



**HAL**  
open science

## The molecular basis of carpel evolution

Ana Clarizza Rivarola Sena

► **To cite this version:**

Ana Clarizza Rivarola Sena. The molecular basis of carpel evolution. Molecular biology. Université de Lyon, 2020. English. ⟨NNT : 2020LYSEN062⟩. ⟨tel-03163956⟩

**HAL Id: tel-03163956**

**<https://theses.hal.science/tel-03163956v1>**

Submitted on 9 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Numéro National de Thèse : 2020LYSEN062

**THESE de DOCTORAT DE L'UNIVERSITE DE LYON**  
opérée par  
**l'Ecole Normale Supérieure de Lyon**

**Ecole Doctorale N° 340**  
**Biologie Moléculaire, Intégrative et Cellulaire (BMIC)**

**Spécialité de doctorat** : Biologie végétale  
**Discipline** : Sciences de la vie et de la santé

Soutenue publiquement le 10/12/2020, par :  
**Ana Clarizza RIVAROLA SENA**

---

**The molecular basis of carpel evolution**  
La base moléculaire de l'évolution du carpelle

---

Devant le jury composé de :

FERRANDIZ MAESTRE, Cristina	Chercheuse senior	Instituto de Biología Molecular y Celular de Plantas, Valencia, Espagne	Rapporteure
BECKER, Annette	Professeure d'université	Universität Justus-Liebig, Giessen, Allemagne	Rapporteure
TREGGAR, James	DR-IRD	IRD-Montpellier	Examineur
ROUDIER, François	Professeur des Universités	RDP-ENS de Lyon	Examineur
SCUTT, Charlie	DR2 CNRS	RDP-ENS de Lyon	Directeur de thèse



# RESUME

## *La base moléculaire de l'évolution du carpelle.*

Le carpelle est l'organe reproducteur femelle des plantes à fleurs. Nous présentons une analyse transcriptomique comparative, focalisant particulièrement sur le développement des tissus femelles reproducteurs, entre la plante modèle établie *Arabidopsis thaliana* et la soeur probable du restant des plantes à fleurs vivantes *Amborella trichopoda*. Des modules de co-expression de gènes ont été d'abord définis et ensuite comparés statistiquement entre espèces sur la base de relations d'orthologie entre tous les gènes dans les deux génomes sous comparaison. Cette étude a révélé des modules génétiques entiers, plutôt que seulement des gènes individuels, qui auraient conservé leurs patrons d'expression depuis le dernier ancêtre commun des plantes à fleurs. D'autres modules d'origine plus récente ont été également mis en évidence, modules qui ont peut-être contribué à une complexification dans la morphologie des carpelles qu'a eue lieu plus tardivement pendant l'évolution des plantes à fleurs. Le deuxième thème majeur de cette thèse concerne l'un des régulateurs de ce deuxième groupe, le facteur de transcription *SPATULA (SPT)* de la famille bHLH. Nous présentons des données suggérant que *SPT* aurait acquis son rôle actuel dans les tissus à l'apex du carpelle chez un ancêtre commun des Brassicacées. Nos données suggèrent également que l'acquisition de ce nouveau rôle dépendait de changements au niveau des séquences cis-régulatrices de *SPT*. Nous décrivons des expériences en cours qui ont pour objectif d'identifier les séquences d'ADN précises responsables pour ce changement évolutive.

# ABSTRACT

## *The molecular basis of carpel evolution.*

The carpel is the female reproductive organ of flowering plants. We present a comparative transcriptomic analysis, focusing in particular on the development of female reproductive tissues, between the established model plant *Arabidopsis thaliana* and the probable sister of all remaining living flowering plants *Amborella trichopoda*. Gene co-expression modules were first defined and then compared statistically between species on the basis of orthology relationships between all genes in the two genomes under comparison. This study revealed entire genetic modules, rather than just individual genes, that appear to have retained their expression patterns since the last common ancestor of living flowering plants. Other modules of more recent origin have also been brought to light, which may have contributed to a complexification in the morphology of carpels that occurred later during the evolution of flowering plants. The second major theme of this thesis concerns one of the regulators of this second group, the transcription factor *SPATULA* (*SPT*) of the bHLH family. We present data suggesting that *SPT* acquired its present role in the tissues at the apex of the carpel in a common ancestor of Brassicaceae. Our data also suggest that the acquisition of this new role depended on changes in the cis-regulatory sequences of *SPT*. We describe ongoing experiments that aim to identify the precise DNA sequences responsible for this evolutionary change.

# ACKNOWLEDGMENTS

If you ever ask me what are my favorite colors, my answer will be: the blue, the green and the flowers. When I was a little girl, I used to collect flowers in my way home back from the school, they were a specific species, the *Ruellia* flowers. Once at home, I used to run to a large flower pot in the shape of a swan that my mom has at the front garden. There I inserted my collected flowers in a different plant that had no flower of its own. In my own way, I was doing plant cross, and got (to my eyes) a successful hybrid.

As I grew, I was always curious of what happened inside the plants, inside the trunks of the trees, and I used to imagine small cities inside them, with tiny little people that work at the leaves and branches, they were highways connecting one city to another, people were always busy, working to make the plant grow. It was my nine-year-old picture of plant regulatory development. And now, that I think about it, it wasn't that far from reality. Only, instead of people, there are genes and transcription factors, always busy, working in these little towns, called cellules, with plasmodesmata that connect one cell to another.

I fell completely and irreversibly in love with plants, in my first practical lecture of plant histology, in college. I could finally see what I had imagined for so long. Over microscopy observation my fascination with plants grew, and there I was combining histology, morphology, taxonomy, anatomy, physiology, cytology and genetics to understand my favorite living beings on earth.

What more could I ask for?

Well, what about me crossing the ocean, landing in France for full four years of a passionate plant biology PhD work? Let's live the dream!

I still remember the first time I saw the RDP web site from Paraguay, there were sparkles on my eyes, I imagine myself working there! And now, here I am, at the end of this magical adventure in the Evo-Devo team, in which I had discovered a new dimension of third favorite color: the flowers. As the end approach, it comes the time to say thank you...

The magic start with BECAL who put at my disposal the pot of gold coins to fly over the Atlantic and to SFERE for the guidance in my path. Also thank you to the ENS of Lyon for helping me to complete the gold needed to live my dreams.

My dreams were fueled by the thoughts and prayers of my family and friends. **Mom** and **Dad**, thank you, your crazy daughter is about to be a Doctor and I have not forgotten to eat healthy during these years. To my sisters **Lily** and **Victoria**, thank you for your support.

My path was made easier by the hand of my greatest loves, **Hector** and **Cynthia**. I am grateful to life for having given me their existence and their company. Magical times require powerful and enlightened beings like you to complete all the journeys. Gracias por su amor.

In this fairy tale, I got to this magical chapter called the PhD thesis, and I want to thank to all the other charmed characters that helped me in this incredible adventure.

I shall start by thanking My Magic Goblin: **Charlie Scutt**, with his Irish flute he got me dancing around carpel origin and evolution, for the last four years. Thank you for your scientific passion, your patience, and thank you for showing me the path of the *Abominable mystery* of Charles Darwin.

Thanks to my Magical Scientific Advisors, **Gwyneth Ingram** and **Francois Parcy** from Comité de Suivi de Thèse, their wisdom has guided my work and has helped me grow as a researcher and as a person. And thank you to my jury members: **Cristina Ferrandiz Maestre**, **Annette Becker**, **James Tregear** and **François Roudier**, who have kindly agreed to review my work.

**Special mentions to:**

All the scientific collaborators of my work. Those from other laboratories (**IPS2-Institute of Plant Sciences Paris-Saclay**; **UMR 1318 INRAe-AgroParisTech**; **Centro Nacional de Biotecnología-Consejo Superior de Investigaciones Científicas**), they have been an indispensable part of my thesis advancement.

My First Master, **Amelie Andres Robin**, she transferred me my first bench-powers, I will always remember her lectures of how to hold a micropipette, how a PCR is made and how to do a miniprep. Thank you Amelie.

My Science Elf, **Aurelie Vialette**, I have an image engraved in my memory, of she asking me to sit at her desk and kindly explaining me how to organize mi scientific work. She was there for me at the beginning and at the end, but even when she moved toward new scientific horizons she was there enlightening me with her advices. Gracias.

My Fairy godmother, **Pierre Chambrier**, instead of a magic wand he has a P10 micropipette, with an unmatched power for molecular biology. Whenever I had a question, he had the answer or the way to find it. Thanks for your magic on the bench and on the dance floor. Merci.

The Wizard of the command lines, **Jérémy Just**, thanks for your time and your patience. Thank you for helping me understand new languages and guiding me in using them.

Many thanks to the rest of the enchanted beings that are part of my team, especially **Michiel** and **Patrice** who have closely contributed to this fairy tale. But I must also thank **Suzanne**, **Marie**, **Lucie**, **Nelu** and **Mathilde**, who throughout these years have accompanied this process in one way or another.

To my magical beings, **GéGé**, **Judit**, **Isa** and **Hervé**: Thank you for those chats at the paillasse, corridor and coffee, thank you for enchanting my working days with your kindness. With the most special mention for Isa and Hervé whose work is and was indispensable for my magical experiments.

To the Elves who have cared for my plants with the greatest of jealousy, **Alexis**, **Patrice** and **Jus**

THANK YOU.

To my Elves from the 017 of M3, **Priscilla** and **Antoine**, thank you for your company, for our talks, cups of tea and your constant encouragement.

**Virginie**, the Fairy of the most ordered protocols of RDP. Thanks for the talks, help and advice during the experiments and for your magical protocols. Merci du fond de mon Coeur!

To the companions and friends of the enchanted village of the Afterwork and the stairs, **Sana, Raquel, Yannik, Jack, Adrien, Julie, Juan, Julie** and **Justine**. Thank you, it has been a pleasure to share with you our dishes, joys and sorrows.

**Raquel**, thank you for being a person of light and giving me your support and your friendship. Voy a extrañar las patatas entre clonaciones, las risas y los bailes al son de los Decadentes en la poyata. ¡GRACIAS por todo guapa!

Finally, thanks to all the **RDP** members, who make the lab a pleasant place to work. MERCI tout le monde!

It is time for me, to undertake new adventures, to search for other treasures, to explore other horizons. I will keep in mind all the people who have surrounded me at this point, but today I want to say: thank you, merci, gracias, aguije.

# TABLE OF CONTENTS

## GENERAL INTRODUCTION

---

The origin of angiosperms: when and where did they appear?	14
What is a flower? -The modern flower and its components.	19
What did the first flower look like, and what were the possible ancestors of angiosperms?	19
A closer look at female reproductive tissues.	23
The ancestral carpel and its origin.	23
Molecular mechanism that control the floral development in extant angiosperms	24
The genetic regulation of female reproductive tissues in <i>Arabidopsis thaliana</i>	29
The Gene Regulatory Network of the ancestral carpel	32
Non-model species for studies of the origin of the flowers	34
Techniques and methods that could help to explain the origin of the flower	39
Unsolved questions of flower origin and evolution	40

## AIMS AND STRATEGIES

---

### RESULTS

---

#### PART I: TRANSCRIPTOMIC DATA FOR CARPEL EVOLUTIONARY STUDIES

---

Aims and strategies of part 1	46
<i>A proof-of-concept using novel transcriptomic data from the early-diverging angiosperm <i>Amborella trichopoda</i> to identify genetic modules conserved over deep evolutionary time.</i>	48
Abstract	49
Introduction	49
Materials and methods	51
Results	53
Discussion	58
References	61
Tables and figures	65
Supplementary information	71

#### PART II: STUDY OF A TRANSCRIPTION FACTOR INVOLVED IN CARPEL

---

Aims and strategies of part 2	74
<i>SPATULA, a bHLH transcription factor, evolution and its role in gynoecium development.</i>	
Abstract	

Introduction	78
Results	80
Discussion	91
Materials and Methods	94
References	99
Tables and figures	104
Supplementary information	120
<b>GENERAL DISCUSSION AND PERSPECTIVES</b>	126
<b>REFERENCES</b>	133

---

# ABBREVIATIONS

AG	AGAMOUS
ALC	ALCATRAZ
AP1	APETALA1
AP3	APETALA3
APB	Active Phytochrome Binding
ARR1	ARABIDOPSIS RESPOSE REGULATOR1
Ath	Arabidopsis thaliana
Atr	Amborella trichopoda
AuxRE	Auxin response element
bHLH	basic Helix-Loop-Helix
CMM	Carpel margin meristem
evo-devo	evolutionary-developmental biology
GRN	Gene Regulatory Network
HEC1-3	HECATE1-3
IND	INDEHISCENT
LBA	Long-branch attraction
LCM	Laser-capture microdissection
MRCA	Most Recent Common Ancestor
MYA	Million years ago
Pab	Picea abies
Pax	Petunia axillaris
PBM	Protein Binding Microarray
PID	PINOID
PIF	Phytochrome-interacting factor
PWMs	Position weight matrices
RSAT	Regulatory Sequence Analysis Tools
SPT	SPATULA
VIGS	Virus-induced gene silencing
WGCNA	Weighted Correlation Network Analysis
WGD	whole genome duplication
Y2H	Yeast two hybrid



*“The rapid development as far as we can judge  
of all the higher plants within recent geological times  
is an abominable mystery”*

Charles Darwin, 1879.

# **GENERAL INTRODUCTION**

---

## Introduction

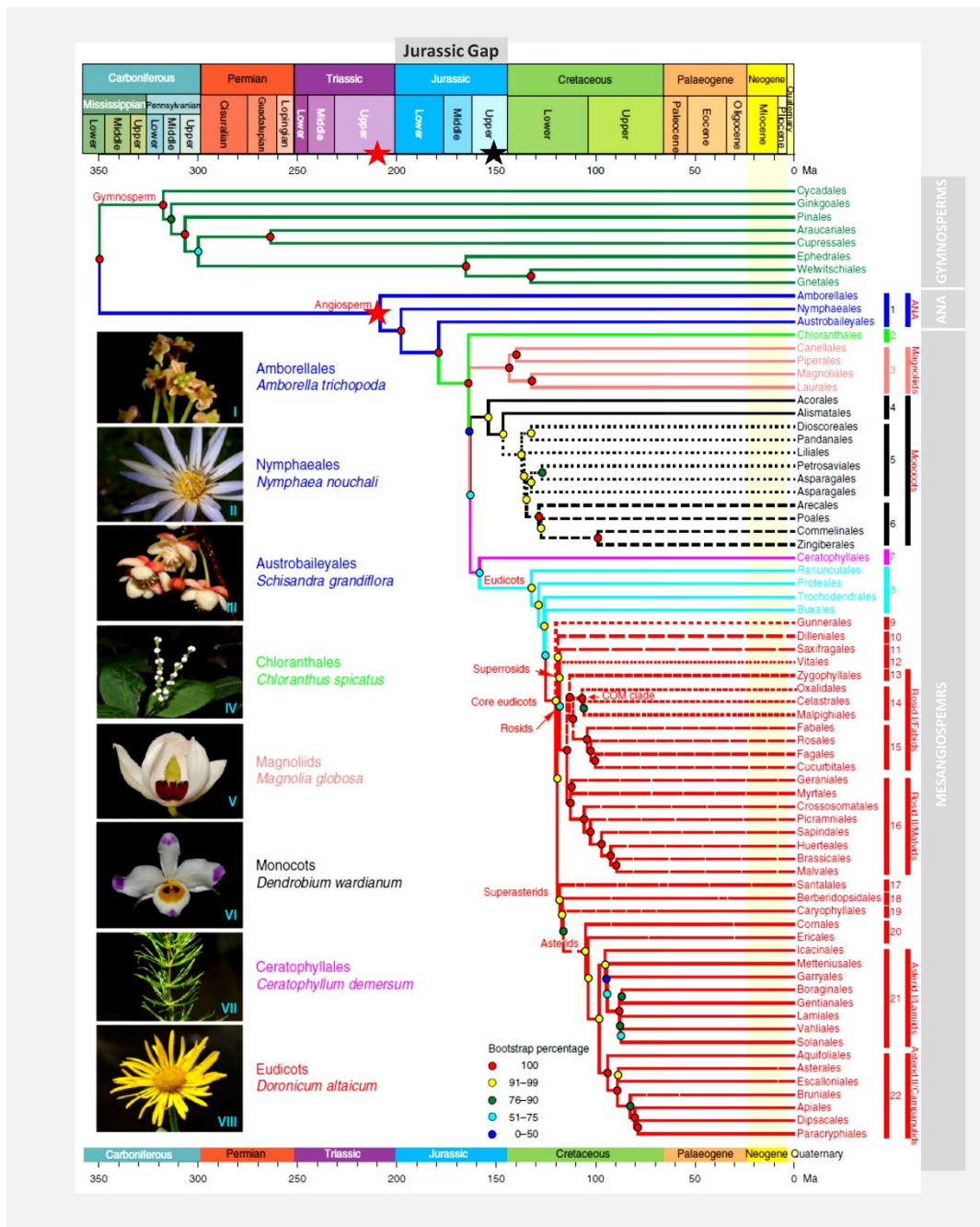
### The origin of angiosperms: when and where did they appear?

