diversity and systematics of the basal eudicots. Bot J Linn Soc. 1999 Aug 1;130(4):305–93.

47. Baskin C, Baskin J. Seeds. San Diego, California: Academic Press; 1998. 666 p.

48. Cheng M, Fry JE, Pang S, Zhou H, Hironaka CM, Duncan DR, et al. Genetic Transformation of Wheat Mediated by Agrobacterium tumefaciens. Plant Physiol. 1997 Nov 1;115(3):971–80.

49. Levée V, Garin E, Klimaszewska K, Séguin A. Stable genetic transformation of white pine (Pinus strobus L.) after cocultivation of embryogenic tissues with Agrobacterium tumefaciens. Mol Breed. 1999 Oct 1;5(5):429–40.

50. Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, Ditta G, et al. MADS-Box Protein Complexes Control Carpel and Ovule Development in Arabidopsis. Plant Cell Online. 2003 Nov 1;15(11):2603–11.

51. Honma T, Goto K. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature. 2001 Jan 25;409(6819):525–9.

52. Melzer R, Verelst W, Theißen G. The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in "floral quartet"-like complexes in vitro. Nucleic Acids Res. 2009 Jan 1;37(1):144–57.

53. Causier B, Castillo R, Zhou J, Ingram R, Xue Y, Schwarz-Sommer Z, et al. Evolution in Action: Following Function in Duplicated Floral Homeotic Genes. Curr Biol. 2005 Aug 23;15(16):1508–12.

54. Davies B, Motte P, Keck E, Saedler H, Sommer H, Schwarz-Sommer Z. PLENA and FARINELLI: redundancy and regulatory interactions between two Antirrhinum MADS-box factors controlling flower development. EMBO J. 1999 Jul 15;18(14):4023–34.

55. Xu H, Li X, Li Q, Bai S, Lu W, Zhang X. Characterization of HoMADS 1 and its induction by plant hormones during in vitro ovule development in Hyacinthus orientalis L. Plant Mol Biol. 2004 May 1;55(2):209–20.

56. Tzeng T-Y, Chen H-Y, Yang C-H. Ectopic Expression of Carpel-Specific MADS Box Genes from Lily and Lisianthus Causes Similar Homeotic Conversion of Sepal and Petal in Arabidopsis. Plant Physiol. 2002 Dec 1;130(4):1827–36.

57. Boss PK, Vivier M, Matsumoto S, Dry IB, Thomas MR. A cDNA from grapevine (Vitis vinifera L.), which shows homology to AGAMOUS and SHATTERPROOF, is not only expressed in flowers but also throughout berry development. Plant Mol Biol. 2001 Mar 1;45(5):541–53.

58. Puranik S, Acajjaoui S, Conn S, Costa L, Conn V, Vial A, et al. Structural Basis for the Oligomerization of the MADS Domain Transcription Factor SEPALLATA3 in Arabidopsis. Plant Cell. 2014 Sep;26(9):3603–15.

59. Schwarz-Sommer Z, Hue I, Huijser P, Flor PJ, Hansen R, Tetens F, et al. Characterization of the Antirrhinum floral homeotic MADS-box gene deficiens: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. EMBO J. 1992 Jan;11(1):251–63.

60. Tröbner W, Ramirez L, Motte P, Hue I, Huijser P, Lönnig WE, et al. GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of Antirrhinum floral organogenesis. EMBO J. 1992 Dec;11(13):4693–704.

61. Wendel J, Greilhuber J, Dolezel J, Leitch IJ. Plant Genome Diversity Volume 1: Plant Genomes, their Residents, and their Evolutionary Dynamics. Springer Science & Business Media; 2012. 282 p.

62. Hittinger CT, Carroll SB. Gene duplication and the adaptive evolution of a classic genetic switch. Nature. 2007 Oct 11;449(7163):677–81.

63. Yellina AL, Orashakova S, Lange S, Erdmann R, Leebens-Mack J, Becker A. Floral homeotic C function genes repress specific B function genes in the carpel whorl of the basal eudicot California poppy (Eschscholzia californica). EvoDevo. 2010 Dec 1;1(1):13.

64. Hands P, Vosnakis N, Betts D, Irish VF, Drea S. Alternate transcripts of a floral developmental regulator have both distinct and redundant functions in opium poppy. Ann Bot. 2011 Jun;107(9):1557–66.

65. Orashakova S, Lange M, Lange S, Wege S, Becker A. The CRABS CLAW ortholog from California poppy (Eschscholzia californica, Papaveraceae), EcCRC, is involved in floral meristem termination, gynoecium differentiation and ovule initiation. Plant J. 2009;58(4):682–93.

66. Lee J-Y, Baum SF, Oh S-H, Jiang C-Z, Chen J-C, Bowman JL. Recruitment of CRABS CLAW to promote nectary development within the eudicot clade. Development. 2005 Nov 15;132(22):5021–32.

67. Sun W, Huang W, Li Z, Lv H, Huang H, Wang Y. Characterization of a Crabs Claw Gene in Basal Eudicot Species Epimedium sagittatum (Berberidaceae). Int J Mol Sci. 2013 Jan 8;14(1):1119–31.

68. Thomas CL, Jones L, Baulcombe DC, Maule AJ. Size constraints for targeting post-transcriptional gene silencing and for RNA-directed methylation in Nicotiana benthamiana using a potato virus X vector. Plant J. 2001 Feb 1;25(4):417–25.

Figure 1: *Thalictrum thalictroides AGAMOUS* **duplicates belong within the C lineage of floral organ identity genes.** Bayesian Analysis of the *AGAMOUS* Subfamily, with expanded sampling of non-core eudicots. Branches with significant posterior probabilities ($\geq$ 0.95) indicated in bold. C and D lineages and the order Ranunculales are shown with brackets. Ranunculaceae *AG (*Ran*AG1* and Ran*AG2)* indicated in green and *Thalictrum thalictroides AG1* and *AG2* indicated with red arrows. Eupteleaceae genes in blue, Lardizabalaceae in magenta, Berberidaceae in red and Papaveraceae in orange. A star shows the duplication event leading to *AG1* and *AG2* in the Ranunculales.

Figure 2: Predicted amino acid alignment of C terminal regions for selected C and D lineage proteins. Sequences arranged phylogenetically, with core eudicots at the top followed by non-core eudicots. All non-core eudicot sequences present in the phylogeny are included in the alignment. Two highly conserved regions, AG Motif I and AG Motif II (15) are indicated by bars at the top. Residues that show 50% chemical conservation are highlighted in black. Residues that show 50% chemical similarity to the consensus residue are highlighted in gray. See Fig. 1 for family color scheme. Arrows show the proteins in this study: *Thalictrum thalictroides* AG1 and AG2.

Figure 3: Expression levels of two C class genes in *T. thalictroides* flowers at two different developmental stages.

A. Young bud dissected from dormant tuber eye representative of the stage at which *in situ* hybridization was performed. Scale bar = 1mm. B D. *ThtAG1* antisense showing expression in carpels and ovules. B. Whole flower. C. Carpel detail. D. Ovule sections with differentiated integuments. E G *ThtAG2* antisense showing expression in carpels and ovules. E. Whole flower. F. Carpel detail. H. Ovule

sections with differentiated integuments. For sense controls, see Supplemental Figure S2. Scale bar = 100 µm. H. Young open flower representative of the stage at which quantitative Real-Time PCR (qPCR) was performed. Scale bar = 1mm. I. Relative expression levels of *Thalictrum thalictroides AGAMOUS 1* (*ThtAG1*, white bars) and *ThtAG2* (gray bars) in dissected sepals, stamens, carpels and ovules (n=6 flowers, 3 plants) by qPCR. Expression calculated using the $2^{-\Delta\Delta Ct}$ method, normalized to *ThtACTIN* and *Elongation Factor 1α (ThtEF1α)*. Mean and standard error bars shown. se = sepal, st = stamen, ca = carpel, i = integument, nu = nucellus, ov = ovule, pl = placenta.

Figure 4: Ovule ontogeny in wild-type *Thalictrum thalictroides*.

A E: Five floral developmental stages, from bud to mature flower. Se = sepal; St= stamen; Ca = carpel; Fi = filament; An = anther. Scale bar = 1 mm. F- J: Confocal images of developing ovules. oi = outer integument, ii = inner integument, es = embryo sac, pn = polar nuclei, sn = secondary nucleus, ec = egg cell, cc = central cell, s = synergid cells, a = antipodal cells, h = hypostase. Scale bar = 100 µm.

Figure 5: Floral phenotypes resulting from virus-induced gene silencing of *ThtAG2* in *Thalictrum thalictroides*.

A. Untreated *T. thalictroides* flower. B. TRV2-*TtPDS-ThtAG2* VIGS-treated flower showing photobleached carpels and stamens, but otherwise normal morphology. Scale bar = 1mm.
C. SEM image of developing ascidiate wild-type carpel, furrow indicated with arrowheads. D. Wild-type carpel with style and stigma. E. Wild-type ovule with two integuments (outer and inner) and an embryo sac (with hypostase below), surrounded by nucellus. E   H. Ovules of flowers treated with TRV2-*TtPDS-ThtAG2*, showing increasing degree of phenotype strength. F. Abnormal ovule lacking an organized embryo sac and with protruding micropylar end emerging from the inner integuments. G-H.

Strong ovule phenotypes lacking embryo sac, with an increased hypostase and a protruding micropylar end elongated into a style-like structure. Scale bar = 100µm. I, K. Optical sections of TRV2-*TtPDS-ThtAG2* treated ovules, showing style-like protrusions. J, L. Three-dimensional reconstructions of the boxed areas in I, K. Furrow-like surfaces are indicated with arrows. M. Mature fruits (achenes) of an untreated flower. N. Partially or fully photobleached mature fruits and atrophied carpels from a plant treated with TRV2-*TtPDS-ThtAG2*. Scale bar = 1mm. O. Mature untreated fruit with pericarp partially removed, showing a normal seed inside. P. Longitudinal section through a similar fruit to show embryo. Scale bar = 1 mm/100 µm. Q-R. Confocal images of aborted ovules within atrophied carpels from VIGS-treated flowers that do not proceed to form a fruit. Scale bar = 100 µm. se = sepal, st = stamen, ca = carpel, oi = outer integument, ii = inner integument, es = embryo sac, nu = nucellus, h = hypostase, sg = stigma, st = style, ov = ovule, o = ovary, stl = style-like structure, a = achene, c = carpel (did not mature), se = seed, pe = pericarp, em = embryo, sc = seed coat, en = endosperm.

Figure 6: Molecular validation of virus-induced gene silencing experiments.

A. RT-PCR showing expression levels of *ThtAG2* and *ThtAG1* in individual flowers from plants treated with TRV2-*TtPDS-ThtAG2* vs. empty TRV2 and untreated controls. Samples are labeled above each lane, corresponding to specific plant (number) and flower (letter). Flower morphology corresponding to each sample shown in the gel can be found in Fig. S4. *ThtACTIN* was used as a loading control. B. Quantification of RT-PCR gel bands in part A using ImageJ. Mean and standard error bars shown for each treatment (individual values can be found in Fig. S5). Asterisk indicates statistical significance in a one-way ANOVA followed by Tukey test ($p < 0.05$), same letters indicate values being compared. Flowers targeted for *ThtAG2* down-regulation by VIGS have a significantly lower average expression of *ThtAG2*, compared to controls. Average expression of *ThtAG1* is not different between treatments.

Figure 7: Yeast two-hybrid analysis of C and E class protein interactions in *Thalictrum thalictroides*.

Growth on Leu/Trp/His-free +AbA medium shown. Cells were spotted in 10-fold serial dilutions from left to right for each interaction. ThtAG1, ThtAG2 and ThtSEP3ΔC were expressed as fusion with the GAL4 activation domain (AD) and the GAL4 DNA binding domain (BD).

Figure 8: Hypothetical reconstruction of the history of gene duplication and loss supporting sub-functionalization of *ThtAG1* and *ThtAG2*.

A. Family-level phylogeny of the order Ranunculales. Presence (filled boxes) or absence (empty boxes) of *AG1*, *AG2* and D lineage orthologs was determined from existing data as depicted in Fig. 1. Missing boxes indicate no data present. Gene loss events were chosen following the most parsimonious explanation. In this scenario the *AG1* ortholog was lost once, the *AG2* ortholog three times and the D lineage ortholog twice. B. Model for the evolution of C and D lineage genes in *Thalictrum,* beginning with a hypothetical common ancestor in the Ranunculales. Three major transitions are depicted, and their phylogenetic locations are also indicated in part A (1-3). 1. C lineage duplication at the base of the Ranunculales. 2. Loss of the D lineage ortholog in the common ancestor of Berberidaceae and Ranunculaceae. 3. Sub-functionalization of *AG2* to ovule identity in *Thalictrum*. Colors of organs indicate function in identity (possibly, but not necessarily matching expression domains).

Table 1: Amino acids exhibiting positive selection within ThtAG2.

Residues of ThtAG2 under positive selection ($\omega>1$) identified by Bayes Empirical Bayes (BEB) analysis, compared to the equivalent amino acid in ThtAG1. Regions and corresponding helices or

motifs are indicated, where applicable. The likelihood of positive selection is listed in the last column; the probability for position 284 is highly significant (**). Positive selection on these sites is shared with *ThhAG2*.

Figure S1: Evidence of episodic selection on the Ranunculales AG phylogeny, by the Branch-Site REL method. Phylogeny constructed by Datamonkey, using a neighbor-joining method, scaled on the expected number of substitutions/nucleotide. Numbers represent individual nodes. Colors indicates strength of selection: red, $\omega > 5$, blue, $\omega = 0$ and grey, $\omega=1$. The width of each color component represents the proportion of sites in the corresponding class. Thicker branches undergoing episodic diversifying selection by the sequential likelihood ratio test at corrected $p \leq 0.05$. The Branch-Site REL method used here indicates positive selection on *RascAG* (p=0.02) and *ThtAG2* (p=0.01).

Figure S2: *In Situ* hybridization sense controls

A C. *ThtAG1* sense controls. A. Whole flower. B. Carpel detail. C. Ovule sections with differentiated integuments. D F. *ThtAG2* sense controls. D. Whole flower. E. Carpel detail. F. Ovule sections with differentiated integuments. Scale bar = 100 μm.

Figure S3: Comparison of surface cells from wild-type carpels, ovules and VIGS-treated carpel-like organs. A. SEM of surface cells of wild-type carpel. B. SEM of surface cells of the outer integument of a wild-type ovule. C. Three-dimensional confocal reconstruction of surface cells of TRV2-*TtPDS-ThtAG2* treated style-like protrusion. Scale bar = 10 μm.

Figure S4: Phenotype of TRV2-*TtPDS-ThtAG2* treated flowers used for RT-PCR in Figure 6 and Figure S5. Plant (number) and flower (letter) indicated. Scale bar = 1mm.

Figure S5: Visual quantification of *ThtAG1* and *ThtAG2* expression in individual flowers targeted for *ThtAG2* gene silencing by VIGS. A. Quantification of *ThtAG2* RT-PCR gel bands from Figure 6 using ImageJ. B. Quantification of *ThtAG1* RT-PCR gel bands from Figure 6 using ImageJ.

Figure S6: Detection of tobacco rattle virus (TRV) transcripts in plants treated by virus-induced gene silencing. Expression analysis of plants treated with TRV2-*TtPDS-ThtAG2* + TRV1, empty TRV2 + TRV1, and untreated controls by RT-PCR. TRV1 and TRV2 were detected in VIGS-treated and empty TRV2-treated samples. Empty TRV2 generates a 390 bp band, while TRV2 containing the *ThtAG2* and *TtPDS* inserts generates a 1244 bp band. Approximate size of bands (in bp) indicated on the right. Plant numbers are labeled above each lane.

