Evolution of the basic helix–loop–helix transcription factor SPATULA and its role in gynoecium development

Ana Rivarola-Sena, Aurélie Vialette, Amélie Andres-Robin, Pierre Chambrier, Loïc Bideau, Jose Franco-Zorrilla, Charles Scutt

Evolution of the basic Helix-Loop-Helix transcription factor SPATULA and its role in gynoecium development

Thank you for agreeing to review this paper for Annals of Botany. The Annals of Botany aims to be among the very top of plant science journals and as we receive over 1000 submissions every year we need to be very selective in deciding which papers we can publish. In making your assessment of the manuscript's suitability for publication in the journal please consider the following points.

Scientific Scope

Annals of Botany welcomes papers in all areas of plant science. Papers may address questions at any level of biological organization ranging from molecular through cells and organs, to whole organisms, species, communities and ecosystems. Its scope extends to all flowering and non-flowering taxa, and to evolutionary and pathology research. Many questions are addressed using comparative studies, genetics, genomics, molecular tools, and modeling.

To merit publication in Annals of Botany, contributions should be substantial, concise, written in clear English and combine originality of content with potential general interest.

- We want to publish papers where our reviewers are enthusiastic about the science: is this a paper that you would keep for reference, or pass on to your colleagues? If the answer is "no" then please enter a low priority score when you submit your report.

- We want to publish papers with novel and original content that move the subject forward, not papers that report incremental advances or findings that are already well known in other species. Please consider this when you enter a score for originality when you submit your report.

Notes on categories of papers:

All review-type articles should be **novel, rigorous, substantial and "make a difference" to plant science**. The purpose is to summarise, clearly and succinctly, the "cutting edge" of the subject and how future research would best be directed. Reviews should be relevant to a broad audience and all should have a **strong conclusion and illustrations** including diagrams.

- *Primary Research* articles should report on original research relevant to the scope of the journal, demonstrating an important advance in the subject area, and the results should be clearly presented, novel and supported by appropriate experimental approaches. The Introduction should clearly set the context for the work and the Discussion should demonstrate the importance of the results within that context. Concise speculation, models and hypotheses are encouraged, but must be informed by the results and by the authors' expert knowledge of the subject.

- *Reviews* should place the subject in context, add significantly to previous reviews in the subject area and moving forward research in the subject area. Reviews should be selective, including the most important and best, up-to-date, references, not a blow-by-blow and exhaustive listing.

- *Research in Context* should combine a review/overview of a subject area with original research, often leading to new ideas or models; they present a hybrid of review and research. Typically a Research in Context article contains an extended Introduction that provides a general overview of the topic before incorporating new research results with a Discussion proposing general models and the impact of the research.

- *Viewpoints* are shorter reviews, presenting clear, concise and logical arguments supporting the authors' opinions, and in doing so help to stimulate discussions within the topic.

- *Botanical Briefings* are concise, perhaps more specialised reviews and usually cover topical issues, maybe involving some controversy.

Original Article

Evolution of the basic Helix-Loop-Helix transcription factor *SPATULA* and its role in gynoecium development.

Ana C. Rivarola-Sena[1], Aurélie C. Vialette[1], Amélie Andres-Robin[1], Pierre Chambrier[1], Loïc Bideau[1], Jose Manuel Franco-Zorrilla[2] and Charles P. Scutt[1]*

[1]Laboratoire Reproduction et Développement des Plantes (CNRS UMR 5667), Ecole Normale Supérieure de Lyon, 69364 Lyon Cedex 7, France ; [2]Centro Nacional de Biotecnología-Consejo Superior de Investigaciones Científicas, C/Darwin3, 28049 Madrid, Spain

For correspondence: charlie.scutt@ens-lyon.fr

Running title: Evolution of SPATULA

Received: 28/03/2024

Returned for revision: 30/07/2024

Editorial decision: 02/08/2024

ABSTRACT

- *Background and Aims*

  SPATULA (SPT) encodes a basic Helix-Loop-Helix transcription factor in *Arabidopsis thaliana* that functions in the development of the style, stigma and replum tissues, all of which arise from the carpel margin meristem (CMM) of the gynoecium. Here, we use a comparative approach to investigate the evolutionary history of *SPT* and identify changes that potentially contributed to its role in gynoecium development.

- *Methods*

  We investigate *SPT's* molecular and functional evolution using phylogenetic reconstruction, yeast-2-hybrid analyses of protein-protein interactions, microarray-based analyses of protein-DNA interactions, plant transformation assays, RNA *in-situ* hybridization, and *in-silico* analyses of promoter sequences.

- *Key results*

  We demonstrate the *SPT* lineage to have arisen at the base of euphyllophytes from a clade of potentially light-regulated transcription factors through gene duplication followed by the loss of an Active Phytochrome Binding (APB) domain. We also clarify the more recent evolutionary history of *SPT* and its paralog *ALCATRAZ (ALC)*, which appear to have arisen through a large-scale duplication within Brassicales. We find that *SPT* orthologs from diverse groups of seed plants share strikingly similar capacities for protein-protein and protein-DNA interactions, and that *SPT* coding regions from a wide taxonomic range of plants are able to complement loss-of-function *spt* mutations in transgenic *Arabidopsis.* However, the expression pattern of *SPT* appears to have evolved significantly within angiosperms, and we identify structural changes in *SPT*'s promoter region that correlate with the acquisition of high expression levels in tissues arising from the CMM in Brassicaeae.

- *Conclusions*

  We conclude that changes to *SPT's* expression pattern made a major contribution to the evolution of its developmental role in the gynoecium of Brassicaeae. By contrast, the main biochemical capacities of SPT, as well as many of its immediate transcriptional targets, appear to have been conserved at least since the base of living angiosperms.

Keywords: *SPATULA, ALCATRAZ,* basic Helix-Loop-Helix, carpel, gynoecium, carpel margin meristem, flower, angiosperm, *Arabidopsis thaliana, Amborella trichopoda, Nymphaea thermarum, Petunia axillaris*

INTRODUCTION

In the syncarpous gynoecium of the model angiosperm *Arabidopsis thaliana* (referred to below as *Arabidopsis),* the margins of the two fused carpels form a Carpel Margin Meristem (CMM) which undergoes cell division to generate the placenta, ovules, replum, style and stigma (Ferrandiz *et al.,* 2010). The carpel margins in the ovary later differentiate to generate a dehiscence zone and separation layer, both of which contribute to fruit dehiscence.

Numerous genes function in tissues that develop from the CMM, including the basic Helix-Loop-Helix (bHLH) transcription factor *SPATULA (SPT)* (Heisler *et al.,* 2001). In strong *spt* loss-of-function mutants, the gynoecium remains unfused at the apex and the septum fails to fully develop (Alvarez and Smyth 1999). Fertility is reduced is *spt* mutants due to a reduction in the size and extent of the pollen transmitting tissue within the style and septum, in which the cells fail to properly elongate. Outside the flower, *SPT* functions in seed dormancy (Penfield *et al.,* 2005), leaf and cotyledon expansion (Josse *et al.,* 2011), and stomatal density (Bernal-Gallardo *et al.,* 2023). In gynoecium and fruit tissues, *SPT* acts in a partially redundant manner with its paralog *ALCATRAZ (ALC;* Rajani and Sundaresan, 2001). *ALC* plays a major role in the development of fruit dehiscence zones, and a lesser role, redundantly with *SPT,* at early stages of gynoecium development (Groszmann *et al.,* 2011).

SPT protein is capable of homodimerization, and also of heterodimerization with ALC (Groszmann *et al.,* 2011), as well as with any of the bHLH factors HECATE1-3 (HEC1-3) (Gremski *et al.,* 2007) that are redundantly necessary for normal style and septum development. SPT can also form heterodimers with INDEHISCENT (IND), a close relative of the redundant HEC proteins (Girin *et al.,* 2011; Ballester *et al.,* 2021). Numerous immediate transcriptional targets of SPT have been identified, including genes known to function in light-regulated pathways (Reymond *et al.,* 2012), or in signaling pathways regulated by auxin (Friml *et al.,* 2004) or cytokinin (Irepan Reyes-Olalde *et al.,* 2017). SPT orthologs are specifically upregulated in gynoecium tissues in other rosids (Tani *et al.,* 2011; Cheng *et al.,* 2022), asterids (Ortiz-Ramirez *et al.,* 2019) and basal eudicot*s* (Zumajo-Cardona *et al.,* 2017). Several gynoecium phenotypes, including defects in carpel fusion, transmitting tissue development and the elongation of stigmatic papillae, are associated with the mutation of *SPT*

homologs in the rosid species *Cucumis sativus* (Cheng *et al.,* 2022). However, functional genetic studies using virus-induced gene silencing in three Solanaceae (asterids) species found no *SPT*-knockdown phenotypes associated with pre-fertilization stages of gynoecium development (Ortiz-Ramirez *et al.,* 2019).

Here, we reconstruct the evolutionary history of the *SPT* lineage using a range of *in silico, in vitro* and *in vivo* approaches. We conclude that the SPT protein has largely conserved its biochemical capacities and downstream pathways, at least since the most recent common ancestor (MRCA) of living angiosperms, but that its role in the development of tissues arising from the CMM in Brassicaeae was acquired more recently and involved changes to *SPT's* expression pattern which correlated with changes in the arrangement of conserved motifs in its promoter sequence.

MATERIALS AND METHODS

Database-searching, phylogenetic reconstruction and genomic sequence analysis.

Genes of interest were identified by tblastn (Altschul *et al.,* 1997) searching of databases listed in SI Figs 1 and 2. Amino-acid alignments were performed using MUSCLE in SeaView (Gouy *et al.,* 2010) and used to generate Maximum Likelihood phylogenies in PhyML (Guindon and Gascuel 2003), employing the LG evolutionary model. Branch support was provided using the aLRT method (Anisimova and Gascuel 2006). Species phylogenies were generated from the APG IV classification (Byng *et al.,* 2016) and gene-tree/species-tree phylogenetic reconciliations were performed on initial phylogenetic reconstructions using Treerecs (Comte *et al.,* 2020). Conserved motifs in promoters and coding sequences were identified using MEME Suite 5.1.1 (http://meme-suite.org/tools/meme) and their similarities to known transcription factor binding sites investigated in the same suite of programs using Tomtom. Gene synteny within Brassicaceae was visualized using PLAZA-Eudicots 5.0.

Plant material and nucleic acid extraction.

Mutant and wild-type *Arabidopsis* seed accessions were obtained from Nottingham *Arabidopsis* Stock Centre (NASC). *Nymphaea thermarum* plants were generously donated by Paula Rudall and Carlos Magdalena, Royal Botanic Gardens-Kew, London. *Petunia axillaris*

material was kindly provided by Michiel Vandenbussche at ENS-Lyon. *Pinus taeda* needle material was a kind gift from Charles Dana Nelson of University of Kentucy, Lexington KY, USA. *Amborella* plants were generously provided by Gildas Gâteblé and Bruno Fogliani (University of New Caledonia). Plants of *Nymphaea caerulea, Selaginella moellendorffii and Picea abies* were obtained from commercial sources. RNA was extracted from flower tissues for RT-PCR amplifications and other procedures using Trizol Reagent (Thermo-Fischer Scientific) according to the manufacturer's instructions. Genomic DNA was extracted from leaf/shoot tissues for PCR amplifications using a Nucleon PhytoPure kit (Thermo-Fischer Scientific).

Genetic transformation.

Arabidopsis thaliana wild-type Col-0 plants, or *spt-11* mutants (Ichihashi *et al.,* 2010) in the Col-0 background, were used in genetic transformation experiments. Plants were grown under long-day conditions (18h light/6h dark cycles) at 22°C. Homozygous *spt-11* mutants to be used in transformation procedures were additionally illuminated using far-red LEDs (24x 1W, 760 nm), from five weeks after germination until seed-set, to generate a low red/far-red light ratio that reduced the negative effect of the *spt-11* mutation on fertility. Plants were transformed using standard "floral dip" procedures (Clough and Bent 1998) and transformants selected on kanamycin (50 µg/L)-containing media.

Protein-binding microarray analyses.

Coding sequences of *SPT* orthologs were inserted into the *pDEST-TH1* expression vector as translational fusions to a Maltose Binding Protein domain, and the resulting plasmid was transferred to *E. coli* BL21 cells (Franco-Zorrilla *et al.,* 2014). Production of recombinant protein was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM) to log-phase cultures and confirmed using SDS-PAGE analysis of cell lysates. The resulting recombinant proteins were analyzed on PBM11 protein-binding microarrays, as described by Godoy *et al.* (2011) and the resulting data analyzed as described by Berger and Bulyk (2009) to produce position weight matrices (PWMs) describing DNA-binding preferences. PWMs were converted to diagrams using ENOLOGOS (Workman *et al.,* 2005). Orthologs of *Arabidopsis* SPT-targets from *Amborella trichopoda* and *Picea abies* were identified by BLAST searching followed by Maximum Likelihood phylogenetic analyses in SeaView

(Gouy *et al.,* 2010). Upstream sequences of targets were obtained from Phytozome for *Arabidopsis* and *Amborella trichopoda* and Congenie for *Picea abies*. The required 3-kb of upstream sequence could be obtained for all but three of the *Picea abies* genes of interest. PWMs were used to scan promoter sequences using MotifScan in RSAT (Nguyen *et al.,* 2018), employing a cut-off score of 6.0. Numbers of promoters containing one or more high-scoring binding sites were compared in R software between sets of putative targets promoters and full genomic promoter sets using the hypergeometric test, as fully described in the notes in SI Tab. 1.

Yeast two-hybrid assays.

Coding sequences of *Amborella trichopoda* genes *AtrSPT* (Tr_v1.0_scaffold00046.26), *AtrHEC1/2* (Tr_v1.0_scaffold00008.223) and the putative *HEC3/IND* ortholog *bHLH87* (Tr_v1.0_scaffold00036.88) were amplified by standard RT-PCR methods employing high-fidelity thermo-stable DNA polymerases. The resulting molecules were inserted into both the *pGBT9* yeast expression vector (Clontech) as a translational fusion with the *GAL4* DNA-binding domain (DBD) and the *pGAD24-GW* expression vector (Clontech) as a translational fusion to the *GAL4* activation domain (AD). The *pGBT9*- and *pGAD24-GW*-derived expression constructs were then transferred by electroporation to cells of *Saccharomyces cerevisiae* strains Y187 and AH109, respectively. Yeast mating and growth on selective media was performed as described by de Bossoreille *et al.* (2018). Background colony growth was reduced by addition of 3-amino-1,2,4-triazole (3-AT) to culture media. Protein dimerization was tested in both directions with respect to the *GAL4* AD and DBD.

GUS reporter gene analyses.

Promoter sequences of *SPT* orthologs were PCR-amplified using high-fidelity thermostable polymerases and primers given in SI Tab. 2. The resulting molecules were inserted by recombination in the *pENTR/D/TOPO* (Invitrogen) vector and sequenced to verify their integrity using automated Sanger DNA sequencing reactions. Promoters were then inserted by Gateway LR (Invitrogen) recombination reactions into the *pKGWFS7.0* GUS-expression vector and the resulting plasmids transferred by electroporation to *Agrobacterium tumefaciens* C58PMP90 cells for plant transformation. Flower tissues of T1 transformants were incubated for 20 min on ice in acetone (90% v/v), rinsed for 10 min in sodium phosphate buffer (0.1 M,

pH 7.0), and then transferred to staining solutions containing 5-bromo-4-chloro-3-indolyl-β-glucuronique acid (X-Gluc, 1 mM), potassium ferrocyanide (0.5 mM), potassium ferricyanide (0.5 mM), EDTA (5 mM), Triton X100 (0.05% v/v) and sodium phosphate buffer (0.1 M, pH 7.0) at 37°C for 24 to 48h, depending on the extent of staining observed. Samples were then rinsed and stored if necessary at 4°C in sodium phosphate buffer (0.1 M, pH 7.0) prior to examination and imagery using a Keyence VHX-900F digital microscope.

Genetic complementation assays.

Coding sequences of interest were amplified by standard RT-PCR procedures employing hi-fidelity, thermo-stable DNA polymerases and primers shown in SI Tab. 2. The resulting DNA molecules were inserted by recombination into the *pENTR/D/TOPO* vector (Invitrogen) and sequenced in automated Sanger sequencing reactions to verify their integrity. Two versions of each of *AthPIF5* and *SmoPIF* were amplified, one of which was truncated at its 5'-end to remove the APB domain while leaving an in-frame initiation codon. The 6.3 kb *AthSPT* promoter fragment (Groszmann *et al.,* 2010) was ligated into the *pENTR5'TOPO* vector (Invitrogen). The *AthSPT* promoter fragment and each coding sequence of interest (separately) were then inserted into the *pK7m24GW* plasmid (Karimi *et al.,* 2007) using a Multisite Gateway LR reaction (Invitrogen) to generate the required plant transformation vectors. Homozygous *spt-11* mutant *Arabidopsis* plants were transformed, as described above. Phenotypes were analyzed in T1 transformants. Numbers of plants in which mutant phenotypes were complemented were analyzed to derive p-values using the two-tailed version of Fisher's Exact Test. Images of transformed plants were obtained using both a Keyence VHX-900F digital microscope, and a HIROX-3000 environmental scanning electron microscope, the latter at a stage temperature of -20°C and a tube-voltage of 10 kV.

RNA in-situ hybridization.

In-situ hybridizations to *Petunia axillaris* and *Nymphaea thermarum* flower buds were performed using the protocol of Morel *et al.* (2018) while those to *Amborella trichopoda* flower buds were performed according to Vialette-Guiraud *et al.* (2011). For both procedures, dioxygenin-labelled riboprobes were prepared from PCR-amplified, full-length coding sequences, as described by Vialette-Guiraud *et al.* (2011). Images were captured using an

Imager-M2 fluorescence microscope (Zeiss) fitted with an AxioCam MRc digital camera (Zeiss).

RESULTS

SPATULA emerged from a clade of potentially light-regulated transcription factors at the base of living euphyllophytes.

The *SPT* lineage was previously concluded to have emerged from a clade of *PHYTOCHROME INTERACTING FACTORs (PIFs)* and to have lost, sometime before the MRCA of living seed plants, a short Active Phytochome Binding (APB) domain (Reymond *et al.,* 2012) through which at least some angiosperm PIFs can physically interact with the active form of the Phytochrome B (phyB) photoreceptor (Shin *et al.,* 2009). We reanalyzed this question using a more extensive taxonomic range of land plants to better situate the loss of the ABP domain. Our phylogenetic analysis of representative species with sequenced genomes (Fig. 1A), incorporating a reconcilation step to improve the accuracy of phylogenetic reconstruction, identified likely orthologs of *SPT*, which also lack an APB domain, in angiosperms, gymnosperms and monilophytes (ferns and their allies). In the lycophyte *Selaginella moellendorffii* and in the bryophyte *Physcomitrium patens,* however, the most closely related genes to *SPT* grouped externally to a clade containing both *SPT* and *PIF* genes from euphyllophytes (monilophytes and seed plants). These results strongly suggest that the *SPT* clade arose by duplication of a *PIF*-like gene in a common ancestor of euphyllophytes, closely followed by the loss of its APB domain.