Angiosperms, or flowering plants, are the main component of most terrestrial ecosystems. They represent more than 95% of the vascular plants existing today (Mandel, 2019). Their morphological, functional and ecological diversity have made angiosperms the dominant group of living plants, with more than 290,000 species described (Ramírez-Barahona et al., 2020). Flowering plants are critical components of aquatic and terrestrial habitats. This clear dominance as a group was established in the Lower Cretaceous, in which the fossil record clearly documents the rapid expansion and diversification of angiosperms.

Nevertheless, there are controversies regarding the dating of the origin of flowering plants. Research based on molecular clock analyses place the origin of angiosperms between 256 and 139 million years ago (MYA) (**Figure 1**), in which the earlier spectrum of dates corresponds to the Jurassic or even the Triassic (Li et al., 2019; Salomo et al., 2017). However, the earliest known fossils with certain affinities to angiosperms date from the early Cretaceous, between 135 and 125 MYA. These fossils are fossilized pollen grains which have been found in or around regions corresponding to the northern parts of the ancient supercontinent of Gondwana (**Figure 2**). However, as mentioned above, the dates of these fossils are not very consistent with that of the most-recent common ancestor (MRCA) of living angiosperms, as determined by molecular-clock based studies, the most recent of which gives a date of 209 MYA (Li et al., 2019).

There is therefore a gap of perhaps about 70 million years (MY) between the earliest known fossil angiosperms and dates obtained using molecular analyses, and this has been called the “Jurassic Gap” (Li et al., 2019; Mandel, 2019). The age of the flowering plant group has been widely discussed over the years, from early studies in the 1960s to the present, but no real consensus has been reached (Coiro et al., 2019; Herendeen et al., 2017). For some authors, the sudden rapid diversification of angiosperms in the lower Cretaceous (**Figure 1**) is an indication of their origin and radiation in earlier periods (Herendeen et al., 2017; Ramírez-Barahona et al., 2020; Smith et al., 2010).

Pollen-wall morphology comprises one of the most useful sets of characters to reconstruct the early evolution of flowering plants. The pollen-fossil record of angiosperms during the Cretaceous period is continuous, while reports of angiosperm pollen prior to this period are very sporadic and highly contested in their possible relationship to angiosperms (Herendeen et al., 2017). Recent studies support the



**Figure 1.** Several molecular clock analyses have situated the origin of angiosperms in the Lower Jurassic (Red stars). However, fossil-based analyses have suggested a date around the upper Jurassic/lower Cretaceous (Black star). The gap between these estimates is called the “Jurassic Gap”. The two major groups of angiosperms (ANA-grade and mesangiosperms) are indicated. (Source: Li et al., 2019).

hypothesis of an early origin of angiosperms, followed by a long period of persistence, before their rapid diversification (Li et al., 2019; Ramírez-Barahona et al., 2020). This greater diversification is located between the middle of the Cretaceous and the end of this period, giving rise to most of the existing lineages (Vea & Grimaldi, 2016). The correspondence of this major expansion of angiosperms with the diversification of insects and other pollinators has been extensively studied, but in addition to this factor, favorable ecological conditions and substantial morphological changes in the first species of flowering plants can be added (de Boer et al., 2012; Ramírez-Barahona et al., 2020).

Three hypotheses have been proposed by (Mandel, 2019), to try to explain the “Jurassic Gap” in the fossil record. The first of these suggests that early angiosperms were rare, and that their fossils are therefore difficult to find and identify. The second hypothesis postulates an environment for early angiosperms that was not very conducive to the preservation of plant fossils. The third hypothesis suggested by (Mandel, 2019) points to the anatomy of the first angiosperms, which might have had very small flowers and fruits, complicating their identification as angiosperms. The closest outgroup to the flowering plants is the remaining seed plants, or living gymnosperms. The separation of the living angiosperm and gymnosperm clades has been dated to around 310 MYA (**Figure 1**), which places a logical upper limit on the date of the MRCA of living angiosperms.

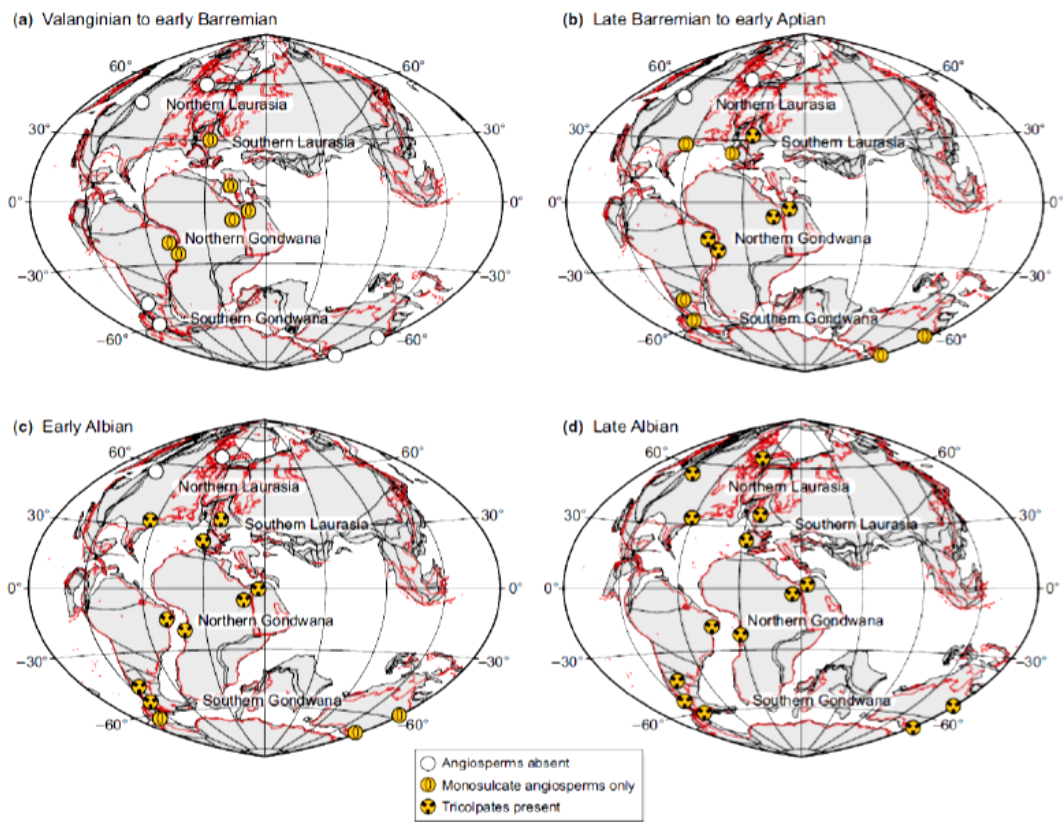
Another point of discussion concerns the geographical origin of angiosperms. The earliest known angiosperm fossils are monoaperturate pollen grains from strata dating from 139–130 MYA, and have been found at locations corresponding to present-day parts of North and Central Africa, the Middle-East, Western Europe and South America (Coiro et al., 2019). Many of these pollen types can be ascribed to particular groups of angiosperms within the living angiosperm clade, and must therefore have originated later than the MRCA of all living angiosperms. The earliest known fossil pollen grains may, therefore, represent an early expansion of angiosperms from their original habitat, rather than the first populations of angiosperms to have existed. Later strata at these same locations, and strata at many other locations extending out of the palaeotropics and into temperate regions, contain the tricolpate pollen (containing three elongated apertures) characteristic of the more recent eudicot group of angiosperms, which seems to have appeared 10–15 MY after the earliest known angiosperm pollen types.

Much of northern Gondwana in the early Cretaceous had a hot and dry climate, and early angiosperm pollen from Congo, Gabon and Brazil have typically been found in association with pollen of xerophytic groups including Cheirolepidiaceae (extinct conifers) and Gnetales (James A. Doyle et al., n.d.)(J. A. Doyle, 1999). However, a band of more humid paleohabitats is known, running from present-day Israel and Egypt to Colombia and Peru (Carvalho et al., 2017). These regions show

evidence of early angiosperm pollen together with pollen and spores of plants typical of humid habitats, including Matoniaceae (ferns) and Araucariaceae (conifers). Therefore, the angiosperms may have originally expanded from humid habitats of equatorial northern Gondwana, first into drier habitats in the surrounding territories, and then within the next 10-15 MY out of northern Gondwana and into temperate zones of both the northern and southern hemispheres. This migration was accompanied by a series of innovations of floral structures in the emerging clades of angiosperms, increasing interactions with pollinating and seed-dispersing agents, and favorable ecological conditions (Endress, 2011; Li et al., 2019). The angiosperms finally came to dominate terrestrial ecosystems in the Paleocene period, which began 66 MYA (Ramírez-Barahona et al., 2020).

From the evolutionary point of view, some of the synapomorphic characters that define angiosperms as a group are: the presence of the carpel (and of the fruit, which is derived from the carpel), ovules with two integuments (except in taxon-specific cases), bisexuality of the reproductive axis, the presence of 4-loculate stamens and double fertilization. In addition to reproductive characters, eco-physiological and vegetative characters particular to flowering plants can also be listed, such as the presence of vessels in the xylem, net-veined leaves and the physiological adaptation to the atmosphere with decreasing concentrations of carbon dioxide, as reviewed by Charles P. Scutt, (2018).

Two large groups of flowering plants can be recognized, the so-called ANA-grade and the mesangiosperms (**Figure 1**) (Group, 2009; Group et al., 2016). Recognized as the most basal group of flowering plants, the ANA-grade is made up of the three orders Amborellales, Nymphaeales and Austrobaileyales, among which *Amborella trichopoda*, the only living species of the order Amborellales, is considered the single sister species of all other living angiosperms (Project, 2013). Mesangiosperms, by contrast, include more than 99.95% of flowering plants, and are divided into four main clades: magnoliids (including Chloranthaceae), monocots, Ceratophyllales and eudicots (**Figure 1**) (Frohlich & Chase, 2007; Group, 2009; Group et al., 2016; Zeng et al., 2014).



**Figure 2.** The angiosperm fossil records from the Cretaceous period suggest a paleotropical origin of flowering plants and a slightly later migration towards temperate regions. (Coiro et al., 2019). Valanginian to Late Albian correspond to Lower Cretaceous subdivisions from 139 MYA to 100 MYA approximatively.

## **What is a flower? -The modern flower and its components.**

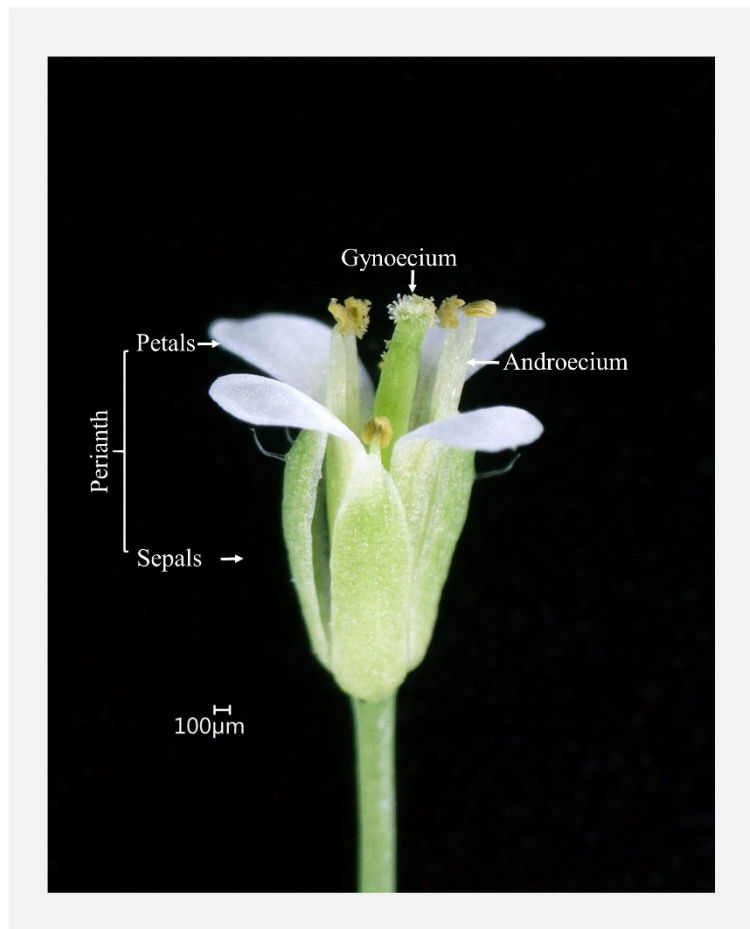
As mentioned above, most of the synapomorphic characters of angiosperms are reproductive characters, which are grouped in the structure called the flower. The flower can be described as a compact reproductive axis, with lateral organs arising in close juxtaposition. Floral organs can be arranged in a spiral phyllotaxy leading to a variable number of organs. However, most flowering plants present relatively fixed numbers and positions of floral organs arranged in whorls. Typically, a bisexual flower has a central gynoecium formed by carpels, surrounded by the androecium of stamens that produce the pollen grains (**Figure 3**). However, there are also angiosperm species bearing unisexual flowers, in which only the gynoecium or androecium develops.

Enclosing the reproductive organs in a typical flower is the perianth, composed of sterile bract-like organs. The perianth is frequently divided into an inner whorl of petals, which may participate in interactions with pollinators, and an outer whorl of sepals, which may function mainly to protect the floral bud (**Figure 3**). Also, in some cases, the perianth may be undifferentiated, in which case its organs are called tepals, and these can be either petals-like or sepal-like.

## **What did the first flower look like, and what were the possible ancestors of angiosperms?**

As the main character that differentiated angiosperms from their precursors, the gymnosperms, a large part of the research effort in plant evolution has been devoted to the reconstruction of the ancestral flower. Over the last few decades, a series of techniques and methods have been developed and combined in order to attempt to reconstruct this structure. These methods include the comparative study of fossils and the morphological and molecular phylogeny of living angiosperms. Sophisticated statistical models can now be applied to the data obtained from comparative studies to reconstruct the flower of the MRCA of flowering plants (De-Paula et al., 2018; Rümpler & Theißen, 2019; Sauquet et al., 2017; Sokoloff et al., 2018).

Since the beginning of the 21st century, botanical experts have tested proposals for the possible structure of the ancestral flower. At the end of the 1990's, the identification of the ANA-grade group greatly facilitated the reconstruction of the MRCA of angiosperms (Rümpler & Theißen, 2019). In the following years, several attempts were made, in most of the cases using phylogenetic analyses based on the principle of parsimony (James A. Doyle, 2008, 2012; Endress & Doyle, 2007, 2009, 2015). The development of



**Figure 3.** *Arabidopsis thaliana* flower, showing the typical bisexual flower with its four whorls, from the center to outside: the gynoecium formed by two congenitally fused carpels, the androecium formed by six stamens, whorls of perianth organs, respectively containing four petals and four sepals.

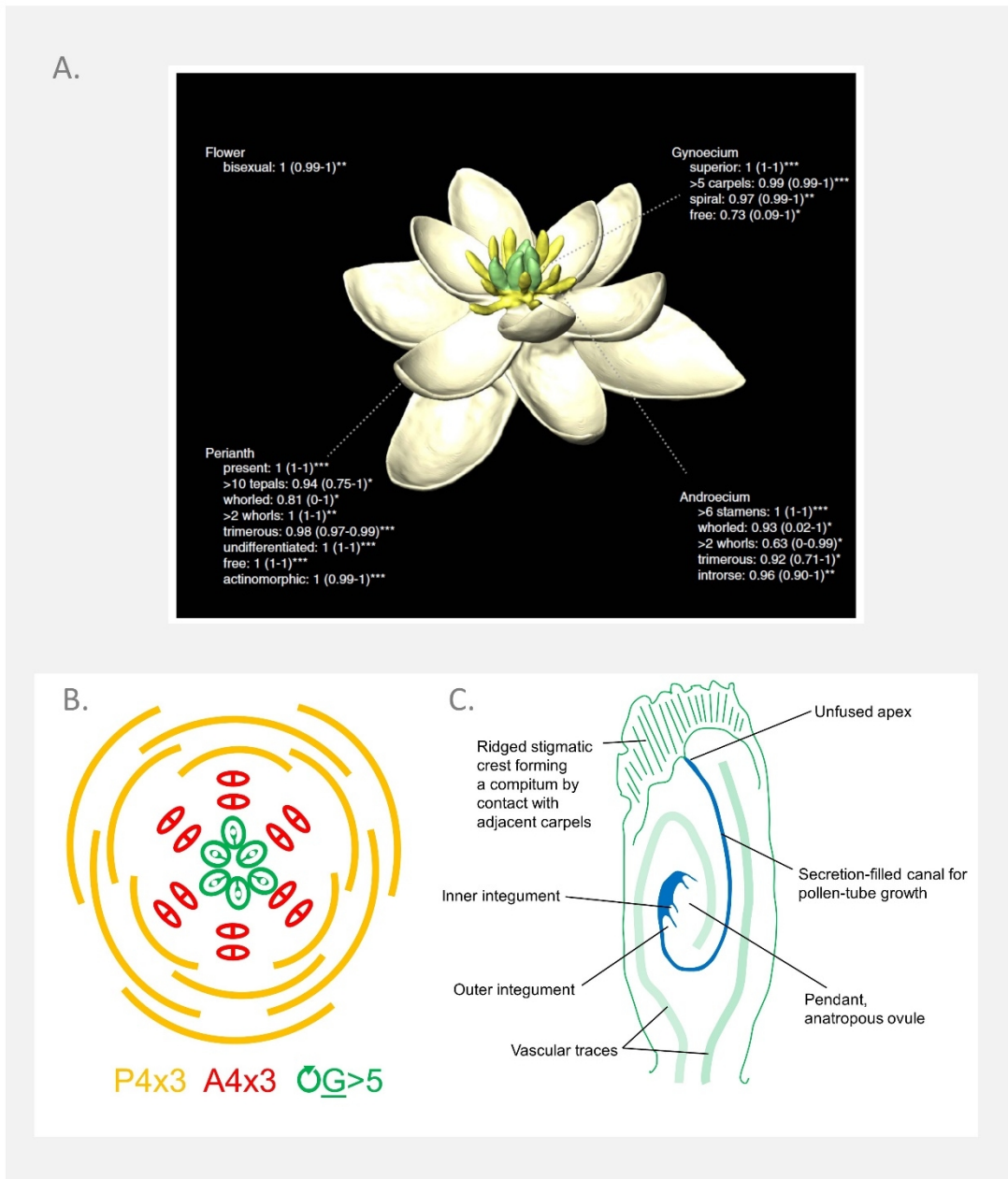
new methods, and the availability of new fossil and molecular data, has further facilitated these reconstructions.