Table S1: List of angiosperm sequences used in the *AGAMOUS* phylogeny.

Species, family, gene name, accession number and relevant publication listed. 1KP project sequences are identified by the first letter of the genus, and first two letters of the specific epithet, followed by the scaffold identification number.

Table S2: List of primers used in cloning and expression analyses.

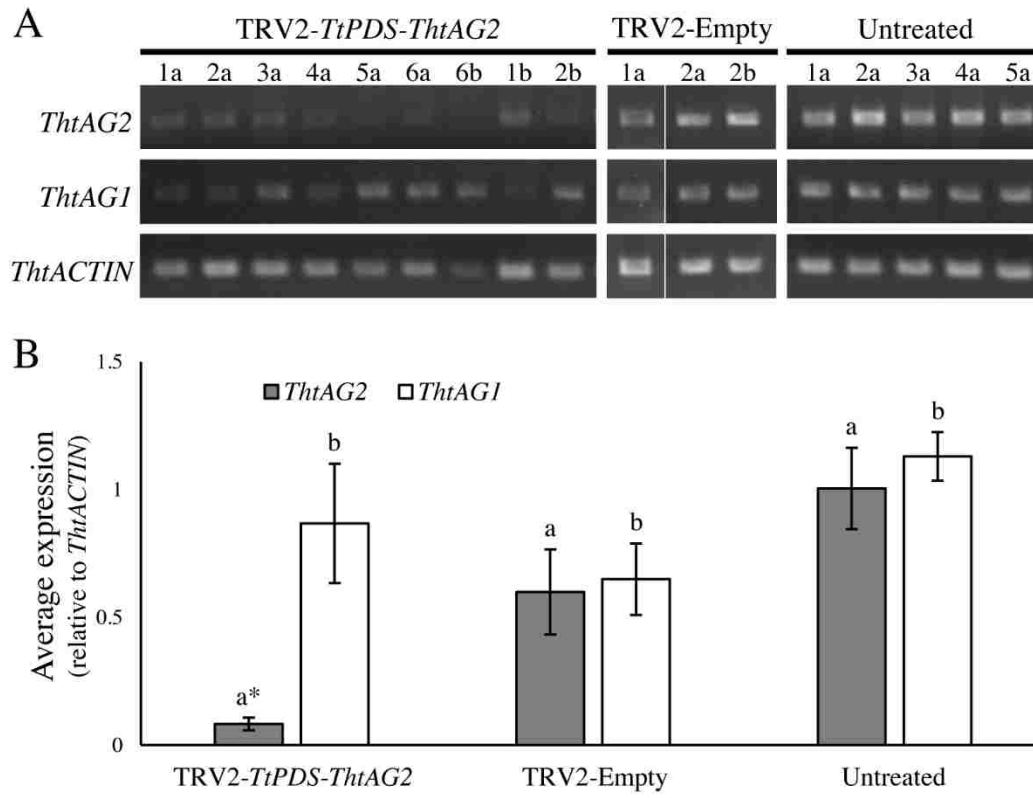Purpose, name and sequence of primers used in this study.
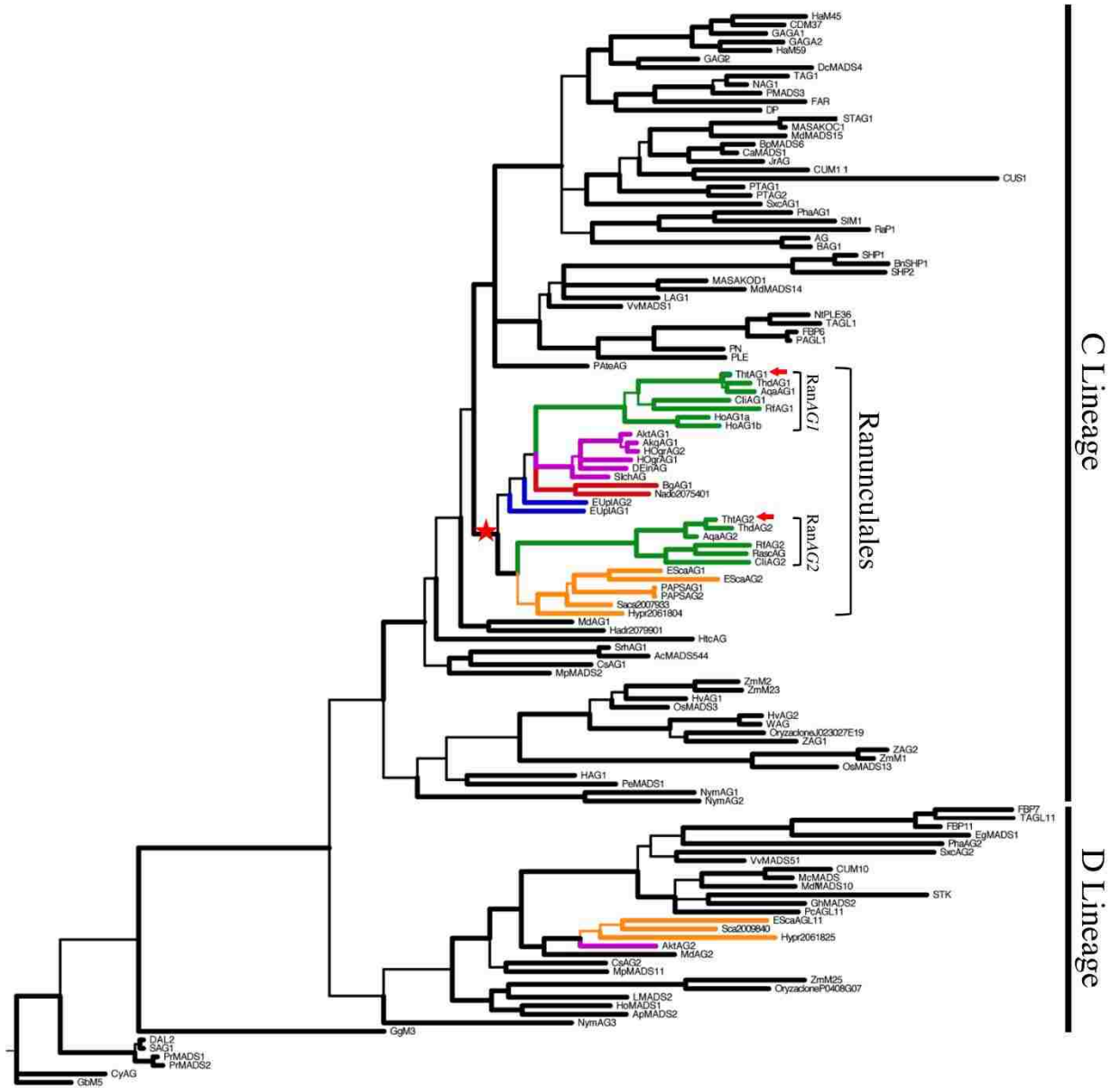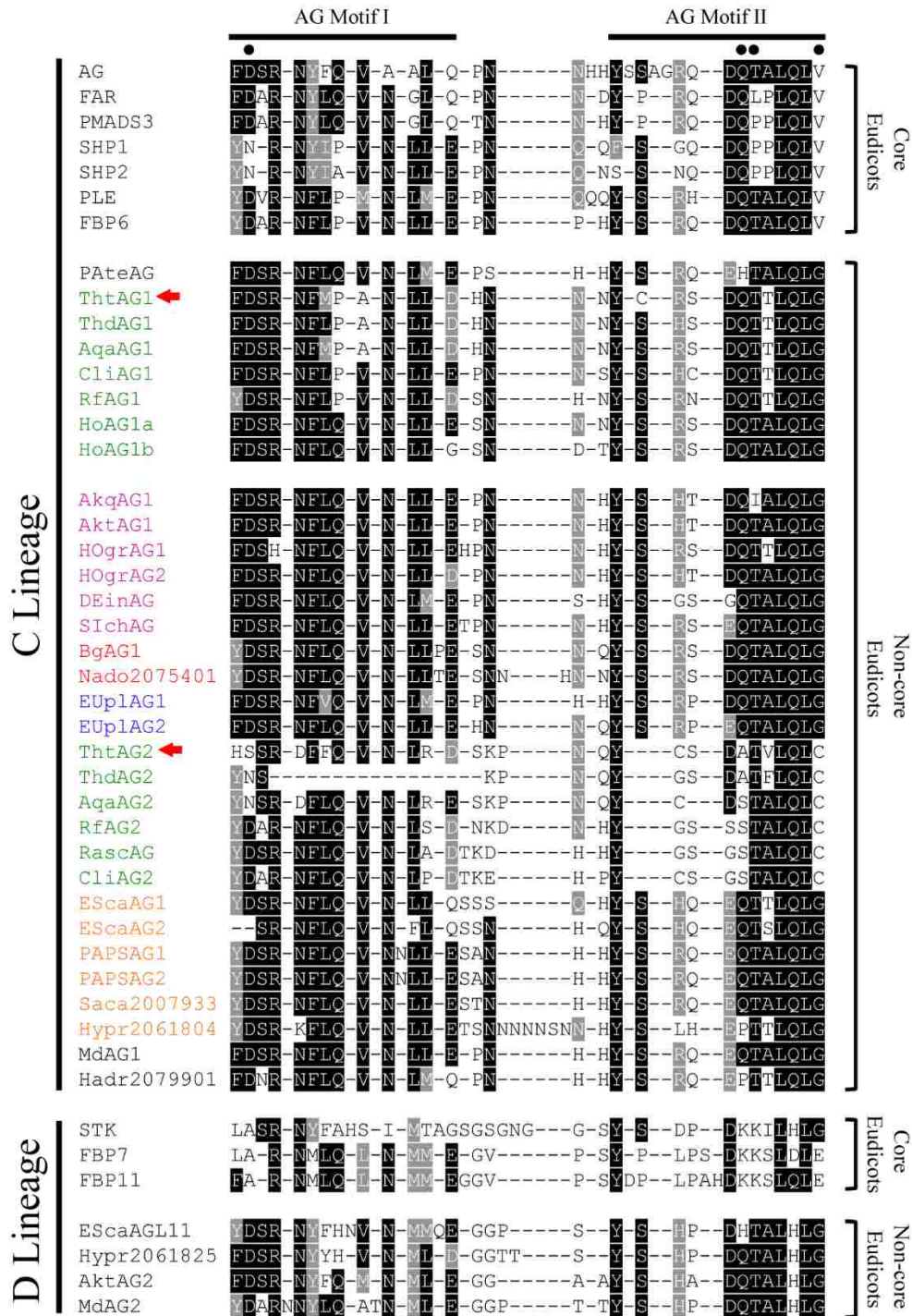
Figure 1
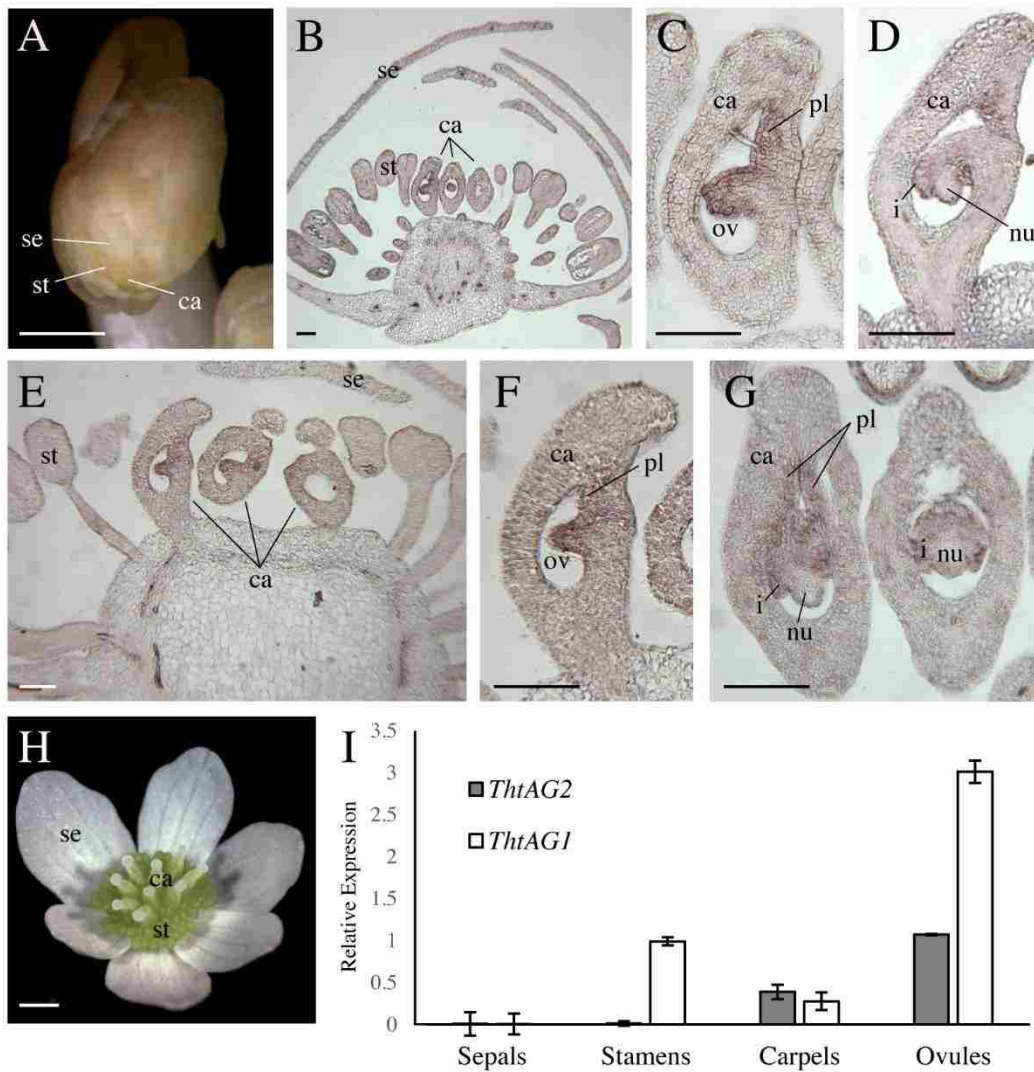
Figure 2
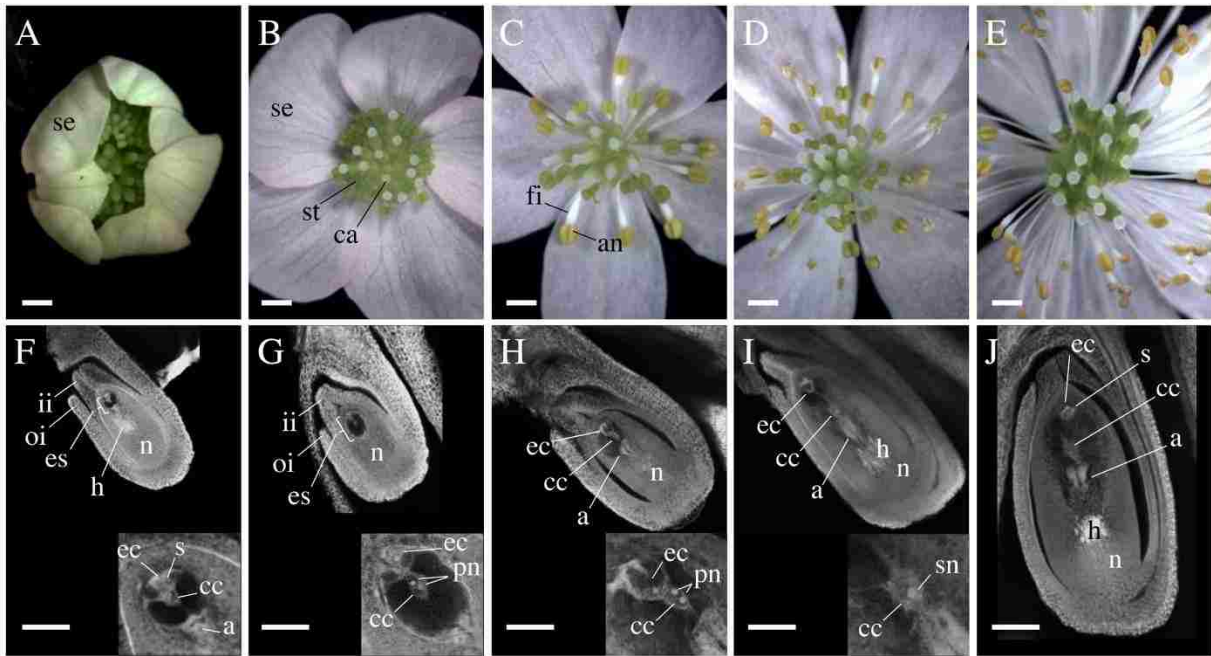
Figure 3
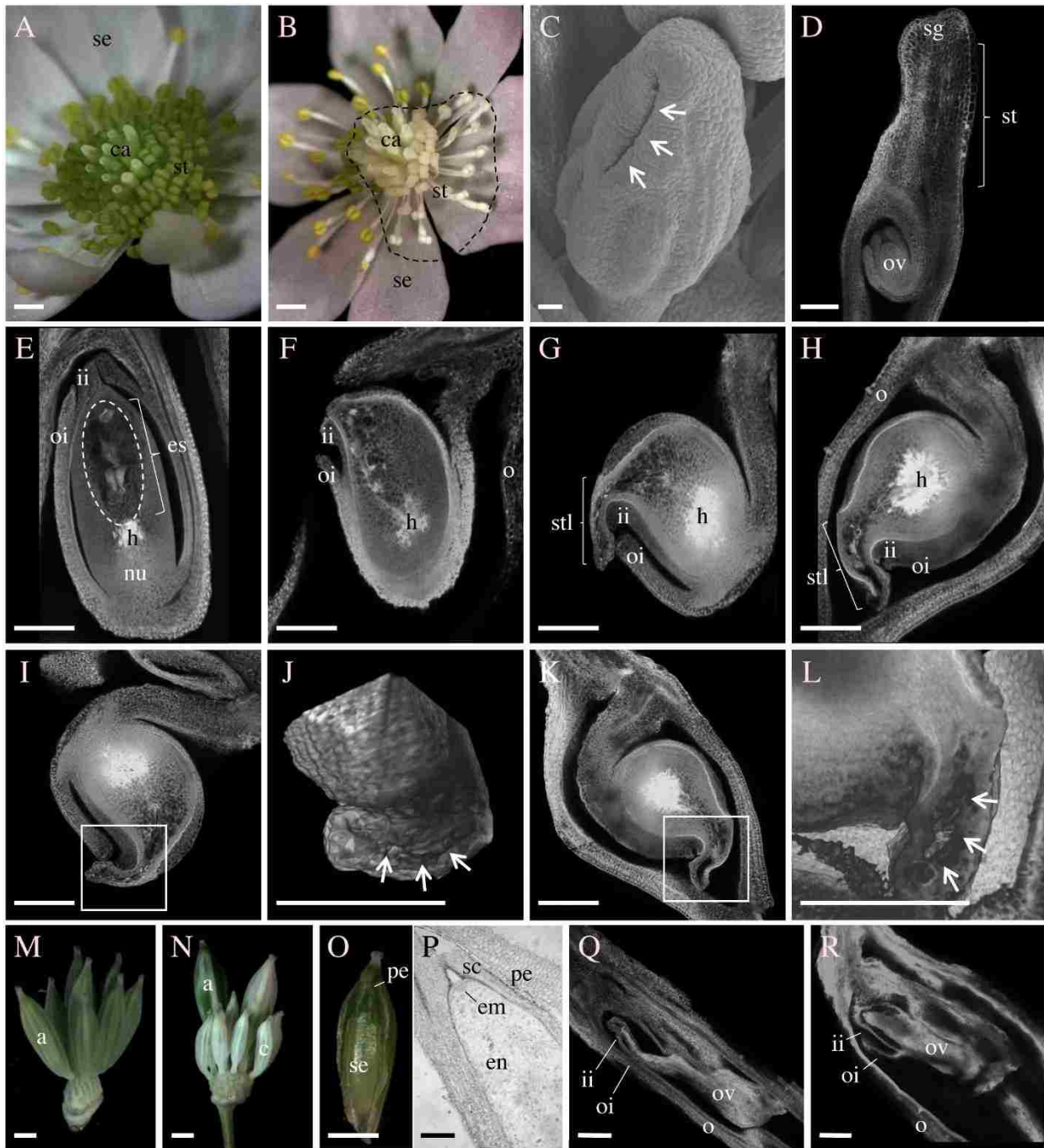
Figure 4