The SPATULA and ALCATRAZ lineages separated within Brassicales, while a second SPATULA-like lineage, present in most core eudicots, was lost at the base of Brassicaeae.

The evolutionary stage at which the *Arabidopsis SPT* and *ALC* lineages separated has been the subject of debate. Groszmann *et al.* (2011) concluded that this duplication occurred within Brassicales. However, two or more *SPT/ALC-like* genes occur in diverse core eudicots (also known as the Pentapetalae), a group which includes the rosids, asterids and Caryophyllales, and Ortiz-Ramirez *et al.* (2018) have proposed an alternative scenario in which the *SPT* and *ALC* lineages separated much earlier, at the base of the core eudicots.

To attempt to resolve this question, we assembled a database broadly covering angiosperms from taxa with sequenced genomes, but also including sequences from transcriptomic databases to increase representation within Brassicales [an approach to this question previously suggested by Pfannebecker *et al.* (2017)] and incorporating a phylogenetic reconciliation step to improve the accuracy of phylogenetic reconstruction. Our phylogeny, summarized in Fig. 2A (with full data given in SI Fig 2), shows *SPT* and *ALC* clades occupying sister positions, both of which contain genes from most Brassicaceae included in the analysis, and also, with good statistical support, from *Gyrostemon ramulosus* (Brassicales, Gyrostemonaceae). The Brassicales *SPT* clade also includes, with good statistical support, two genes from *Taranaya hassleriana* (Brassicales, Cleomaceae). *SPT* genes from Brassicales taxa outside of the clade containing Brassicaceae, Cleomaceae and Gyrostemonaceae are found as a closely-related grade of lineages, with overall statistical support for a Brassicales *SPT+ALC* clade of 92%. Externally to the Brassicales *SPT+ALC* clade are found *SPT* genes from other rosids, asterids and Caryophyllales, this overall core eudicot *SPT* clade having 98% statistical support.

A second major clade of core eudicot *SPT*-like genes in our phylogeny, with 96% statistical support, includes sequences from a wide range of core eudicots, though Brassicales are represented in this clade only by *Taranaya hassleriana*. *SPT* genes from ANA-grade angiosperms and monocots form a grade of sequences that diverge basally within the overall phylogeny. Basal eudicot and *Ceratophyllum dimersum SPT* sequences occur in clades arising after the ANA-grade lineages, which occupy sister positions to one or both of the core eudicot *SPT* and *SPT*-like clades described above. The *SPT* lineage in all taxa other than core eudicots is indicated as *PaleoSPT* in Figs. 2A and B.

Taken together, these results indicate that the *SPT* and *ALC* lineages present in *Arabidopsis* and other Brassicaceae were derived from a duplication event that took place within Brassicales, more precisely in a common ancestor of Gyrostemonaceae, Cleomaceae and Brassicaceae. A second lineage of *SPT*-like genes is widely present in core eudicots, but was probably lost at the base of Brassicaceae, as it cannot be found in any Brassicaceae with sequenced genomes, including the most basally diverging genus *Aethionama*, but is still present in *Taranaya hassleriana* of Cleomaceae. The *paleoSPT* lineage, in basal eudicots, *Ceratopyllum dimersum,* monocots and ANA-grade angiosperms is pro-orthologous to both

the *SPT* and *SPT*-like lineages of core eudicots. These conclusions are illustrated in a simplified version of the species phylogeny used in phylogenetic reconciliation (Fig. 2B).

We also examined the microsynteny associated with *SPT* and *ALC* genes in three selected Brassicaceae species. The genomic regions analysed showed highly conserved synteny, with six orthogroups in common occurring among the fifteen genes up- or downstream of *SPT* and *ALC* in *Arabidopsis*, *Capsella rubella* and/or *Cardamine hirsuta* (SI Fig. 3). These data indicate the duplication that generated *SPT* and *ALC* in Brassicaceae to have been large-scale in nature. Given that this duplication occurred in a common ancestor of Gyrostemonaceae, Cleomaceae and Brassicaceae, it probably, as proposed by Groszmann *et al.* (2011), corresponded to the At-β whole genome duplication (Mabry *et al.,* 2020).

Groszmann *et al.* (2008) identified two domains outside the bHLH domain that were found to support *SPT* functions in transgenic *Arabidopsis*: an amphipathic helix and an acidic domain. These domains were stated to be specific to SPT/ALC proteins from eudicots, being absent in SPT proteins from monocots. We observe, however, that the acidic domain identified by Groszmann *et al.* (2008) is present in most of the SPT proteins included in our phylogenetic analyses, including those from basally diverging angiosperms and gymnosperms (SI Fig. 4A). This domain is also apparent in the most closely related PIF protein in the lycophyte *Selaginella moellendorffii*. The amphipathic helix identified by Groszmann *et al.* (2008) is not, by contrast, evident in SPT orthologs from outside core eudicots (SI Fig. 4B).

The protein-protein and protein-DNA interactions of SPATULA are conserved in widely diverged angiosperms and gymnosperms.

To investigate the evolutionary conservation of SPT's biochemical properties, we first used yeast-2-hybrid analysis to determine whether the SPT protein from the basal-most living angiosperm, *Amborella trichopoda (*AtrSPT*,* protein identifier = ATR0800G167), was capable of homodimerization, and/or of heterodimerization with the HEC-like proteins from the same species, as is *Arabidopsis* SPT. *Amborella trichopoda* contains two *HEC*-like genes, one of which, *AtrHEC1/2,* is the likely pro-ortholog of *Arabidopsis HEC1* and *HEC2*, while the other, *bHLH87,* is the likely pro-ortholog of *Arabidopsis HEC3* and *IND* (Pabon-Mora *et al.,* 2014). The results of our analysis (Fig 3) show that *Amborella trichopoda* SPT is able to

homodimerize, and to heterodimerize with both of the *Amborella trichopoda* HEC/IND-like proteins, strongly suggesting that these dimerization capacities of *Arabidopsis* SPT have been conserved in both the *Arabidopsis* and *Amborella* lineages since the MRCA of living angiosperms.

We then made an inter-species comparison of immediate transcriptional targets of SPT, this time extending our analysis to cover the gymnosperm *Picea abies.* Recombinant SPT proteins from *Arabidopsis, Amborella trichopoda* and *Picea abies* were analyzed using protein-binding microarrays (Godoy *et al.,* 2011) to determine their *in-vitro* DNA-binding preferences. For each species, position weight matrices (PWMs) were derived from the three highest-scoring single oligonucleotides identified (SI Fig. 5). Twenty-six genes have been experimentally identified as immediate targets of *Arabidopsis* SPT (Girin *et al.,* 2011; Reymond *et al.,* 2012; Irepan Reyes-Olalde *et al.,* 2017), most of which possess canonical G-boxes (CACGTG) in their putative promoter regions. In agreement with this observation, the consensus sequence (showing the most probable nucleotide at each position) derived from each of the nine PWMs obtained was also found to contain a canonical G-box.

We used the three top-scoring PWMs from each species to analyse *in-silico* up to 3 kb of 5'-flanking sequences from the known SPT-targets in *Arabidopsis* and their putative orthologs from *Amborella trichopoda* and *Picea abies.* Full results of these analyses are shown in SI Tab. 1 and summarized in Tab. 1. Interestingly, almost one-third (30/95) of high-scoring SPT-binding sites detected (SI Tab. 1) did not include a canonical G-box. Eighteen of 24 putative *Arabidopsis* target promoters analyzed were found to contain at least one high-scoring binding site, while the equivalent figures were 13/22 for *Amborella trichopoda* and 10/13 for *Picea abies* (not including genes for which truncated sequences only were available). Statistical analyses indicated the significant overrepresentation of putative target promoters by factors of 4.83 ($p = 6.00x10^{-7}$ and 2.66 ($p = 1.5x10^{-2}$) in *Arabidopsis* and *Amborella trichopoda*, respectively (SI Tab. 3). This comparison was not possible for *Picea abies* due to the lack of 5'-flanking sequence data covering the entire genome, though we note that a higher proportion of putative targets analyzed contained high-scoring sites in *Picea abies* than in *Amborella trichopoda,* indicating the likelihood of a statistically significant enrichment also in *Picea abies.*

These data suggest that numerous immediate targets of SPT in *Arabidopsis* have been conserved, at least since the MRCA of living angiosperms, and probably since that of living seed plants. These targets include several genes involved in shade-avoidance and auxin-related processes (Reymond *et al.,* 2012). Two further plant hormone-related SPT targets identified in *Arabidopsis*, *PID* (Girin *et al.,* 2011) and *ARR1* (Irepan Reyes-Olalde *et al.,* 2017), were confirmed to be likely conserved as SPT targets in *Amborella trichopoda* (SI Tab. 1), though this confirmation was not possible in *Picea abies* due to a lack of sequence data from the upstream regions of the relevant genes.

SPATULA orthologs from a wide range of vascular plants are able to rescue Arabidopsis spatula mutants.

The conservation of the *in-vitro* biochemical properties of *Arabidopsis* SPT with orthologous proteins from *Amborella trichopoda* and *Picea abies* led us to ask whether the *in-vivo* activities of SPT might be similarly conserved between distant taxa. To investigate this possibility, we expressed in transgenic *Arabidopsis spt-11* loss-of-function mutants the coding sequences of a range of *SPT* orthologs from several angiosperms and gymnosperms, as well as the most closely-related coding sequence from the lycophyte *Selaginella moellendorffii.* We also included in these studies *Arabidopsis PIF5,* which along with its paralog *PIF4* was shown by Reymond *et al.* (2012) to be capable, under low red/far-red light-ratio conditions, of complementing the *spt* loss-of-function phenotype in the *Arabidopsis* gynoecium. Two versions of each of Smo*PIF* from *Selaginella moellendorrfii* and *AthPIF5* from *Arabidopsis* were included in these studies, one of which in each case, respectively termed *ΔSmoPIF* and *ΔAthPIF5,* had been truncated to remove its APB domain. All coding sequences tested were expressed from the fully functional 6.3-kb promoter sequence of *Arabidopsis SPT, as* defined by Groszmann *et al.* (2010).

The *Arabidopsis spt-11* mutant shows reduced carpel fusion over the style and stigma region, fails to generate a septum in the upper portion of the gynoecium, and after fertilization produces shorter fruits containing fewer seeds (Fig. 4A-C). At least 16 independent transformants were examined for each transgene construct tested, and untransformed mutants were grown for comparison in each batch of plants studied. All the constructs tested were

able, in at least some of the transgenic lines analyzed, to fully restore fusion of the gynoecium apex (Fig 4D-N, Tab. 2). There was no significant difference at the $p<0.05$ level in the efficiency of the constructs tested, compared to the *AthSPT::AthSPT* positive control construct, with the exceptions of: *pSPT::PIF5, pSPT::SmoPIF* and *pSPT:: ΔSmoPIF,* which proved less efficient than the others. The proportion of lines restored was higher in the N-terminally deleted versions *AthSPT:: ΔSmoPIF* and *AthSPT:: ΔAthPIF5* than in their respective full-length counterparts, which may indicate that the APB domain rendered the full-length versions of these transcription factors susceptible to a degree of negative regulation via phyB.

The constructs tested varied more markedly in their capacity to fully restore fruit-size and fertility (Fig. 5, Tab. 2). Accordingly, *AthSPT::AthALC, AthSPT::NcaSPT (*containing the *SPT* coding sequence from *Nymphaea caerulea)* and *AthSPT::PaxSPT-L (*containing the *SPT-like* coding sequence from *Petunia axillaris)* all complemented the *Arabidopsis spt-11* fruit-size phenotype with no significant difference in efficiency (at $p<0.05$) from the positive control construct. The constructs *AthSPT::AtrSPT, AthSPT::AthPIF5* and *AthSPT::ΔAthPIF5* complemented the *spt-11* fruit-size phenotype with intermediate efficiency (p-values between 0.001 and 0.05), while the remaining constructs tested showed very low efficiency compared to the positive control ($p<0.001$), with no complementation in any transformed line generated using *AthSPT::PabSPT (*containing the *SPT-like* coding sequence from *Picea abies).* Again, the constructs containing N-terminally deleted coding sequences *AthSPT::ΔAthPIF5* and *AthSPT::ΔSmoPIF* appeared more efficient at complementing the *spt-11* mutant fruit phenotype than the respective full-length versions, though a lack of significant difference in efficiency to the control construct was observed in the former case only. Taken globally, these results indicate the *in-vivo* biological activity of the *SPT* coding sequence in the *Arabidopsis* gynoecium and fruit to be conserved with widely diverged proteins from the *SPT* clade and even, to some extent, with *Arabidopsis PIF5* and with a *PIF*-like gene from the lycophyte *Selaginella moellendorffii.*

SPT is upregulated in the gynoecium in diverse angiosperms, though its precise expression pattern varies considerably between taxa.

We investigated *SPT* expression patterns by RNA *in-situ* hybridization in one asterid species and in two members of the ANA grade of basally diverging angiosperms. *Petunia axillaris* (asterids, Solanales, Solanaceae) contains one gene from each of the core-eudicot *SPT* and *SPT-like* clades identified in the analysis shown in Fig. 2, referred to here as *PaxSPT* (gene identifier = Peaxi162Scf00450g00124) and *PaxSPT-L* (gene identifier = Peaxi162Scf00503g00311). *PaxSPT* is expressed in the L1 cell-layer of both the floral meristem and stamen primordia (Fig. 6A-B), with a peak of expression also at the centre of the early gynoecium in the region in which the placenta will form (Fig. 6B-C). *PaxSPT* is likewise very highly expressed in the L1 cell-layer of the inflorescence meristem and in deeper cell layers in the central zone of this structure (Fig. 6A). At later developmental stages (Fig. 6D-F), high *PaxSPT* expression persists in the placenta and developing ovules, and is also clearly observed in the L1 cell-layer of the developing stigma, petals and stamens. Weaker expression of *PaxSPT* is apparent around the loculi of the anthers and in the stigma and gynoecium wall. *PaxSPT-L* is less strongly and widely expressed than *PaxSPT* (Fig. 6G-L), but likewise shows clear expression in the L1 cell-layer of the gynoecial primordium and at later stages in the L1 cell-layer of the upper style and stigma. Expression of *PaxSPT* and *PaxSPT-L* in the stigma resembles the expression of their likely orthologs in several other species of Solanaceae (Ortiz-Ramirez et al 2019).

The single *SPT* gene in the ANA-grade angiosperm *Amborella trichopoda, AtrSPT* (gene identifier = ATR0800G167), is expressed in the stamens of male flower buds, particularly in the zones that will form the anther loculi (Fig. 7A-B). In female buds, *AtrSPT* is expressed very strongly in carpel primordia and moderately in young tepals (Fig. 7C). Strong *AtrSPT* expression is also observed at later developmental stages (Fig. 7D-E) in the ovule, and in the adaxial tissues of the carpel wall that line the route of pollen tube growth from the stigmatic crest to the micropyle of the ovule. *NthSPT* (gene identifier = NYTH01284), the *SPT* ortholog from the ANA-grade angiosperm *Nymphaea thermarum,* is expressed during the formation of the anther loculi (Fig. 7F) and in the L1 cell-layer of all floral organs (Fig. 7F-G). *NthSPT* is also strongly expressed adaxially in the gynoecium and in the placenta and developing ovules (Fig. 7F-G).

Promoters of SPT orthologs show diverse activities in transgenic Arabidopsis.

The varying gynoecium expression patterns of *SPT* orthologues within angiosperms (c.f. Figs 6 and 7 with published data from Heisler et al., 2001; Zumajo-Cardona et al., 2017; Ortiz-Ramirez et al. 2019; and Chen et al., 2022) may reflect differing anatomical arrangements between taxa and/or differing transcriptional pathways upstream of *SPT*, but may also be due to differences in the cis-regulatory regions of the *SPT* orthologs themselves. To test the latter possibility, we made a comparison in transgenic *Arabidopsis* of the activity of *SPT* promoters from *Petunia axillaris* (genomic identifier = Peaxi162Scf00450g00124)*, Nymphaea caerulea* (scaffold 10, genomic location 28015916-28018936; genomic identifier = Nca_g63293), *and Pinus taeda* (scaffold 99620, genomic location 131404-135986; coding sequence identifier = PITA_20563)*,* all driving a GUS reporter gene. Promoter activities were compared to a positive-control construct containing the 6.3-kb upstream sequence of *Arabidopsis SPT* (Groszmann *et al.,* 2010). A similar length of upstream-flanking sequence was taken from the *SPT* ortholog tested from *Petunia axillaris.* Two versions of the putative *SPT* promoter from *Nymphaea caerulea* were tested, one of which included the entire upstream intergenic region of 1.4 kb, while the other included this sequence as well as part of the adjacent upstream gene, measuring 6.2 kb in total. At the time at which these experiments were performed, only ~1.7 kb of sequence upstream of the *Pinus taeda SPT* ortholog was available from genomic databases*,* and so this relatively short sequence was tested as the putative *SPT* promoter.

The *AthSPT::GUS* control construct generated very high reporter activity in the stylar transmitting tissue, extending down into the replum within the ovary (Fig. 8A-B, Tab. 3), in good agreement with the known expression pattern of *AthSPT* (Heisler *et al.* 2001; Groszmann *et al.,* 2010). A reporter construct containing the promoter region of *PaxSPT* from *Petunia axillaris,* by contrast, showed high expression in valve tissues and no visible expression in the style (Fig. 8C-D, Tab. 3). The putative promoter region of *PtaSPT* from the gymnosperm *Pinus taeda* generated very high reporter activity in sepals and at the base of pedicels, but no expression in the gynoecium (Fig. 8E-F, Tab. 3). It should be noted, however, that the upstream DNA fragment tested was short in this case, and may not have included all of the cis-acting sequences present in the native gene. Finally, both versions of the orthologous promoter from the *NcaSPT* gene of the basal angiosperm *Nymphaea caerulea,* a close relative of *N. thermarum* (as used in Fig. 7) from the same subgroup of African

waterlilies, generated a low level of reporter activity in anthers, but none in the gynoecium (Fig. 8G-H, Tab. 3).

None of the heterologous *SPT* promoters tested generated a native *SPT*-like expression pattern in transgenic *Arabidopsis*, though all of the angiosperm promoters tested, with the exception of *N. caerulea*, showed some activity in the gynoecium. These data are consistent with possible evolutionary changes to *SPT* promoter activity, in addition to other factors, as mentioned above, which may have contributed to the diversification of *SPT* expression patterns within angiosperms.

Comparison of SPT promoters in core eudicots reveals novel motifs and structural rearrangements.