In the reconstruction presented by Sauquet et al. (2017) (**Figure 4. A and B**), the ancestral flower is presented as a bisexual and actinomorphic flower, with a superior gynoecium composed of five free carpels with spiral insertion. Six or more stamens are present, arranged in two whorls, and surrounded by a whorled perianth of tepals, arranged in trimers. This model has, however, been criticized because of a switch it proposes in floral organ phyllotaxy between the androecium and gynoecium: such a change in is not found in extant angiosperms (Sokoloff et al., 2018). Also, the reconstructions of Sauquet et al. (2017) assessed character states independently from each other, and so may have produced a list of characters that were mutually incompatible (De-Paula et al., 2018).

Reyes et al., 2018, postulated that the high heterogeneity of extant flowers may have an impact in the reconstruction of ancestral states of characters, and that this effect is mostly observed in perianth organs. Despite these various caviats, it may be concluded, from a synthesis of studies, that flowers in the MRCA living angiosperms were probably small, protogynous and bisexual. Their gynoecium was most likely superior and composed of five or more carpels. These flowers probably had an undifferentiated perianth of tepals, it is possible that the androecium and the perianth were arranged in two or more trimerous whorls. The first flowering plant were probably insect-pollinated, and provided pollen as a reward.

Another aspect of the origin of angiosperms to be addressed concerns the gymnosperm group from which the angiosperms emerged. The patchy nature of the relevant fossil record, and the absence of close living relatives of the angiosperms, have been major aspects contributing to this problem. In the “anthophyte hypothesis”, the results of morphological cladistic analyses had suggested the gymnosperm groups closest to angiosperms were living Gnetales and mesozoic (extinct) Bennettitales. However, with the inclusion of molecular analyses from all major groups of living seed plants, this hypothesis (at least regarding Gnetales) was largely refuted (James A. Doyle, 2012).

Morphological data from both living and extant group have, however, permitted the tentative placement of extinct gymnosperms on a molecular phylogenetic backbone of living seed plants (**Figure 5**) (James A. Doyle, 2008, 2012). From these studies, the two extinct taxa shown to be the closest possible angiosperm stem-lineage relatives are Caytoniales and Bennettitales (**Figure 5**). *Caytonia* was the first member of Caytoniales to be discovered. *Caytonia* fossils represents female reproductive structures, while



**Figure 4 . A.** Reconstruction of the ancestral flower by Sauquet et al., (2017). **B.** Floral diagram and floral formula of the ancestral angiosperm based on Sauquet et al., (2017) and other references from Charles P. Scutt (2018). **C.** *Trimenia* an ANA-grade carpel anatomy that helps to visualize the ancestral carpel structure from Charles P. Scutt (2018).

the pollen-producing structures of *Caytonanthus* and the leaves of *Sagenopteris* may correspond to the same species.

*Caytonia* consists of an axis bearing laminate cupules, each of which contains multiple ovules. These cupules are folded from the tip to the base and could be compared to the ovule-containing outer integument in angiosperms. However, homology between structures in *Caytonia* and the angiosperm carpel is more difficult to define (**Figure 5**) (James A. Doyle, 2008). Some interpretations propose that the angiosperm carpel derives from the female rachis in a *Caytonia*-like ancestor. However, Bennettitales, which have no outer integument or carpel-like organs are also considered strong candidates for a close relationship to angiosperms. Bennettitales present the angiosperm-like features of net-veined leaves, a bisexual reproductive axis (in some species), nonsaccate pollen grains, and a perianth-like, which gives the reproductive axis a flower-like appearance (**Figure 5**). Another shared character with angiosperms is the capacity of Bennettitales to synthesize oleananes, which are terpenoid compounds, which are so stable they can persist in fossils (Taylor et al., 2006).

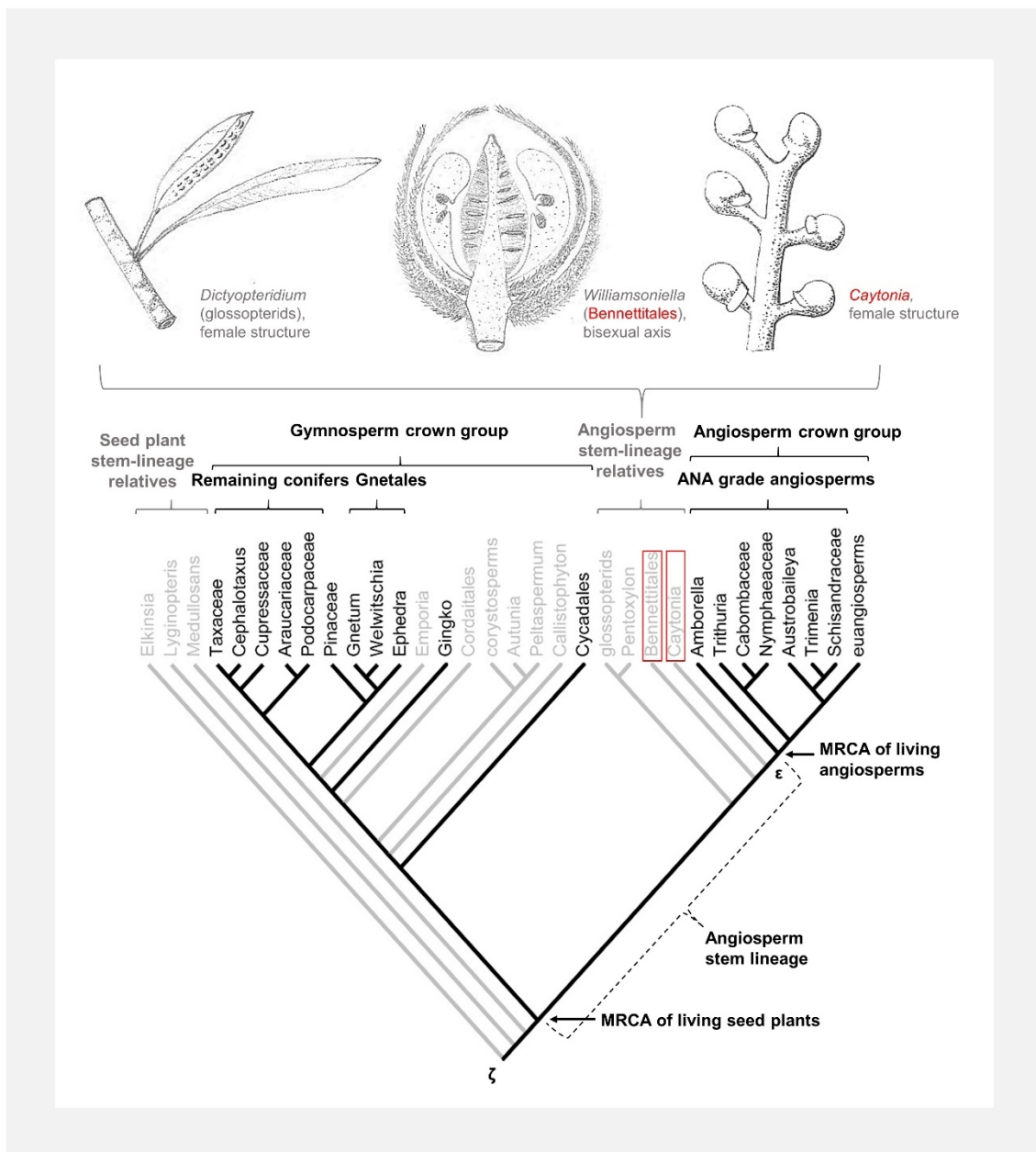
### **A closer look at female reproductive tissues.**

Moving on from the origin, appearance and ancestors of the first flowering plants, I will focus in this work mainly on the evolution of the female part of the flower. Situated at the center of the typical flower, the gynoecium is composed of one or several carpels. The carpel has probably conferred on angiosperms several major advantages in the process of sexual reproduction. These functions include the mechanical and biochemical protection of the ovules, the reception of pollen grains, and pollen-tube guidance. Carpel tissues also act as a selective barrier to fertilization. After fertilization, the wall of the ovary within the carpel develops further to form the fruit, which protects the seeds and then facilitates their dispersion.

### **The ancestral carpel and its origin.**

The gynoecium of the flower in the MRCA of living angiosperms was most probably composed of separate, ascidiate (bottle-shaped) carpels. The carpel apex was probably not completely fused by cellular structures, but instead had a restricted entrance in the form of an aperture or canal. From this canal, substances would have been secreted to close the apex of the carpel and facilitate the entry of pollen tubes (Endress & Doyle, 2015).

The stigmatic surfaces of the ancestral carpel were probably covered by multicellular striations that may have been physically in contact among adjacent carpels, forming an intragynoecial compitum to allow the

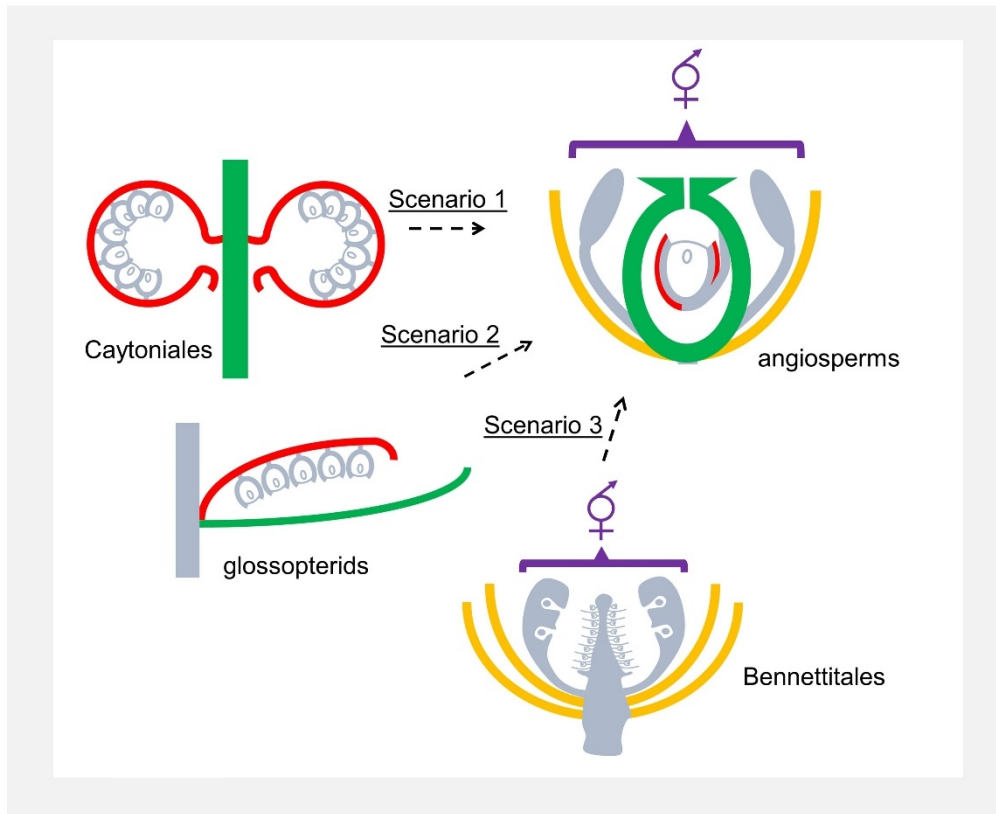


**Figure 5** Based on Doyle, (2012) phylogeny of seed plants, including extant and extinct species. Showing possible angiosperm stem-lineage relatives. Fossil taxa are in gray and the closest possible stem-lineage relatives to angiosperms are indicated in red (from Charles P. Scutt, 2018).

exchange of pollen tubes (Endress & Doyle, 2015). The ancestral carpel probably enclosed a single pendant ovule, or a small number of such ovules. These ovules were likely bitegmic and contained, in addition to the female gametophyte, a large nucellus, or female-derived storage tissue. The fruits of the MRCA of living angiosperms were probably indehiscent (Becker, 2020; Endress & Doyle, 2015). To get an idea of the appearance of the ancestral carpel, it may be useful to look at ANA-grade carpels, such as those of *Trimenia* (Austrobaileyales), which conserve most of the inferred features of the ancestral carpel (**Figure 4. C**).

The carpel is considered as a modified leaf that has closed around the ovules, but this structure is probably more directly related to a sporophyll: a specialized leaf bearing one or more sporangia. If the carpel is considered as related to the (female) megasporophyll (though this is not the case in at least one hypothesis by M. W. Frohlich, (2003), then the female cone scales of conifers, for example, would be considered homologous to the carpel. Depending on which fossil candidate is taken as the possible ancestor of flowering plants, different possible anatomical precursors to the carpel can be recognized. (Doyle, 2008)(**Figure 6**) identified three groups as possible close stem-relatives of angiosperms, and these are related to three different evolutionary scenarios for the origin of the carpel and outer integument of angiosperms. In the first of these scenarios, *Caytonia*-like gymnosperms are considered as potential ancestors of angiosperms, and in this case the outer integument would have evolved from the cupule by a reduction in ovule-number to one, leaving a single ovule now enclosed in two tissue layers, corresponding to the two integuments of angiosperms. If *Caytonia*-like gymnosperms were ancestral to angiosperms, it has been hypothesized that the female rachis then expanded and folded around the bitegmic ovules to form the carpel.

In a second evolutionary scenario, a glossopterid-type of female reproductive axis is considered as a possible progenitor to the outer integument and carpel of angiosperms. In this scenario, the cupule became the outer integument, as in the first scenario above, relating to *Caytonia*. However, in the case of glossopterids, the cupules emerge from a subtending bract, and it this bract that is considered to have possibly closed around the cupule/bitegmic ovule to form the carpel. One difficulty for this scenario relates to the date of the extinction of glossopterids, which may have preceded the origin of flowering plants (James A. Doyle, 2008). Indeed, glossopterids have also been suggested as possible ancestors of Caytoniales (including *Caytonia*). However, as there is still much uncertainty in the date of origin of angiosperms (as discussed above), glossopterids remain contenders as possible angiosperm ancestors. The third alternative scenario, based on the analyses of (Doyle, 2008), postulates that Bennettitales were the



**Figure 6.** Three alternative scenarios for the origin of angiosperm female reproductive structures, including the outer integument of the ovules, the carpel, the bisexual axis and the vegetative perianth. Color codes are used to indicated homologous characters. (from Charles P. Scutt, 2018).