Figure 5

Figure 6

Figure 7

AD \ BD	ThtAG1	ThtAG2	ThtSEP3ΔC
ThtAG1			
ThtAG2			
ThtSEP3ΔC			

Figure 8
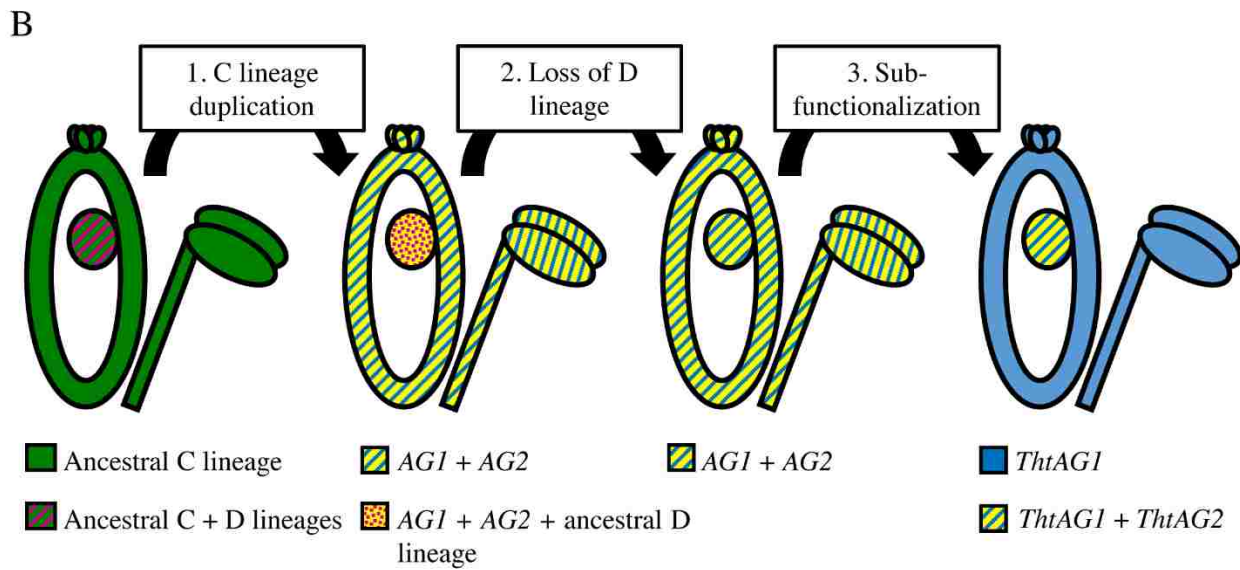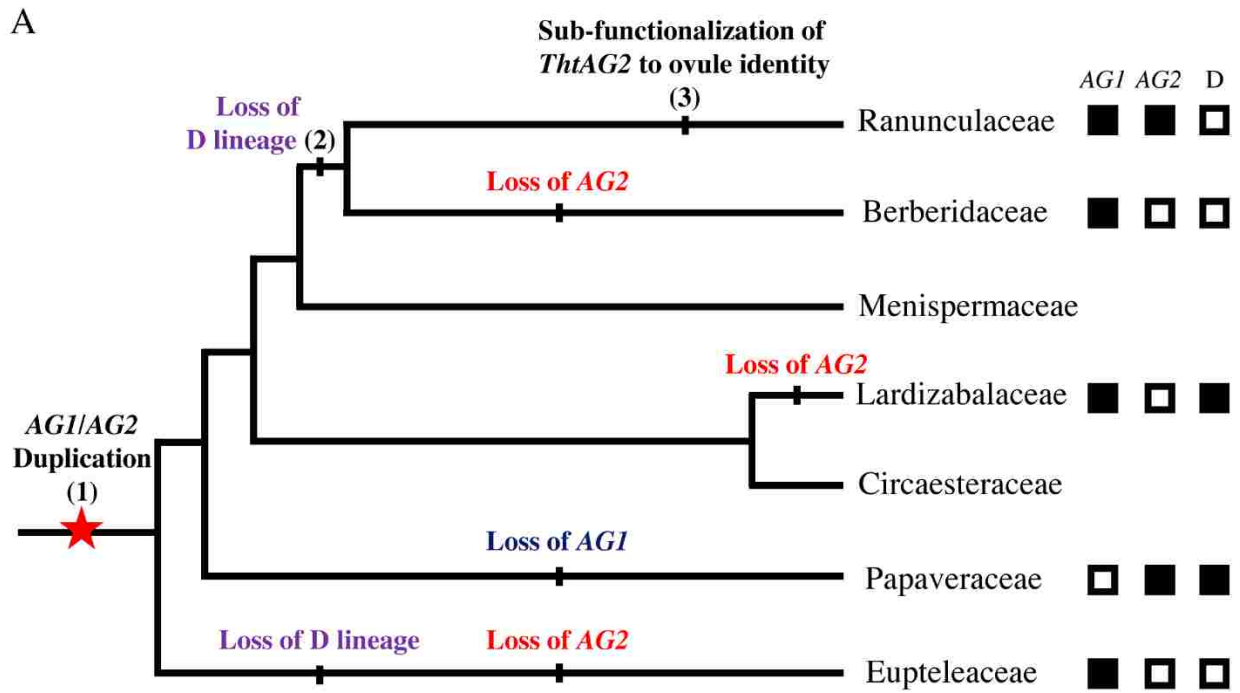
Table 1