To attempt to identify elements in *SPT* promoters whose presence/absence or different spatial arrangements correlates with inter-species differences in *SPT* expression patterns, we compared promoters from 20 species of core eudicots, including seven species of Solanaceae (asterids), five species of Brassicaceae (rosids), and eight species from diverse families of other rosids. We identified ten motifs shared between two or more of the species under comparison, six of which contain consensus sequences known to interact with specific classes of transcription factors, including members of the bHLH, TCP, WRKY, HD-ZIP and C2C2 zinc-finger families (SI Fig. 7). The Brassicaceae species analyzed showed striking similarity in the spatial arrangement of conserved motifs (Fig. 9). Groszmann *et al.* (2010) previously compared three Brassicaceae *SPT* promoters and defined from these eight conserved regions. Most of the ten motifs identified in the present work fit within the Regions 1, 2, 5 and 6 identified by Groszmann *et al.* (2010), as shown in SI Fig 6. However, Motifs 4, 9 and 10 are newly identified sequences that occur in a well-conserved cluster between the Regions 5 and 6 defined by Groszmann *et al.* (2010).

We also found Solanaceae *SPT* promoters to share numerous motifs conserved with Brassicaceae. For example, the *Petunia axillaris SPT* promoter shares seven of the ten motifs identified in *Arabidopsis*, lacking only Motifs 2, 7 and 9 (Fig. 9). Groszmann *et al.* (2010) used a series of deletions of the *Arabidopsis SPT* promoter to define the functions of conserved promoter regions. Though some additional enhancers and suppressors were found

in upstream regions, reporter expression in the transmitting tissue and replum was lost when the *SPT* promoter was shortened from -260 bp to -180 bp, thus removing a motif containing a G-box (CACGTG) that partially overlapped an auxin response element (AuxRE, TGTCTC). A similar combined G-box/AuxRE motif present at -68 bp in the *Arabidopsis SPT* promoter was found by Groszmann *et al.* (2010) to be necessary for reporter-gene expression in the silique dehiscence zones. These hybrid motifs identified by Groszmann *et al.* (2010) were identified in the present study as Motifs 1 and 3 (Fig. 9). These two motifs were found to participate in an ordered cluster of Motifs 1-6-7-3 (from upstream to downstream), situated near the transcriptional start site of almost all Brassicaceae *SPT* genes (Fig. 9). Interestingly, three of these four motifs, Motifs 1, 3 and 6, are also found in a closely clustered formation in four of the five Solanaceae *SPT* promoters analyzed. However, this cluster occurs in Solanaceae *SPT* promoters in the order 6-1-3 (rather than the 1-6-(7)-3 order of Brassicaceae). This Solancaeae cluster is also typically augmented by the presence of Motif 4 near its proximal end. Motif 4 is universally present too in the Brassicaeae promoters analyzed, but in much more variable positions, and mostly far upstream of the well-conserved cluster of Motifs 1-6-7-3 near the transcriptional start site. Motif 7, which is present in Brassicaceae but absent in Solanaceae *SPT* promoters, resembles a typical HD-ZIP motif (consensus = cAATnATTG). The absence of this motif in Solanaceae therefore correlates with an absence of *SPT*-expression in the style, and while the *PaxSPT* promoter from the Solanaceae species *Petunia axillaris* is capable of directing strong GUS-reporter expression in the gynoecium of transgenic *Arabidopsis* (Fig. 8), such expression is absent from the style.

Taken together, these data show an underlying conservation of *SPT* promoter composition within core eudicots, particularly in terms of the presence/absence of conserved motifs. However, the relative positions of the motifs present varies highly between Brassicaceae (rosids) and Solanaceae (asterids). In particular, Brassicaceae *SPT* promoters contain a highly conserved order of ten conserved motifs, including a tight cluster of four motifs (Motifs 1-6-7-3) in the proximal promoter region. This juxtaposition of motifs is partially conserved in the Brassicales species *Carica papaya*, which contains three of the five conserved Brassicaeae motifs (1-3-7) near its transcriptional start site. Within this cluster, Motif 7 appears particularly interesting as it is conserved in Brassicales while absent from most other rosids and Solanaceae (asterids).

DISCUSSION

The SPATULA lineage likely arose from a clade of light-regulated transcription factors in a common ancestor of euphyllophytes.

In the present work, we have shown that the *SPT* lineage likely originated through the duplication of a *PIF*-like gene in a common ancestor of euphyllophytes (seed-plants plus ferns and their allies; Fig. 1). In the *SPT* lineage, the APB motif of the ancestral *PIF* gene was lost, which appears to be a defining feature of the *SPT* clade. Several *Arabidopsis* PIF proteins, including PIF1, PIF3, PIF4 and PIF5, are known to interact via their APB domain with the active Pfr form of the phyB photoreceptor and are consequently targeted for degradation in the proteasome (Shin et al., 2009). As PIF genes from the moss *Physcomitrium patens* and the liverwort *Marchantia polymorpha* are capable of rescuing the *Arabidopsis pif1 pif3 pif4 pif5* quadruple loss-of-function mutant (Inoue *et al.,* 2016; Xu and Hiltbrunner, 2017; Possart *et al.,* 2017) it seems that the PIF-phytochrome interaction has been conserved from an early stage in land plant evolution. Therefore, it seems reasonable to postulate that the PIF transcription factor that was ancestral to SPT in a common ancestor of euphyllophytes was similarly regulated by phytochrome, and that the loss of its APB domain caused this factor to escape from direct phytochrome-mediated light-regulation.

The SPATULA and ALCATRAZ lineages of Arabidopsis were generated in a large-scale duplication within Brassicales.

A further phylogenetic analysis in the present work has confirmed the conclusion of Groszmann *et al.* (2011) that the gene duplication leading to the *SPT* and *ALC* lineages of *Arabidopsis* took place within Brassicales, and probably corresponded to the At-β whole genome duplication. Our study reveals a complex history of gene duplications and losses in the angiosperm *SPT* clade (Fig. 2). In particular, a duplication near the base of core eudicots appears to have generated paralogous *SPT* and *SPT*-like lineages, both of which have persisted in many rosid and asterid taxa. One of these lineages (*SPT*) further duplicated within Brassicales to generate the *SPT* and *ALC* lineages present in Brassicaceae, Gyrostemonaceae and Cleomaceae, while the other (*SPT-like*) was lost at the base of Brassicaceae. An analysis of the genomic regions surrounding the *SPT* and *ALC* loci in Brassicaceae shows a strong conservation of synteny (SI Fig. 3), in agreement with the possibility that these lineages separated at the At-β event. A corollary of these conclusions is that genes named

"ALCATRAZ" from species outside Brassicaceae and closely related families should not be considered as direct orthologs of *ALC in Arabidopsis;* these genes appear to be equally closely related to both *SPT* and *ALC* in *Arabidopsis.*

SPATULA proteins show largely conserved biochemical properties between widely diverged land-plant groups.

In this work, we used yeast-2-hybrid analyses, protein-binding microarrays and plant transformation experiments to assess the degree to which SPT proteins from different plant lineages show similar biochemical properties. Our yeast-2-hybrid experiments (Fig.3) showed that the capacity of SPT to form homodimers, and to form heterodimers with transcription factors of the HEC/IND clade, are conserved between the *Arabidopsis* and *Amborella* lineages and therefore probably since the MRCA of living flowering plants. Protein-binding microarray assays furthermore showed that DNA-binding preferences of SPT proteins have been largely conserved since the MRCA of living seed plants (SI Fig. 5), while complementary *in-silico* analyses showed a very significant conservation of putative SPT-binding sites in orthologs of SPT-targets from the basal angiosperm *Amborella trichopoda* and the gymnosperm *Picea abies* (Tab. 1). Finally, *in-vivo* studies showed that *SPT* coding sequences from *Amborella trichopoda* and *Picea abies* were capable of almost perfectly replacing *SPT* function in the gynoecium in transgenic *Arabidopsis*, while sequences as far removed as the *PIF*-like ortholog of *SPT* from the lycophyte *Selaginella moellendorffii* could partially complement the same mutant phenotypes (Figs 4 and 5). These results indicate a high conservation of the biochemical properties of SPT proteins, at least since the MCRA of living seed plants.

Differential recruitment of SPATULA and its downstream pathways has likely contributed to the generation of morphological biodiversity in the angiosperm gynoecium, including the characteristic gynoecium structure of Brassicaceae.

SPT in *Arabidopsis* has been shown to control developmental processes through two distinct mechanisms. Firstly, *SPT* regulates tissue patterning in the gynoecium through the control of auxin and cytokinin signaling. This mechanism involves the direct transcriptional regulation of *PID* (Girin *et al.,* 2011) and *ARR1* (Irepan Reyes-Olalde *et al.,* 2017), and both of these SPT-targets have been shown in the present work to be likely conserved, at least since the

MRCA of living angiosperms (Tab. 1). The second major mechanism through which *SPT* regulates developmental processes occurs through the direct transcriptional control of genes involved in cell elongation, many of which also act in the process of shade avoidance in vegetative tissues. These cell-elongation/shade avoidance targets have also been shown, in the present work, to be likely conserved, at least since the MRCA of living seed plants (Tab. 1, SI Tab. 1).

Despite the apparent conservation of the downstream pathways associated with *SPT*, the expression patterns of *SPT* homologs and their associated mutant phenotypes, where known, vary markedly among angiosperms. Therefore, distinct patterns of recruitment of *SPT* and its downstream pathways to developmental processes represent a potentially important evolutionary mechanism for the generation of developmental biodiversity, particularly in organ systems whose anatomy varies considerably between angiosperm taxa, such as the gynoecium. Different patterns of deployment of *SPT* homologs and their associated downstream pathways may have been generated during gynoecium evolution by changes to pathways acting upstream of *SPT*, by changes to the *SPT* promoter itself, and/or by changes to the expression of *SPT*'s conserved co-factors such as the *HEC/IND* genes.

The gynoecia of all angiosperms examined to date show high levels of *SPT* expression in ovule tissues, and also in the placenta, where present. However, the functional significance of this expression is unknown, as no loss-of-function mutations in *SPT* homologs have yet revealed a phenotype in these tissues. The mutants analysed in this context include the *spt alc* double mutant of *Arabidopsis*, which shows wild-type ovule development (Groszmann *et al.,* 2011). However, multiple mutants of *Arabidopsis* that include loss-of-function alleles of both *SPT* and *ALC* together with those of related PIF genes such as *PIF1, PIF3, PIF4 and* PIF5 (Shin *et al.,* 2009; Leivar *et al.,* 2009), some of which are known to share immediate transcriptional targets with *SPT* (Reymond *et al.,* 2012), have yet to be constructed.

In ANA-grade angiosperms, the stigmatic surface typically bears multicellular striations (Endress and Igersheim 2000), or occasionally hairs (Taylor and Williams 2012), which function to help immobilize pollen grains. However, in more recently-evolved angiosperm groups, elongated unicellular stigmatic papillae generally play this role. High levels of *SPT*

expression have been found in the stigmatic papillar cells of the rosid *Cucumis sativus* (Cheng *et al.,* 2022) and, in this study (Fig. 6), the asterid *Petunia axillaris.* A similar pattern of *SPT* expression at the apex of the stigma was previously observed in two further Solanceae species: *Solanum lycopersicum and Capsicum annuum* (Ortiz-Ramirez *et al.,* 2019). A double knock-out mutant of both *SPT* homologs present in *Cucumis sativus* shows a lack of elongation in stigmatic papillae (Cheng *et al.,* 2022), and though gene-knockdowns using virus induced gene silencing failed to show a similar phenotype in Solanceae species, it has been acknowledged that this may have been due to incomplete effects of the knock-down procedure (Ortiz-Ramirez *et al.,* 2019). The conservation of *SPT* expression in the L1 cell-layer of the stigma between certain asterids and rosids suggest that this mechanism may have been conserved in the lineages concerned at least since the MRCA of core eudicots. The function of *SPT* in cell elongation of stigmatic papillae is therefore likely to be relatively ancient in angiosperms, though less ancient than the expression of this gene in ovules and placentae.

In most ANA-grade angiosperms, pollen tubes grow through secretion-filled canals or apertures to reach the ovules, though a short pollen transmitting tissue is present just below the stigmatic surface in Nymphaeaceae (Williams *et al.,* 2010). By contrast, in most eudicots, an extensive pollen transmitting tissue is present in the stigma, style and ovary. The development and cellular composition of transmitting tissue appears, however, to have undergone multiple changes during eudicot evolution (Lora *et al.,* 2016; Gotelli *et al.,* 2017), and consequently this tissue may be heterologous, or only partially homologous, between distantly related taxa. High levels of *SPT* expression are found in the stylar transmitting tissue of rosids including *Arabidopsis* (Heisler *et al.,* 2001) and *Cucumis sativus* (Cheng *et al.,* 2022), but not in Solanaceae (asterids) (Fig 6; Ortiz-Ramirez *et al.,* 2019), or in the basal eudicot *Bocconia frutescens* (Zumajo-Cardona *et al.,* 2017). Recruitment of *SPT* to form the stylar transmitting tract may, therefore, have occurred in a common ancestor of eurosids 1 (including *Cucumis*) and eurosids 2 (including *Arabidopsis).*

In Brassicaceae, the gynoecium of two fused carpels typically emerges from the centre of the floral meristem, while the stigma, style, replum, placentae and ovules later arise from the CMM that forms at the margins of these organs. A similar anatomical arrangement is present in some Cleomaceae (Watson and Dallwitz, 1992 onwards), the sister family to Brassicaceae.

However, in somewhat more distant families of Brassicales (see Fig. 2B) such as Gyrostemonaceae, the anatomical arrangement of the gynoecium differs markedly. In *Gyrostemon ramulosus*, for example (see Fig. 2B for the phylogenetic placement of this species), no replum is present, and the placenta forms on the flanks of the floral meristem, facing the incompletely fused carpels (Hufford 1996). *SPT* is highly expressed in all the tissues that emerge from the CMM in *Arabidopsis*, and contributes to the development of these both by regulating tissue patterning via hormone signaling (Girin *et al.,* 2011; Irepan Reyes-Olalde *et al.,* 2017) and by positively regulating cell extension, notably in the style and replum (Reymond *et al.,* 2012). Elements of the characteristic gynoecium and fruit structure present in *Arabidopsis* may thus have emerged through changes to *SPT*'s expression pattern in a common ancestor of Brassicaceae and Cleomaceae, sometime after the gene duplication that separated the *SPT* and *ALC* lineages (Fig. 2B). In this work, we show a high conservation of motifs of potential functional importance between *SPT* promoters from a wide range of core eudicots including both asterids and rosids (Fig. 9). However, the juxtaposition of these conserved motifs varies markedly between taxa. A highly conserved order of ten motifs, including a tight cluster of four motifs in the proximal promoter region, correlates with and may have been important for the origin of the particular *SPT* expression pattern found in Brassicaceae. In particular, Motif 7 within this cluster (Fig 10 and SI Fig 7), which contains a likely HD-ZIP binding site, was found to be largely specific to Brassicales.

ACKNOWLEGEMENTS

ACRS was supported by a BECAL studentship from the Government of Paraguay. We thank Patrice Morel and Virginie Battu for help and advice with *in-situ* hybridization procedures. We thank Gildas Gâteblé, Bruno Fogliani, Paula Rudall, Carlos Magdalena, Charles Dana Nelson and Michiel Vandenbussche for generously supplying plant material. The RDP Laboratory is financially supported by its parent organisations: CNRS, INRAe, ENS-Lyon and Université Claude Bernard-Lyon. Work in Centro Nacional de Biotecnología was supported Spanish MICIN grant PID2020-119451GB-100. We acknowledge the use of microscopy facilities of the common imagery platform of the UAR344 federation of bioscience laboratories in Lyon-Gerland.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/aob and consist of the following.

Figure S1. Supplemental data to the phylogenetic analysis of the SPT/PIF clade in representative land plants shown in Fig. 1.

Figure S2. Supplemental data to the phylogenetic analysis of the SPATULA clade in angiosperms shown in Fig. 2.

Figure S3. Microsynteny associated with SPT and ALC loci in selected Brassicaceae species.

Figure S4. Amino-acid alignments showing a conserved acidic domain and a partially conserved amphipathic helix in SPATULA proteins and related molecules.

Figure S5. Top-scoring position weight matrices for SPATULA proteins from Arabidopsis thaliana, Amborella trichopoda and Picea abies.

Figure S6. Comparison of conserved regions and motifs in core-eudicot SPT promoters identified in the present study and by Groszmann et al.(2010).

Figure S7. Conserved motifs in upstream regions of core-eudicot SPATULA orthologs identified using MEME.

Table S1. High-scoring SPATULA binding sites identified using MotifScan.

Table S2. Primers used in the PCR amplification of promoters and coding sequences.

Table S3. Statistical analysis of SPATULA target genes and sites.

LITERATURE CITED

Altschul SF, Madden TL, Schaffer AA, *et al.* 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

Alvarez J, Smyth DR. **1999**. CRABS CLAW and SPATULA, two Arabidopsis genes that control carpel development in parallel with AGAMOUS. *Development* **126**: 2377–2386.

Anisimova M, Gascuel O. **2006**. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology* **55**: 539–552.

Ballester P, Martinez-Godoy MA, Ezquerro M, *et al.* 2021. A transcriptional complex of NGATHA and bHLH transcription factors directs stigma development in Arabidopsis. *Plant Cell* **33**: 3645–3657.

Berger MF, Bulyk ML. **2009**. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols* **4**: 393–411.

Bernal-Gallardo JJ, Zuniga-Mayo VM, Marsch-Martinez N, de Folter S. **2023**. Novel Roles of SPATULA in the Control of Stomata and Trichome Number, and Anthocyanin Biosynthesis. *Plants-Basel* **12**: 596.

de Bossoreille S, Morel P, Trehin C, Negrutiu I. **2018**. REBELOTE, a regulator of floral determinacy in Arabidopsis thaliana, interacts with both nucleolar andnucleoplasmic proteins. *Febs Open Bio* **8**: 1636–1648.

Byng JW, Chase MW, Christenhusz MJM, *et al.* 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181**: 1–20.

Cheng Z, Song X, Liu X, *et al.* 2022. SPATULA and ALCATRAZ confer female sterility and fruit cavity via mediating pistil development in cucumber. *Plant Physiology* **189**: 1553–1569.

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

Comte N, Morel B, Hasic D, *et al.* 2020. Treerecs: an integrated phylogenetic tool, from sequences to reconciliations. *Bioinformatics* **36**: 4822–4824.

Endress PK, Igersheim A. **2000**. Gynoecium structure and evolution in basal angiosperms. *International Journal of Plant Sciences* **161**: S211–S223.

Ferrandiz C, Fourquin C, Prunet N, *et al.* 2010. Carpel Development In: Kader JC, Delseny M, eds. *Advances in Botanical Research, Vol 55*.1–73.

Franco-Zorrilla JM, Lopez-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R. **2014**. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 2367–2372.