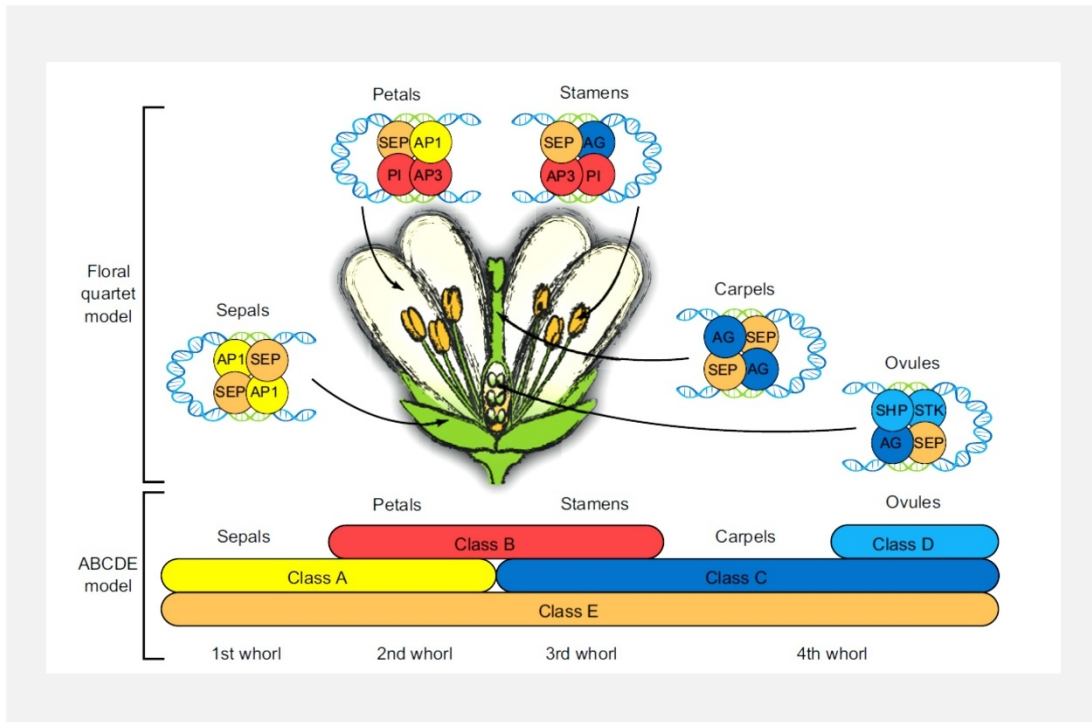
ancestors of angiosperms. These plants have megasporophylls attached to a central rachis, which might be homologous to the carpel, though they have no cupule-like organs of possible homology to the angiosperm outer integument (**Figure 6**).

## **Molecular mechanism that control the floral development in extant angiosperms**

Despite its great morphological diversity, the flower is, in general, composed of four types of organs arranged in whorls. Floral-organ identity is specified by homeotic genes which function in a genetic model termed the ABC model (Coen & Meyerowitz, 1991). This model was extended to the ABCDE model, which notably incorporates a further class of genetic component, the E-class, which was found to act together with the other floral homeotic genes (**Figure 7**) (Pelaz et al., 2000; Theissen & Saedler, 2001). In the ABC model, A-class genes are required to specify sepal identity, A and B-class genes together establish petal identity, B and C-class genes together establish stamen identity, and C-class genes specify carpel identity. The E-class are genes necessary to support the activities of the A, B and C class genes (**Figure 7**), while D-class genes were originally identified as necessary for specifying ovule development (Colombo et al., 1995). However, D-class genes are closely related to C-class genes and may have overlapping activities with these, depending on the species analyzed. (Soltis et al., 2007; Thomson & Wellmer, 2019).

Almost all homeotic genes of the ABCDE model are MIMIC-type MADS-domain transcription factors, which contain four conserved domains, among which the MADS (M) domain is the most highly conserved region and is associated with DNA-binding activity, dimerization and nuclear localization (Theissen et al., 2016; Thomson & Wellmer, 2019). In *Arabidopsis thaliana*, the A-class genes are *APETALA1* (*AP1*) and *APETALA2* (*AP2*), the B-class genes are *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), the C-class gene is *AGAMOUS* (*AG*), the D-class genes are *SEEDSTICK* (*STK*), *SHATTERPROOF1* and *2* (*SHP1*, *SHP2*), and finally *SPETALLATA* 1-4 (*SEP1-4*) encode E-class genes (Becker & Ehlers, 2016). Of the above-listed genes, only *AP2* is not a member of the MADS-box family, but is instead the founder member of the *AP2*-domain family of transcription factors (Okamuro et al., 1997).

The floral quartet model explains how the homeotic genes of the ABCDE model interact with downstream target genes to control the formation of different floral organs (**Figure 7**), as reviewed by (Thomson & Wellmer, 2019). In addition to regulating diverse downstream targets, genes of the ABCDE model also establish an extensive genetic cross talk and regulate their own expression (Becker & Ehlers, 2016).



**Figure 7.** The ABCDE model of *Arabidopsis* flower development showing homeotic genes that act to specify floral organ identity. The A-class genes are *APETALA1* (*AP1*) and *APETALA2* (*AP2*), the B-class genes are *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), the C-class gene is *AGAMOUS* (*AG*), the D-class genes are *SHATTERPROOF1*, and 2 (*SHP1*, *SHP2*) and *SEEDSTICK* (*STK*), and finally *SPETALLATA 1-4* (*SEP1-4*) encodes E-class genes (from Theißen et al., 2016)

The ABCDE model has only been directly investigated in a few model angiosperms (Becker & Ehlers, 2016; Litt & Kramer, 2010). However, *in-situ* hybridization assays, protein-DNA interaction data, and functional analyses suggest conserved mechanism of class B and C genes between extant gymnosperm species and flowering plants (Moyroud et al., 2017; Wang et al., 2010; Winter et al., 2002; Zeng et al., 2014).

## **The genetic regulation of female reproductive tissues in *Arabidopsis thaliana***

Evolutionary changes within the clades of MADS-box genes of the ABCDE model may have been a key step for the origin of the flower, and of the carpels within it (Litt & Kramer, 2010; Charles P. Scutt, 2018; Charlie P. Scutt et al., 2006). Here, I first describe the genetic interactions that control carpel development, focusing initially on *Arabidopsis thaliana*. Stages of flower development referred to are from (Smyth et al., 1990).

The gynoecium of *A. thaliana* is composed of two congenitally fused carpels that develop in the center of the flower. These are the last organs formed by the Floral Meristem (FM), which goes on to lose its meristematic activity at Stage 6 (Zúñiga-Mayo et al., 2019). The female developmental program starts at Stage 3, when the C-class *AG* factor is activated by *WUSCHEL* (*WUS*) and *LEAFY* (*LFY*). *AG* is expressed in the FM and is required to repress *WUS* activity. At the Stages 4-5, *REBELOTE* (*RBL*), *ULTRAPETALAI* (*ULT1*) and *SQUINT* (*SQN*) promote *AG* expression, specifically in the fourth whorl. This expression is essential to guarantee FM termination by continuously repressing *WUS* (Zúñiga-Mayo et al., 2019).

*AG* interacts with *SEPALLATA1-4* proteins (*SEP1-4*) to specify carpel identity (Pfannebecker et al., 2017a). After carpel identity is determined by *AG* expression, several genes of several classes, most of which encode transcription factors, act to regulate the developmental program of the *A. thaliana* gynoecium. A meristematic zone appears at the margins of the carpel tissues to produce new tissues in the medial region of the gynoecium. This meristem is termed the Carpel Margin Meristem (CMM) (Reyes-Olalde et al., 2013). Tissues that are very important for sexual reproduction derive from this meristem (Reyes-Olalde et al., 2017).

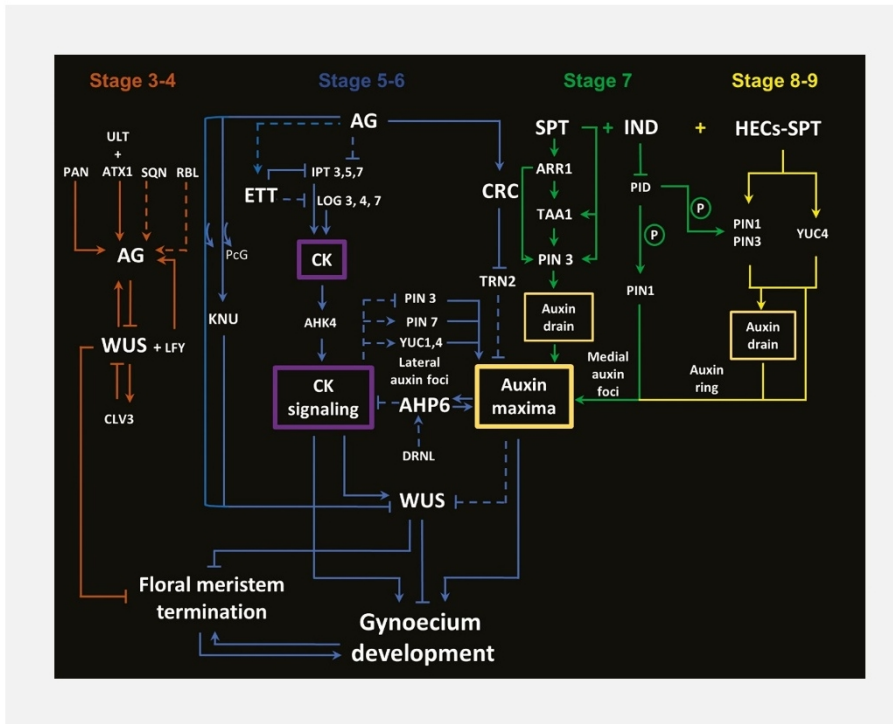
Besides the genes involved in gynoecium development, phytohormones also play an essential role. These include auxin and cytokinin, which together promote the development medial domain of the gynoecium (Reyes-Olalde et al., 2017). Mainly following the recent review of (Zúñiga-Mayo et al., 2019), I present below a summary of the regulation of gynoecium development in *A. thaliana*.

From Stages 3- 5, *AG* expression is necessary to determine gynoecium/carpel identity and the termination of the FM. *AG* coordinates both of these processes through the regulation of crosstalk between cytokinin and auxin, repressing *WUS* both directly and indirectly through *KNUCKLES* (*KNU*). At Stage 5, there is a peak of cytokinin at the apex of the gynoecium, while auxin is concentrated in two lateral foci in the apical cells. These patterns of hormone distribution are coordinated by the action of the flavin monooxygenases *YUCCA1 and 4* (*YUC1/4*), and of the auxin efflux carriers *PINOID1, 3 and 7* (*PIN1, 3 and 7*). Active repression of cytokinin also occurs through *ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEIN6* (*AHP6*) (**Figure 8**).

*ARF3/ETTIN* (*AUXIN RESPONSE FACTOR3, ETT*) is activated by both *AG* and auxin. Between Stages 5 and 7, *ETT* is involved in FM termination and gynoecium development. At this stage, it controls valve elongation through the positive regulation of pectin methylesterase (PME) activity in the cell wall (Andres-Robin et al., 2018). *ETT* also represses *WUS* activity and the expression of genes encoding cytokinin biosynthesis. At Stage 6, the gynoecium begins to grow apically. *AG* activates the expression of *CRABS CLAW* (*CRC*) a YABBY transcription factor, involved in lateral polarity in the carpels (J. Alvarez & Smyth, 1999). At this stage, *WUS* repression is reinforced by *CRC* and its target gene *TORNADO2* (*TRN2*) (**Figure 8**).

At Stage 7, *SPATULA* (*SPT*) a bHLH transcription factor, is expressed in the medial domain of the gynoecium to ensure proper development of the medial tissues. From Stages 7-9, *SPT* acts with other bHLH family members *INDEISCENT* (*IND*) and *HECATE1- 3* (*HEC1-3*). Style and stigma development are controlled by SPT-IND protein dimers through the regulation of auxin distribution. Meanwhile, *HEC1- 3* act redundantly to control median tissue development. At Stage 10, the CMM is thought to no longer exist (**Figure 8**).

This is a very quick and general review of early gynoecium development, as there are many other genes that interact in the control of *A thaliana* carpel development. For example, more than 80 genes are believed to regulate the development of the CMM (Reyes-Olalde et al., 2013). It is important to highlight that *Arabidopsis* does not necessarily typify the gynoecium regulatory network of all angiosperms. Indeed, the carpel regulatory networks of other core eudicots, such as Solanaceae, shows important differences to that of *A thaliana* (Ortiz-Ramírez et al., 2018).



**Figure 8.** Genes involved in gynoecium development of *Arabidopsis thaliana* flowers from Stages 3-9. Key: *WUSCHEL* (*WUS*), *LEAFY* (*LFY*), *AGAMOUS* (*AG*), *ULTRAPETALA1* (*ULT1*), *SQUINT* (*SQN*), *REBELOTE* (*RBL*), *KNUCKLES* (*KNU*), *YUCCA1* and *4* (*YUC1/4*), *PINOID1 3* and *7* (*PIN1, 3* and *7*), *ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEIN6* (*AHP6*), *AUXIN RESPONSE FACTOR3/ETTIN* (*ARF3/ETT*), *CRABS CLAW* (*CRC*), *TORNADO2* (*TRN2*), *SPATULA* (*SPT*), *INDEISCENT* (*IND*), *HECATE1, 2* and *3* (*HEC1, 2* and *3*) (from Zúñiga-Mayo et al., 2019).

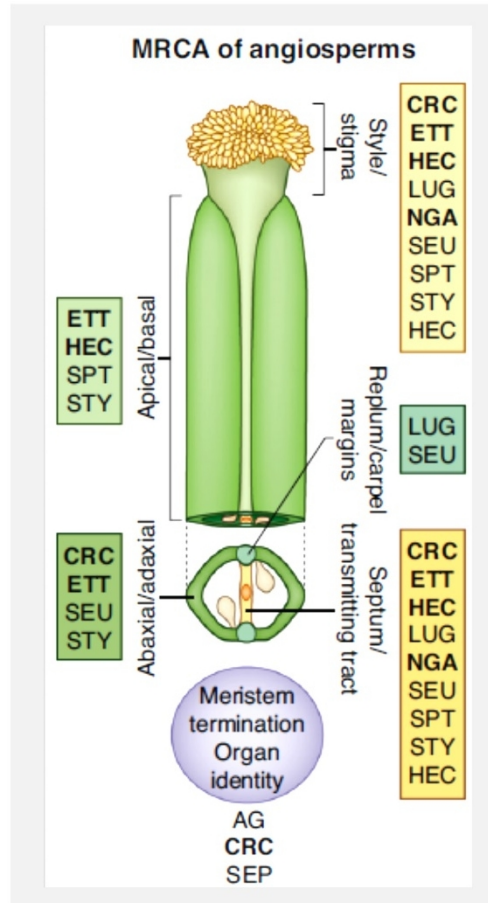
## The Gene Regulatory Network of the ancestral carpel

The *A thaliana* gynoecium is highly complex and contains two fused carpels, a dry stigma, a false septum and transmitting tract. ANA-grade species, representing the basal-most living angiosperms, mostly have unfused ascidiate carpels, and no transmitting tract or false septum. The ancestral carpel described above, similar to those of present-day basal angiosperms, may have had a limited GRN when compared with present-day *A thaliana*, partly because of the absence in the former of several specialized tissues described in the latter (Pfannebecker et al., 2017b).

Through the phylogenetic reconstruction of transcription factors (Pfannebecker et al., 2017a, 2017b) have estimated the age of gene families involved in carpel development. According to these authors, lineages including *AG* and *SEP* were already present in the MRCA of seed plants, whereas other regulators were already present in the MRCA of all land plants. The lineages encoding the bHLH transcription factors *SPT* and *HEC*, present also in living gymnosperms, were probably therefore present in the MRCA of living seed plants.

Becker A (2020) reviewed carpel development regulators, and presented a list of genes that were most probably active in the carpel regulatory network in the MRCA of flowering plants (**Figure 9**). According to this work, *AG*- and *SEP*-like genes were present in the ancestral flower and specified carpel identity, together with *CRC*. In early angiosperms, adaxial-abaxial polarity was already under the control of *ETT* and *CRC*, in combination with other transcription factors. *ETT*, *SPT* and *HEC* probably acted together to control apical/basal polarity in the carpel. The regulation of median and apical tissues was probably brought about through the action of *ETT*, *LEUNIG (LUG)*, *SEUSS (SEU)*, *STYLISH (STY)* and *NGATHA (NGA)*, in combination with *CRC* and the bHLH factors *SPT* and *HEC* (**Figure 9**).

The above description represents a simplified regulatory network of the ancestral carpel. To improve our knowledge in this domain, it will be necessary to include data from more species that are strategically placed across seed plant evolution. The next section of this introductory chapter, therefore, concerns species which occupy key phylogenetic positions for the reconstruction of carpel development mechanisms from the earliest stages of angiosperm evolution.



**Figure 9.** Transcription factors and regulators of carpel development that may have been present in the carpel regulatory network of the MRCA of flowering plants, from Becker A, (2020). Key: *AGAMOUS* (*AG*), *CRABS CLAW* (*CRC*), *SPETALLATA* (*SEP*), *ETTIN* (*ETT*), *SEUSS* (*SEU*), *STYLISH* (*STY*), *LEUNIG* (*LUG*), *SPATULA* (*SPT*), *HECATE* (*HEC*), *NGATHA* (*NGA*).

## Non-model species for studies of the origin of the flowers

In addition to the work carried out on model species, taxa from strategically-placed groups should be included in studies aimed at elucidating the origin of the carpel. In particular, gymnosperms and basal angiosperms could help to elucidate the genetic regulation of reproductive tissues in the MRCA of flowering plants by comparing the molecular networks that control reproductive tissue development in these two groups.

Scutt, CP. 2019, has made a summary of the requirements for potential new model species to study early angiosperm evolution. Such models should be amenable to standard molecular biology procedures and functional genetic approaches, and have a genome size that facilitates sequencing and assembly tasks. Further requirements include easy access to biological material at all developmental stages for laboratory work, small size at maturity, self-fertility, a short generation time, and the production of copious amounts of seed. Below, I summarize the features of groups of model and non-model species that are strong candidates in the light of the above-listed requirements.