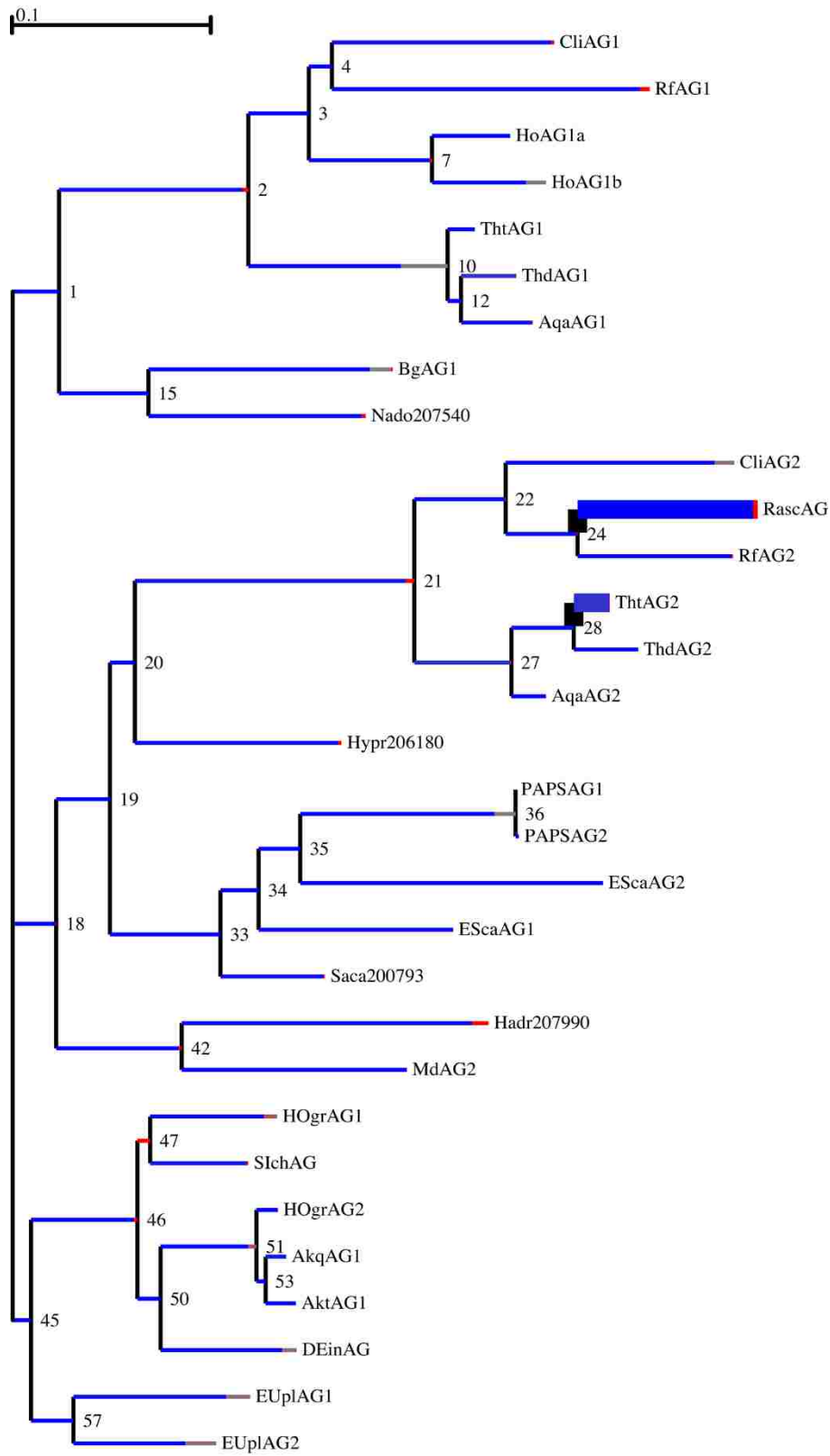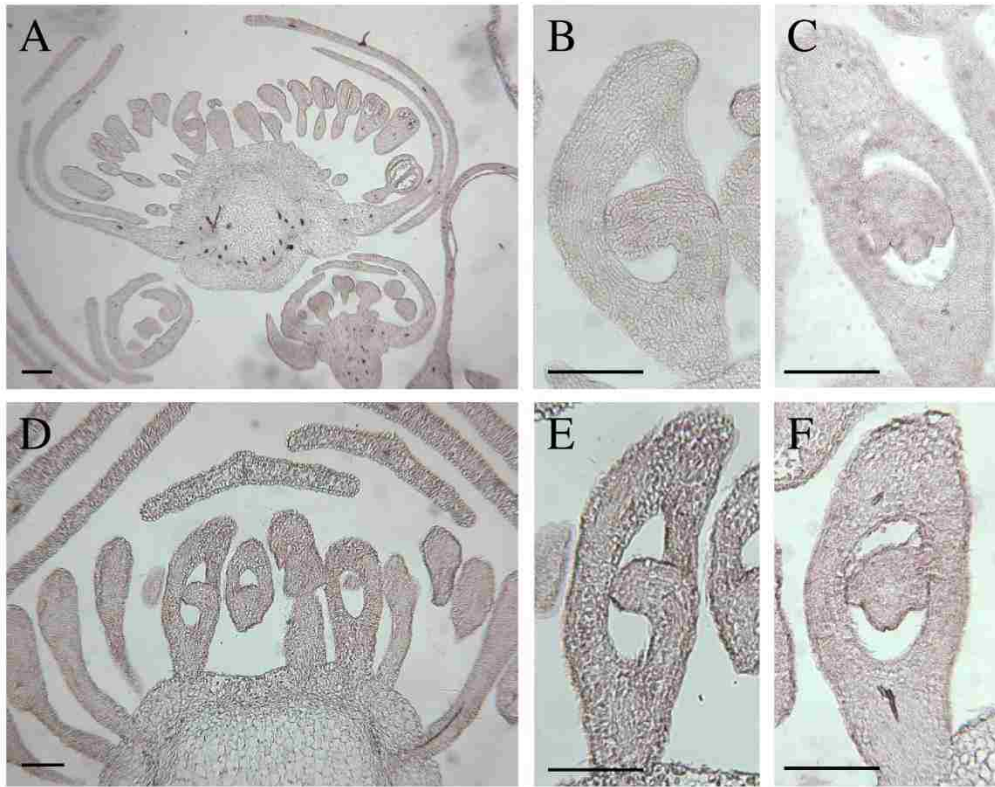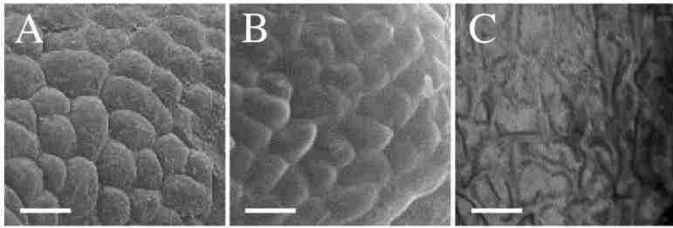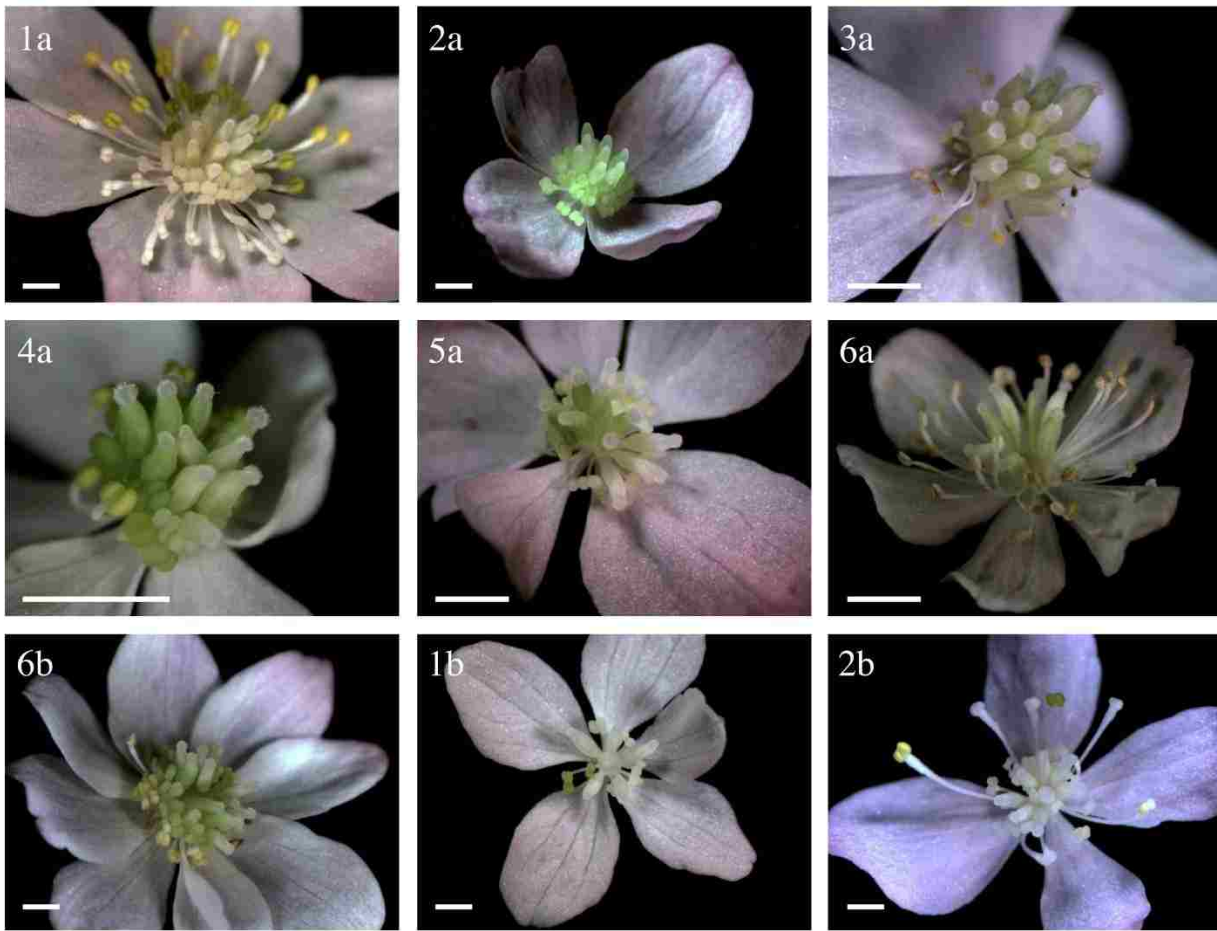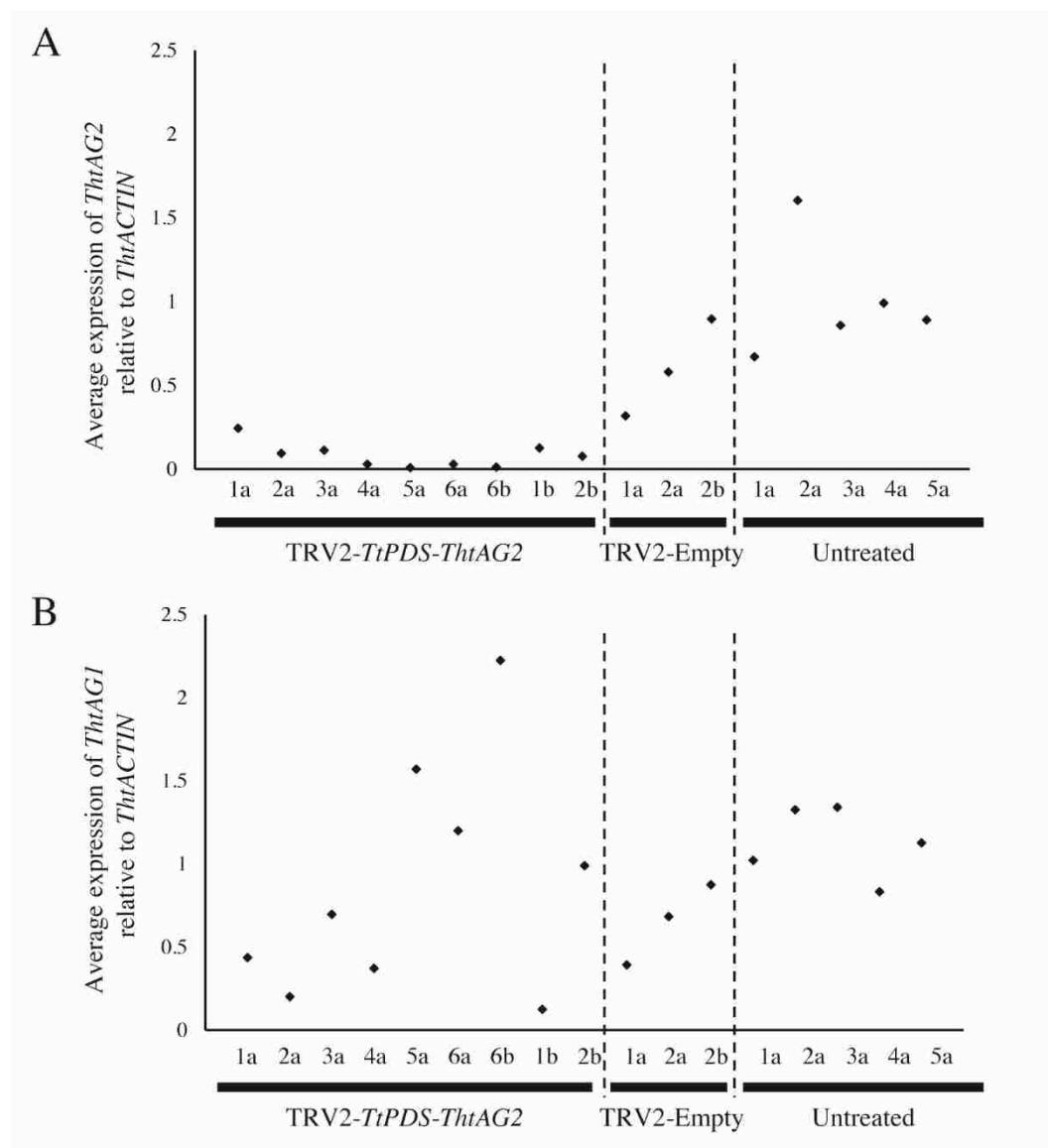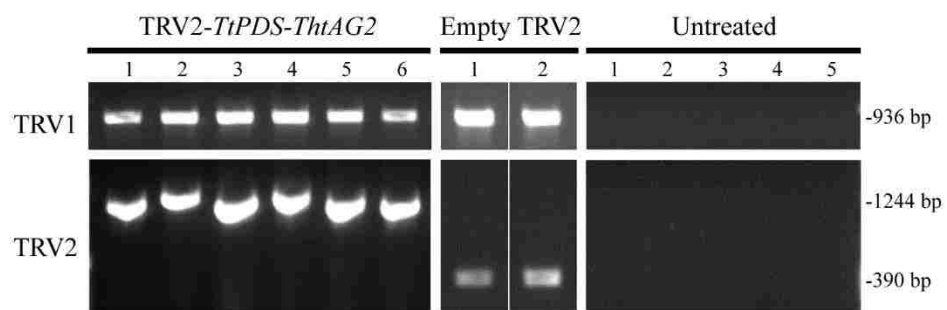| | Amino Acid | | | | |
Position	ThtAG1	ThtAG2	Domain (MIKC)	Feature	Probability of $\omega>1$
145	**L -** hydrophobic	**T -** polar	K	Helix 2	0.859
197	**F -** hydrophobic	**H -** polar	C	-	0.867
198	**D -** charged	**S -** polar	C	AG Motif 1	0.856
203	**M -** polar	**F -** hydrophobic	C	AG Motif 1	0.607
211	**N -** polar	**K -** charged	C	-	0.998**

Figure S1

Figure S2

Figure S3

Figure S4

Figure S5

Figure S6

Table S1

Species	Family	Locus	GenBank Accession	Reference
Akebia trifoliata	Lardizabalaceae	AktAG1	AY627635	Liu et al. 2010
		AktAG2	AY627629	Liu et al. 2010
Eupetelia pleiosperma	Eupteleaceae	EUplAG1	GU357452	Liu et al. 2010
		EUplAG2	GU357453	Liu et al. 2010
Pachysandra terminalis	Buxaceae	PAteAG	GU357456	Liu et al. 2010
Sinofranchetia chinensis	Lardizabalaceae	SIchAG	JQ806398	Hu et al. 2012
Holboellia grandiflora	Lardizabalaceae	HOgrAG1	JQ806406	Hu et al. 2012
		HOgrAG2	JQ806407	Hu et al. 2012
Decaisnea insignis	Lardizabalaceae	DEinAG	JQ806403	Hu et al. 2012
Ranunculus scleratus	Ranunculaceae	RascAG	AB473879	Unpublished
Thalictrum thalictroides	Ranunculaceae	ThtAG1	JN887118	Galimba et al. 2012
		ThtAG2	KM656465	This Paper
Thalictrum hernandezeii	Ranunculaceae	ThhAG2		This Paper
Papaver somniferum	Papaveraceae	PapsAG-1	GU123602	Hands et al. 2011
		PapsAG-2	GU123603	Hands et al. 2011
Anemone hupenhensis	Ranunculaceae	Ahu2138307		1KP Project
		Ahu2138876		
Hypecoum procumbens	Papaveraceae	Hpr2061804		1KP Project
		Hpr2061825		
		Hpr2061825		
Sanguinaria canadensis	Papaveraceae	Sca2007933		1KP Project
		Sca2009840		
Nandina domestica	Berberidaceae	Ndo2075401		1KP Project
Hakea drupaceae	Proteaceae	Hdr2079901		1KP Project

Table S2

Primer Purpose	Name	Sequence
ThtAG2 cloning	TdAG2_full_F	5' CGA AGA AGA CTA CTC TCC CAC AAC 3'
	TdAG2_full_R	5' TGA GCC TCC TAT TAA CTC TGC AAT C 3'
5' RACE of *ThdAG2*	GSP1	5' TAA CAC AGT TGG AGG ACA G 3'
	GSP2	5' TTG CGT CTG AGC AGT ATT GGT TCG G 3'
VIGS cloning	TthAG2_fwd_Xbal	5' GCT CTA GAA GAG AAC AGA TAA CTG CAA AC 3'
	TthAG2_rev_BamH1	5' ATG GAT CCC ACA TTA TCT GTT AAG CCA TAT 3'
Molecular validation by RT-PCR	TthAG2 RT for	5' TAC TCT GCG AAG CCG AGA TTG C 3'
	TthAG2 RT rev	5' GCA ATC TCG GCT TCG CAG AGT A 3'
Presence of TRV1	pTRV1_fwd	5' CTT GAA GAA GAA GAC TTT CGA GT CTC 3'
	pTRV1_rev	5' GTA AAA TCA TTG ATA ACA ACA CAG ACA AAC 3'
Presence of TRV2	OYL195	5' GGT CAA GGT ACG TAG TAG AG 3'
	OYL198	5' CGA GAA TGT CAA TCT CGT AGG 3'
Y2H cloning	ThtAG1 Primers	5' TAG CGC GAA TT CAT GGG AAG GGG AAA GAT TGA AA 3'
		5' AAT CGC GGA TCC CAA AAG CTT TCA CAG AAT ATC ACC 3'
	ThtAG1 Primers	5' AGT GTA GAA TTC AGC CAT GGG AAG GGG AAA GAT TGA A 3'
		5' ACT GCA GTC GAC CAC AGA ATA TCA CCC AAG TTG 3'
	ThtAG2 Primers	5' CAC GCG GAA TTC ATG GGA AGA GGA AAG ATT GAA ATC 3'
		5' CGT TAG GGA TCC CTA AGT AGT TAT TAG GCA GT TCT AC 3'
	ThtSEP3 Primers	5' CCT AGT GAA TTC ATG GGA AGA GGA AGA GTT GAG T 3'
		5' AAG CTT GGA TCC GTC GAC TTA ATT TGG TTG GCT ACC TTC TCC 3'
Expression by qPCR	TthAG1-fwD qPCR	5' AGT CTC TCA GCA ATC TCA ATA TCA GGG 3'
	TthAG1-rev-qPCR	5' GCC CTG AGA TAC TTG TTA TCA GTC TGC 3'
	TthAG2 qPCR for	5' TAT CTC AGA GAA CAG ATA ACT GCA AAC 3'
	TthAG2 qPCR rev	5' AGT AAG TTA ACA CAG TTG GAG GAC AGT 3'
In Situs	TthAG1ISFor2	5' AAT TAA CCC TCA CTA AAG GGA GCA GAA GAG GGA AAT AGA CTT 3'
	TthAG1ISRev3	5' TAA TAC GAC TCA CTA TAG GGT TAT CTA ACC AGA CAA AAT GCC 3'
	TthAG2ISFor2	5' AAT TAA CCC TCA CTA AAG GGA TTG GCC GAG ATC GAG TAT ATG 3'
	TthAG2ISRev3	5' TAA TAC GAC TCA CTA TAG GGT TTG ATA GAA AGT CTT ACT TCA 3'

Gene duplication and neo-functionalization in the *APETALA3* lineage of floral organ identity genes

Kelsey D. Galimba, Jesús Martínez-Gómez, Verónica S. Di Stilio

Introduction

The ABCE model of flower development predicts how combinations of classes of transcription factors determine floral organ identity. A class genes specify sepals; A and B class, petals; B and C class, stamens; and C class, carpels and floral meristem determinacy (1–3). The E class was discovered later as a group of proteins that interact with the ABC proteins in all floral organs to form the DNA binding tetramers necessary for full functionality (4–6). Duplications in these floral organ identity genes have been proposed as drivers for the speciation of the angiosperms (7)

In the model angiosperm *Arabidopsis thaliana,* the B class consists of two genes, *APETALA3* (*AP3*) and *PISTILLATA (PI).* Both are necessary, in combination with the E class genes *SEPALLATA1-4 (SEP1-4*), for petal and stamen identity in the second and third whorls (6,8). When B class function is lost, stamen and petal primordia are homeotically converted into carpels and sepals, respectively. This function is highly conserved throughout the angiosperm phylogeny, occurring in core and non-core eudicots as well as monocots (9–11). Even so, there are numerous examples of B class genes being expressed in other plant organs (e.g. root nodules), suggesting that they may have adopted novel roles in a number of lineages (11).