Friml J, Yang X, Michniewicz M, *et al.* 2004. A PINOID-dependent binary switch in apical-basal PIN polar targeting directs auxin efflux. *Science* **306**: 862–865.

Girin T, Paicu T, Stephenson P, *et al.* 2011. INDEHISCENT and SPATULA Interact to Specify Carpel and Valve Margin Tissue and Thus Promote Seed Dispersal in Arabidopsis. *Plant Cell* **23**: 3641–3653.

Godoy M, Franco-Zorrilla JM, Perez-Perez J, Oliveros JC, Lorenzo O, Solano R. **2011**. Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors. *Plant Journal* **66**: 700–711.

Gotelli MM, Lattar EC, Zini LM, Galati BG. **2017**. Style morphology and pollen tube pathway. *Plant Reproduction* **30**: 155–170.

Gouy M, Guindon S, Gascuel O. **2010**. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution* **27**: 221–224.

Gremski K, Ditta G, Yanofsky MF. **2007**. The HECATE genes regulate female reproductive tract development in Arabidopsis thaliana. *Development* **134**: 3593–3601.

Groszmann M, Bylstra Y, Lampugnani ER, Smyth DR. **2010**. Regulation of tissue-specific expression of SPATULA, a bHLH gene involved in carpel development, seedling germination, and lateral organ growth in Arabidopsis. *Journal of Experimental Botany* **61**: 1495–1508.

Groszmann M, Paicu T, Alvarez JP, Swain SM, Smyth DR. **2011**. SPATULA and ALCATRAZ, are partially redundant, functionally diverging bHLH genes required for Arabidopsis gynoecium and fruit development. *The Plant Journal* **68**: 816–829.

Groszmann M, Paicu T, Smyth DR. **2008**. Functional domains of SPATULA, a bHLH transcription factor involved in carpel and fruit development in Arabidopsis. *The Plant Journal* **55**: 40–52.

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

Heisler MGB, Atkinson A, Bylstra YH, Walsh R, Smyth DR. **2001**. SPATULA, a gene that controls development of carpel margin tissues in Arabidopsis, encodes a bHLH protein. *Development* **128**: 1089–1098.

Hufford L. **1996**. Developmental morphology of female flowers of Gyrostemon and Tersonia and floral evolution among Gyrostemonaceae. *American Journal of Botany* **83**: 1471–1487.

Ichihashi Y, Horiguchi G, Gleissberg S, Tsukaya H. **2010**. The bHLH Transcription Factor SPATULA Controls Final Leaf Size in Arabidopsis thaliana. *Plant and Cell Physiology* **51**: 252–261.

Inoue K, Nishihama R, Kataoka H, *et al.* 2016. Phytochrome Signaling Is Mediated by PHYTOCHROME INTERACTING FACTOR in the Liverwort *Marchantia polymorpha*. *Plant Cell* **28**: 1406–1421.

Irepan Reyes-Olalde J, Zuniga-Mayo VM, Serwatowska J, *et al.* 2017. The bHLH transcription factor SPATULA enables cytokinin signaling, and both activate auxin biosynthesis and transport genes at the medial domain of the gynoecium. *Plos Genetics* **13**: e1006726.

Josse E-M, Gan Y, Bou-Torrent J, *et al.* 2011. A DELLA in Disguise: SPATULA Restrains the Growth of the Developing Arabidopsis Seedling. *Plant Cell* **23**: 1337–1351.

Karimi M, Bleys A, Vanderhaeghen R, Hilson P. **2007**. Building blocks for plant gene assembly. *Plant Physiology* **145**: 1183–1191.

Leivar P, Tepperman JM, Monte E, Calderon RH, Liu TL, Quail PH. **2009**. Definition of Early Transcriptional Circuitry Involved in Light-Induced Reversal of PIF-Imposed Repression of Photomorphogenesis in Young *Arabidopsis* Seedlings. *Plant Cell* **21**: 3535–3553.

Lora J, Hormaza JI, Herrero M. **2016**. The Diversity of the Pollen Tube Pathway in Plants: Toward an Increasing Control by the Sporophyte. *Frontiers in Plant Science* **7**: 107.

Mabry ME, Brose JM, Blischak PD, *et al.* 2020. Phylogeny and multiple independent whole-genome duplication events in the Brassicales. *American Journal of Botany* **107**: 1148–1164.

Morel P, Heijmans K, Ament K, *et al.* 2018. The Floral C-Lineage Genes Trigger Nectary Development in Petunia and Arabidopsis. *Plant Cell* **30**: 2020–2037.

Nguyen NTT, Contreras-Moreira B, Castro-Mondragon JA, *et al.* 2018. RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Research* **46**: W209–W214.

Ortiz-Ramirez CI, Giraldo MA, Ferrandiz C, Pabon-Mora N. **2019**. Expression and function of the bHLH genes ALCATRAZ and SPATULA in selected Solanaceae species. *Plant Journal* **99**: 686–702.

Ortiz-Ramirez CI, Plata-Arboleda S, Pabon-Mora N. **2018**. Evolution of genes associated with gynoecium patterning and fruit development in Solanaceae. *Annals of Botany* **121**: 1211–1230.

Pabon-Mora N, Wong GK-S, Ambrose BA. **2014**. Evolution of fruit development genes in flowering plants. *Frontiers in Plant Science* **5**: 300.

Penfield S, Josse EM, Kannangara R, Gilday AD, Halliday KJ, Graham IA. **2005**. Cold and light control seed germination through the bHLH transcription factor SPATULA. *Current Biology* **15**: 1998–2006.

Pfannebecker KC, Lange M, Rupp O, Becker A. **2017**. Seed Plant-Specific Gene Lineages Involved in Carpel Development. *Molecular Biology and Evolution* **34**: 925–942.

Possart A, Xu T, Paik I, *et al.* 2017. Characterization of Phytochrome Interacting Factors from the Moss *Physcomitrella patens* Illustrates Conservation of Phytochrome Signaling Modules in Land Plants. *Plant Cell* **29**: 310–330.

Rajani S, Sundaresan V. 2001. The Arabidopsis myc/bHLH gene ALCATRAZ enables cell separation in fruit dehiscence. *Current Biology* **11**: 1914-1922.

Reymond MC, Brunoud G, Chauvet A, *et al.* 2012. A Light-Regulated Genetic Module Was Recruited to Carpel Development in Arabidopsis following a Structural Change to SPATULA. *Plant Cell* **24**: 2812–2825.

Shin J, Kim K, Kang H, *et al.* 2009. Phytochromes promote seedling light responses by inhibiting four negatively-acting phytochrome-interacting factors. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 7660–7665.

Tani E, Tsaballa A, Stedel C, *et al.* 2011. The study of a SPATULA-like bHLH transcription factor expressed during peach (Prunus persica) fruit development. *Plant Physiology and Biochemistry* **49**: 654–663.

Taylor ML, Williams JH. **2012**. Pollen tube development in two species of *Trithuria* (Hydatellaceae) with contrasting breeding systems. *Sexual Plant Reproduction* **25**: 83–96.

Vialette-Guiraud ACM, Alaux M, Legeai F, *et al.* 2011. Cabomba as a model for studies of early angiosperm evolution. *Annals of Botany* **108**: 589–598.

Williams JH, McNeilage RT, Lettre MT, Taylor ML. **2010**. Pollen tube growth and the pollen-tube pathway of *Nymphaea odorata* (Nymphaeaceae). *Botanical Journal of the Linnean Society* **162**: 581–593.

Workman CT, Yin YT, Corcoran DL, Ideker T, Stormo GD, Benos PV. **2005**. enoLOGOS: a versatile web tool for energy normalized sequence logos. *Nucleic Acids Research* **33**: W389–W392.

Xu T, Hiltbrunner A. **2017**. PHYTOCHROME INTERACTING FACTORs from *Physcomitrella patens* are active in *Arabidopsis* and complement the *pif* quadruple mutant. *Plant Signaling & Behavior***12**: e1388975.

Zumajo-Cardona C, Ambrose BA, Pabon-Mora N. **2017**. Evolution of the SPATULA/ALCATRAZ gene lineage and expression analyses in the basal eudicot, Bocconia frutescens L. (Papaveraceae). *Evodevo* **8**: 5.

LEGENDS TO FIGURES

Figure 1. Phylogenetic analysis of the *SPT/PIF* clade in major land plant taxa. **A**. Phylogenetically reconciled tree with aLRT branch support values indicated at nodes. **B.** Species phylogeny used in phylogenetic reconciliation. **C.** Alignment of Active Phytochome Binding (APB) domains (lost from the *SPT* sub-clade) in representative sequences from the phylogeny. The correspondence of gene names to species names, the amino-acid alignment and the phylogeny from (A) before reconciliation are given in SI Fig. 1.

Figure 2. Phylogenetic analysis of the *SPATULA* clade in angiosperms**. A.** Summary of reconciled phylogeny with two sub-clades shown in full (boxes). LRT branch support values are indicated at nodes. **B.** Schematic species phylogeny showing the positions of selected gene duplications and losses. The correspondence of gene names to species names, the full species phylogeny used in phylogenetic reconciliation, the amino-acid alignment and the full phylogeny before and after reconciliation are given in SI Fig 2.

Figure 3. Yeast-2-hybrid assays of interactions involving *Amborella trichopoda SPT* and *HEC/IND* orthologs. **A.** Interaction matrices on positive-control media (left panel) and selective media (right panel) containing 3-AT (1 mM). **B. S**ummary of interactions. EV = empty vector (negative control on selective medium); BD = DNA-binding domain, AD = activation domain.

Figure 4. Complementation of the *Arabidopsis thaliana spt-11* carpel-apex mutant phenotype using coding sequences of selected *SPT* orthologs and related *PIF* genes. **A.** Wild-type carpel apex. **B.-C.** *spt-11* style and stigma. **D.** *spt-11* mutants transformed with *AthSPT* (positive control). **E. – N.** s*pt-11* mutants showing rescued carpel development following transformation using the constructs: (E) *pPST::AtrSPT,* (F) *pSPT::NcaSPT, (*G) *pSPT::PabSPT,* (H) *pSPT::PaxSPT,* (I) *pSPT::PIF5,* (J) *pSPT:: ΔPIF5,* (K) *pSPT::SmoPIF,* (L) *pSPT:: SmoΔPIF.* (M) *pSPT::AthALC,* (N) *pSPT::PaxSPT-L.*

Figure 5. Complementation of the *Arabidopsis thaliana spt-11* silique mutant phenotype using coding sequences of selected *SPT* orthologs and related *PIF* genes. **A.** *AthSPT* (positive control). **B. – K.** *spt-11* mutants showing complementation after transformation with: (B) *pPST::AtrSPT,* (C) *pSPT::NcaSPT,* (D) *pSPT::PaxSPT.* (E) *pSPT::PIF5* (F) *pSPT::ΔPIF5.* (G) *pSPT::SmoPIF,* (H) *pSPT::SmoΔPIF.* (I) *pSPT::AthALC,* (J) *pSPT::PaxSPT-L,* (K) *with pSPT::PabSPT* (partial rescue only). Untransformed siliques are shown at top-left.

Figure 6. Expression of *SPT* homologs in longitudinal sections of sequential stages in *Petunia axillaris* flower bud development. **A.-F.** *PaxSPT* **G.-L.** *PaxSPT-L.* Key: fm = floral

meristem g = gynoecium, im = inflorescence meristem, o = ovules, p = petal, pl = placenta, pp = petal primordium, sp = sepal primordium, st = stamen, stp = stamen primordium.

Figure 7. Expression of *SPT* orthologs from ANA-grade angiosperms. **A.-B.** *Amborella trichopoda* male flower buds, **C.-E.** *Amborella trichopoda* female flower buds, **F.-G.** *Nymphaea thermarum* flower buds. Key: c = carpel, cp = carpel primordium, o = ovule, op = ovule primordia, p = petal, st = stamen, sc = stigmatic crest t = tepal, tp = tepal primordium.

Figure 8: GUS reporter expression in the *Arabidopsis* gynoecium using promoters of *SPT* orthologs from various angiosperms and gymnosperms. **A**.-**B.** *Ath_pSPT* from *A. thaliana* (positive control). **C**.-**D.** *Pax_pSPT* from *Petunia axillaris,* **E.-F.** *Pta_pSPT* from *Pinus taeda,* **G.** *Nca_pSPT* 1401-bp fragment from *Nymphaea caerulea,* **H.** *Nca_pSPT* 6277-bp fragment *from Nymphaea cerulea.*

Figure 9. Structural analysis of upstream sequences from 20 core-eudicot *SPT* orthologs. Of the ten conserved motifs detected (SI Fig. 7), only Motifs 1 and 8 were present in all sequences. Potri = *Populus trichocarpa,* Rco = *Ricinus communis,* Sapur = *Salix purpurea ,* Glyma = *Glycine max,* MDP = *Malus domestica ,* Csa = *Cucumis sativus,* Egra = *Eucalyptus grandis ,* Capa = *Carica papaya ,* Ath = *Arabidopsis thaliana ,* Araha = *Arabidopsis halleri ,* Aly = *Arabidopsis lyrata ,* Bostr = *Boechera stricta ,* Bra = *Brassica rapa ,* Carub = *Capsella rubella ,* Esal = *Eutrema salsugineum ,* Paxi = *Petunia axillaris ,* Ntab = *Nicotiana tabacum ,* Nsyl = *Nicotiana sylvestris ,* Cann = *Capsicum annuum ,* Slyc = *Solanum lycopersicum.*

Table 1. Summary of the high-scoring SPT-binding sites in putative target promoters in *Arabidopsis, Amborella trichopoda* and *Picea abies*. Full data are given in SI Tab. 1.

		Arabidopsis thaliana		***Amborella trichopoda***		***Picea abies***	
Orthogroup		Gene ID	No. sites	Gene ID	No. sites	Gene ID	No. sites
1	phyB	AT2G18790.1	5	AmTrV1scaffold00003.45	0	MA_10435530g0010	1
2	bHLH Gp. 15	AT1G02340.1	1	AmTrV1scaffold00039.9	0	MA_26114g0010	1
		AT2G46970.1	2				
3	KDR	AT1G26945.1	0	AmTrV1scaffold00010.380	2	MA_9119217g0010	1
4	HDZip3	AT2G44910.1	4	AmTrV1scaffold00111.49	0	MA_57689g0010	0
		AT3G60390.1	1				
		AT4G16780.1	1				
		AT4G17460.1	2				
5	AFB1	AT4G03190.1	0	AmTrV1scaffold00016.85	2	MA_14836g0010	2
6	Aux/IAA20	AT2G46990.1	2	AmTrV1scaffold00045.141	4	no clear ortholog	ND
7	SAUR14	AT4G38840.1	2	AmTrV1scaffold17.24	1	MA_10431311g0020	0
				AmTrV1scaffold00245.2	0		
8	BR6OX2	AT3G30180.1	0	AmTrV1scaffold00047.151	0	MA_31668g0010	2
				AmTrV1scaffold00047.152	0		
9	AFP3/4	AT3G02140.1	3	AmTrV1scaffold00013.244	12	MA_2575g0010	8
				AmTrV1scaffold00154.36	1		
10	ATL5	AT3G62690.1	1	AmTrV1scaffold00021.115	2	no clear ortholog	ND
11	RING	AT1G19310.1	3	AmTrV1scaffold00059.139	0	MA_127251g0010, trunc.	ND
12	DIT2.1	AT5G64290.1	0	AmTrV1scaffold00008.134	0	MA_31952g0020	2
13	AtBHLH149	AT1G09250.1	0	AmTrV1scaffold00003.223	2	MA_9135164g0010	1
14	GT-3a	AT5G01380.1	4	AmTrV1scaffold00048.218	1	MA_7129732g0010	1
15	FLP1	AT4G31380.1	2	AmTrV1scaffold00057.274	1	no clear ortholog	ND
16	AtCXE6	AT1G68620.1	3	AmTrV1scaffold00010.6	1	MA_42231g0010	0
17	Triacylglycerol lipase	AT5G24200.1	0	AmTrV1scaffold00021.257	0	no clear ortholog	ND
18	Expressed protein	AT1G16850.1	1	no clear ortholog	ND	no clear ortholog	ND
19	PID	AT2G34650.1	2	AmTrV1scaffold00092.12	2	MA_935763g0010, trunc.	ND
20	ARR1	AT3G16857.2	1	AmTrV1scaffold00057.85	1	MA_8982282g0010, trunc.	ND

Table 2. Complementation of *Arabidopsis spt-11* mutants using diverse coding sequences under the control of the *Arabidopsis SPT* promoter.

Construct	Total number of transformants analysed	Number of transformants showing fused (wild-type) gynoecium	Percentage of transformants showing fused (wild-type) gynoecium	p-value (compaison with positive control construct)	Number of transformants showing wild-type fruit length	Percentage of transformants showing wild-type fruit length	p-value (comparison with positive control construct)
pSPT::AthSPT (pos. control)	21	21	100	N/A	18	85.7	N/A
pSPT::AtrSPT	25	20	80	0.0536	11	44	**0.0054**
pSPT::NcaSPT	23	21	91.3	0.4894	18	78.3	0.701
pSPT::PabSPT	25	21	84	0.1142	0	0	**<0.0001**
pSPT::PaxSPT	22	19	86.4	0.2326	4	18.2	**<0.0001**
pSPT::AthPIF5	25	19	76	**0.0247**	13	52	**0.026**
pSPT:: ΔAthPIF5	22	20	90.9	0.4884	13	59.1	0.0883
pSPT::SmoPIF	25	12	48	**<0.0001**	8	32	**0.0003**
pSPT:: ΔSmoPIF	19	15	78.9	**0.0424**	6	31.6	**0.0009**
pSPT::AthALC	20	19	95	0.4878	16	80	0.6965
pSPT::PaxSPT-L	16	16	87.5	1	13	81.3	1

P-values of <0.05, indicating constructs that gave a significantly lower level of phenotypic complementation than the positive control, are underlined.