**Gymnosperms.** As the closest living relatives of the angiosperms, the gymnosperms form an important external reference point to study the evolutionary processes that led to the origin of the flower. Even if many of the traits of the typical gymnosperm reproductive axis differ from those of angiosperms, numerous orthologs of floral development genes are known to be conserved in gymnosperms. Analyses of the molecular mechanisms of male and female cone development in gymnosperms are possible using a number of methods. The techniques available include *in situ* hybridization, *in-vitro* studies of the protein-protein and protein-DNA interactions involving transcription factors, and *in-vivo* functional analyses by the complementation of model-angiosperm mutants by genetic transformation using gymnosperm coding sequences. Also very useful is the increasing availability of gymnosperm transcriptomic data, in particular datasets from reproductive structures.

However, most of the data currently available from gymnosperms are from a few species of high economic interest such as the conifers *Picea abies* and *Pinus taeda*, and do not cover the whole range of living gymnosperms. Another problem in working with gymnosperm species concerns sampling, including the slow growth of many gymnosperm reproductive tissues, and difficulties accessing cones that develop at tens of meters from the ground among dense and fragile branches. Also gymnosperms tend to have very large genomes, which complicates genome sequencing and assembly. However, some of these problems may be solved, notably with the emergence of new methods and decreasing costs of sequencing, and the use of gymnosperms that have either shorter developmental cycles like

*Welwitschia mirabilis* and *Ephedra spp*, or which produce cones at accessible levels, like the *acrocona* mutant of *P. abies*.

**ANA-grade angiosperms:** The orders Amborellales, Nymphaeales and Austrobaileyales form the three most basal lineages of flowering plants, termed the ANA grade. Amborellales contains the single living species *Amborella trichopoda*, considered as probable sister to all other extant angiosperms. The second earliest diverging ANA-grade order Nymphaeales regroups three families: Nymphaeaceae (water lilies, with a cosmopolitan distribution), Cabombaceae and Hydatellaceae (aquatic and semiaquatic herbs with more restricted distributions). The Austrobaileyales order is the probable third earliest-diverging order of the ANA grade, and includes approximately 100 woody plant species in three families.

*Amborella trichopoda* (**Figure 10. A and B**) is a key taxon for reconstructions of the GRN in early flowers, due to both its unique phylogenetic position and the absence of any whole-genome duplications in its lineage since the MRCA of living angiosperms. The genome of *Amborella* is therefore also an important reference to study genome evolution among angiosperms. *A trichopoda* is an understory shrub, endemic to sub-tropical cloud forests of New Caledonia. It is a dioecious species, whose flowers have an undifferentiated perianth of tepals. *Amborella* female flowers have free, ascidiate carpels, each of which produces a secretion that fills an apical canal to provide a route for pollen-tube growth. Each carpel also has a large stigmatic surface for pollen contact. Inside each *Amborella* carpel is a single, pendent, bitegmic ovule. The *Amborella* ovule possesses a unique eight-celled embryo sac that, after double fertilization, generates a diploid zygote and triploid endosperm.

Morphological analyses of *Amborella* have helped establish a list of plesiomorphic features in angiosperms. Interestingly, some differences in *Amborella* compared to other ANA-grade angiosperms, such as its near orthotropous ovule symmetry, lead to ambiguities in reconstructions of the MRCA of extant flowering plants. Molecular analyses such as *in situ* hybridization can be performed in *Amborella*, and this technique has been extensively used to study the conservation of expression of flower development genes. However, its dioecious breeding and woody habit make *Amborella* a poorly adapted species for functional-genetic studies.

Several species of Nymphaeales have been used as evolutionary developmental models, including cultivars of *Nuphar* and *Nymphaea*. However, the most promising Nymphaeales model for functional-genetic studies is perhaps *Nymphaea thermarum*. This species was endemic to Rwanda, but is now considered extinct in the wild. *N thermarum* (**Figure 10. C and D**), is the smallest water lily, the diameter

of adult plants being 10-20cm, and that of mature, open flowers being up to 2 cm. It is a perennial herb whose life cycle from seed to flowering takes 2-3 months. It can be easily cultivated in a warm, humid and well-lit environment, and grown in pots of wet compost, rather than completely immerse in water.

The *N. thermarum* flower has whorled phyllotaxy. Its perianth has two sets of tepals: four greenish outer tepals and six to eight whitish inner tepals. The gynoecium of *N. thermarum* is comprised of seven to nine basally-fused carpels with post-genital fusion at the carpel margins. Within each carpel, a high number of bitegmic, anatropous ovules develop. These ovules have four-celled embryos sacs which, after double fertilization, produce a diploid embryo and a diploid endosperm. *N. thermarum* is self-compatible, and each fruit typically produces around 150 seeds.

Currently, the scientific literature on *N. thermarum* is limited. However, this species possesses several advantages as a potential model ANA-grade angiosperm: the small size of its flowers makes it practical for *in situ* hybridization and other microscopic procedures, its diploid genome is one of the smallest within basal angiosperms, and its seeds can easily be sterilized for *in-vitro* culture. The development of tissue culture and protocols for stable genetic transformation might be possible in this species, and it might also be possible also to generate seed-based mutant collections, depending on the (currently unknown) suitability of *N. thermarum* seed for long-term storage.

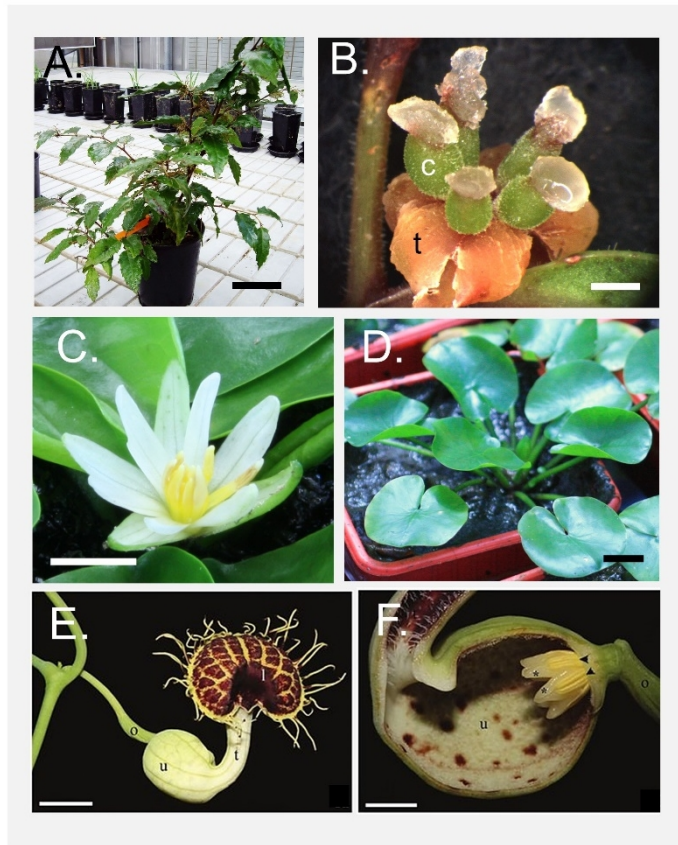
The Austrobaileyales order includes plants with various breeding systems, such as a mixture of monoecy and dioecy in *Schisandra chinensis*, protogyny in *Austrobaieleya scandens* and self-incompatibility in *Illicium floridanum* and *Trimenia mooreii* (Bernhardt et al., 2003; THIEN et al., 1983). Flowers of Austrobaileyales mostly possess a spiral phyllotaxy with an undifferentiated perianth of tepals. The typical Austrobaileyales gynoecium consists of unfused carpels (except partially fused in *Illicium*), each of which has a secretion-filled apical canal for pollen tube growth. Inside each carpel a single or a low number of bitegmic ovules develop.

Species of Austrobaileyales have been extensively incorporated in analyses for morphological reconstructions of the MRCA of extant angiosperm. However, their woody habit makes them poorly adapted as potential molecular-genetic models, and in consequence this group has not been extensively used for molecular studies of flower evolution.

**Magnoliids.** The magnoliids are a large clade of basal angiosperms which emerged shortly after the ANA-grade lineages. There are five orders of magnoliids regrouping around 11000 species, including both herbaceous and woody plants. Herbaceous magnoliids may make more practical models for the study of the origin of angiosperms than the woody species of Amborellales and Austrobaileyales, or

the aquatic herbs of Nymphaeales, within the ANA grade. The phylogenetic position of magnoliids make this clade also an important external reference point for the origin of later-emerging clades of angiosperms, such as the eudicots and monocots.

*Aristolochia fimbriata* is a promising model magnoliid. Plants of this species are small, have a rapid life cycle, are self-compatible and possess a small genome. *A. fimbriata* is a vine that attract its pollinators (flies) into a chamber formed by a bilaterally symmetrical perianth of fused sepals (**Figure 10. E and F**). Bilateral symmetry has arisen several times independently in angiosperms, but this may have occurred first inside the magnoliid clade. The presence of an inferior ovary, and the stamens fused to the style to form a gynostemium, are further distinct floral features of *A. fimbriata*. *A. fimbriata* has important advantages for laboratory work as it can be transformed using *Agrobacterium tumefaciens*, and is readily amenable to standard molecular procedures such as *in situ* hybridization.



**Figure 10.** Non model species of interest for evo-devo studies of angiosperm origin and evolution. A and B, *Amborella trichopoda*. C and D *Nymphaea thermarum*. E and F, *Aristolochia fimbriata*. c, carpel; t, tepals; l, limb; o, ovary; t, tube; u, utricles. (A-D from Scutt C.P., 2019; E and F from Pabón-Mora et al., 2015).

## Techniques and methods that could help to explain the origin of the flower

Several approaches are needed to elucidate the origin of the flower and the angiosperms. Recent developments offer the possibility of better understanding this major event in plant evolution. Here, I describe the techniques that are of particular importance for these studies:

1. **Next Generation Sequencing (NGS):** The increasing efficiency and decreasing costs of NGS facilitate whole genome and transcriptome sequencing. The problem of genome assembly in the absence of physical linkage maps and genetic markers etc, is compensated now by certain recent NGS methods that provide reads extended to tens of kilobases. These methods will be very useful for the sequencing of basal angiosperms and gymnosperms, which in general have large genomes.
2. **Bioinformatics and statistical analyses:** the development of new bioinformatics and statistical packages could help comparisons of the expression profiles of genes in model and non-model species. Also, such methods could facilitate the analysis of genetic modules involved in flower development across the flowering plants.
3. **Synteny and phylogeny:** The reconstruction of the structure of ancestral genomes from those of their living descendants (Murat et al., 2017) is now possible due to progress in methods based on the combined analysis of synteny and phylogeny. It may soon be possible to reconstruct, in considerable detail, the genome of the ancestor of living angiosperms, and even that of their more distant ancestor from before the epsilon whole genome duplication event (Jiao et al., 2011).
4. **Functional studies:** Recent advances in gene editing should significantly facilitate functional studies, allowing the direct study of the function of developmental regulators in basal angiosperms and gymnosperms. Well-adapted model ANA-grade angiosperms could be adopted to study the function of genes that were crucial genes to the origin of the angiosperms. Strong candidates for such models are the recently discovered *Nymphaea thermarum*, and possibly also certain *Trithuria spp.* (both from Nymphaeales), if transformation methods can be developed for these species.
5. **In-vitro and heterogeneous in-vivo methods:** Questions in evolution and development of the flower could be answered by combining methods that measure protein-DNA or protein-protein interactions with genomic-scale analyses and modelling approaches. Such approaches could be used to describe the networks that control reproductive development at the post-transcriptional, transcriptional and epigenetic levels. Such methods, which do not depend on

functional-genetic capabilities, can help to elucidate the evolution of developmental regulatory mechanisms in non-model plants (Viallette-Guiraud et al., 2016a) that are important for the study of angiosperm origin.

6. **“Protein Resurrection”**: ancestral sequence reconstruction from key stages in plant evolution, can be used to study the biophysical and biochemical properties of ancestral molecules that form part of the regulatory network of the reproductive axis.
7. **Paleobotany**: it would be very useful to discover further fossils of potential stem lineage-relatives of the angiosperms, and also of early flowers. The latter may have been preserved as mesofossils (fossils of a few millimeters in diameter) which can be generated by events such as forest fires which generate “coalified” specimens (Schonenberger, 2005). Recently developed tomographic methods could help examine anatomical details at the cellular and even sub-cellular level in such fossils (von Balthazar et al., 2007).

## Unsolved questions of flower origin and evolution

We are still far from a complete knowledge of the GRN that was present in the ancestral flower or from answering the question of how the flowering plants first appeared. There remains some controversy on the characteristics of the ancestral flower, and much uncertainty concerning the gymnosperm group from which the flowering plants emerged. Also, little is known of the molecular changes that were required for the origin of angiosperms.

In this work, I focus on the origin and evolution of female reproductive tissues of angiosperms. The central question I address is: “*what is the molecular basis of the evolution of the carpel?*” In attempting to answer this question, I have divided my work into two parts, each of them responding to a more specific question derived from the above central question. The questions, aims and strategies of my work are specified in the next section.



# **AIMS AND STRATEGIES**

---

The carpel is one of the main characters that differentiates angiosperms from its closest sister clade, the gymnosperms. This thesis is centered on the study on the evolution of the carpel and other female reproductive tissues. Its main aim is to help elucidate the Gene Regulatory Network (GRN) that controlled female reproductive development in the MRCA of living flowering plants.

In this work, I have attempted to answer the following two specific questions:

1. *Is it possible to identify groups of co expressed genes involved in the development of the carpel and ovule, which have been conserved from the MRCA of living flowering plants?*
2. *When was the transcription factor SPATULA recruited to its function in the carpel, and what were the molecular changes that occurred to integrate SPT into this molecular pathway?*

These questions involved two different approaches, and as a consequence, I will present my PhD work in two parts:

**PART I:** in which I analyze transcriptomic data, combining our own unpublished datasets with data available from other work, to explore a new method for inter-species analysis using the R-package, WGCNA. The initial aim of this work was to identify genetic modules that control female reproductive development and are conserved between the ANA-grade angiosperm *Amborella trichopoda* and the core-eudicot model angiosperm *Arabidopsis thaliana*.

**PART II:** in which I centered the analyses on one transcription factor that is heavily involved in *A. thaliana* gynoecium development. In this work, I analyzed the coding and upstream sequences of the bHLH transcription factor *SPATULA* and its orthologues in diverse plant species. *In-vitro* and *in-vivo* experiments were performed to study the interactions and functions of *SPT* across flowering plants and beyond. These studies aimed to determine the molecular changes which were responsible for the acquisition of *SPT*'s role in the *A. thaliana* carpel, and when, during angiosperm evolution, these changes occurred.

# RESULTS

---

# **PART I: TRANSCRIPTOMIC DATA FOR CARPEL EVOLUTIONARY STUDIES**

## Aims and strategies of Part 1

Basal angiosperms of the ANA-grade provide key morphological, ecological and molecular data for the evolutionary study of angiosperms. The comparative analysis of distantly related species such as the eudicots and the ANA-grade provides enables us to infer the ancestral states of different characters in the first flowering plants.

This chapter explores the application of interspecific transcriptomic analysis, in order to answer the following question: *Is it possible to identify groups of co-expression of genes involved in the development of the carpel and ovule that have been conserved from the most recent common ancestor of flowering plants?*

As mentioned above, the genome of *Amborella trichopoda* is an important reference for carpel regulatory networks in early angiosperms, and can help us understand the evolution of the molecular mechanisms that lay behind the appearance of new structures, such as the carpel, in early flowering plants. In this work, novel transcriptomic data of female reproductive tissues from *A. trichopoda*, produced through laser microdissection and RNA-seq technology, were analyzed.



**A proof-of-concept using novel transcriptomic data from the early-diverging angiosperm *Amborella trichopoda* to identify genetic modules conserved over deep evolutionary time.**

Rivarola Sena AC<sup>1\*</sup>, Andres-Robin A<sup>1\*</sup>, Vialette AC<sup>1</sup>, Just J<sup>1</sup>, Alexandra Launay-Avon<sup>2</sup>, Dubreucq B<sup>3</sup> and Scutt CP<sup>1#</sup>

1. Laboratoire Reproduction et Développement des Plantes, Univ. Lyon, ENS de Lyon, UCB Lyon-1, CNRS, INRA, F-69342 Lyon, France
2. IPS2-Institute of Plant Sciences Paris-Saclay, Bâtiment 630, rue de Noetzlin, 91190, Gif-sur-Yvette, France.
3. UMR 1318 INRAe-AgroParisTech, Route de Saint Cyr, 78026 Versailles cedex, France

\*Contributed equally

#For correspondence: [charlie.scutt@ens-lyon.fr](mailto:charlie.scutt@ens-lyon.fr)

Keywords: *Amborella trichopoda*, *Arabidopsis thaliana*, WGCNA, gene-clustering, orthogroup, angiosperm, flower, carpel

## Abstract

Gene-expression clustering from transcriptomic data can help elucidate the gene regulatory networks (GRNs) that control developmental processes in model organisms. In parallel, transcriptomic data are becoming available from an ever-increasing number of species that occupy key phylogenetic positions, of potentially great use to evo-devo studies aimed at reconstructing the evolution of developmental GRNs. However, such large-scale datasets are often sub-optimally exploited in evo-devo studies, in part because few methods are yet available to compare gene expression dynamics between distantly-related species on a genomic scale, rather than on a simple gene-by-gene basis. Here, we use laser-capture microdissection (LCM) to contribute the first detailed transcriptomic study of female flower development in the likely sister to all other living angiosperms, *Amborella trichopoda*. We use the data obtained to address two questions: Firstly, can transcriptomic data be used to identify entire genetic modules whose expression in floral tissues has been conserved over deep evolutionary time? Secondly, to what extent can transcriptomic datasets, generated using diverse methods in independent studies, be usefully compared in plant evo-devo analyses? We use a gene-expression clustering method, followed by a custom procedure to compare the resulting genetic modules on the basis of orthology relationships at the genomic scale between *Amborella* and the molecular-genetic model angiosperm *Arabidopsis thaliana*. Using this approach, we have succeeded in identifying 13 pairs of modules that show both a highly significant intersection in orthogroup-content and congruent expression dynamics between the two species under comparison. We conclude that transcriptomic datasets resulting from independent studies can be used efficiently to identify genetic modules conserved at least since the most recent common ancestor of living angiosperms, which is believed to have lived at least 149 million years ago.