After an ancient duplication leading to the *AP3* and *PI* lineages, and before the diversification of the order Ranunculales, two duplication events led to three paralogous lineages of *AP3*: *AP3-I*, *AP3-II* and *AP3-III* (Fig. 1A). *Thalictrum thalictroides* (Ranunculaceae) is a member of the Ranunculid lineage that is sister to all other eudicots (12,13). This species has one *PI* (*ThtPI)* and three *AP3* orthologs (*ThtAP3-1*, *ThtAP3-2a* and *ThtAP3-2b)* (14,15). *ThtAP3-1* belongs to the *AP3-I* clade, while *ThtAP3-2a* and *ThtAP3-2b* belong to the *AP3-II* clade and are products of a more recent duplication that likely occurred in the common ancestor of *Thalictrum* and *Aquilegia* (16). *ThtAP3-2a* has a premature stop codon, causing a truncation of the protein affecting conserved C terminal motifs: half of the "PI motif-derived" and all of the "PaleoAP3" motif are missing in this gene (Fig. 1B) (15). The three proteins share a 45% identity, while ThtAP3-2a and ThtAP3-2b share a 76% identity, at the amino acid level. No *AP3-III* paralog has been identified so far in the genus *Thalictrum*, either by PCR (15,17), or, more recently, by our BLAST search of available transcriptomes of *T. thalictroides* and *T. hernandezii* (1KP and BGI, 2015) or a draft genome of *T. thalictroides* (BGI, 2015). Functional studies of the *Aquilegia AP3-III* ortholog show sub-functionalization of this gene to petal identity (16), and this function appears to be conserved throughout the order (16,17). The lack of an *AP3-III* ortholog in *Thalictrum* correlates with the absence of petals in this genus.

Although *T. thalictroides* lacks petals from the onset of floral meristem initiation, its sepals have a number of features typically associated with petals; they are white or pink (instead of photosynthetic and green), relatively large (longer than stamens and broad), and they contain papillate cells in the upper epidermis (18). In addition to having morphological traits typically associated with petals, *T. thalictroides* sepals express all B class genes (15). Down-regulation of *ThtPI* results in the complete homeotic conversion of stamens into carpels, and does not change the morphology of sepals, except for causing a small amount of green streaking (19). Petals may be interpreted as adaptive structures that

attract biotic pollinators and this function has been transferred to other organ types multiple times during angiosperm evolution (20). In fact in the Ranunculaceae, petal loss has evolved independently multiple times (17). In a number of distantly-related species, including *Aristolochia* (21), ectopic expression of B class genes has been proposed as causative of ectopic petaloidy. In the ranunculid *Aquilegia*, the *AqAP3*s are not involved in sepal identity or papillate epidermal cells, although their role in other petaloid characteristics such as color cannot be discarded (16,22)

The goal of this study was to conduct a functional analysis of the three *AP3* orthologs in *T. thalictroides*, a ranunculid with representatives of the paleoAP3 lineage of B class genes. We aimed to determine the degree of redundancy vs. divergence amongst these paralogs, while comparing their role to that of euAP3 lineage B class genes. Finally, we wanted to investigate whether one or more of these loci are involved in ectopic petaloidy of sepals of the insect pollinated species *T. thalictroides*. We addressed these questions by characterizing gene function through targeted gene silencing, and by testing protein-protein interactions by yeast-two hybrid analyses. Our results illustrate partial redundancy of the *ThtAP3s*, with deep conservation of B class function in stamen identity, as well as a novel role in ectopic petaloidy.

Materials and Methods

Plant Materials

T. thalictroides bare root plants were purchased from nurseries and grown in the University of Washington (UW) Greenhouse under ambient conditions from mid-February through mid-May. A voucher specimen for this species had been previously deposited at the UW Herbarium (WTU 376542).

Viral Induced Gene Silencing (VIGS)

Two constructs of different lengths were used to silence each gene individually; longer constructs were used due to evidence from a similar, concurrent experiment that short constructs were not sufficiently down-regulating gene expression. The shorter TRV2-*ThtAP3-1* construct was prepared using a 342 bp fragment comprising the C terminal region (195 bp) and 3'UTR (147 bp) of *ThtAP3-1*, while the longer *ThtAP3-1* construct consisted of a 429 bp fragment, comprising an extra 87 bp of the C terminal region. The short TRV2-*ThtAP3-2a* construct consisted of a 216 bp fragment comprising the C terminal region (16 bp) and 3'UTR (200 bp) of *ThtAP3-2a,* while a longer *ThtAP3-2a* construct (427 bp) included an extra 211 bp of the C terminal region. The short TRV2-*ThtAP3-2b* (224 bp) comprised the C terminal region (69 bp) and 3'UTR (155 bp), while the longer version (408 bp) contained an extra 184 bp of the C terminal region. No regions selected for silencing had more than thirteen continuous homologous base pairs in common with other genes. All *ThtAP3* fragments were amplified by PCR from a representative clone using primers with added BamHI and KpnI restriction sites (Table S1), followed by digestion, sticky end ligation to linearized TRV2 and transformed into *Agrobacterium tumefaciens* strain GV3101. A double construct consisting of the concatenated fragments of the short versions of *ThtAP3-2a* and *ThtAP3-2b* and a triple construct consisting of short version for all three *ThtAP3*s were similarly prepared.

A total of sixty *T. thalictroides* tubers that had been kept in soil in the dark at 4°C for 8 weeks, were treated with each construct as described previously (23,24). Briefly, a small incision was cut in the tubers near the bud using a clean razor blade, they were then submerged in infiltration medium containing appropriate *Agrobacterium* cultures (TRV1 and one of the TRV2 constructs), and infiltrated in a chamber under full vacuum (−100 kPa) for 10 min. Fifty plants were infiltrated with empty TRV2 vector (mock-treated control, to detect background viral effects). All plants were transferred to the

greenhouse following infiltration and were grown together with forty untreated plants under equal conditions. Leaves and flowers arose approximately two weeks later. Flowers from all treatments were observed under a Nikon SMZ800 dissecting scope, twenty seven flowers were photographed using a Q-Imaging MicroPublisher 3.3 digital camera or a Canon PowerShot SD890 IS digital camera and flash-frozen for RNA processing. Three mock-treated flowers and six untreated flowers were similarly processed as controls.

Molecular Validation of VIGS Lines

Young open flowers were processed to determine presence of TRV1 and TRV2 constructs, and expression levels of B class genes. Total RNA was extracted using TRIzol (Invitrogen, Life Technologies, CA), following manufacturer instructions and subsequently DNased using amplification grade DNase (Invitrogen, Life Technologies, CA). First strand cDNA synthesis was carried out on 1 μg of total mRNA using iScript (Bio-Rad, CA), following manufacturer instructions, and diluted 2.5-fold for use in quantitative RT-PCR (qPCR).

To test for the presence of the TRV1 and TRV2 viral RNA in treated plants, RT-PCR was carried out with GoTaq (Promega, WI) on 1μl of cDNA using TRV1- and TRV2-specific primers (Table S1) for 30 cycles at 51°C annealing temperature. Locus-specific primers, designed to amplify regions distinct from those used in the VIGS constructs, were used to test expression levels of *ThtAP3-1*, *ThtAP3-2a, ThtAP3-2b* and *ThtPI* by qPCR (Table S1). Each qPCR reaction contained 15μl iQ SYBR Green Supermix (Bio-Rad, CA), 12.2 μl H_2O, 0.9 μl each of forward and reverse primer and 1μl of cDNA. Samples were amplified on an MJ Research Chromo4 Detector for 45 cycles, in triplicate, including a no-template control. Cycling conditions were: 94ºC for 10 minutes and 45 cycles of 94ºC for 30 seconds, 53ºC for 30 seconds and 72ºC for 30 seconds. Relative expression was calculated using the

$2^{-\Delta\Delta Ct}$ method (25), and normalized against *ThtACTIN* and *ThtEF1α (Elongation Factor 1α)*. Since B class gene expression levels between untreated (n=6) and TRV2-treated flowers (n=3) were not significantly different in an unpaired t test (p>0.05), both controls were combined for further analyses. Statistical significance of differences in gene expression was determined using one-way ANOVA among the four treatment groups (controls (n= 9) and VIGS of *ThtAP3-1* (n=8), *ThtAP3-2a* (n=10) and *ThtAP3-2b* (n=6)), followed by post-hoc Tukey Honest Significant Different (HSD) contrasts in JMP (SAS, NC).

Yeast Two-Hybrid Assays

cDNA sequences encoding proteins for *ThtAP3-1*, *ThtAP3-2a. ThtAP3-2b* were cloned into pGADT7 and pGBKT7 vectors using the In-Fusion HD cloning kit (Clontech, CA) and custom primes (Table S1). Plasmid pGADT7 and pGBKT7 containing cDNA sequence for *ThtSEP3ΔC* was previously generated and kindly provided for this study by Rainer Melzer and Günter Theißen (23). *ThtSEP3 ΔC* is a C terminal-deleted version of *ThtSEP3* lacking the last 195 bp of the coding sequence, used previously to avoid auto activation (23,26).

Yeast two-hybrid assays were carried out using Matchmaker Gold Yeast Two-Hybrid System (Clontech, CA). Cells were co-transformed and plated on Leu/Trp-free media to select for diploid colonies. Single colonies were serially diluted 10-fold to 1:10,000 in water and 5 μL of each dilution were plated on Leu/Trp/His-free media and Leu/Trp/His-free media supplemented with Aurebasidin A (AbA) to test for protein interaction, and on Leu/Trp-free media to control for yeast growth. Interactions were scored and photographed using a Canon PowerShot SD890 IS digital camera after 4 days of incubation at 28°C.

Results

Quantification of *ThtAP3* gene expression in untreated plants

In order to determine expression levels of B class genes in wild-type flowers at the stage when VIGS-treated flowers are collected, we performed qPCR on six young open flowers from three plants (Fig. 3). *ThtAP3-1* and *ThtAP3-2a* expression levels were statistically similar, and *ThtAP3-2b* expression was significantly higher than the other two ($p < 0.01$). *ThtPI* was expressed at the highest level of all B class genes ($p \leq 0.001$).

Targeted gene silencing of *AP3* orthologs in *T. thalictroides*

***ThtAP3-1* VIGS**

In flowers treated with the *ThtAP3-1* constructs (n=8) a significant down-regulation of *ThtAP3-1* was observed. On average, treated flowers exhibited a 3.8-fold decrease in *ThtAP3-1* expression ($p = 0.02$) (Fig. 5 A). Expression levels of other B class genes in *ThtAP3-1*-treated flowers were statistically similar to controls (Fig 5 B - D). Flowers presented a variety of abnormal phenotypes, including small extra sepals in place of outer stamens (Fig. 2 B), narrow sepals (Fig. 2 C), small sepals with patches of green (photosynthetic) tissue (Fig. 2 F, H, I), and chimeric organs in place of outer stamens (Fig. 2 D - G) that were sepal-like, with a stigma or an anther sac on the distal tip (Fig. 2 E, G). In summary, the down-regulation of *ThtAP3-1* altered petaloidy of the sepals, reducing their overall size, making them narrow and causing portions to become photosynthetic. Down-regulation of *ThtAP3-1* also affected stamen identity, causing the development of either small sepals or chimeric organs in place of outer stamens.

114

ThtAP3-2a VIGS

Neither short nor long *ThtAP3-2a* constructs (n=10) succeeded in down-regulating *ThtAP3-2a* in any of the plants that were verified for the presence of TRV1 and TRV2 virus. Flowers treated with the *ThtAP3-2a* constructs showed, on average, a 2.3-fold increase in *ThtAP3-2a* expression, which was not statistically different from controls ($p = 0.05$) (Fig. 5 B), and also had on average, statistically higher levels (1.6-fold) of *ThtAP3-1* expression than controls ($p = 0.03$) (Fig. 5 A). Flowers treated with *ThtAP3-2a* constructs presented abnormal phenotypes consisting of chimeric organs similar to those seen in flowers treated with *ThtAP3-1* (Fig. 2 J, M), as well as lobed sepals (Fig. 2 J, K, L). In summary, treatment with *ThtAP3-2a* constructs did not succeed in down-regulating *ThtAP3-2a*, and resulted in unexpected up-regulation of *ThtAP3-1*.

ThtAP3-2b VIGS

Flowers treated with the longer *ThtAP3-2b* construct (n=6) showed, on average, 1.4-fold lower expression of *ThtAP3-2b* compared to controls ($p = 0.62$) (Fig. 5 C). All other B class gene expression levels were similar to controls (Fig. 5 A, B, D). These flowers showed small sepals in place of outer stamens (Fig. 2 N, O), green sections in sepals (Fig. 2 P, Q), lobed sepals (Fig. 2 P, Q), and chimeric organs (Fig. 2 R, S), similar to those seen in other treatments. *ThtAP3-2b* expression levels varied greatly among *ThtAP3-2B* treated flowers (Fig. 5 C), and flowers with green sections or chimeric organs (Fig. 2 P - S) had lower levels of *ThtAP3-2b* expression than those with extranumerary, small sepals (Fig. 2 N, O) or small amounts of sepal lobing. For example, flower 10.1 (Fig. 2 P) had 4.1-fold and flower 7.1 (Fig. 2 R) had 5.4-fold lower expression of *ThtAP3-2b* than controls (Fig. 4 H, I), while

flower 3.2 (Fig. 2 N) had the same amount of *ThtAP3-2b* expression (Fig. 4 G). Overall, when all

phenotypes were considered together and averaged, it appears that we did not succeed in down-

regulating *ThtAP3-2b* to a significant level. However, upon closer inspection, individual flowers with

defects in stamen identity (chimeric organs) or petaloidy of sepals (lobing or green sections) exhibited

greater down-regulation of *ThtAP3-2b* than flowers with extra sepals or slight lobing. In summary,

inspection of individual flowers reveals a correlation between the intensity of the phenotypes observed

and the degree of down-regulation of *ThtAP3-2b* achieved by VIGS.

*ThtAP3*s interact in novel ways at the protein level

In order to test the interactions of the B class and E class proteins, we performed yeast two-

hybrid (Fig. 6). Our analyses revealed that both ThtAP3-1 and ThtAP3-2a could dimerize with the

typical B class partner ThtPI. All three ThtAP3 proteins were also able to dimerize with each other:

ThtAP3-1 interacted with ThtAP3-2a and ThtAP3-2b, ThtAP3-2ba interacted with ThtAP3-2b. ThtAP3-

1 alone was able to homodimerize, and all four B class genes, ThtAP3-1, ThtAP3-2a, ThtAP3-2b and

ThtPI, were able to interact with the E class gene, ThtSEP3. Control transformations with empty vectors

did not produce growth, ruling out autoactivation in any of the B or E class proteins analyzed .