Table 3. Gus reporter assays in transgenic *Arabidopsis*

Gene	Coordinates of promoter fragment tested relative to translational initiation codon	No. of transformants	No. of transformants showing GUS-staining	Tissues showing GUS staining
Arabidopsis SPT (positive control)	-6253 to - 1 bp	17	14	sepals, petals, carpel margins, ovary wall, style, stigma
Nymphaea carulea SPT	-1401 to -1 bp	17	1	young floral buds
Nymphaea carulea SPT	-6277 to -1 bp	25	7	young floral buds, anthers
Petunia axillaris SPT	-5006 to -1 bp	19	11	sepals, ovary wall
Pinus taeda SPT	-1659 to -1 bp	25	18	sepals, petals, base of pedicel

A

bHLH065 PIF5 (AthPIF5)
bHLH009 PIF4
Atr00039.9
bHLH015 PIF1
TBA00015182
bHLH124 PIL1
bHLH132 PIF6
bHLH008 PIF3
Atr00045.145
bHLH024 SPT(AthSPT)
ALC (AthALC)
Atr00046.26 (AtrSPT)
MA 26114g0010 (PabSPT)
TBA00009751
Afi s0093.g043373
Scu s0050.g013831
At5g61270 PIF7
Atr00110.8
Atr00106.97
bHLH016 At4g00050
CMI00011119
TBA00020365
MA 29186g0010
Afi s0079.g038322
Scu s0187.g025097
Smo114141 (SmoPIF)
Pp3c2 8960V3.1 BUA
Pp3c1 38820V3.1 BUA
Pp3c14 23520V3.15 BUA
Pp3c17 21890V3.1 BUA

100 77 92 78 92 99 82 16 85 60 94 95 83 96 0 98 70 97 97 95 0 85 91 95 99 0 48 0 98

Euphyllophyte remaining PIF clade

APB domain

DNA-binding domain

Euphyllophyte SPT clade

Euphyllophyte PIF7 clade

Lycophyte PIF clade

Bryophyte PIF clade

0.5

B

Amborella trichopoda
Arabidopsis thaliana
Picea abies
Taxus baccata
Cycas micholtzii
Salvinia cucullata
Azolla filiculoides
Selaginella moellendorffii
Physcomitrium patens

C

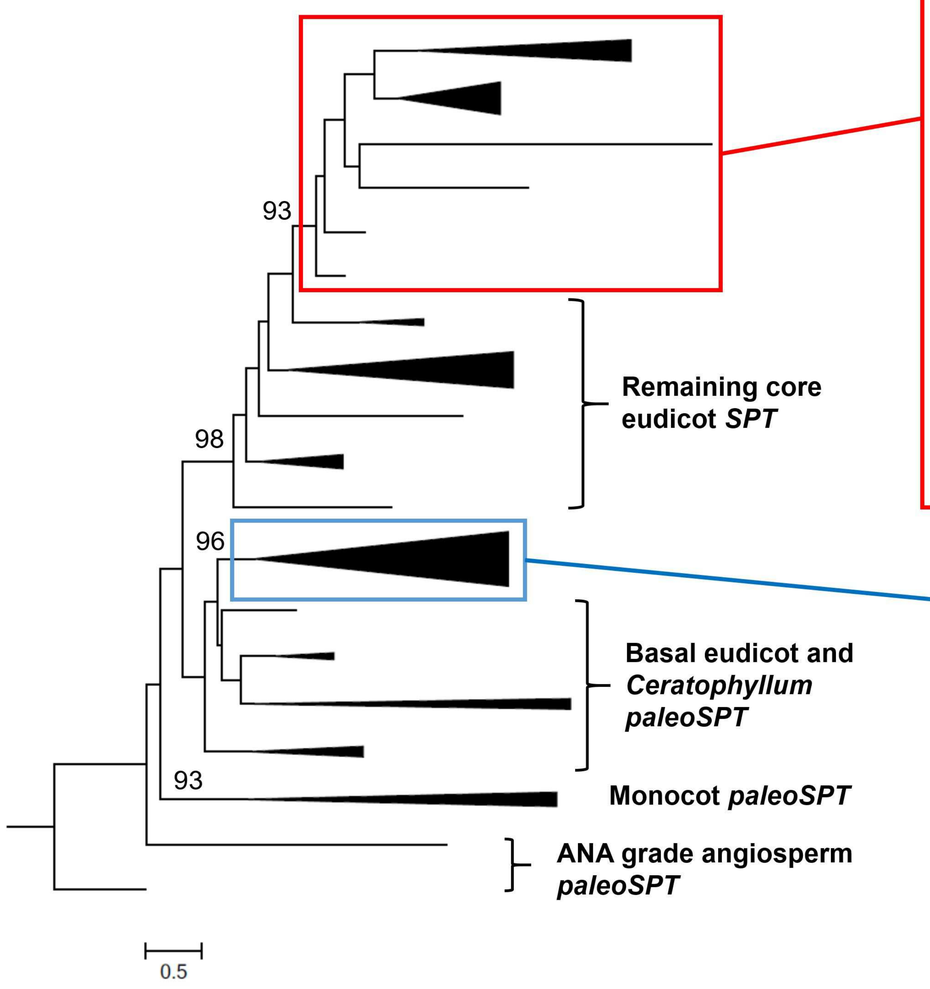
APB domain

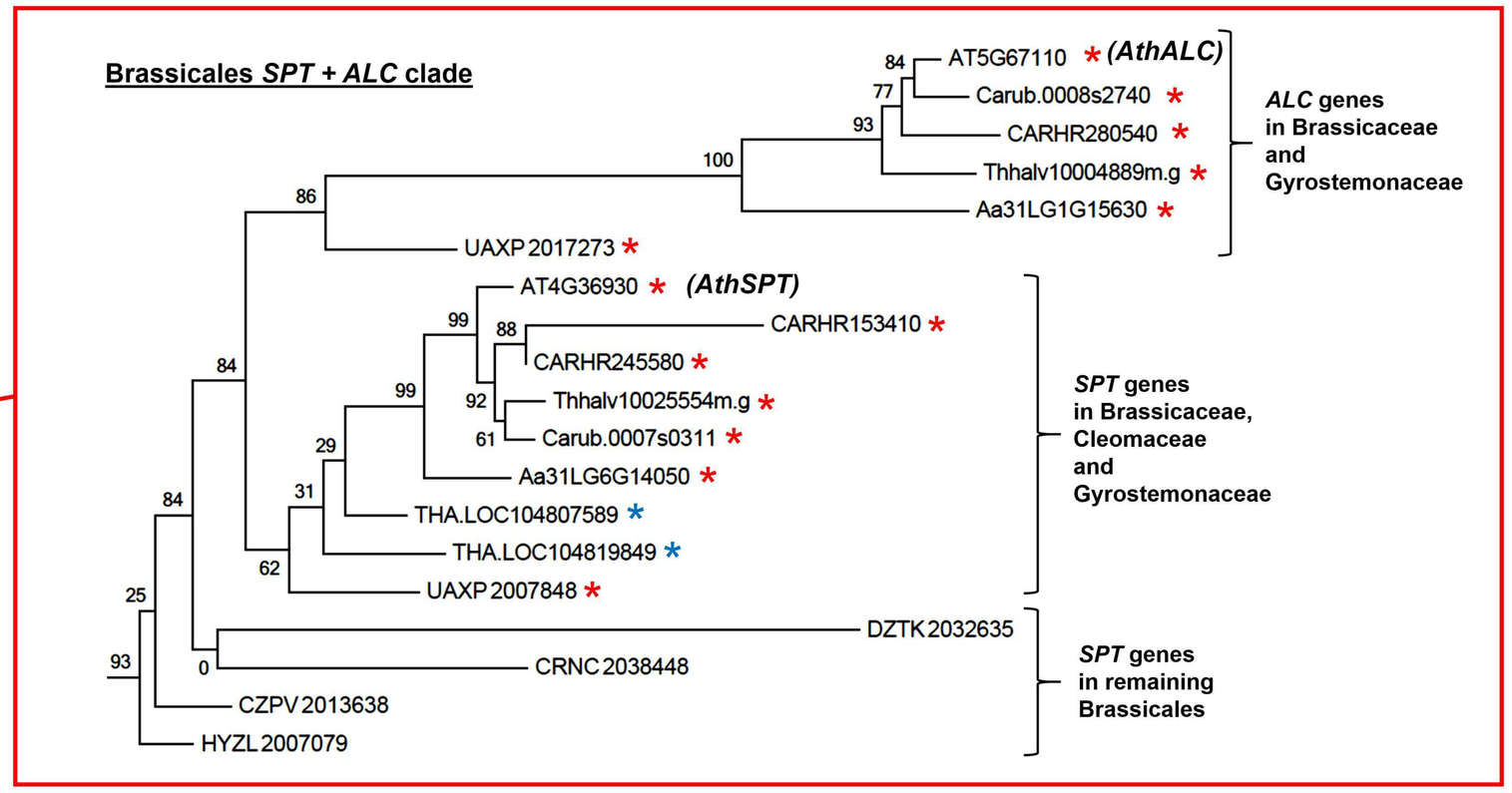
```
bHLH065_PIF5          IRP EDELVELLWRDGQV--VLQSQ ARR
bHLH009_PIF4          IRP QDELVELLWRDGQV--VLQSQ THR
Atr00039.9            APS EDELVELVWQDGQV--MMQKT SRR
bHLH015_PIF1          MGE DDDLMELLWQNGQVVVQNQRL HTK
bHLH124_PIL1          KLK DEDYMELVCENGQILAKIRRP KNN
bHLH132_PIF6          RLS DQEYMELVFENGQI--LAKGQ RSN
bHLH008_PIF3          SPP VDEVVELVWENGQI--STQSQ SSR
Atr00045.145          KWP DHEFVELLWENGQI--VLQGQ TSR
At5g61270_PIF7        -MS NYGVKELTWENGQL--TVHGL GDE
Atr00110.8            SIS DYHVAELTWENGHL--AMHGL ESV
Atr00106.97           LES DYQVAELTWENGHL--TMHGL GPP
bHLH016_At4g00050     PIL DYEVAELTWENGQL--GLHGL G--
CMI00011119           VMI KANLHTLGIVSGPT--NVQRV RQH
TBA00020365           VPD FDQVTELTWKDGHL--AMQEL GSR
MA_29186g0010         MPD YEQLAELIWKDGHP--AMQDL GSR
Afi_s0079.g038322     FVP DRDFEELCWEDGQLLVVMQDL GTT
Scu_s0187.g025097     VPD YNSFEELCYEDGRLL-VMQCL GSK
Smo114141             KFC ENDFSELIWTSDGSLGVMQVG GNK
Pp3c2_8960V3.1_BUA    VPD QDTLEAGRIDNGHI--VMQGQ TSK
```

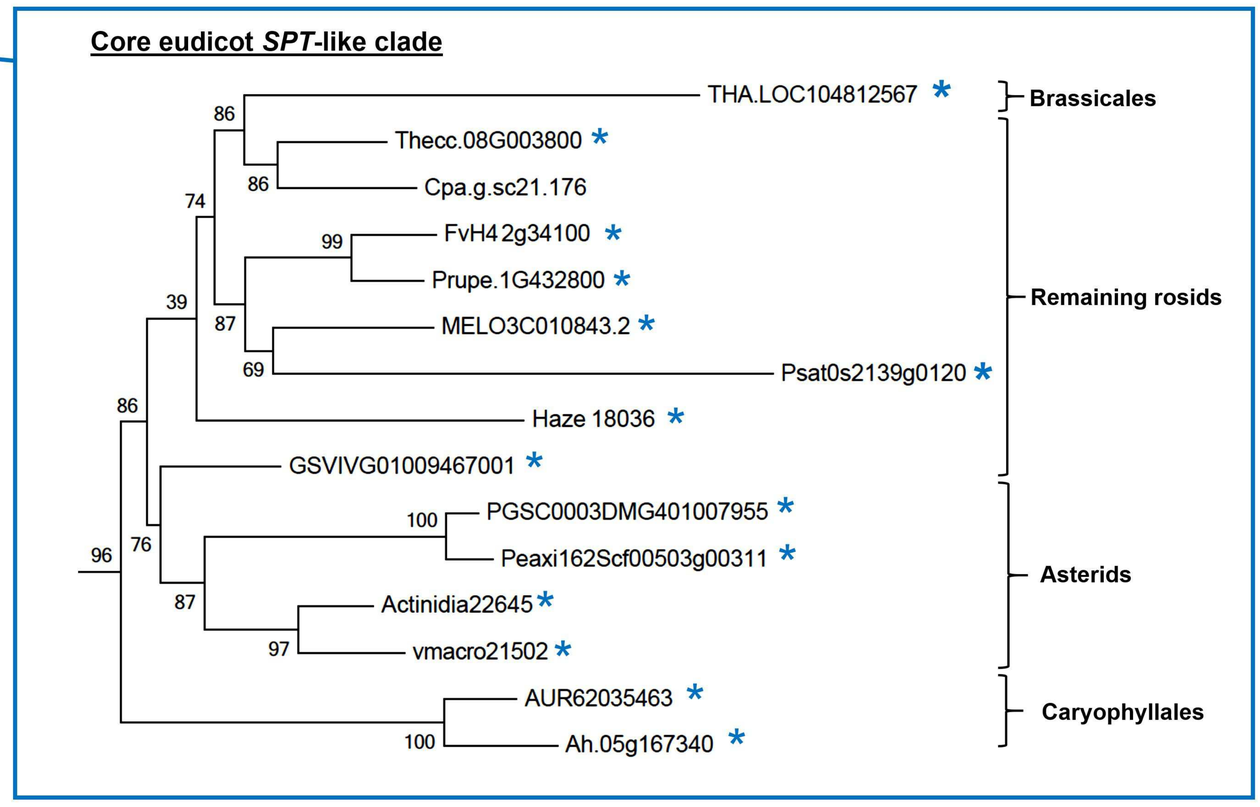
● Loss of APB domain in *SPT* lineage

angiosperms
gymnosperms
monilophytes
lycophytes
bryophytes

A



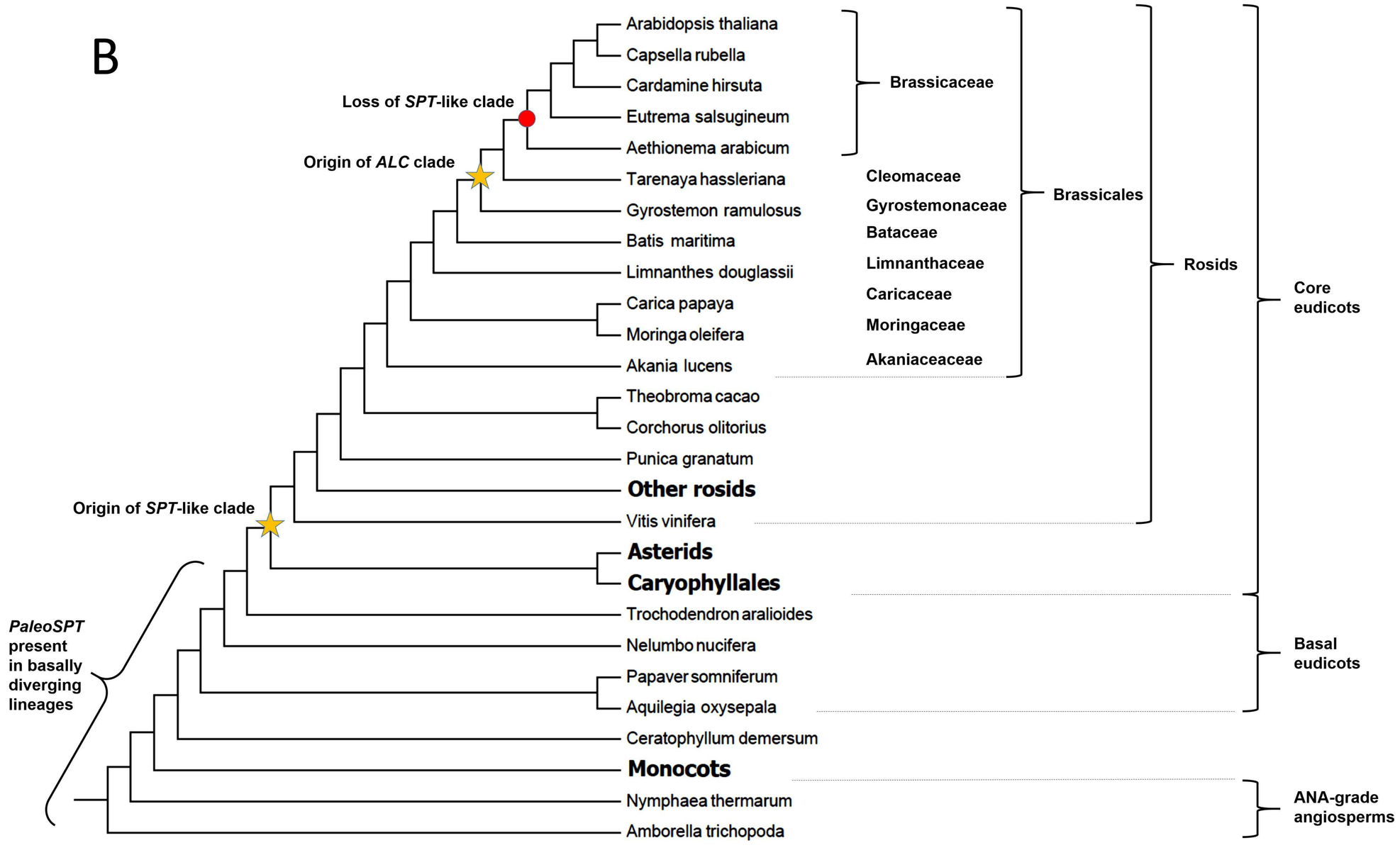




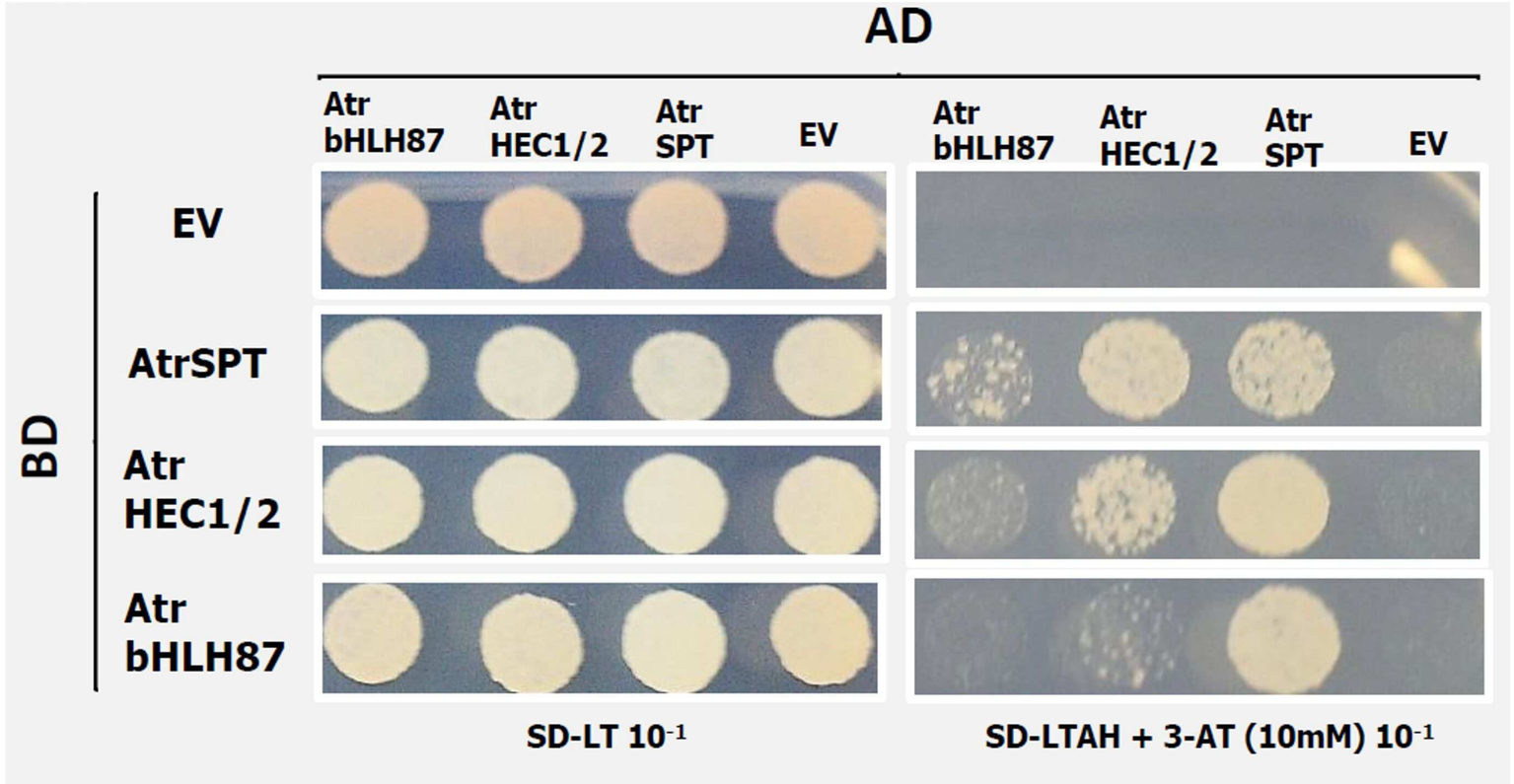
* Genes with probable paralogs between *SPT* and *ALC* clades (within Brassicales)