## Introduction

A major branch of evolutionary-developmental biology (evo-devo) uses inter-species comparisons within an established phylogenetic framework to reconstruct the evolution of the gene regulatory networks (GRNs) that control organismal development. In parallel, gene clustering methods have been widely used to define GRNs within individual species, and in some cases to make comparisons of networks between closely related species (Morandin et al., 2016; Muntane et al., 2017; Yu et al., 2020). However, few studies to date have used clustering methods to compare gene expression between highly

diverged species, as is often necessary in evo-devo studies that aim to answer macro-evolutionary questions.

The origin of the angiosperm flower, and of its numerous unique organ systems and processes, is one of the major unresolved questions of macro-evolutionary biology. To address this question from a molecular angle, numerous studies have compared flower-development regulators from model angiosperms such as *Arabidopsis thaliana* with their orthologs from members of the three earliest-diverging “ANA-grade” angiosperm orders: Amborellales, Nymphaeales and Austrobaileyales (reviewed by Scutt, 2018). Indeed, many of these studies have focused on *Amborella trichopoda*, the only living representative of Amborellales and probable sister to all other living flowering plants (Stevens, 2001). *Amborella* is a scrambling, dioecious shrub, endemic to the sub-tropical rainforests of New Caledonia. Flowers of female *Amborella* individuals contain spirally arranged organs that make up a perianth of typically 7-8 tepals, surrounding a gynoecium of 5(-6) unfused carpels (Endress & Igersheim, 2000b). One or two staminodes (sterile, stamen-like organs) may also be present in *Amborella* female flowers, inserted between the tepals and carpels. The undifferentiated perianth of the *Amborella* flower is a probable pleisiomorphic feature of angiosperms, while the carpels of its gynoecium also contain a number of likely pleisiomorphies, including the presence of a compitum, generated through physical contact between the stigmatic surfaces of adjacent carpels, which facilitates the exchange of pollen tubes and thus increases the efficiency of fertilization. *Amborella* carpels also show the probable pleisiomorphies of an ascidiate (vase-like) shape; ridged, multicellular protrusions on the stigmatic surface; a secretion-filled apical canal for pollen tube growth; and a single, pendent, bitegmic ovule.

The *Amborella* nuclear genome measures approximately 810 Mb (Albert et al., 2013). Its detailed analysis reveals clear traces of the At- $\epsilon$  whole genome duplication (WGD), which is believed to have preceded the radiation of living flowering plants (Yuannian Jiao et al., 2011), but provides no evidence of any subsequent WGD events along the *Amborella* lineage. The availability of flower tissue-specific transcriptomic resources in *Amborella* has until now been limited, though a detailed study of gene expression in the *Amborella* egg apparatus was recently published (Flores-Tornero et al., 2019).

Here, we contribute the first detailed transcriptomic study of *Amborella* female flower development, performed using Laser Capture Microdissection (LCM). As a pilot study for the application of gene-expression clustering methods to early flower evolution, we have applied weighted gene co-expression network analysis (WGCNA; Langfelder & Horvath, 2008) to our *Amborella* datasets, and in parallel to reasonably comparable published datasets from the molecular-genetic model angiosperm

*Arabidopsis thaliana*. We have developed a simple work-flow and computer codes to compare the results of gene-expression clustering between distantly related species on the basis of gene-orthology relationships at the genomic scale. Using these procedures, we have succeeded in identifying 13 pairs of genetic modules that show both a highly significant intersection in orthogroup-content and congruent expression dynamics between *Amborella* and *Arabidopsis* flowers, strongly suggesting their conservation since the most recent common ancestor (MRCA) of living angiosperms.

## **Materials and methods**

### Plant material

Seedlings of *Amborella trichopoda* (Bail.) were generously provided by Gildas Gâteblé and Bruno Fogliani (IAC, St Michel, New Caledonia) and grown to maturity in Lyon in a greenhouse under conditions of ~70% relative humidity, 18-30°C, and natural daylight attenuated by artificial shade-covering. Flower buds for LCM procedures were harvested from five-year-old plants during peak flowering season in the autumn. These materials were used to initiate the workflow indicated in Fig. 1.

### Laser Capture Microdissection

*Amborella* female flower buds were fixed, embedded, sectioned and subjected to LCM procedures on a PALM Micro-Beam system (Zeiss) as described by Sakai et al. (2018). Approximately 10 tissue sections were combined for each sample. RNA was extracted from these using an Arcturus Pico RNA Extraction Kit (Thermo Fisher Scientific). RNA integrity was assessed using a 2100 BioAnalyzer (Agilent Technologies) and samples showing RNA Integrity Numbers (RINs) of >7.0 were processed for library construction.

### Library construction, sequencing, read mapping and data normalization

*Amborella* libraries were constructed using a cDNA SMARTer Ultra-Low Kit (Illumina) and pooled in groups of five using a barcoding system to be processed in three runs of Hi-seq2000 (Illumina) high-throughput sequencing. Read quality was assessed using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and reads were then mapped to the *Amborella* V1 complete genome sequence (Albert et al., 2013) using HISAT2 (Sirén et al. 2014). Read-counts were performed using FeatureCounts (Liao et al., 2014) and the resulting count data normalized using DEseq2 (Love et al., 2014).

### Gene expression clustering

Gene clustering was performed in parallel using the WGCNA package (Langfelder & Horvath, 2008) on *Amborella* female flower datasets generated in the present study and on published *Arabidopsis* flower datasets (Yuling Jiao & Meyerowitz, 2010). Clustering was performed using the WGCNA automated method, and default settings including the selection of an unsigned network topology. After scrutiny of plots of Scale Free Model Topology Fit against Scale Independence, a Soft Thresholding Power of 7 was chosen for clustering procedures in both species, and genetic modules were generated by cutting dendrograms at a cut-height of 0.5. Eigengene values for these modules, generated in WGCNA, were averaged over biological replicates and converted to a green-red color scale in Microsoft Excel for use in schematic representations.

### Inter-species comparison of gene-expression modules and statistical analysis

A global comparison of gene orthology between the *Amborella* and *Arabidopsis* genomes (SI Tab 1) was generated using BIOMART in PHYTOZOME (<https://phytozome.jgi.doe.gov/pz/portal.html>). This comparison classes genes from each species into orthogroups, each of which has a unique identifier and is putatively derived from a single ancestral gene that was present in the MRCA of *Amborella* and *Arabidopsis* (i.e. in the probable MRCA of all living angiosperms). Orthogroups are classed in this analysis as “one-to-one”, “one-to-many”, “many-to-one” or “many-to-many”, depending on whether gene or genome duplication events in one or both plant lineages have, since their separation, expanded the orthogroup in question.

The *Amborella* and *Arabidopsis* modules generated using WGCNA were processed into suitable lists using Perl-Code-1 (SI Methods), and converted into their respective orthogroup identifiers using Perl-Code-2 (SI Methods), which also takes as input a mapping file in .csv format, derived from the gene-orthology relationships obtained using Phytozome. These mapping files each contained a list of *Amborella* or *Arabidopsis* (as appropriate) gene identifiers in the first column, and the corresponding orthogroup identifiers in the second. The lists of orthogroups in each module, generated using Perl-Code-2, were compared pairwise between species using Perl-Code-3 (SI Methods). Perl-Code-3 is set up such that the *Amborella* module is the second argument used in each pairwise comparison, which is significant as this argument represents the trials used to find matches in the list provided by the first (*Arabidopsis*) argument. Accordingly, in Perl-Code-3, an *Amborella* module containing N members of a given orthogroup will find N corresponding elements in an *Arabidopsis* module that contains one or more members of that same orthogroup. Output from Perl-Code-3 comprises lists of the orthogroups

present in all pairwise module-intersections between the species analyzed. This output was used as input for Perl-Code-4 (SI Methods), which returns the number of orthogroups in each list.

The statistical significance of the intersection between each pair of modules was calculated using the hypergeometric test. The number of sample successes ( $q$ ) in this test is represented by the number of orthogroup members in the intersection between the pair of modules under consideration (from the output of Perl-Code-4). Population successes ( $m$ ) are represented by the number of unique orthogroups (i.e. counting two or more identical entries as one) in the entire *Arabidopsis* transcriptome that are also present in the *Amborella* module under consideration. Population failures ( $n$ ) thus correspond to the total number of unique orthogroups in the *Arabidopsis* transcriptome, minus the number of population successes ( $m$ ). The number of trials ( $k$ ) corresponds to the total number of orthogroups (i.e. counting  $N$  identical entries as  $N$ ) in the *Amborella* module under consideration. The upper-tail, cumulative  $p$ -value was calculated in this way for each pairwise module comparison between *Amborella* and *Arabidopsis* using the R code:

```
>phyper(q-1, m, n, k, lower-tail=FALSE);
```

The *Amborella* and *Arabidopsis* orthologs occupying the intersections between highly conserved modules of interest from the two species were then compiled in lists, together with their orthogroup IDs, using Perl-Code-5 (SI Methods), and *Arabidopsis* gene annotations for one gene from each orthogroup (obtained from Yuling Jiao & Meyerowitz, 2010) added using Perl-Code-2 on appropriately formatted input files.

## Results

Laser Capture Microdissection proves an efficient method to define the floral transcriptome of the early-diverging angiosperm *Amborella trichopoda*.

We have used LCM and RNA-seq procedures to generate transcriptomic datasets from two stages of *Amborella* female flower development as defined by Buzgo et al. (2004). The stages samples correspond to (1) Stage 5-6, and (2) a stage between Stage 7 and that of the mature female flower (female anthesis), referred to here as “Late Stage 7”. At Stage 5-6, the young tepals are expanding laterally and the carpels are developing from a solid primordium into a hollow structure that will later enclose the ovule. At this stage, young carpels are clearly distinguishable from any staminodes that may also be present. Ovule initiation occurs at Stage 7, and by Late Stage 7, ovule development is

well underway, while distinct regions of the carpel corresponding to the ovary wall and carpel apex are also clearly apparent.

We sampled two tissues from Stage 5-6, corresponding to young tepals and carpels (Figs 2A and B). From Late Stage 7, we sampled three tissues corresponding to the developing ovule, ovary wall and carpel apex (Figs 2C and D). All samples were obtained in triplicate, making a total of 15 samples for library construction and paired-end read sequencing. Between 89% and 96% of reads generated from the 15 samples could be mapped to the *Amborella* genome sequence, giving an average of  $2.62 \times 10^7$  mapped reads per sample. The datasets obtained (SI Tab 2, normalized counts) constitute the first detailed transcriptomics resource from flower tissues of an ANA-grade angiosperm.

#### Orthogroup-based comparisons identify 24 highly significant intersections between genetic modules from *Amborella* and *Arabidopsis*.

We performed clustering analyses in WGCNA on the *Amborella* female flower datasets generated in this study and on RNA-seq datasets (Yuling Jiao & Meyerowitz, 2010) from two stages of *Arabidopsis* flower development, as defined by Smyth et al. (1990). These *Arabidopsis* datasets represent the transcriptome (i.e. the population of mRNAs attached to ribosomes) from three overlapping regions of flower buds in which are expressed, respectively, the A, B- and C-class MADS-box floral homeotic regulators: *APETALA1* (*AP1*, expressed in whorls 1-2), *APETALA3* (*AP3*, expressed in whorls 2-3) and *AGAMOUS* (*AG*, expressed in whorls 3-4).

Our clustering analyses, using identical parameters for each species, yielded totals of 14 and 13 modules in *Amborella* and *Arabidopsis* flowers, respectively, not including a “grey” module of unassigned genes in each case (SI Tab 3). The remaining modules are labelled in WGCNA using a fixed series of RGB colour-names (see key in SI Tab 4) in descending order of gene-content (so modules bearing the same colour-name may be of similar size, but are not necessarily closely related between species). Module eigengenes (Langfelder & Horvath, 2007), calculated in WGCNA, were averaged between biological replicates (SI Tab 4) and plotted on a red/green color-scale for the schematic diagrams in Fig 3. Each eigengene is the theoretical gene which typifies its module, the value for each tissue/sample being the first principal component of the variance in gene expression. High positive eigengene values thus represent high expression, while high negative values represent low expression. As we used an unsigned clustering procedure, modules may contain some genes that show inverted expression dynamics compared to their module eigengene (as well as many others whose expression dynamics correlate positively with those of the module eigengene). Such unsigned

networks may be particularly useful to identify negative as well as positive regulators in GRNs of interest.

We used a custom procedure to compare the contents of the genetic modules identified in WGCNA (SI Tab 3) by first converting their *Amborella* or *Arabidopsis* gene-identifiers to orthogroup-identifiers generated in PHYTOZOME (SI Tab 1). We then compared all modules pairwise between species and analyzed their intersections (SI Tab 5) using the hypergeometric test to produce a list of 182 upper-tail p-values representing the probability of each positive (over-represented) intersection arising by chance (SI Tab 6). In Fig 3, we illustrate relationships between modules with a highly significant p-value of  $<0.01$ , which includes a total of 24 pairwise intersections (dotted lines). At this level of statistical significance, a few modules appear to be not closely related to any module in the other species under comparison, including the closely similar red and tan modules of *Arabidopsis*. These two modules both show high eigengene expression in the *AP3* (B-function)-expression domain, which includes the developing stamens. Stamen or staminode tissues were not sampled from *Amborella* in our study, which may explain the lack of a closely similar *Amborella* module to these two *Arabidopsis* modules.

Several other cases exist in our study in which only one of two very similar modules in one species shows a highly significant ( $p<0.01$ ) intersection with a module from the other. Examples of this phenomenon include the *Amborella* tan and green modules, both of which are characterized by high eigengene expression in the ovary wall. Of these two modules, only Atr\_tan appears closely related ( $p=0.0028$ , SI Tab 6) in orthogroup-content to Ath\_magenta, which shows high eigengene expression in the central zone of the *Arabidopsis* flower bud at Stage 6-7 and is thus coherent with the expression of the Atr\_tan and Atr\_green eigengenes. However, the full list of module-comparisons (SI Tab 6) indicates that Atr\_green is also related to Ath\_magenta with the moderately low p-value of  $p=0.092$ . Indeed, it appears that, in most cases of multiple modules in one species that show very similar expression dynamics, the main distinction between these modules resides in subtle differences between biological replicates, rather than more pronounced differences between the tissues or stages analyzed. In the case of Atr\_tan and Atr\_green, for example, the former shows a particularly high eigengene value for the first biological replicate of the ovary wall tissue, while the latter shows a similarly high eigengene value for the second ovary wall replicate (SI Tab 4).

Several of the modules that show highly significant ( $p<0.01$ ) positive intersections in our study do so with more than one module from the other species under comparison. This is the case, for example, for the Atr\_turquoise module, which shares a close relationship with five different *Arabidopsis* modules (Fig. 2, SI Tab 6).

Thirteen highly significant module intersections show congruent expression dynamics, suggesting their conservation since the MRCA of living angiosperms.

It is not trivial to compare gene-expression dynamics between such widely diverged angiosperm species as *Amborella* and *Arabidopsis*. Such comparisons are affected by problems of defining organ and tissue homology (Roux et al., 2015). For example, *Amborella* contains an undifferentiated perianth of tepals, while *Arabidopsis* contains a differentiated perianth containing both a calyx of sepals and a corolla of petals. Furthermore, the petals of core eudicots such as *Arabidopsis* may be “andropetals”, derived from sterilized former stamens, whereas petaloid perianth organs in more basally-diverging angiosperm taxa such as *Amborella* are usually interpreted as “bracteopetals”, derived from leaf-like organs (De Craene & Brockington, 2013). Furthermore, it is not clear that all tissue-types in the complex syncarpous gynoecium of *Arabidopsis* possess a counterpart in the much simpler, unfused carpels of *Amborella* (Endress & Igersheim, 2000a). Flower development between *Amborella* and *Arabidopsis* also exhibits heterochrony, making it very difficult to designate equivalent stages of flower bud development between these two species. Most notably, *Amborella* floral organ primordia are produced along a continuous spiral, whereas the primordia of each whorl of the *Arabidopsis* flower are initiated simultaneously.