Discussion

This study aimed to determine the function of three *AP3* orthologs in *T. thalictroides*, asking

whether they have remained redundant or have diverged in function and whether they are responsible for

ectopic petaloidy of sepals. We addressed these questions by characterizing gene function through

targeted gene silencing, and by testing protein-protein interactions by yeast-two hybrid analysis. We

achieved robust down-regulation of one paralog, *ThtAP3-1* and intermittent down-regulation of *ThtAP3-2b,* with a clear correlation to phenotypes affecting stamen and sepal development. In addition, all of the *ThtAP3*s were shown to interact with each other and with the E class gene *ThtSEP3* at the protein level, providing a potential mechanism for *ThtAP3* to function in novel ways. For these reasons, we conclude that both *ThtAP3-1* and *ThtAP3-2b* function in stamen identity, and have also adopted a new role in ectopic petaloidy of sepals. Finally, we propose that the truncated gene *ThtAP3-2a* acts as a dominant negative, based on the combined results from VIGS (where other *ThtAP3*s were upregulated) and the positive interactions with the protein products of other B class genes.

Due to data gathered from a similar, concurrent experiment, we had reason to believe that our shorter *ThtAP3* constructs (216 - 342 bp) were not sufficiently down-regulating our target genes. For this reason, we designed longer constructs and repeated the VIGS experiment. We found that indeed, longer constructs (408 - 429 bp) down-regulated the targeted genes more efficiently both for *ThtAP3-1* and *ThtAP3-2b*. This was unexpected, given that a homologous length of 23 bp has been reported to be sufficient for silencing in other systems (27). Another unexpected result was the up-regulation of all B class genes, including *ThtAP3-2a,* in plants treated with both short and long *ThtAP3-2a* constructs, that resulted in sepal lobing and stamen chimeras. This result suggests that fixed ratios of *ThtAP3* transcripts (and proteins) are necessary for the correct development of sepals and stamens(28). We hypothesize that ThtAP3-2a is actually non-functional, but is still important to stamen and sepal development because it acts in an inhibitory fashion, outcompeting the functional proteins ThtAP3-1 and ThtAP3-2b. Under this scenario, down-regulation of *ThtAP3-2a* at an early stage would cause an increased proportion the level of functional tetramers containing ThtAP3-1 or ThtAP3-2b, therefore promoting the expression of all four B class genes, since *AP3* and *PI* maintain their own expression in a positive feedback loop (29,30). Additionally, VIGS-treated flowers often consist of a mix of silenced and non-silenced regions due to

the chimeric nature of the process, and we frequently observe only one to a few organs per flower with abnormal morphology. It is likely that due to our sampling technique (processing the entire flower for RNA), we are not able to detect gene down-regulation that may be occurring only in a small portion of the flower. Flowers were collected at a young, open stage, which was the earliest developmental stage at which the observation of all organs was possible. This necessary delay in collection may also be a potential reason for higher than expected gene expression levels in flowers treated by VIGS. *AP3* orthologs are expressed very early in flower development, in stamen and petal primordia where they confer organ identity, but they are also responsible for later developmental events, such as petal expansion in a number of species (reviewed in 31).

Phenotypes resulting from VIGS could be organized into two broad categories: those affecting sepal morphology and those affecting stamen morphology. Abnormal sepal morphology occurred in flowers from all treatments. Down-regulation of *ThtAP3-1* was shown to cause reduction in size, narrowing and a greening of sepals, compared to wild type (Fig. 2 C - I). In TRV2-*ThtAP3-2a* treated flowers, which exhibited higher levels of all B class genes, we observed lobing of sepals (Fig 2. J, M). Down-regulation of *ThtAP3-2b* also affected sepal morphology and was shown to cause lobing and greening (Fig 2. P, Q). The second category of phenotypes affected stamen morphology, which again, was present in each treatment. In flowers down-regulated for *ThtAP3-1*, we observed stunted stamens and conversions of outer stamens to chimeric sepal-stamen or sepal-carpel like organs (Fig. 2 B, D-G). Down-regulation of *ThtAP3-2b* also affected stamen morphology, as indicated by the conversion of stamens to chimeric organs (Fig. 2 N-O, R-S). Flowers treated with TRV2-*ThtAP3-2a* had the unexpected up-regulation of the targeted gene and also showed conversion of stamens to chimeric organs, which may be explained by the timing of VIGS, RNA extraction techniques or protein ratios necessary for development (see above). The presence of different phenotypes between flowers treated

with the same construct (e.g. narrow sepals vs. small, greenish sepals in *ThtAP3-1*-treated flowers) could potentially be a result of different levels of development reached before silencing occurred. This phenomenon has been previously reported in *T. thalictroides*, as this species pre-forms flowers prior to emerging from the soil (23).

Taken together, these phenotypes indicate that the *ThtAP3* genes are involved in both stamen identity and the ectopic petaloidy of sepals. However, down-regulation of the *ThtAP3* paralogs individually does not result in complete homeotic conversion of stamens into carpels or a complete loss of petaloidy. This is unlike *ThtPI,* which causes the complete loss of stamen identity and a full conversion of stamens into carpels when down-regulated (19). This apparent incomplete conversion could be attributed to partial redundancy among the three genes, or possibly only among *ThtAP3-1* and *ThtAP3-2b*, which is supported by the presence of similar phenotypes in all treatments. We therefore suggest that these paralogs together exhibit a conserved, ancestral role in stamen development, but have also taken on a new role in sepal petaloidy. This is unlike the closely related species, *Aquilegia*, in which three *AqAP3* paralogs have clearly sub- and neo-functionalized. *Aquilegia* has three *AP3* paralogs, *AqAP3-1, AqAP3-2* and *AqAP3-3* in the I, II and III clades, respectively (Fig. 1A). *AqAP3-3* is responsible for petal identity (16) and *AqAP3-2* is responsible for stamen identity, representing a sub-functionalization of the canonical B class functions (32). *AqAP3-1* is involved in the identity of the staminodia, a novel sterile organ type located between the second and third whorls and likely evolved from stamens (32). Further studies in *T. thalictroides* will concentrate on silencing the *ThtAP3*s with double (*ThtAP3-1+ThtAP3-2b*) and triple (*ThtAP3-1+ ThtAP3-2a+ThtAP3-2b*) constructs, to determine whether they are indeed redundant.

Our yeast two-hybrid assays indicate that the ThtAP3s form novel protein-protein interactions. Eu*AP3* and most paleo*AP3* proteins from core eudicots and monocots do not homodimerize and have

been shown to obligately interact with PI in order to function. (33–37). However, *AP3* and *PI* orthologs from early-diverging angiosperms both homo- and heterodimerize with each other, and are also able to weakly interact with other floral MADS-box genes (38). While the ability to heterodimerize has been maintained throughout angiosperm evolution, B class protein homodimerization is much less conserved. AP3 homodimerization was probably lost very early in angiosperm evolution (before the eudicot-monocot split), while PI homodimerization was likely lost later during evolution, but before the diversification of the eudicots (38). Protein products of *AP3* orthologs in *Aquilegia (AqvAP3-2 and AqvAP3-3)* have been shown to homodimerize, yet the three *AqvAP3* paralogs do not dimerize with each other (39). In *Thalictrum*, we have shown that ThtAP3-1 is able to homodimerize, and additionally, all three ThtAP3 proteins are able to interact with each other. All four B class gene products can also interact with ThtSEP3. Taken together, our results suggest that the ThtAP3s could potentially be forming protein tetramers that consist of only ThtAP3s and E class genes. This hypothetical tetramer would differ from the SEP3-SEP3/AP3-PI that has been shown to form preferentially over other B class/E class combinations in *Arabidopsis* (40). We therefore propose that these distinctive protein interactions may provide a mechanism for the novel functionality of the *ThtAP3s*.

Conclusions

Although gene duplication has long been recognized as an important contributor to the evolution of biological complexity (41,42), functional studies showing the fates of duplicated developmental genes are still limited. Here, we analyze the function of duplicated *AP3* orthologs and find that they show deep conservation of the stamen-identity role, and also have a novel role in ectopic petaloidy. We also find that they appear to have remained at least partially redundant in both of these functions. Analyses such

as this can ultimately help to determine the potential role of duplicated floral organ identity genes in angiosperm diversification.

References

1. Bowman JL, Smyth DR, Meyerowitz EM. Genetic interactions among floral homeotic genes of Arabidopsis. Development. 1991 May 1;112(1):1–20.

2. Coen ES, Meyerowitz EM. The war of the whorls: genetic interactions controlling flower development. Nature. 1991;353(6339):31–7.

3. Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H. Genetic Control of Flower Development by Homeotic Genes in Antirrhinum majus. Science. 1990 Nov 16;250(4983):931–6.

4. Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. The SEP4 Gene of Arabidopsis thaliana Functions in Floral Organ and Meristem Identity. Curr Biol. 2004 Nov 9;14(21):1935–40.

5. Honma T, Goto K. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature. 2001 Jan 25;409(6819):525–9.

6. Theißen G, Saedler H. Plant biology: Floral quartets. Nature. 2001 Jan 25;409(6819):469–71.

7. Airoldi CA, Davies B. Gene Duplication and the Evolution of Plant MADS-box Transcription Factors. J Genet Genomics. 2012 Apr 20;39(4):157–65.

8. Jetha K, Theißen G, Melzer R. Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. Nucleic Acids Res. 2014 Sep 29;42(17):10927–42.

9. Kim S, Yoo M-J, Albert VA, Farris JS, Soltis PS, Soltis DE. Phylogeny and diversification of B-function MADS-box genes in angiosperms: evolutionary and functional implications of a 260-million-year-old duplication. Am J Bot. 2004 Dec 1;91(12):2102–18.

10. Litt A, Kramer EM. The ABC model and the diversification of floral organ identity. Semin Cell Dev Biol. 2010 Feb;21(1):129–37.

11. Zahn LM, Leebens-Mack J, dePamphilis CW, Ma H, Theissen G. To B or Not to B a Flower: The Role of DEFICIENS and GLOBOSA Orthologs in the Evolution of the Angiosperms. J Hered. 2005 May 1;96(3):225–40.

12. Maia VH, Gitzendanner MA, Soltis PS, Wong Gane Ka-Shu, Soltis DE. Angiosperm Phylogeny Based on 18S/26S rDNA Sequence Data: Constructing a Large Data Set Using Next-Generation Sequence Data. Int J Plant Sci. 2014 Jul 1;175(6):613–50.

13. Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, et al. Angiosperm phylogeny: 17 genes, 640 taxa. Am J Bot. 2011 Apr 1;98(4):704–30.

14. Kramer EM, Stilio Verónica S. Di, Schlüter PM. Complex Patterns of Gene Duplication in the APETALA3 and PISTILLATA Lineages of the Ranunculaceae. Int J Plant Sci. 2003 Jan 1;164(1):1–11.

15. Di Stilio VS, Kramer EM, Baum DA. Floral MADS box genes and homeotic gender dimorphism in Thalictrum dioicum (Ranunculaceae) – a new model for the study of dioecy. Plant J. 2005 Mar 1;41(5):755–66.

16. Sharma B, Guo C, Kong H, Kramer EM. Petal-specific subfunctionalization of an APETALA3 paralog in the Ranunculales and its implications for petal evolution. New Phytol. 2011 Aug 1;191(3):870–83.

17. Zhang R, Guo C, Zhang W, Wang P, Li L, Duan X, et al. Disruption of the petal identity gene APETALA3-3 is highly correlated with loss of petals within the buttercup family (Ranunculaceae). Proc Natl Acad Sci. 2013 Mar 26;110(13):5074–9.

18. Di Stilio VS, Martin C, Schulfer AF, Connelly CF. An ortholog of MIXTA-like2 controls epidermal cell shape in flowers of Thalictrum. New Phytol. 2009 Aug 1;183(3):718–28.

19. LaRue NC, Sullivan AM, Di Stilio VS. Functional recapitulation of transitions in sexual systems by homeosis during the evolution of dioecy in Thalictrum. Plant Evol Dev. 2013;4:487.

20. Cronk QCB, Bateman RM, Hawkins JA. Developmental Genetics and Plant Evolution. CRC Press; 2002. 568 p.

21. Jaramillo MA, Kramer EM. APETALA3 and PISTILLATA homologs exhibit novel expression patterns in the unique perianth of Aristolochia (Aristolochiaceae). Evol Dev. 2004 Dec;6(6):449–58.

22. Jaramillo MA, Kramer EM. Molecular evolution of the petal and stamen identity genes, APETALA3 and PISTILLATA, after petal loss in the Piperales. Mol Phylogenet Evol. 2007 Aug;44(2):598–609.

23. Galimba KD, Tolkin TR, Sullivan AM, Melzer R, Theißen G, Di Stilio VS. Loss of deeply conserved C-class floral homeotic gene function and C- and E-class protein interaction in a double-flowered ranunculid mutant. Proc Natl Acad Sci U S A. 2012 Aug 21;109(34):E2267–75.

24. Di Stilio VS, Kumar RA, Oddone AM, Tolkin TR, Salles P, McCarty K. Virus-induced gene silencing as a tool for comparative functional studies in Thalictrum. PloS One. 2010;5(8):e12064.

25. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods. 2001 Dec;25(4):402–8.

26. Causier B, Davies B. Analysing protein-protein interactions with the yeast two-hybrid system. Plant Mol Biol. 2002 Dec 1;50(6):855–70.

27. Thomas CL, Jones L, Baulcombe DC, Maule AJ. Size constraints for targeting post-transcriptional gene silencing and for RNA-directed methylation in Nicotiana benthamiana using a potato virus X vector. Plant J. 2001 Feb 1;25(4):417–25.

28. Geuten K, Viaene T, Irish VF. Robustness and evolvability in the B-system of flower development. Ann Bot. 2011 Jun;107(9):1545–56.

29. Goto K, Meyerowitz EM. Function and regulation of the Arabidopsis floral homeotic gene PISTILLATA. Genes Dev. 1994 Jul 1;8(13):1548–60.

30. Jack T, Brockman LL, Meyerowitz EM. The homeotic gene APETALA3 of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens. Cell. 1992 Feb 21;68(4):683–97.

31. Dornelas MC, Patreze CM, Angenent GC, Immink RGH. MADS: the missing link between identity and growth? Trends Plant Sci. 2011 Feb;16(2):89–97.

32. Sharma B, Kramer E. Sub- and neo-functionalization of APETALA3 paralogs have contributed to the evolution of novel floral organ identity in Aquilegia (columbine, Ranunculaceae). New Phytol. 2013 Feb 1;197(3):949–57.

33. Moon YH, Jung JY, Kang HG, An G. Identification of a rice APETALA3 homologue by yeast two-hybrid screening. Plant Mol Biol. 1999 May;40(1):167–77.

34. Riechmann JL, Krizek BA, Meyerowitz EM. Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. Proc Natl Acad Sci U S A. 1996 May 14;93(10):4793–8.

35. Schwarz-Sommer Z, Hue I, Huijser P, Flor PJ, Hansen R, Tetens F, et al. Characterization of the Antirrhinum floral homeotic MADS-box gene deficiens: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. EMBO J. 1992 Jan;11(1):251–63.

36. Tröbner W, Ramirez L, Motte P, Hue I, Huijser P, Lönnig WE, et al. GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of Antirrhinum floral organogenesis. EMBO J. 1992 Dec;11(13):4693–704.

37. Vandenbussche M, Zethof J, Royaert S, Weterings K, Gerats T. The Duplicated B-Class Heterodimer Model: Whorl-Specific Effects and Complex Genetic Interactions in Petunia hybrida Flower Development. Plant Cell. 2004 Mar 1;16(3):741–54.