* Genes with probable paralogs between *SPT* and *SPT-like* clades (within core eudicots)

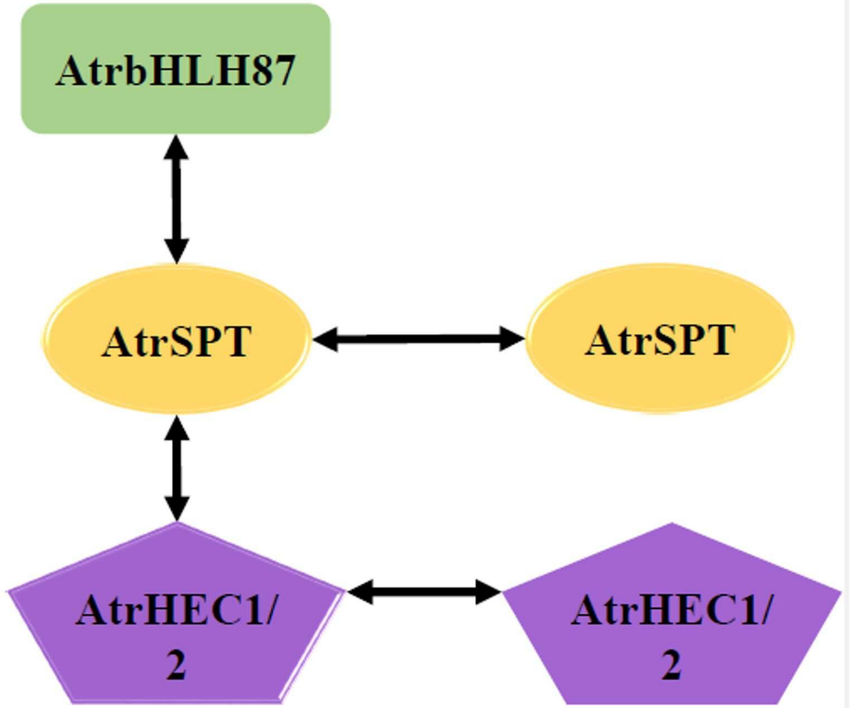
B

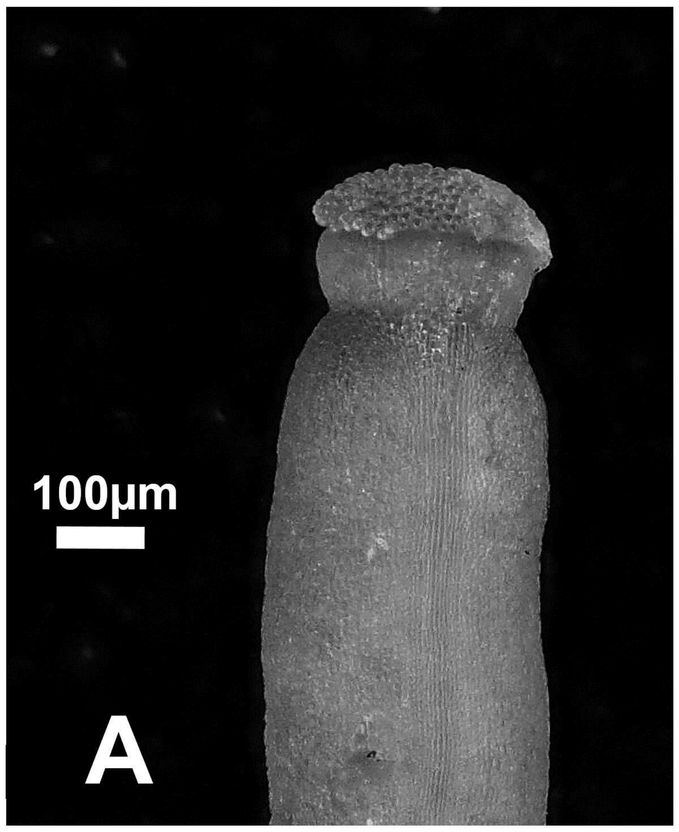


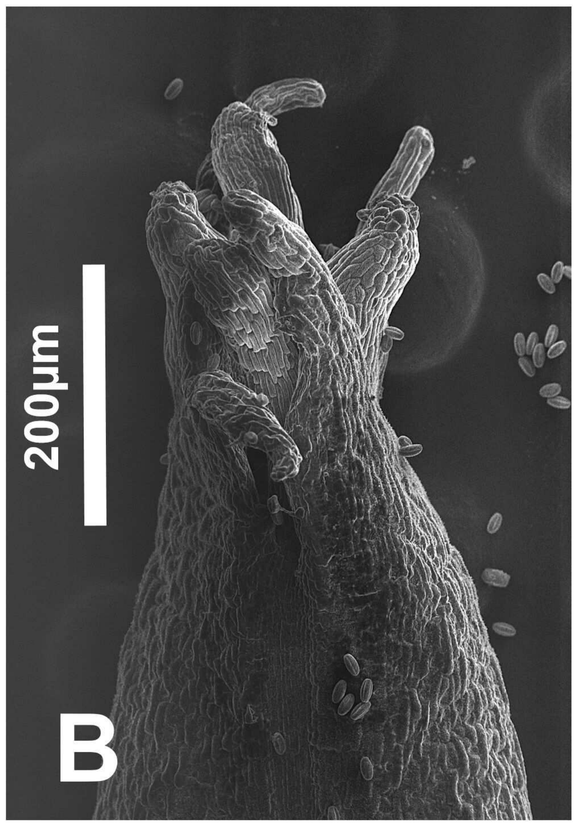
A

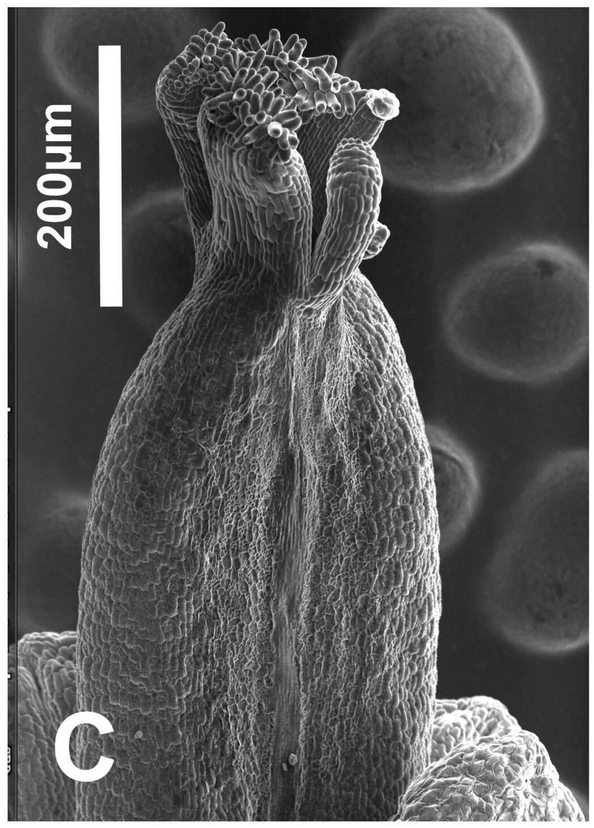

AD
Atr bHLH87
Atr HEC1/2
Atr SPT
EV
Atr bHLH87
Atr HEC1/2
Atr SPT
EV
BD
EV
AtrSPT
Atr HEC1/2
Atr bHLH87
SD-LT 10-1
SD-LTAH + 3-AT (10mM) 10-1

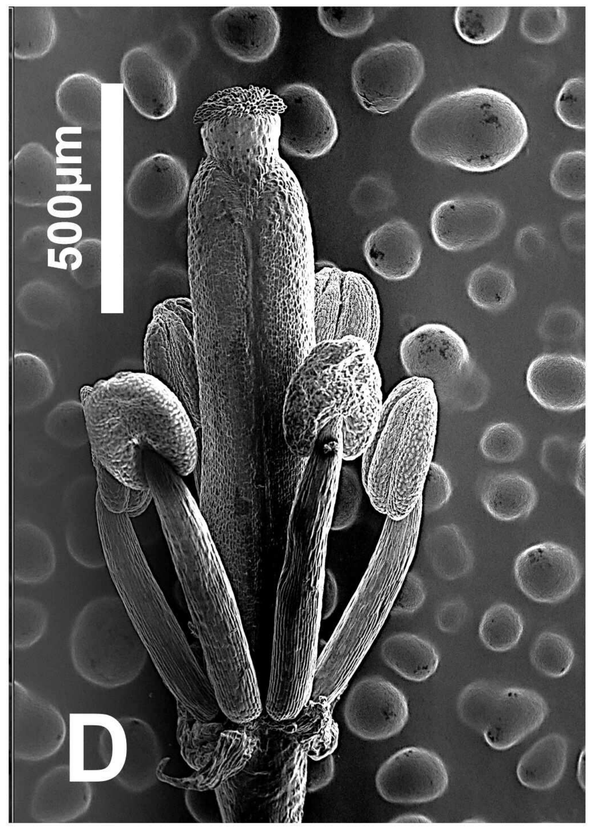
B

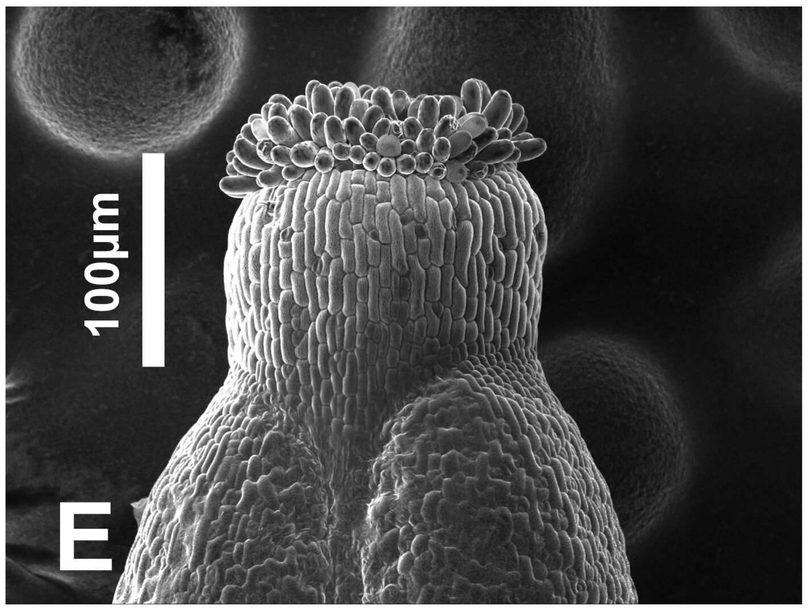

AtrbHLH87
AtrSPT
AtrSPT
AtrHEC1/ 2
AtrHEC1/ 2

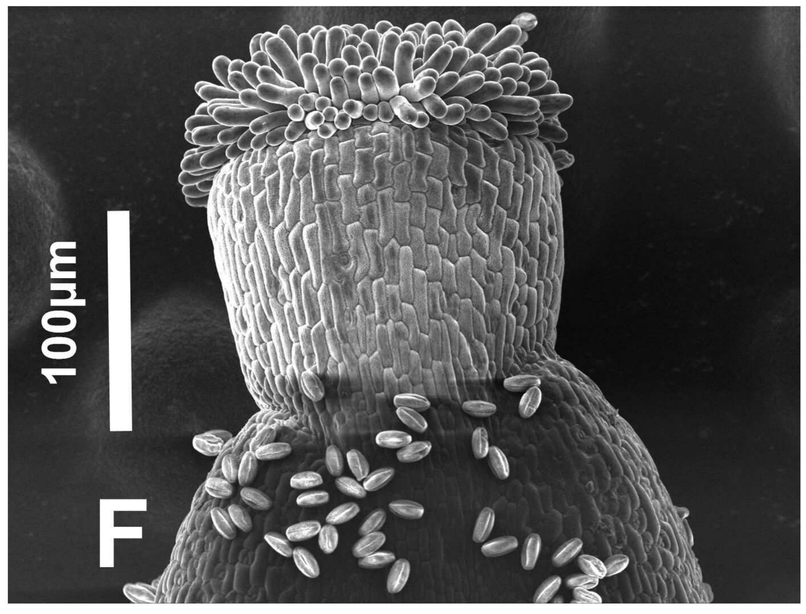

100µm
A

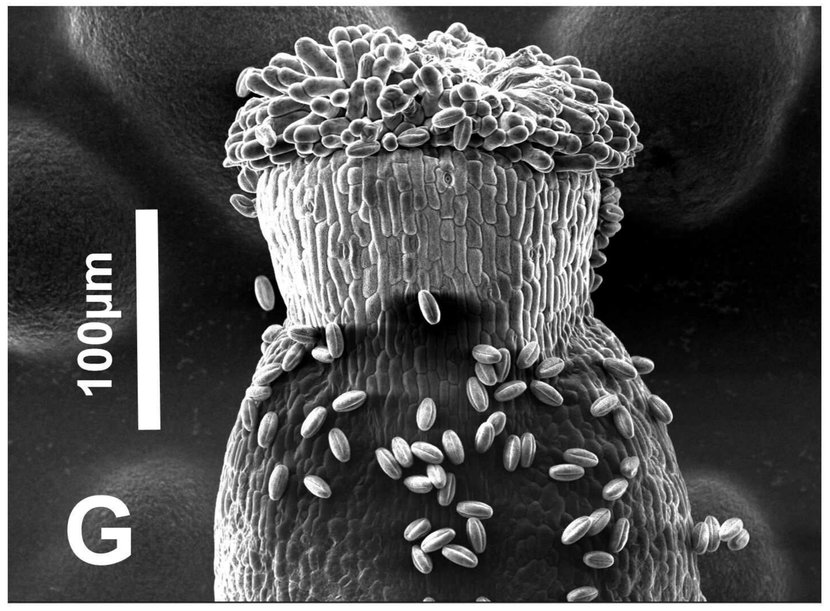

200µm
B

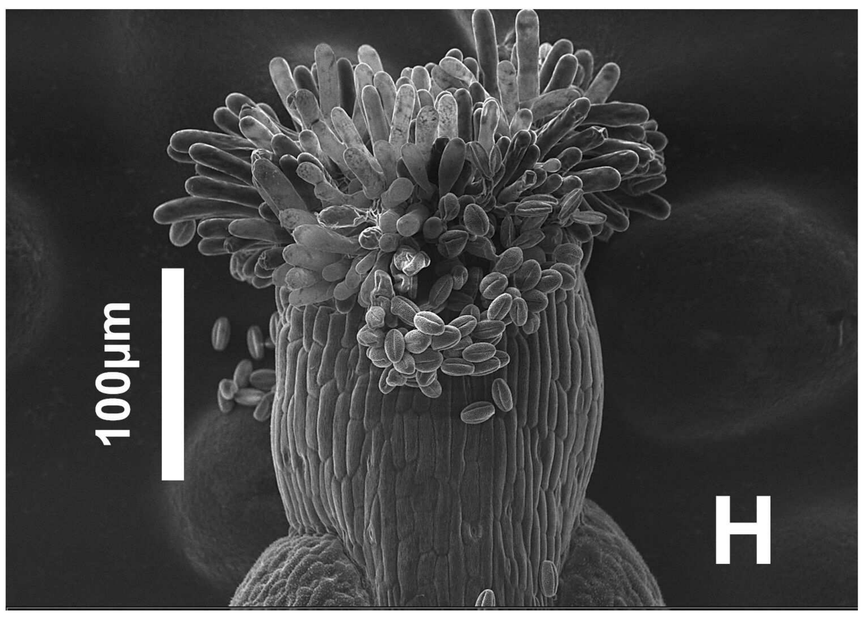

200µm
C

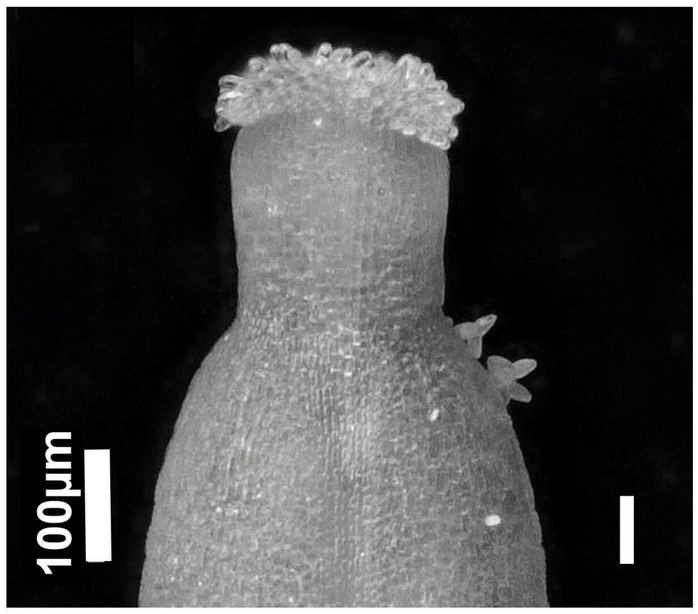

500µm
D

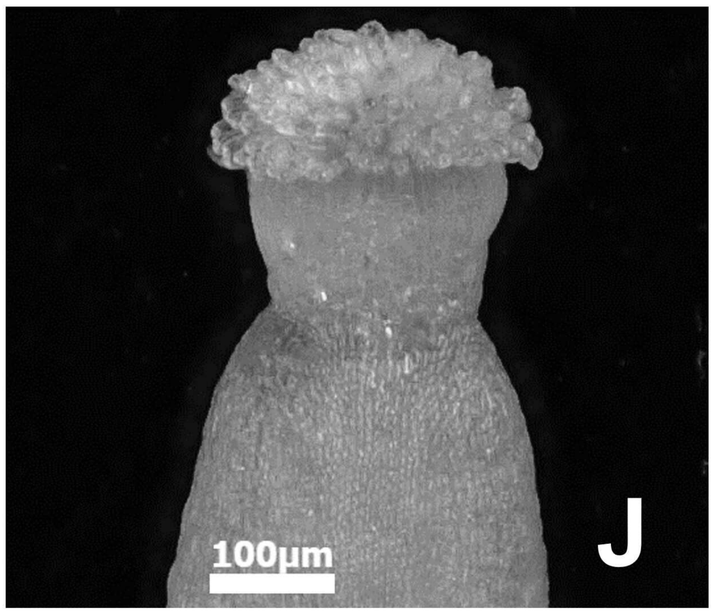

100µm
E

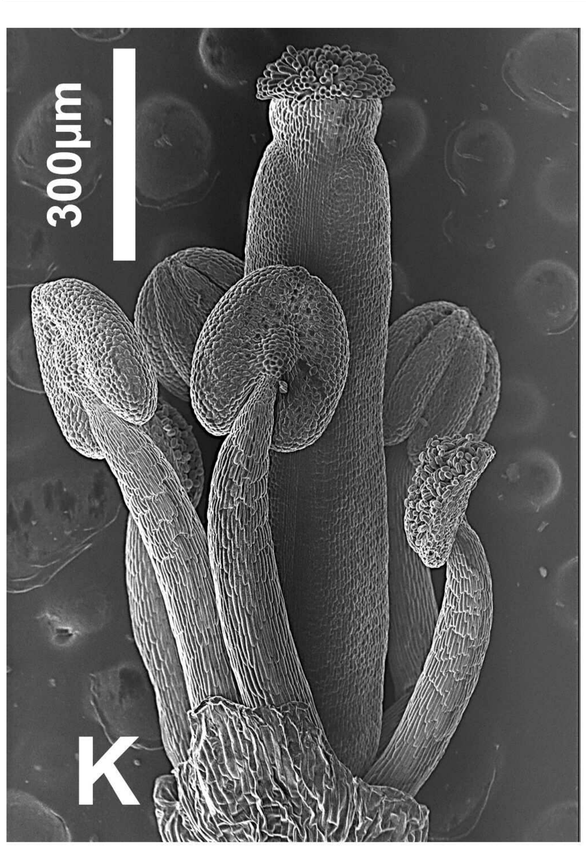

100µm
F

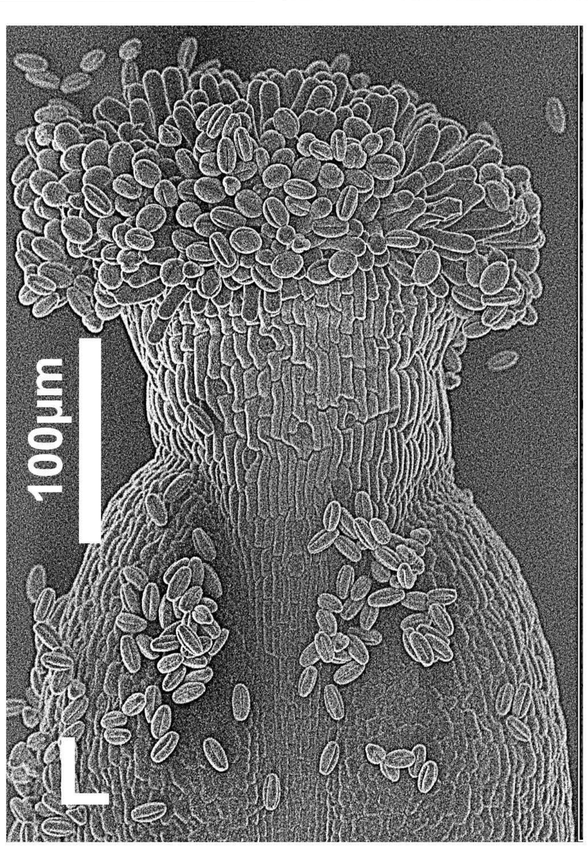

100µm
G

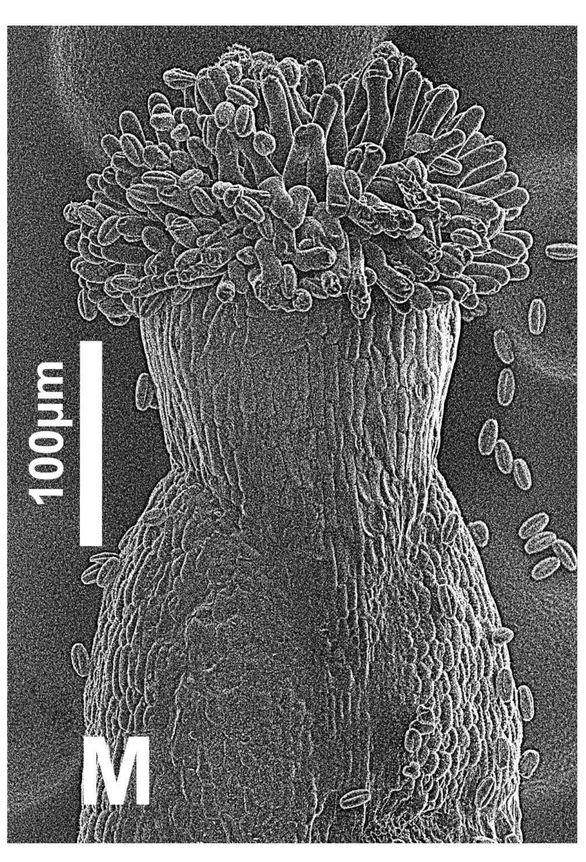

100µm
H

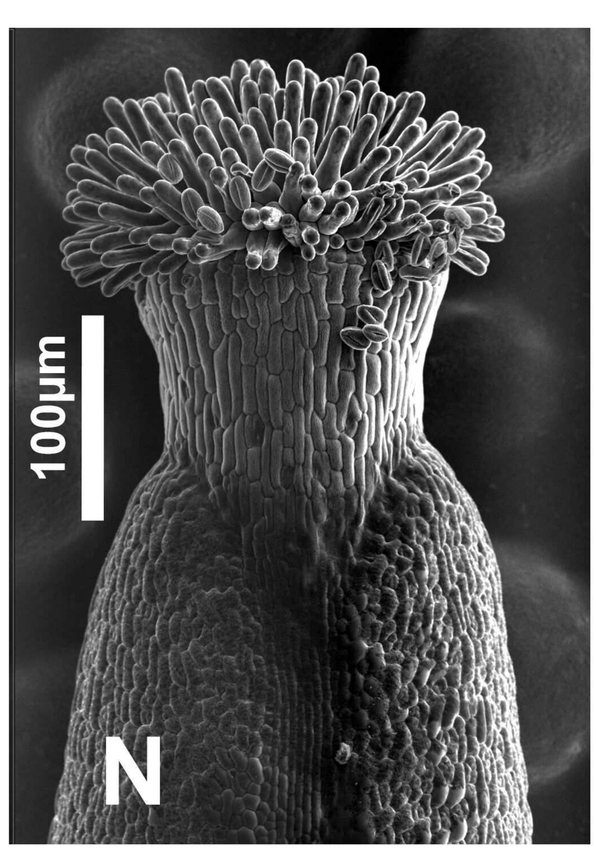

100µm
I

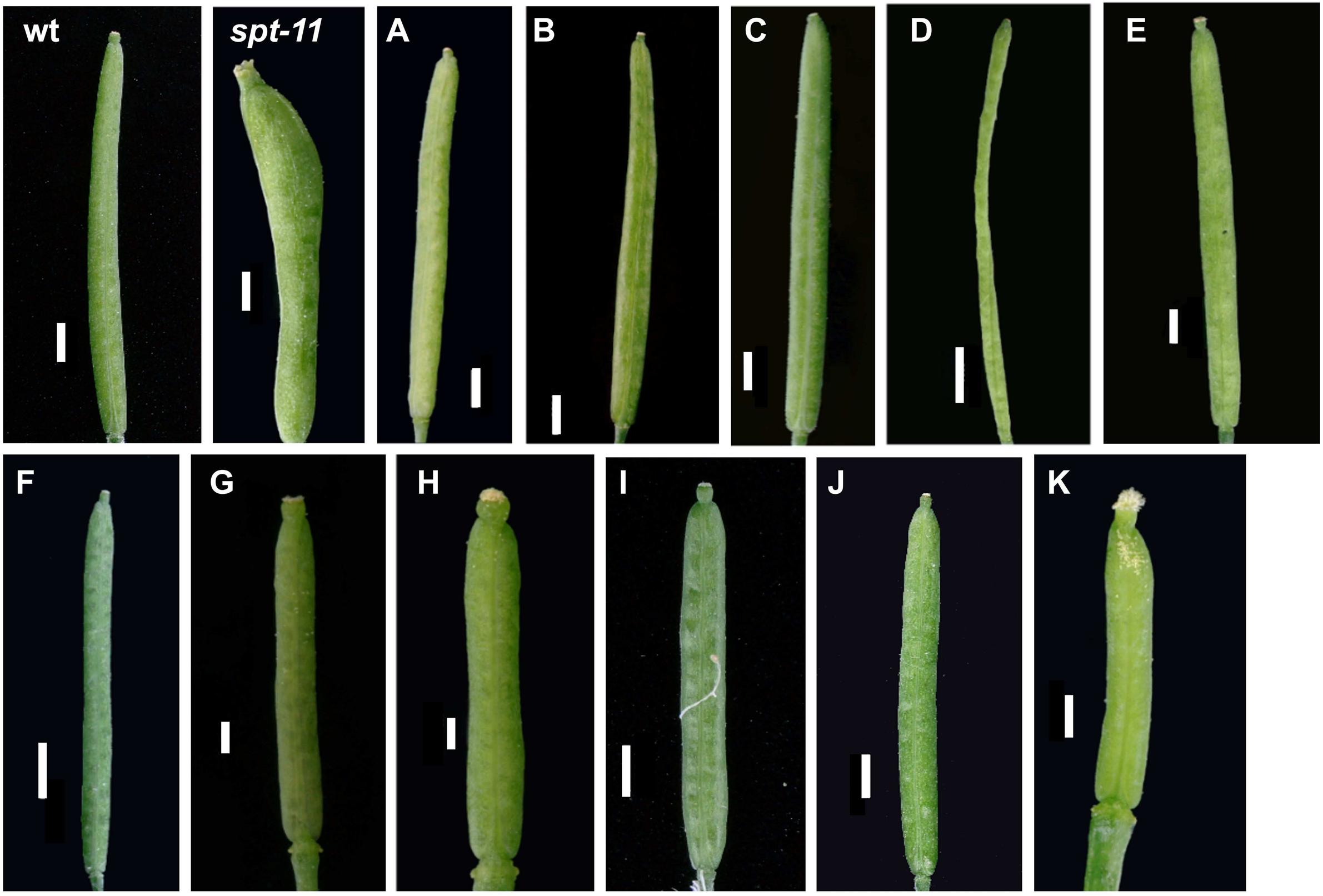

100µm
J

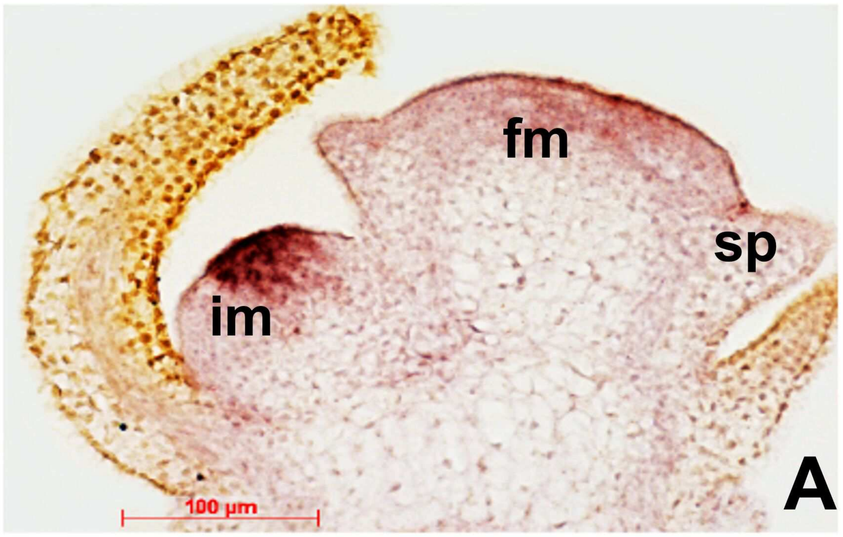

300µm
K

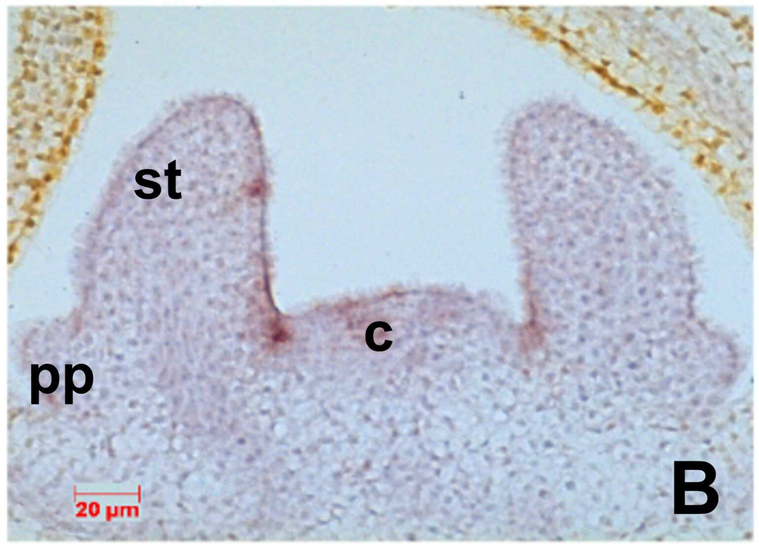

100µm
L

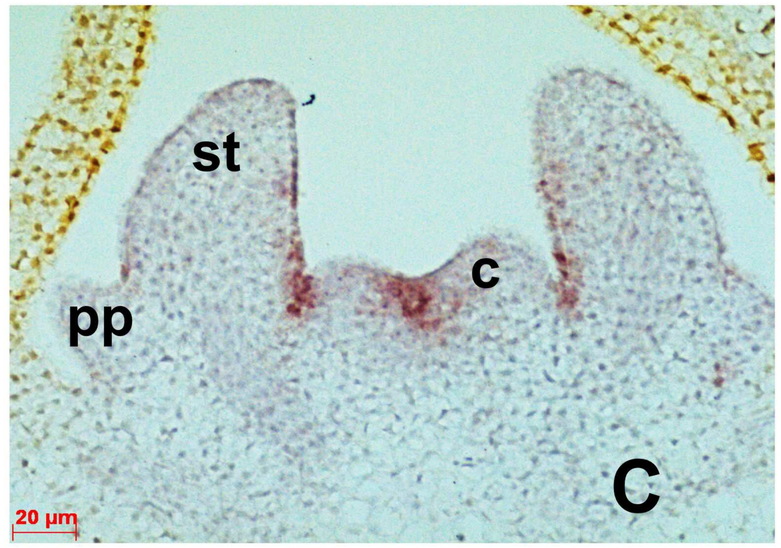

100µm
M

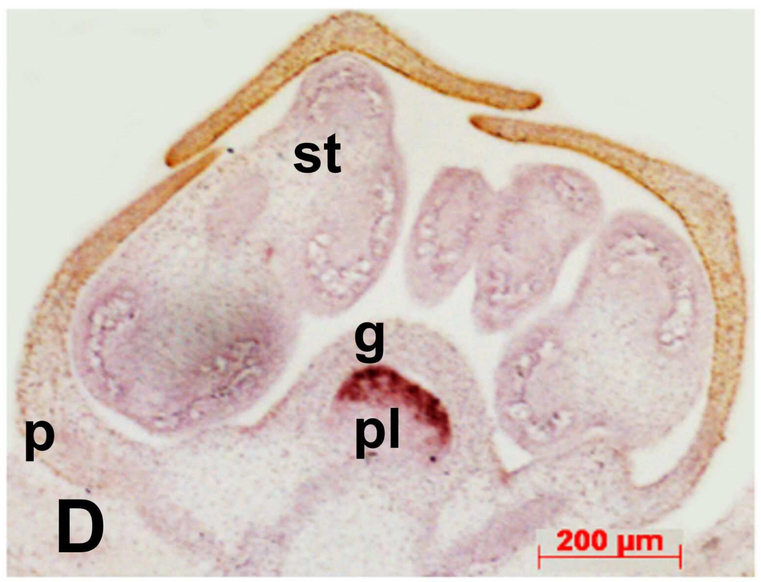

100µm
N

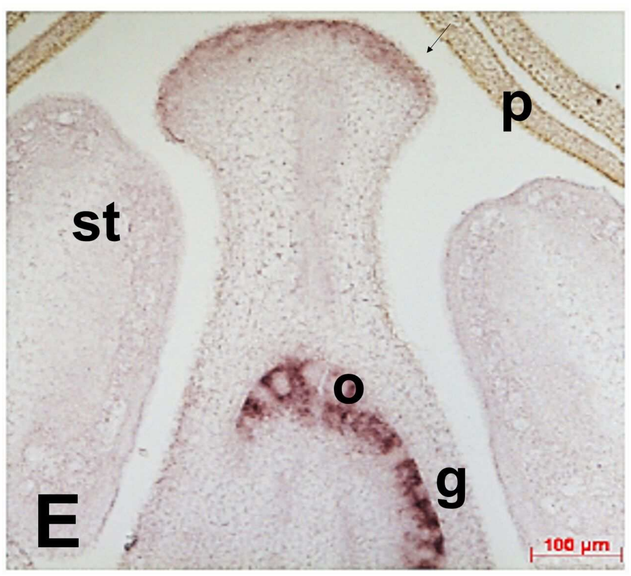

wt
spt-11
A
B
C
D
E
F
G
H
I
J
K

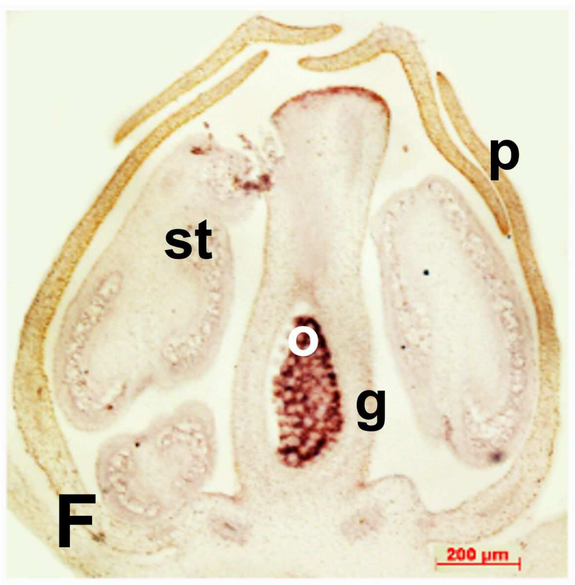

fm
sp
im
100 µm
A

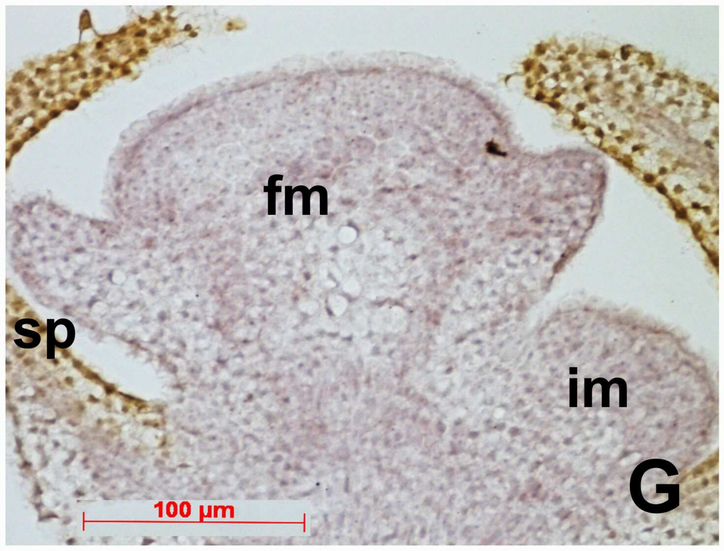

st
c
pp
20 µm
B

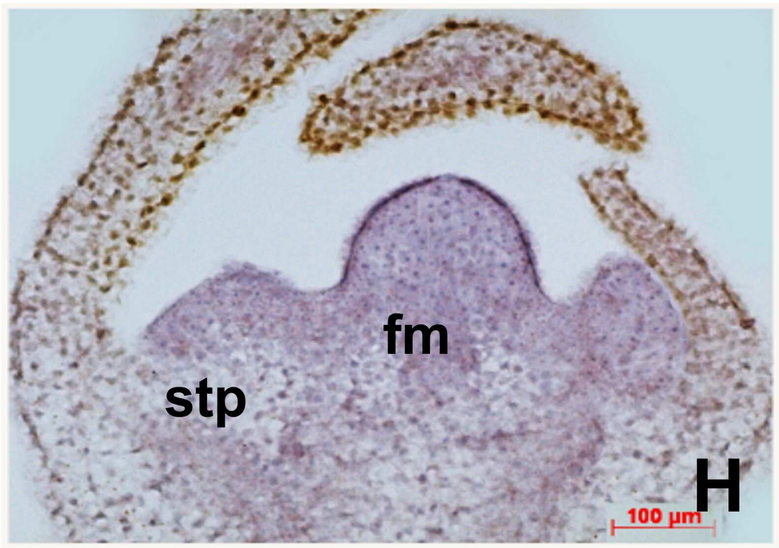

st
c
pp
20 µm
C