38. Melzer R, Härter A, Rümpler F, Kim S, Soltis PS, Soltis DE, et al. DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. Ann Bot. 2014 Jun 5;mcu094.

39. Kramer EM, Holappa L, Gould B, Jaramillo MA, Setnikov D, Santiago PM. Elaboration of B Gene Function to Include the Identity of Novel Floral Organs in the Lower Eudicot Aquilegia. Plant Cell Online. 2007 Mar 1;19(3):750–66.

40. Melzer R, Theißen G. Reconstitution of "floral quartets" in vitro involving class B and class E floral homeotic proteins. Nucleic Acids Res. 2009 May;37(8):2723–36.

41. Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. Genetics. 1999 Apr 1;151(4):1531–45.

42. Ohno S. Evolution by gene duplication. Springer-Verlag; 1970. 184 p.

Figure Captions

Figure 1. Simplified B class gene phylogeny illustrating the relationship of *Thalictrum thalictroides* loci to the broader history of duplications within this gene lineage. A. An ancient duplication within angiosperms lead to the *AP3* and *PI* lineages (black star), the latter with one representative from *T. thalictroides* (*ThtPI*). In the core eudicots, a more recent duplication led to the eu*AP3* and *TM6* lineages (white star). In the paleo*AP3* lineage, two Ranunculales-wide duplication events (red stars) led to the three paralogous clades *AP3-I* (including *ThtAP3-1*)*, AP3-II* and *AP3-III*. A more recent wide duplication led to two *T. thalictroides* genes in the *AP3-II* clade: *ThtAP3-2a* and *ThtAP3-2b*. No *AP3-III* genes have been recovered from *Thalictrum. T. thalictroides* gene names are indicated with red text. Simplified phylogeny modified from (16,22). B. Simplified schematic of the three ThtAP3 protein products, showing MIKC regions. ThtAP3-2a is truncated by an early stop codon, which results in the loss of the PaleoAP3 motif and a portion of the PI motif (15).

Figure 2. Phenotypes resulting from treatment with *ThAP3-1, ThtAP3-2a* and *ThtAP3-2b* VIGS constructs. A. Untreated *T. thalictroides* flower. B - I. Flowers treated with *ThtAP3-1* constructs. B. Flower treated with short *ThtAP3-1* construct. Arrow indicates sepal in place of outer stamen. C. Flower treated with long *ThtAP3-1* construct, with narrow sepals. D. Flower treated with long *ThtAP3-1* construct, with chimeric organ in place of outer stamen boxed. E. Detail of sepal-stamen chimera in D, indicated with arrow. F. Flower treated with long *ThtAP3-1* construct, with small, curved sepals and chimeric organ in place of outer stamen boxed. G. Detail of sepal-carpel chimera in F, stigma indicated with arrow. H. Flower treated with long *ThtAP3-1* construct, with small, curved, greenish sepals. I. Detail of sepal in H, showing small green section on distal tip. J-M. Flowers treated with *ThtAP3-2a* constructs. J. Flower treated with short *ThtAP3-2a* construct, with lobed sepal indicated with arrow and

chimeric organ in place of outer stamen boxed. K. Detail of sepal-carpel chimera in J, with stigma indicated by arrow. L. Flower treated with long *ThtAP3-2a* construct, with chimeric organ in place of outer stamen, indicated by arrow. M. Flower treated with long *ThtAP3-2a* construct, with lobed sepal indicated by arrow. N – S. Flowers treated with *ThtAP3-2b* constructs. N. Flower treated with short *ThtAP3-2b* construct, with sepal in place of outer stamen boxed. O. Detail of extra sepal in N. P. Flower treated with long *ThtAP3-2b* construct, with lobed sepal green sepal, boxed. Q. Detail of lobed green sepal in P. R. Flower treated with long *ThtAP3-2b* construct, with chimeric organ in place of outer stamen boxed. S. Detail of sepal-stamen chimera in R. S = short construct, L = long construct. Numbers indicate plant number followed by flower number. Scale bar = 1mm.

Figure 3. Quantification of B class gene expression in untreated plants. Quantitative RT-PCR (qPCR) showing relative average expression levels of *ThtAP3-1, ThtAP3-2a, ThtAP3-2b, and ThtP1* in untreated young, open *T. thalictroides* flowers (n = 6). Expression calculated using the $2^{-\Delta\Delta Ct}$ method, normalized to *ThtACTIN* and *Elongation Factor 1α (ThtEF1α).* Mean and standard error bars shown. *ThtAP3-2b* expression is significantly higher than *ThtAP3-1 and ThtAP3-2a* (p < 0.01). *ThtPI* expression is significantly higher than each of the *ThtAP3s* (p ≤ 0.001). Letters indicate statistical significance in a one-way ANOVA followed by Tukey test (p < 0.01), same letters indicate no statistical difference.

Figure 4. Molecular validation of individual *T. thalictroides* flowers treated with *ThAP3-1, ThtAP3-2a* and *ThtAP3-2b* VIGS constructs. A. Relative average expression levels of *ThtAP3-1, ThtAP3-2a, ThtAP3-2b, and ThtP1* in untreated young, open *T. thalictroides* flowers (n = 6), as shown in Fig. 3. Mean and standard error bars shown. B - I. Relative average expression levels in VIGS-treated flowers from Fig. 2. B - D. *ThtAP3-1*-treated flowers, showing decreasing levels of *ThtAP3-1*, indicated

by red arrows. E - F. *ThtAP3-2A* treated flowers from Fig. 2, showing elevated levels of *ThtAP3-2a*, indicated by red arrows. G - H. *ThtAP3-2B* treated flowers from Fig. 2, showing decreasing levels of *ThtAP3-2b*, indicated by red arrows. Expression calculated using the $2^{-\Delta\Delta Ct}$ method, normalized to *ThtACTIN* and *Elongation Factor 1α (ThtEF1α)*.

Figure 5. Average molecular validation of *T. thalictroides* flowers treated with *ThAP3-1, ThtAP3-2a* and *ThtAP3-2b* VIGS constructs. Quantitative RT-PCR (qPCR) showing relative average expression levels of *ThtAP3-1, ThtAP3-2a, ThtAP3-2b, and ThtP1* in each of four VIGS treatments: Controls (untreated n=6 plus TRV2-empty n=3), *ThtAP3-1* long construct (n=6), *ThtAP3-2a* long construct (n=6) and *ThtAP3-2b* long construct (n=6), normalized with the housekeeping genes *ThtACTIN* and *ThtEEF1*. Mean and SE are shown; different letters indicate a significant differences in a one-way ANOVA with Tukey contrasts ($p \leq 0.05$). A. The expression of *ThtAP3-1* is significantly lower in *ThtAP3-1*-treated flowers than controls ($p \leq 0.05$). B. The expression of *ThtAP3-2a* is significantly higher in *ThtAP3-2A* treated flowers than controls ($p \leq 0.05$). C. The expression of *ThtAP3-2b* is not statistically different than controls in *ThtAP3-2B* treated flowers. The expression of *ThtAP3-2b* is significantly higher in *ThtAP3-2A* treated flowers than other treatments ($p \leq 0.05$), but not controls. D. Levels of *ThtPI* are statistically similar among treatments. Letters indicate statistical significance in a one-way ANOVA followed by Tukey test ($p < 0.05$), same letters indicate no statistical difference.

Figure 6. Yeast two-hybrid assays. Interactions between different ThtAP3 proteins, ThtPI and ThtSEP3 were determined with the yeast two-hybrid system. Colony growth on selective Leu/Trp/His-free + AbA medium is shown. Yeast cells were spotted in 10-fold serial dilution (from left to right) for

each interaction tested. All proteins were expressed as fusion with the GAL4 activation domain (AD) and the GAL4 DNA binding domain (BD).

Supplemental Figure Captions

Figure S1. Detection of tobacco rattle virus (TRV) transcripts in plants treated by virus-induced gene silencing. Expression analysis of plants treated with untreated controls, empty TRV2 + TRV1, TRV2-*ThtAP3-1* short (S) + TRV1, TRV2-*ThtAP3-1* long (L) + TRV1, TRV2-*ThtAP3-2a* S + TRV1, TRV2-*ThtAP3-2a* L + TRV1, TRV2-*ThtAP3-2b* S + TRV1, TRV2-*ThtAP3-2b* L + TRV1 by RT-PCR. TRV1 and TRV2 were detected only in VIGS-treated and empty TRV2-treated samples. Empty TRV2 generates a 390 bp band, while TRV2 containing the *ThtAP3-1 S* = 732 bp*, ThtAP3-1 L* = 819 bp*, ThtAP3-2a S* = 606 bp*, ThtAP3-2a L* = 817 bp, *ThtAP3-2b S* = 614 bp and *ThtAP3-2b L* = 798 bp. Approximate size of bands (in bp) indicated on the right. Treatments and plant numbers are labeled above each lane.
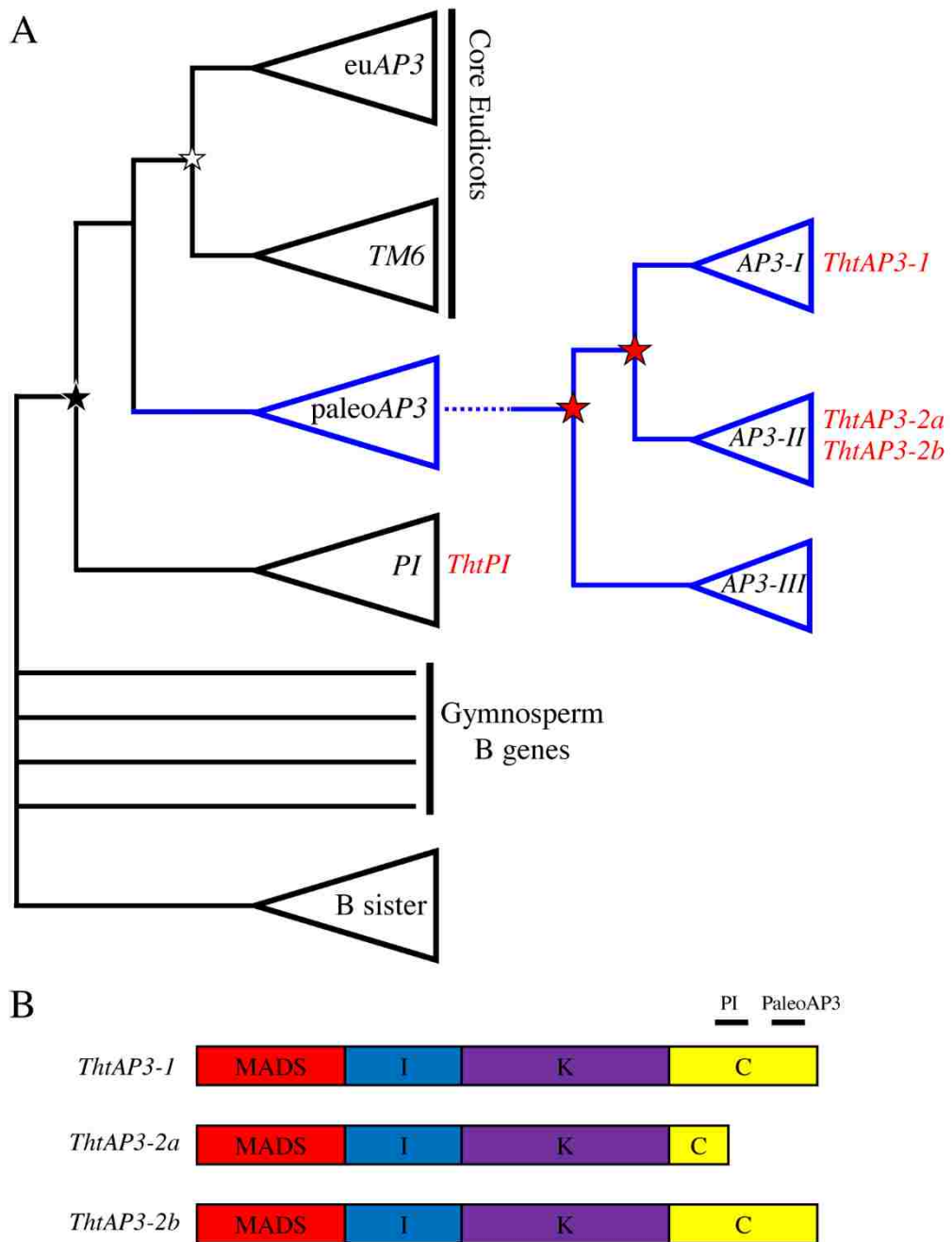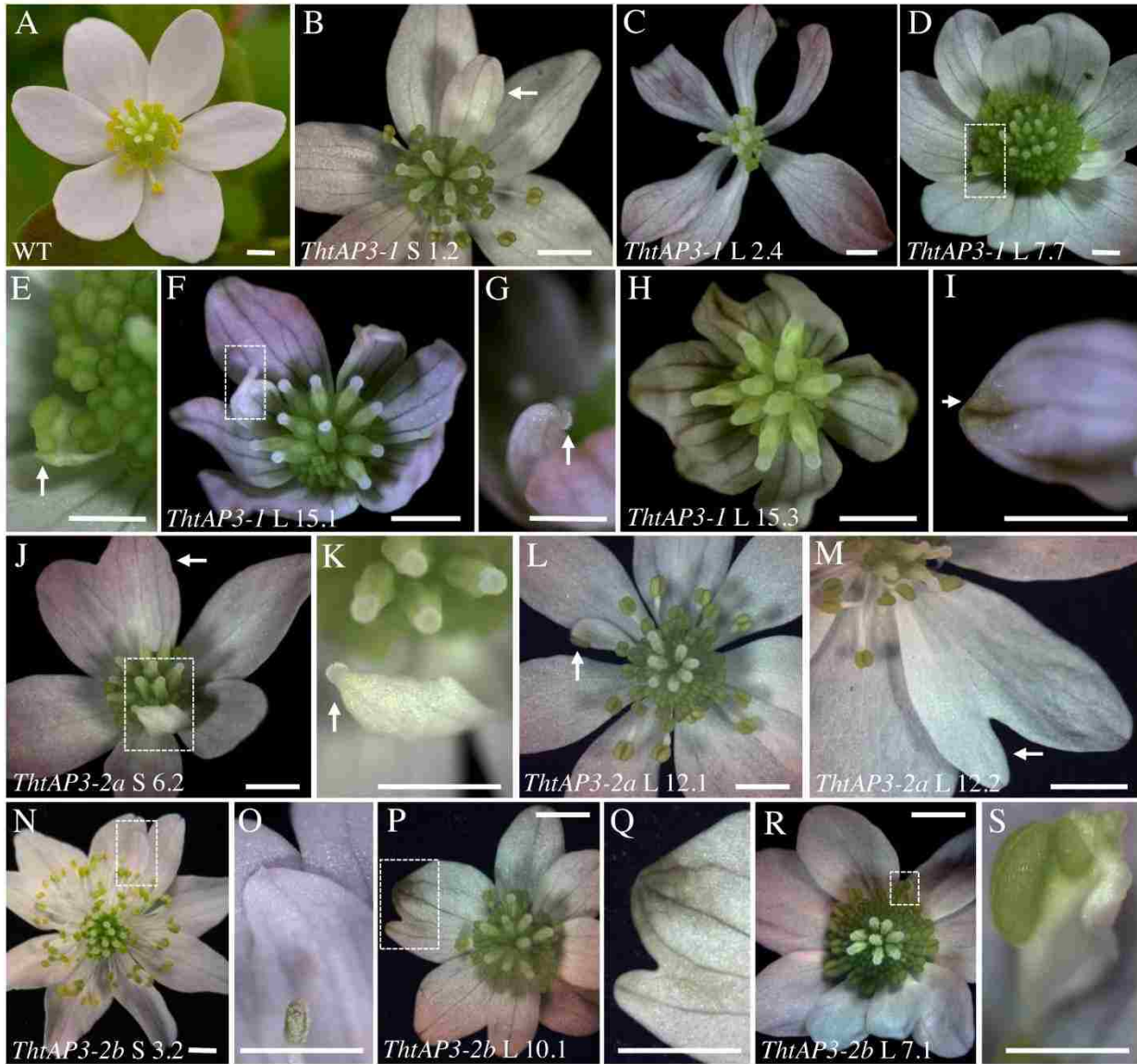
Table S1.

Purpose, name and sequence of primers used in this study.
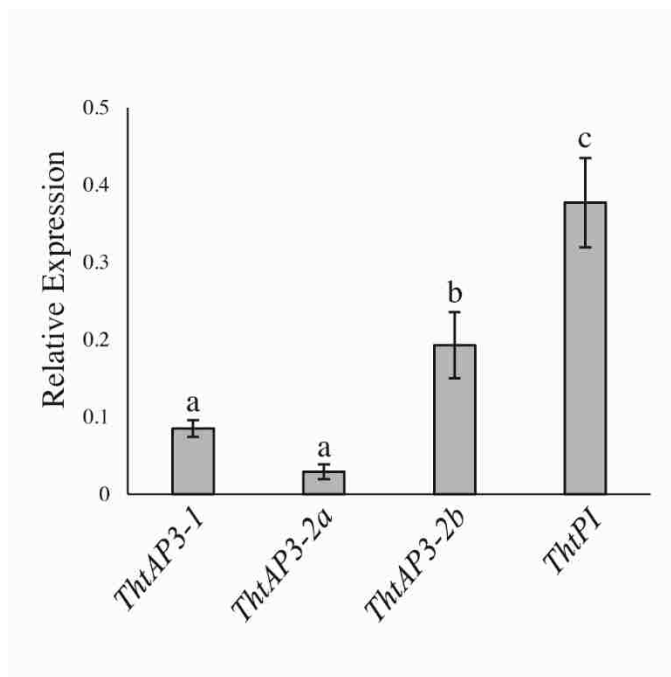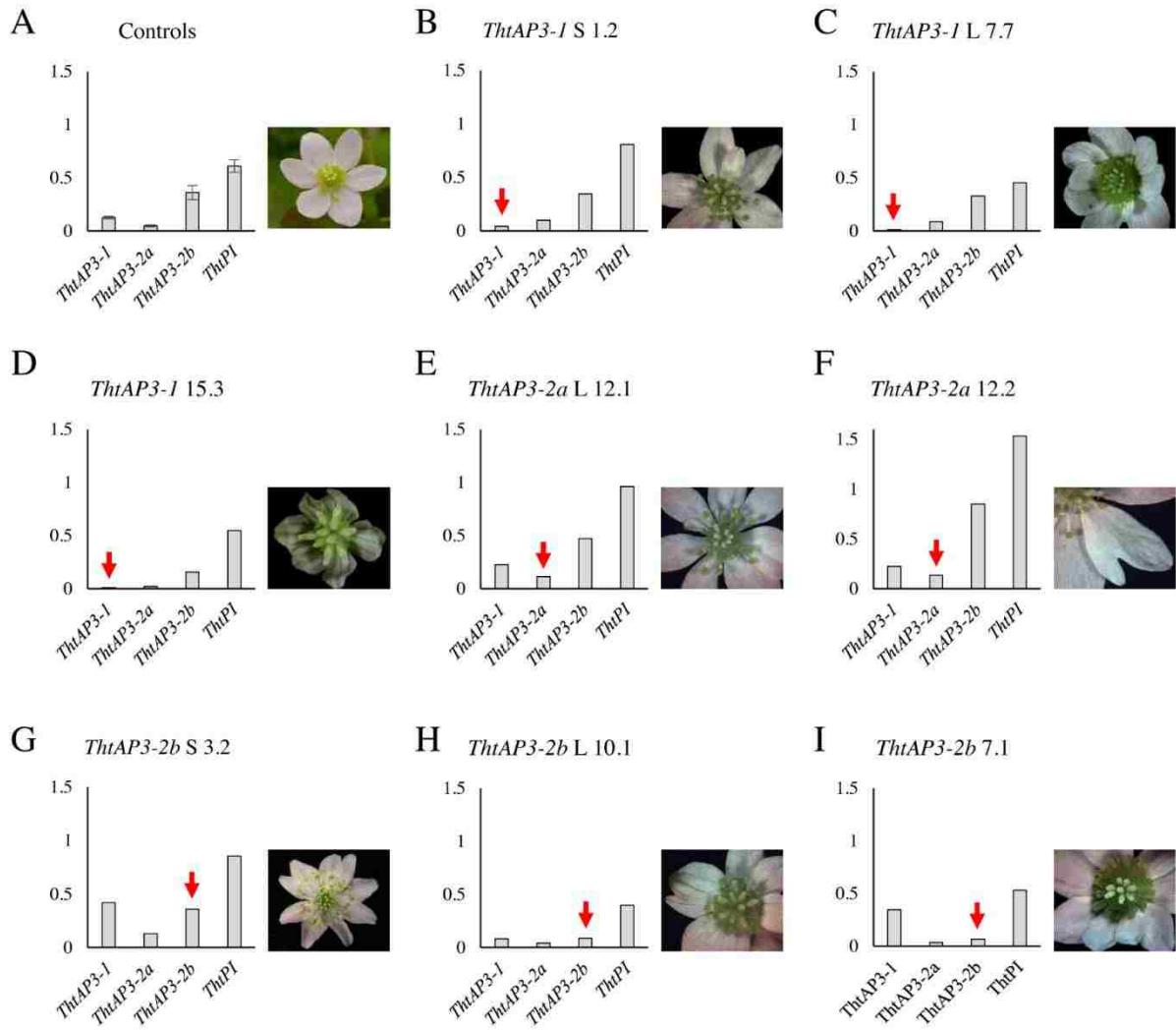
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

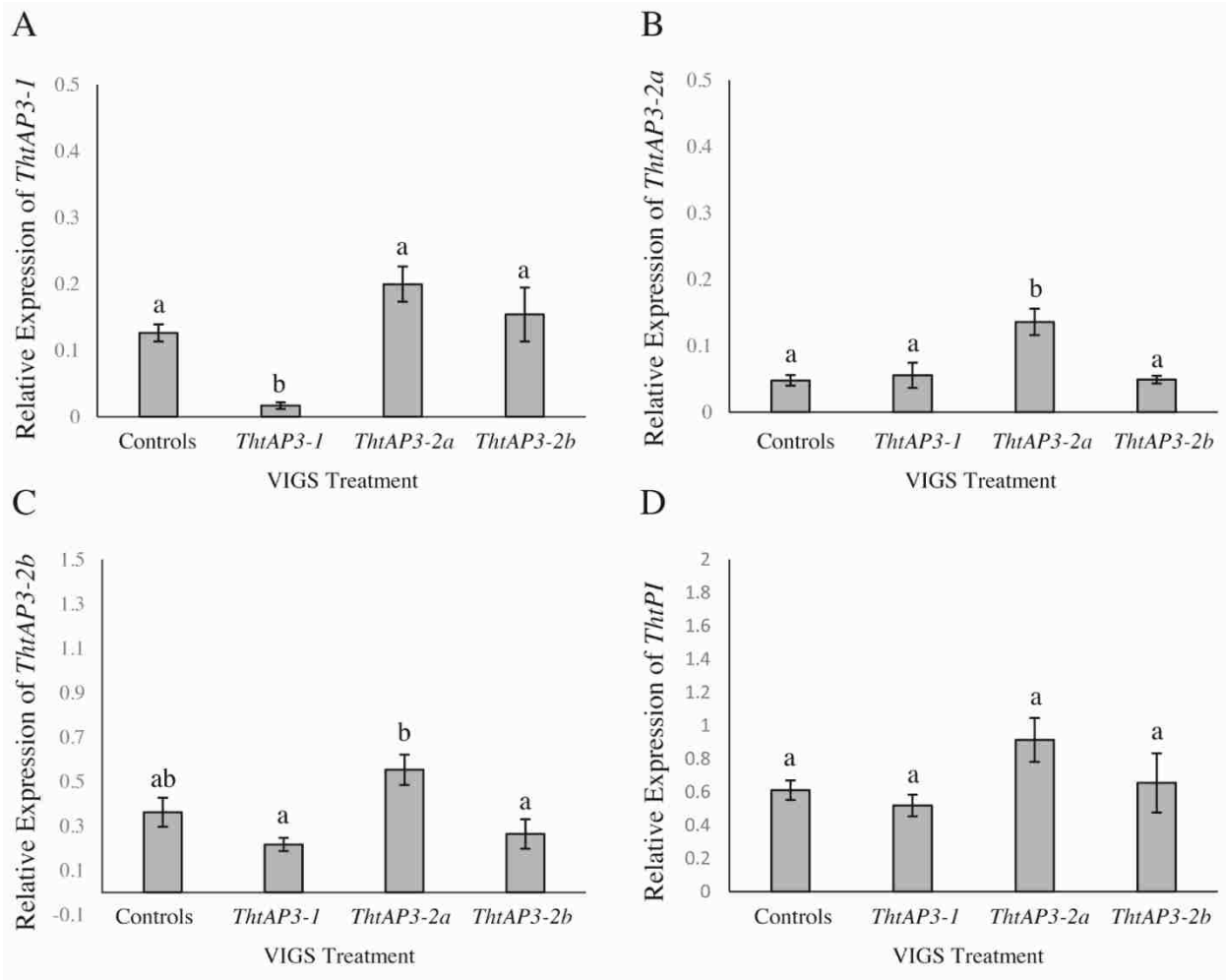Figure 6

AD \ BD	ThtAP3-1	ThtAP3-2a	ThtPI	ThtSEP3ΔC	pGADT7
ThtAP3-1					
ThtAP3-2a					
ThtAP3-2b			na		
ThtPI			na		
ThtSEP3ΔC					
pGBKT7					

AD \ BD	ThtAP3-1	ThtAP3-2a	TthPI	TthSEP3ΔC	pGADT7
ThtAP3-1	++	++	++	++	+
ThtAP3-2a	-	+	++	++	-
ThtAP3-2b	++	++	na	++	-
TthPI	++	++	na	-	-
TthSEP3ΔC	-	-	++	++	-
pGBKT7	-	-	-	-	-

Supplemental Figure S1
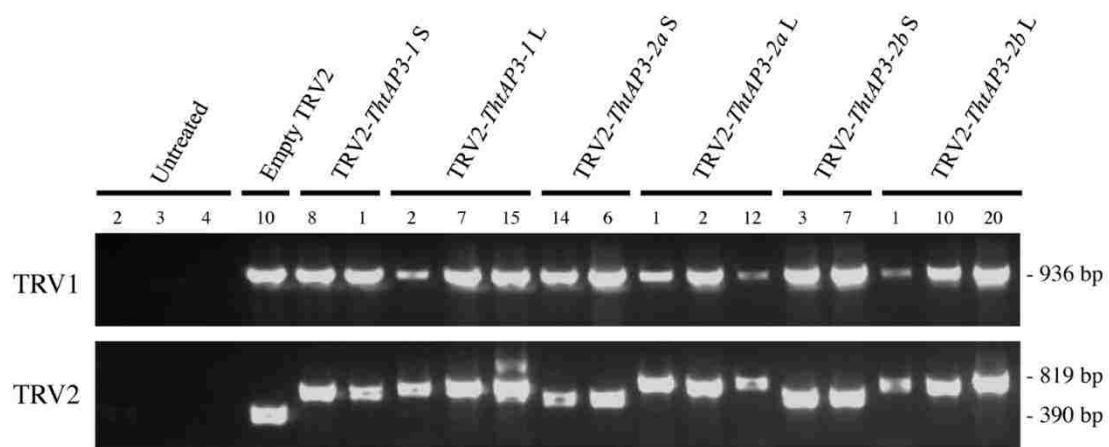
Primer Purpose	Name	Sequence
VIGS cloning		
Short	*TthAP3-1* F BamHI	5' TAG GGA TCC ATG CAT AAT GAT CTA CTG C 3'
Long	*TthAP3-1* F BamHI	5' TAG GGA TCC GTG TTC GTA ATC GAA AGT TT 3'
	TthAP3-1 R KpnI	5' GAT GGT ACC GCA GAA CAC ATA CAA GTT A 3'
Short	*TthAP3-2a* F BamHI	5' TAG GGA TCC GCA ACC ACA GCT GTA GCC C 3'
Long	*TthAP3-2a* F BamHI	5' TAG GGA TCC GTG TTC GTA ATC GAA AGT TT 3'
	TthAP3-2a R KpnI	5' GAT GGT ACC GCA GAA CAC ATA CAA GTT A 3'
Short	*TthAP3-2b* F BamHI	5' TAG GGA TCC GCA ACC AAA GCT GCA GTC C 3'
Long	*TthAP3-2b* F BamHI	5' TAG GGA TCC TAC TAC TAG TAC TGA CAC TT 3'
	TthAP3-2b R KpnI	5' GAT GGT ACC GTT CAT ACA AAA GAT CCA A 3'
Presence of TRV1	pTRV1_fwd	5' CTT GAA GAA GAA GAC TTT CGA AGT CTC 3'
	pTRV1_rev	5' GTA AAA TCA TTG ATA ACA ACA CAG ACA AAC 3'
Presence of TRV2	OYL195	5' GGT CAA GGT ACG TAG TAG AG 3'
	OYL198	5' CGA GAA TGT CAA TCT CGT AGG 3'
Y2H cloning	TthAP3-1_Y2H_F	5'- CAT GGA GGC CGA ATT CAT GGG GAG AGG AAA GAT TGA GAT C -3'
	TthAP3-1_Y2H_R	5'- GCA GGT CGA CGG ATC CTC AAC CTA ATC GAA GAC CCT CGA A -3'
	TthAP3-2a/b_Y2H_F	5'- CAT GGA GGC CGA ATT CAT GGG GAG AGG AAA GAT TGA GAT T -3'
	TthAP3-2a_Y2H_R	5'- GCA GGT CGA CGG ATC CTT ACA GCT GTG GTT GCA AAC CGA A -3'
	TthAP3-2b_Y2H_R	5'- GCA GGT CGA CGG ATC CTC AAA CAA GAC TTA AGC CAT ATG A -3'
	TthPI_Y2H_F	5'- CAT GGA GGC CGA ATT CAT GGG AAG AGG TAA GAT TGA GAT C -3'
	TthPI_Y2H_R	5'- GCA GGT CGA CGG ATC CCT ATT TTT CCT GTA AAT TAG GCT G -3'
Expression by qPCR	*ThtAP3-1*-for-qPCR	5' GGG TTG GTG GTG AAG ATC TGA GTG ATA 3'
	ThtAP3-1-rev-qPCR	5' ATC AAC TAG TGC ATA TGG TTC CTC AAG 3'
	ThtAP3-2A for-qPCR	5' GCA AGA TAA TCT GCA GAA GCT GAA GGA 3'
	ThtAP3-2A rev-qPCR	5' ACG TAG CAG ACC GTT ATA GGT CAC TTG 3'
	ThtAP3-2B for-qPCR	5' GCA AGA TAA TCT GCA GAA GCT GAA GGA 3'
	ThtAP3-2B rev-qPCR	5' ACG TAG CAG ACC GTT ATA GGT CAC TTG 3'
	ThtPI-for-qPCR	5' TAC AAG TAT GCT AGA GGA AGA GAA CAA 3'
	ThtPI-rev-qPCR	5' AGA TCT ATT CAT TAT AAG GCT CAT GGT 3'