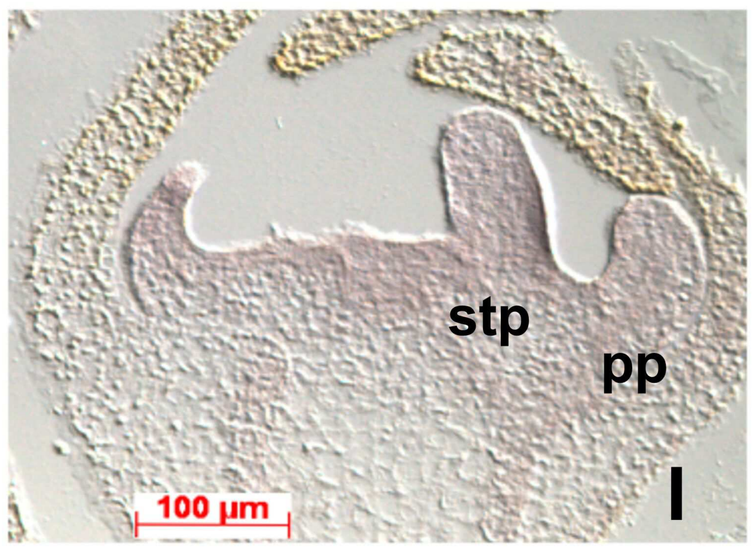

st
g
p
pl
D
200 µm

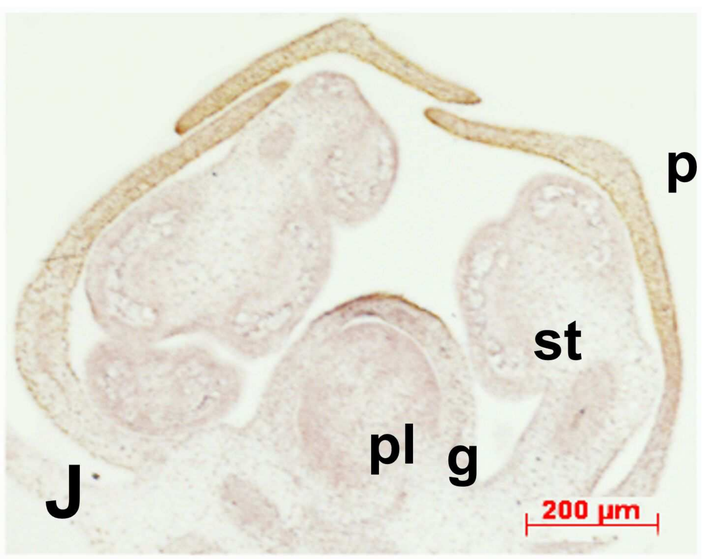

p
st
o
g
E
100 µm

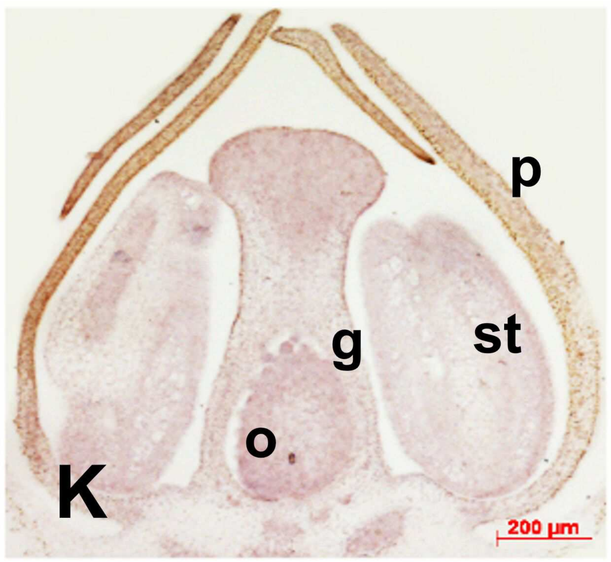

p
st
o
g
F
200 µm

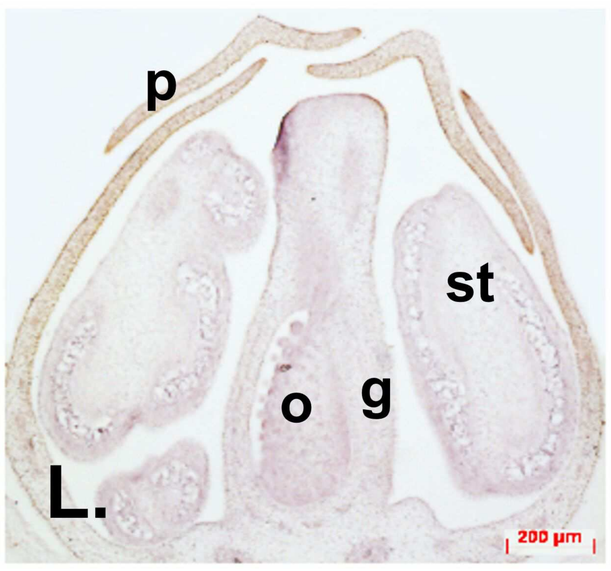

fm
sp
im
100 µm
G


fm
stp
100 µm
H

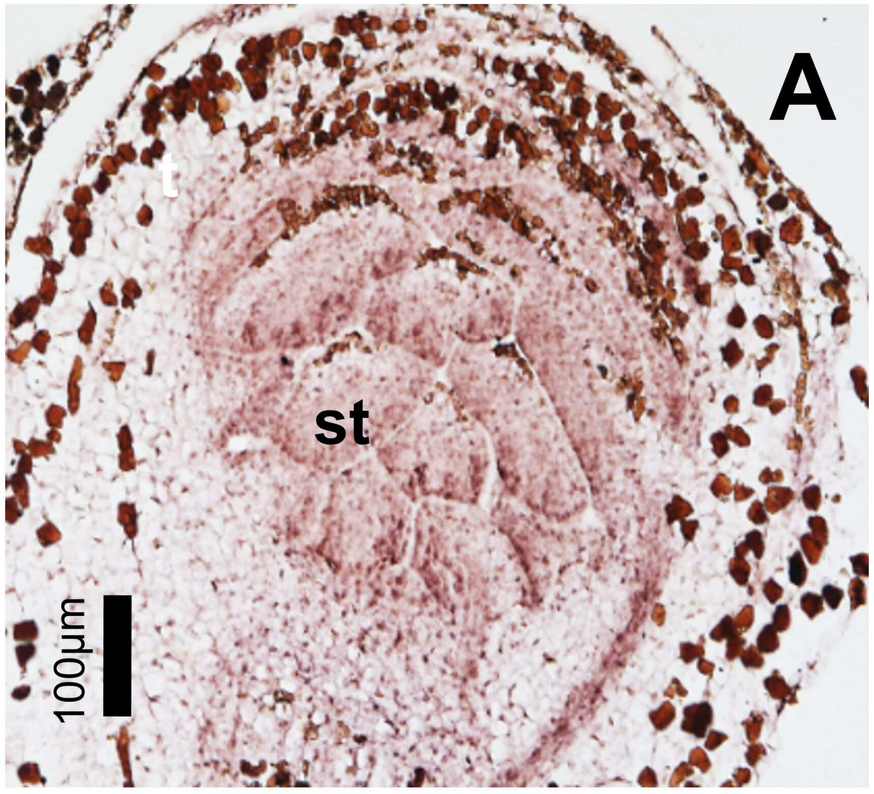

stp
pp
100 µm
I


p
st
pl
g
J
200 µm

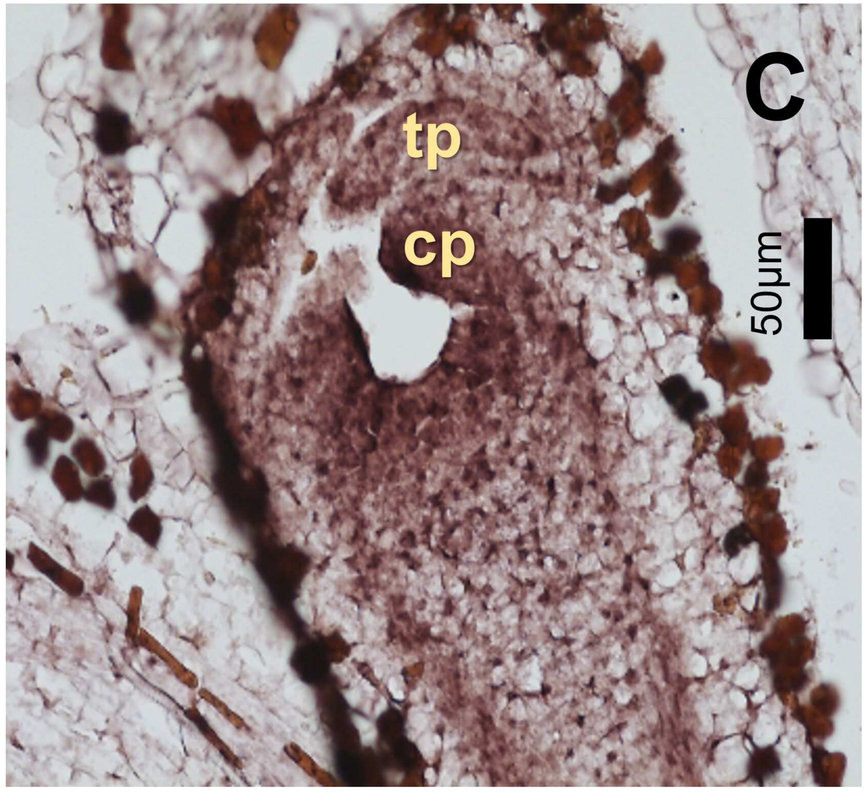

p
g
st
o
K
200 µm

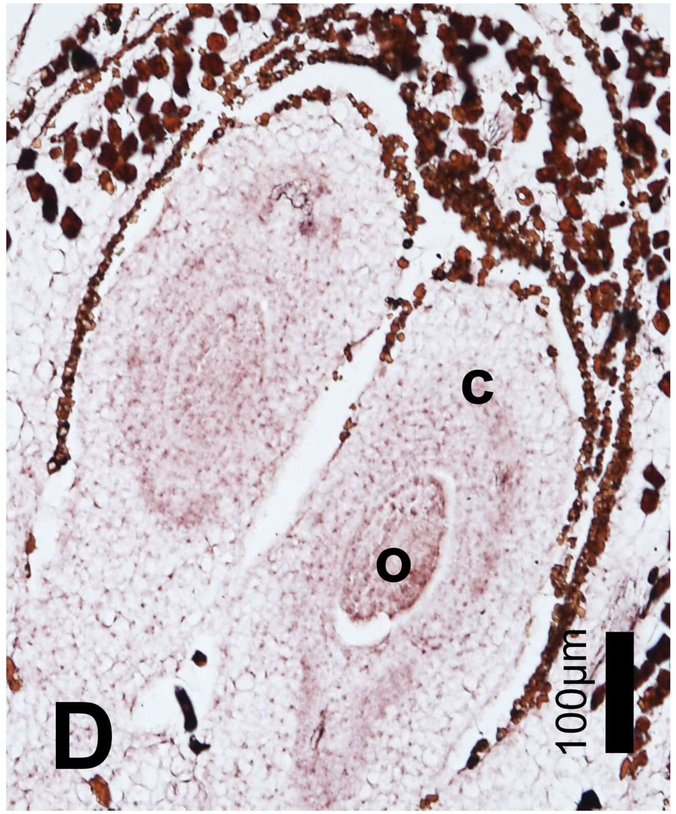

p
st
o
g
L.
200 µm

.

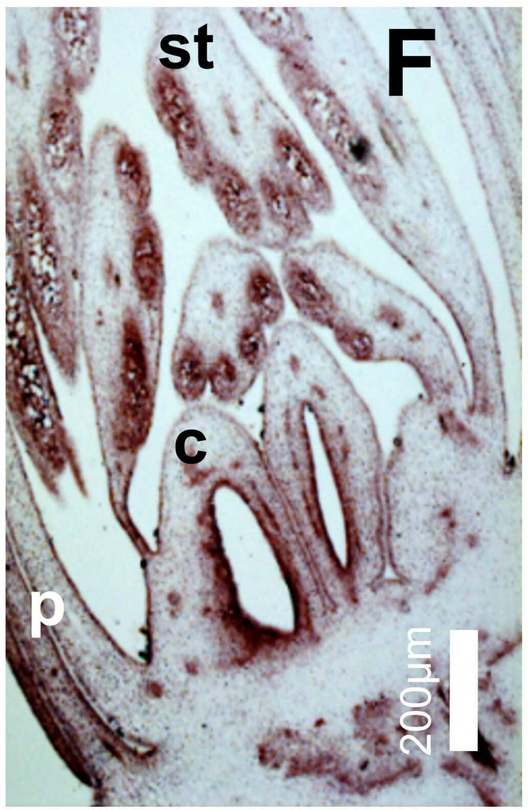

A
t
st
100µm

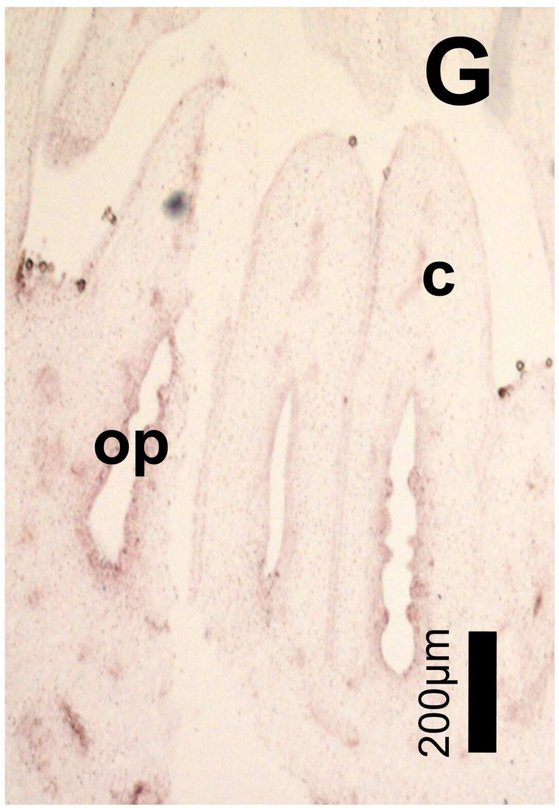

B
st
200µm


C
tp
cp
50µm


c
o
D
100µm


sc
o
c
E
200µm


st
F
c
p
200µm


G
c
op
200µm

A
1000µm

B
1000µm

C
1000µm

D
400µm

E
500µm

F
400µm

G
500µm

H
500µm

Name

Motif Locations

Other rosids

Brassicaceae

Solanaceae

Potri_pSPT

Rco_pSPT

Sapur_pSPT

Glyma_pSPT

MDP_pSPT

Csa_pSPT

Egra_pSPT

Capa_pSPT

Ath_pSPT

Araha_pSPT

Aly_pSPT

Bostr_pSPT

Bra_pSPT

Carub_pSPT

Esal_pPST

Paxi_pPST

Ntab_pSPT

Nsyl_pPST

Cann_PST

Slyc_pSPT

-5000pb

ATG

Motif 1

Motif 2

Motif 3

Motif 4

Motif 5

Motif 6

Motif 7

Motif 8

Motif 9

Motif 10