Despite these various differences, there are clear indications that several of the highly significant module-intersections identified in the current work derive from the conservation of gene expression. We accordingly observe that, in each module identified in the present study from either *Amborella* or *Arabidopsis*, one or two tissues/stages show relatively high eigengene values of  $>0.25$ , and this information can be used to broadly characterize the expression dynamics of the module. Using this information, 13 of the 24 highly significant intersections identified through statistical testing appear to show congruent expression patterns between *Amborella* and *Arabidopsis*, and these link together all but six of the modules (three from each species) for which highly significant intersections ( $p < 0.01$ ) were found (Tab 1). These 13 intersections contain a total of 1809 orthogroups, which appear to have conserved important elements of their expression patterns, as typified by their module eigengenes, since the MRCA of living angiosperms. Full data on these 13 sets of conserved orthologs are presented in SI Tab. 7, including annotation information for the first-listed *Arabidopsis* gene from each orthogroup.

Orthologs from conserved modules show several different types of relationship with each other and with their module eigengenes.

We cannot summarize here the expression characteristics of all the genes of the 1809 orthogroups from intersecting modules listed in SI Tab 7, other than through their module eigengenes, which are shown in Fig 3. However, it is important to note that several different types of relationship can be found between orthologs from intersecting modules, and between those orthologs and their respective module eigengenes. In Fig. 4, we illustrate the main types of relationship observed using three examples from each of two module intersections: Atr\_blue vs. Ath\_blue, both of whose eigengenes are highly expressed in female reproductive tissues (Fig. 4A-C); and Atr\_turquoise vs. Ath\_black, both of whose eigengenes are highly expressed in perianth tissues (Fig 4D-F).

Orthologous genes encoding a DEAD-box RNA helicase (Fig 4A) show high expression in young carpel and ovule tissues in *Amborella*, and in the *AG*-expressing central zone of the *Arabidopsis* flower at both of the developmental stages analyzed. These patterns are therefore in close agreement with the Atr\_blue and Ath\_blue module eigengenes (Fig. 3). Similarly, orthologous WUSCHEL-like homeobox genes (Fig. 4B) show highly congruent expression patterns in the two species under study, and with their respective module eigengenes. However, orthologous GATA-family transcription factors that group within the Atr\_blue and Ath\_blue modules (Fig. 4C) show an apparently inverse set of expression dynamics to those of the examples given in Figs 4A and B. This orthogroup clearly fits within the Atr\_blue and Ath\_blue modules as a negatively-regulated component, and was presumably included by WGCNA in the same intersecting modules as the previous two examples (4A and B) due to the unsigned nature of the clustering procedure.

The Atr\_turquoise and Ath\_black modules both show high eigengene values in perianth tissues, including both of the developmental stages analyzed in *Arabidopsis* (Fig. 3). Similarly to the examples shown in Figs 4A and B, orthologs in these two modules, encoding a B7-domain transcription factor (Fig. 4D) and a circadian-clock regulated transcript (Fig. 4E), show expression patterns very close to each other, and to their respective module eigengenes. By contrast, a third example of an orthogroup shared between the Atr\_turquoise and Ath\_black modules, encoding a ENTH-domain protein (Fig. 4F), reveals a novel set of expression relationships in which the *Amborella* ortholog shows a similar expression pattern to its Atr\_turquoise module eigengene, but the *Arabidopsis* ortholog shows a broadly opposite expression pattern to those of both the *Amborella* ortholog and the Ath\_black module eigengene. Thus, whereas in the first five examples described above (Figs 4A-E), the orthologs of each orthogroup examined show similar expression patterns in their respective species, in this last example

the two orthologous genes examined show broadly opposite patterns, the *Amborella* gene being highly expressed only in the perianth, while the *Arabidopsis* gene is highly expressed in all domains of the floral bud except the perianth. As for the example shown in Fig 4C, this *Arabidopsis* ortholog was presumably included in the same module as genes with an opposite expression profile due to the unsigned nature of the clustering procedure. In this case however, and in contrast to that shown in Fig 4C, it appears there has been a change in the expression pattern of one or other of the two orthologs under consideration since the MRCA of living angiosperms.

## Discussion

### What is modularity and why is it so important?

Modularity is a key concept of evolutionary-developmental biology (evo-devo). However, the definition given of a “module” may not always correspond perfectly between evolutionary morphologists and developmental biologists (Mabee, 2006). Evolutionary morphologists tend to think of modules as elements of an organism in which internal connectivity is greater than external connectivity. Such modules, which might for example correspond to a particular organ system or tissue, may be able to undergo evolutionary change without substantial change to the rest of the organism. Developmental biologists, by contrast, think increasingly in terms of Gene Regulatory Networks in which a module might represent a particular set of genes that can be regulated together to produce a given developmental effect, perhaps analogous to a computer subroutine that can be invoked to perform a precise task from within any part of the overall program.

Clearly, profound links must exist between morphological and molecular modules, and it is perhaps surprising that relatively few clear examples can be cited to link evolutionary change to modular processes at both the genetic and morphological levels. One possible reason for this, particularly in plant biology, is that large-scale gene expression datasets from a wide phylogenetic range of species have only become available over the last few years. Indeed, even though many plant transcriptomes have now been assembled (Leebens-Mack et al., 2019), the availability of large-scale gene-expression datasets that cover different organs, tissues and stages continues to lag considerably behind in all but a few model species. It is likely, however, with techniques such as the LCM method employed in the present work, and also the very promising single-cell RNA-sequencing technologies (Frank & Scanlon, 2015; Ryu et al., 2019) that detailed large-scale expression data will become increasingly available from a much wider range of plant species. It will consequently become increasingly important to devise methods to compare global expression data between species, including those

separated by great evolutionary distances, which is one of the reasons we decided to perform the present pilot study.

This study shows that conserved gene-expression modules can be identified over large phylogenetic distances.

In this work, we show that it is possible to meaningfully compare the results of independently-performed gene-expression clustering analyses between plant species that are believed to have diverged at least 149 MYA (Barba-Montoya et al., 2018). By comparing gene-expression modules on the basis of gene orthology, we have been able to identify 13 highly significant intersections between modules whose eigengenes show congruent expression patterns between the species under comparison. Accordingly, the members of the 1809 orthogroups contained in these 13 intersections appear to have conserved their expression profiles in both the *Amborella* and *Arabidopsis* lineages since the MRCA of living angiosperms. These results represent an initial proof-of-concept, answering the first of the questions we asked at the outset of this work concerning the feasibility of comparing gene-expression clustering data over a vast phylogenetic distance.

Analyses of genome structure indicate the *Amborella* genome to have undergone no whole genome duplication since the At- $\epsilon$  event, believed to have occurred before the radiation of living angiosperms (Yuannian Jiao et al., 2011). In the *Arabidopsis* genome, by contrast, there is evidence of one genome triplication (At- $\gamma$ ) and two genome duplications (At- $\alpha$  and At- $\beta$ ) that have taken place since the At- $\epsilon$  event (Yuannian Jiao et al., 2011). These multiple, large-scale duplication events in the *Arabidopsis* lineage inevitably render gene-orthology relationships with *Amborella* more complex than with, for example, a recently diverged member of Brassicaceae. Nonetheless, the use of a genome-wide analysis of orthology relationships has permitted, in the present study, a meaningful comparison of global transcriptome data between species whose lineages diverged at the dawn of the angiosperm clade.

The majority of the conserved orthogroup members identified in the present work show congruent expression patterns between the two species under comparison. However, inversions of expression dynamics appear to have occurred between orthologs in a few cases (e.g. Fig. 4F), which may point to a switch in an upstream regulatory component in one or other of the plant lineages concerned. These examples of converse expression patterns between orthologs may merit further attention, as some of them might be indicative of important evolutionary changes to upstream developmental regulators.

### How useful are datasets derived from independent transcriptomic analyses?

Clearly, it is very useful for practitioners of evo-devo to be able to use published datasets from other workers, in addition to their own novel data. Accordingly, the second question we asked at the outset of this study concerned the usefulness of comparing datasets from different sources, which by definition were not generated as part of a unified experimental design. The answer to this question, from the results presented here, seems to be “both yes and no”. We used partially similar flower RNA-seq datasets from *Amborella* (this study) and *Arabidopsis* (Yuling Jiao & Meyerowitz, 2010), which differed in the exact tissues and stages sampled and in the experimental approach used (i.e. LCM-dissection of discrete tissues/organs in *Amborella*, compared to sampling of overlapping tissue zones in *Arabidopsis*). As stated above, this approach was nevertheless successful in identifying 13 cases of apparent large-scale conservation between genetic modules. However, a further 11 pairs of highly significant module intersections were identified for which the corresponding module eigengenes did not show congruent expression patterns between the species analyzed. In addition, many of the module-intersections identified in our study showed deviations from a 1:1 relationship between species. It is likely that two major factors contributed to the above two effects: (1) we used datasets produced in independent studies, in which the tissues and developmental stages sampled corresponded only partially between species. Secondly, real differences between the species under study, resulting from evolutionary processes since the separation of their lineages, may also have contributed to changes in gene expression patterns and/or to a departure from a 1:1 correspondence between genetic modules.

It seems clear that the best way to identify, with reasonable confidence, real, intrinsic differences between gene-expression clustering data from different species will be to minimize variation between datasets that are due to different sampling strategies and/or experimental methods. This caveat means that, to reach solid conclusions about evolutionary change as well as evolutionary conservation, it will be very important to compare datasets that are as closely matched as possible. This stringent requirement could, however, still permit the use of data obtained from other studies. In particular, the general applicability of detailed gene expression data might be facilitated by the adoption by the scientific community of data standards that maximize equivalence and comparability between datasets.

## Author contributions

ACRS performed most of the bioinformatics work and co-wrote the manuscript, AAR performed the LCM work, ACV collaborated with ACRS in read-mapping and figure preparation, JJ provided advice and supervision in bioinformatics work, ALA performed the library construction and high-throughput sequencing, BD supervised the LCM work, and CPS coordinated the project and co-wrote the manuscript.

## Acknowledgements

This work was supported by ANR grant “ORANGe” (2013-18) and through core funding from CNRS, INRAe, ENS-Lyon and Université Claude-Bernard Lyon-1. ACRS was supported by a BECAL studentship from the Government of Paraguay. We thank Gildas Gâteblé and Bruno Fogliani (IAC, St Michel, New Caledonia) for generously supplying plant material.

## References

- Albert, V. A., Barbazuk, W. B., dePamphilis, C. W., Der, J. P., Leebens-Mack, J., Ma, H., Palmer, J. D., Rounsley, S., Sankoff, D., Schuster, S. C., Soltis, D. E., Soltis, P. S., Wessler, S. R., Wing, R. A., Albert, V. A., Ammiraju, J. S. S., Barbazuk, W. B., Chamala, S., Chanderbali, A. S., ... Tomsho, L. (2013). The Amborella Genome and the Evolution of Flowering Plants. *Science*, 342(6165), 1467-+. <https://doi.org/10.1126/science.1241089>
- Barba-Montoya, J., dos Reis, M., Schneider, H., Donoghue, P. C. J., & Yang, Z. (2018). Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *New Phytologist*, 218(2), 819–834. <https://doi.org/10.1111/nph.15011>
- Buzgo, M., Soltis, P. S., & Soltis, D. E. (2004). Floral developmental morphology of *Amborella trichopoda* (Amborellaceae). *International Journal of Plant Sciences*, 165(6), 925–947. <https://doi.org/10.1086/424024>
- De Craene, L. P. R., & Brockington, S. F. (2013). Origin and evolution of petals in angiosperms. *Plant Ecology and Evolution*, 146(1), 5–25. <https://doi.org/10.5091/plecevo.2013.738>
- Endress, P. K., & Igersheim, A. (2000a). Gynoecium structure and evolution in basal angiosperms. *International Journal of Plant Sciences*, 161(6), S211–S223. <https://doi.org/10.1086/317572>

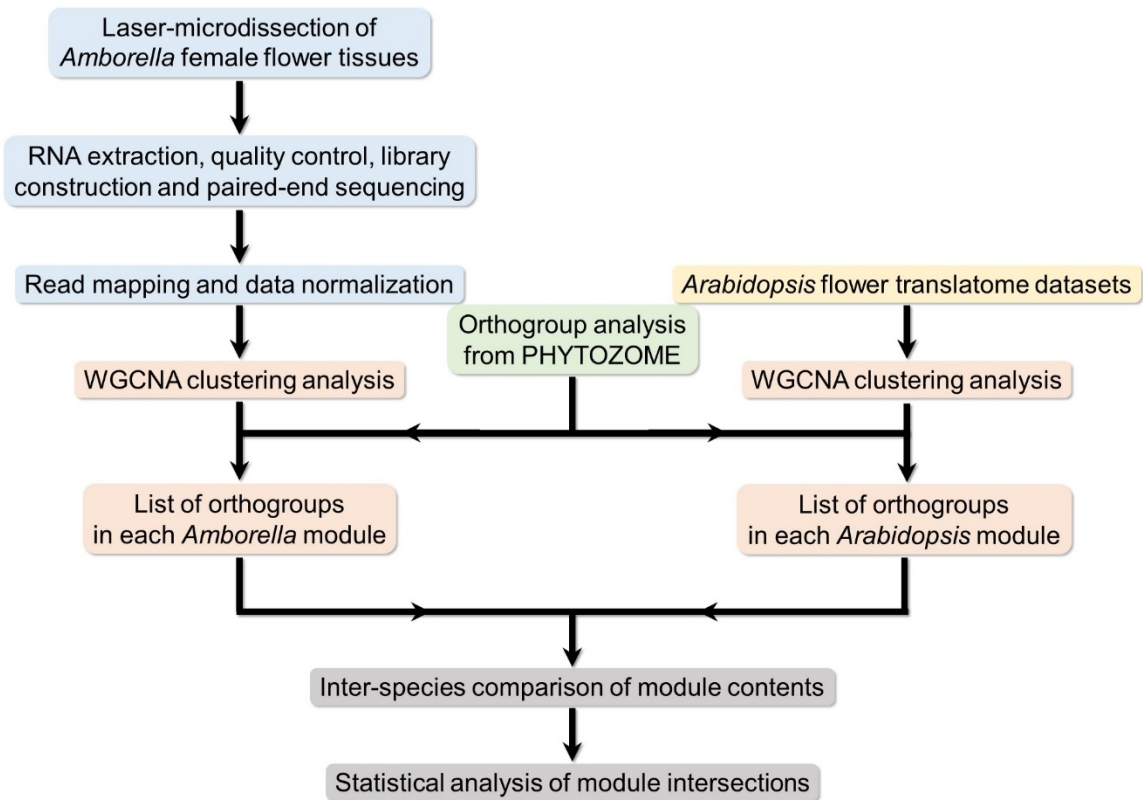
- Endress, P. K., & Igersheim, A. (2000b). Reproductive structures of the basal angiosperm *Amborella trichopoda* (Amborellaceae). *International Journal of Plant Sciences*, *161*(6), S237–S248. <https://doi.org/10.1086/317571>
- Flores-Tornero, M., Proost, S., Mutwil, M., Scutt, C. P., Dresselhaus, T., & Sprunck, S. (2019). Transcriptomics of manually isolated *Amborella trichopoda* egg apparatus cells. *Plant Reproduction*, *32*(1), 15–27. <https://doi.org/10.1007/s00497-019-00361-0>
- Frank, M. H., & Scanlon, M. J. (2015). Cell-specific transcriptomic analyses of three-dimensional shoot development in the moss *Physcomitrella patens*. *Plant Journal*, *83*(4), 743–751. <https://doi.org/10.1111/tpj.12928>
- Jiao, Yuannian, Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., & dePamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, *473*(7345), 97–U113. <https://doi.org/10.1038/nature09916>
- Jiao, Yuling, & Meyerowitz, E. M. (2010). Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Molecular Systems Biology*, *6*, 419. <https://doi.org/10.1038/msb.2010.76>
- Langfelder, P., & Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. *Bmc Systems Biology*, *1*, 54. <https://doi.org/10.1186/1752-0509-1-54>
- Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *Bmc Bioinformatics*, *9*, 559. <https://doi.org/10.1186/1471-2105-9-559>
- Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., Porsch, M., Quint, M., Rensing, S. A., Soltis, D. E., Soltis, P. S., Stevenson, D. W., Ullrich, K. K., Wickett, N. J., DeGironimo, L., ... Wong, G. K.-S. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, *574*(7780), 679–+. <https://doi.org/10.1038/s41586-019-1693-2>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>

- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Mabee, P. M. (2006). Integrating evolution and development: The need for bioinformatics in evo-devo. *Bioscience*, *56*(4), 301–309. [https://doi.org/10.1641/0006-3568\(2006\)56\[301:IEADTN\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2006)56[301:IEADTN]2.0.CO;2)
- Morandin, C., Tin, M. M. Y., Abril, S., Gomez, C., Pontieri, L., Schiott, M., Sundstrom, L., Tsuji, K., Pedersen, J. S., Helanterä, H., & Mikheyev, A. S. (2016). Comparative transcriptomics reveals the conserved building blocks involved in parallel evolution of diverse phenotypic traits in ants. *Genome Biology*, *17*, 43. <https://doi.org/10.1186/s13059-016-0902-7>
- Muntane, G., Santpere, G., Verendeev, A., Seeley, W. W., Jacobs, B., Hopkins, W. D., Navarro, A., & Sherwood, C. C. (2017). Interhemispheric gene expression differences in the cerebral cortex of humans and macaque monkeys. *Brain Structure & Function*, *222*(7), 3241–3254. <https://doi.org/10.1007/s00429-017-1401-7>
- Roux, J., Rosikiewicz, M., & Robinson-Rechavi, M. (2015). What to compare and how: Comparative transcriptomics for Evo-Devo. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, *324*(4), 372–382. <https://doi.org/10.1002/jez.b.22618>
- Ryu, K. H., Huang, L., Kang, H. M., & Schiefelbein, J. (2019). Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. *Plant Physiology*, *179*(4), 1444–1456. <https://doi.org/10.1104/pp.18.01482>
- Sakai, K., Taconnat, L., Borrega, N., Yansouni, J., Brunaud, V., Paysant-Le Roux, C., Delannoy, E., Magniette, M.-L. M., Lepiniec, L., Faure, J. D., Balzergue, S., & Dubreucq, B. (2018). Combining laser-assisted microdissection (LAM) and RNA-seq allows to perform a comprehensive transcriptomic analysis of epidermal cells of Arabidopsis embryo. *Plant Methods*, *14*, 10. <https://doi.org/10.1186/s13007-018-0275-x>
- Scutt, C. P. (2018). The origin of angiosperms. *Evolutionary Developmental Biology: A Reference Guide*, 1–20.
- Sirén, J., Välimäki, N., & Mäkinen, V. (2014). Indexing graphs for path queries with applications in genome research. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *11*(2), 375–388. <https://doi.org/10.1109/TCBB.2013.2297101>

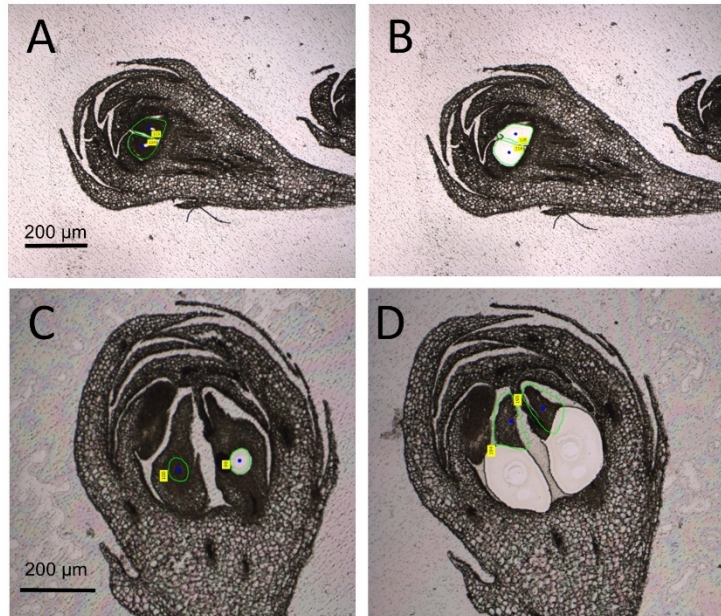
- SMYTH, D., BOWMAN, J., & MEYEROWITZ, E. (1990). EARLY FLOWER DEVELOPMENT IN ARABIDOPSIS. *Plant Cell*, 2(8), 755–767.
- Stevens, P. F. (2001). Angiosperm Phylgeny Website. Version 14, July 2017. [Http://Www.Mobot.Org/MOBOT](http://www.Mobot.Org/MOBOT).
- Yu, Y., Hu, H., Doust, A. N., & Kellogg, E. A. (2020). Divergent gene expression networks underlie morphological diversity of abscission zones in grasses. *New Phytologist*, 225(4), 1799–1815. <https://doi.org/10.1111/nph.16087>

Table 1. Summary of highly significant positive module intersections (p< 0.01) between WGCNA clustering analyses of *Amborella* and *Arabidopsis*, ranked by p-value.

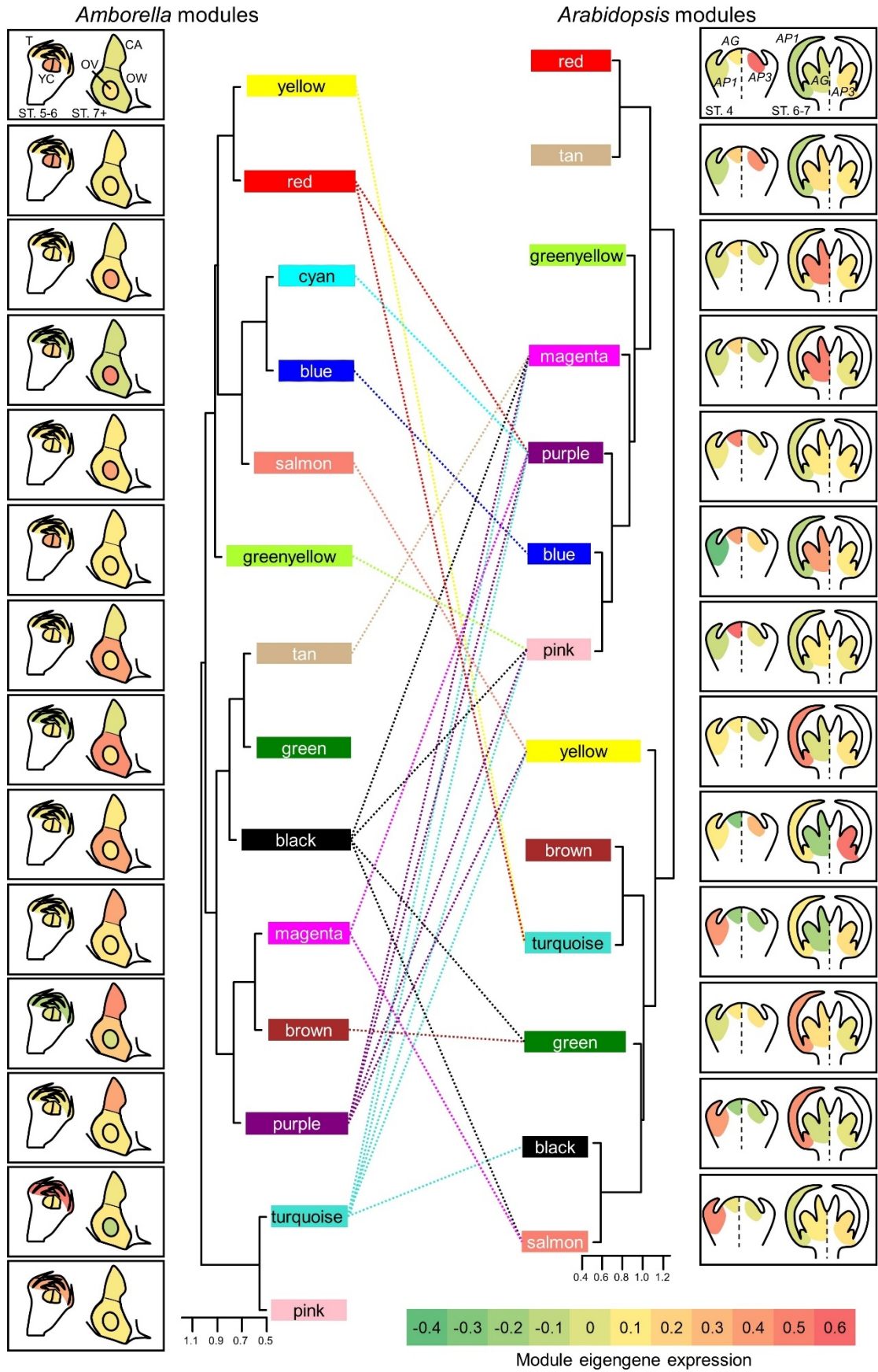
<i>Arabidopsis</i> module	<i>Amborella</i> module	No. of common orthologs	p-value	<i>Arabidopsis</i> module eigengene expression >0.25	<i>Amborella</i> module eigengene expression >0.25	Eigengene expression congruent between species?
Ath_turquoise	Atr_yellow	365	2.73E-12	AP1 St. 4	Carpel St. 5-6	N
Ath_black	Atr_turquoise	245	6.32E-08	AP1 St. 4, AP1 St. 6	Tepal St. 5-6	Y
Ath_pink	Atr_turquoise	163	6.22E-07	AGs4	Tepal St. 5-6	N
Ath_magenta	Atr_turquoise	124	3.70E-06	AGs6	Tepal St. 5-6	N
Ath_purple	Atr_red	23	1.45E-05	AGs4	Carpel St. 5-6	Y
Ath_yellow	Atr_turquoise	223	0.000102605	AP1s6	Tepal St. 5-6	Y
Ath_blue	Atr_blue	1310	0.00012267	AGs4, AGs6	Ovule St. L7	Y
Ath_purple	Atr_magenta	11	0.000266973	AGs4	Ovary St. L7	Y
Ath_magenta	Atr_black	11	0.000419318	AGs6	Ovary St. L7	Y
Ath_purple	Atr_turquoise	87	0.000534048	AGs4	Tepal St. 5-6	N
Ath_pink	Atr_greenyellow	10	0.000622646	AGs4	Carpel St. 5-6	Y
Ath_salmon	Atr_black	7	0.000622887	AP1s4	Ovary St. L7	N
Ath_purple	Atr_purple	12	0.000670362	AGs4	Ovary St. L7	Y
Ath_magenta	Atr_purple	14	0.001492205	AGs6	Ovary St. L7	Y
Ath_turquoise	Atr_red	202	0.001722988	AP1s4	Carpel St. 5-6	N
Ath_pink	Atr_black	12	0.001746861	AGs4	Ovary St. L7	Y
Ath_yellow	Atr_salmon	6	0.001839423	AP1s6	Ovule St. L7	N
Ath_green	Atr_black	18	0.002593529	AP1s6	Ovary St. L7	N
Ath_magenta	Atr_tan	6	0.002927334	AGs6	Ovary St. L7	Y
Ath_salmon	Atr_magenta	7	0.003261332	AP1s4	Ovary St. L7	N
Ath_yellow	Atr_purple	21	0.0086795	AP1s6	Ovary St. L7	N
Ath_purple	Atr_cyan	2	0.00931224	AGs4	Ovule St. L7	Y
Ath_pink	Atr_purple	15	0.009741206	AGs4	Ovary St. L7	Y
Ath_green	Atr_brown	135	0.009886491	AP1s6	Carpel apex St. L7	N



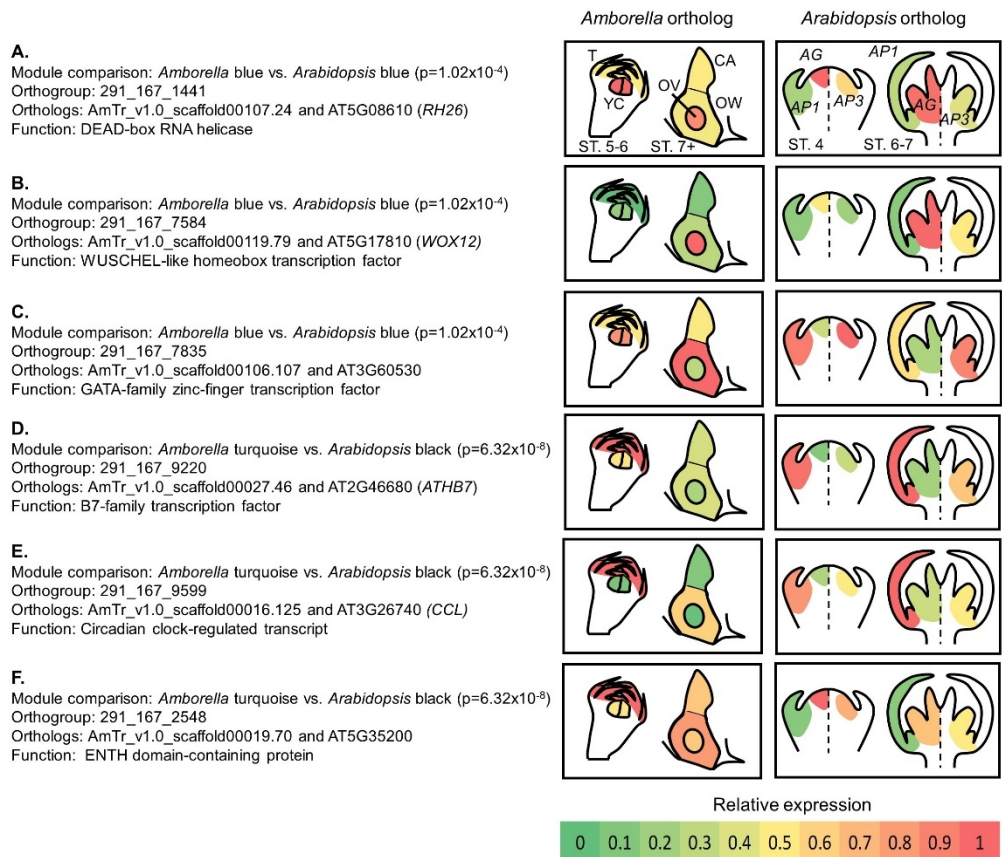
**Figure 1.** Workflow summary.



**Figure 2.** Laser Capture Microdissection of *Amborella* female flower buds showing: (A and B) microdissection of carpels at Stage 5-6 and (C and D) the sequential dissection of ovule, ovary and stigmatic tissues at Late Stage 7.



**Figure 3.** Parallel dendrograms of *Amborella* and *Arabidopsis* gene expression modules identified in WGCNA analyses. Module-eigengene expression is summarized schematically on a red/green colour-scale from full data given in SI Tab. 4. Dotted lines represent positively correlated intersections between modules at  $p < 0.01$ . ST. = stage, YC = young carpel, CA = carpel apex, OV = ovule, OW = ovary wall, T = tepal, AG, AP1 and AP3 = expression domains of *AGAMOUS* APETALA1 and APETALA3, respectively.



**Figure 4.** Examples of pairs of orthologs from highly significant module intersections between: (A-C) the *Amborella* blue and *Arabidopsis* blue modules, and (D-F) the *Amborella* turquoise and *Arabidopsis* black modules. Normalized gene expression levels, averaged across replicates, are shown schematically on a red-green colour scale, relative to the highest average value in each analysis. Full normalized expression data are given in SI Tab. 2. ST. = stage, YC = young carpel, CA = carpel apex, OV = ovule, OW = ovary wall, T = tepal, AG, AP1 and AP3 = expression domains of *AGAMOUS*, *APETALA1* and *APETALA3*, respectively.

## **Supplementary information**

### **SI Tables**

[SI Tab 1. Table of gene-orthology between Amborella and Arabidopsis](#)

[SI Tab 2. Normalized counts from RNA-seq analyses of laser-microdissected female Amborella flower tissues](#)

[SI Tab 3. Contents of gene-expression modules in Amborella and Arabidopsis following WGCNA analyses](#)

[SI Tab 5. Lists of orthogroups in intersections between modules in pairwise comparisons \(Arabidopsis vs. Amborella\)](#)

[SI Tab 6. Hypergeometric test of significance of between-species module intersections.](#)

[SI Tab 7. Module intersections Ath vs. Atr](#)

### **SI Methods**

[SI Methods 1. PERL\\_CODES](#)



**PART II: STUDY OF A TRANSCRIPTION  
FACTOR INVOLVED IN CARPEL DEVELOPMENT**

## Aims and strategies of this part

The following manuscript aims to reconstruct the evolutionary history of the gene *SPATULA* (*SPT*), which encodes a bHLH transcription factor that plays important roles in gynoecium development and other developmental processes in the molecular-genetic model angiosperm *Arabidopsis thaliana*. In the *Arabidopsis* gynoecium, *SPT* is known for its role in the development of the replum, which develops from the carpel margin meristem (CMM), and the style and stigma which develop at the apex of the gynoecium.

At the outset of this doctoral work, *SPT* was one of the best-characterized regulators of gynoecium development in *Arabidopsis*. It was known to have undergone a structural change before the origin of flowering plants, which may have been a prerequisite for the evolution of its role in the gynoecium. Coding sequences of *SPT* orthologs from asterids such as tomato were known to be able to rescue *spt* mutants in transgenic *Arabidopsis*, and *SPT* orthologs were known to be upregulated in gynoecium tissues of widely diverged angiosperms. All the above lines of evidence suggested that *SPT* was recruited very early to carpel development and that this factor accordingly formed part of the basic Gene Regulatory Network (GRN) of the carpel over 145 MYA in the first angiosperms.

We set out to test the above hypothesis using a large range of studies performed on living species chosen for their informative phylogenetic positions, including several basal angiosperms and the eudicot model *Petunia*, which is extensively studied in our research group. To study the pre-angiosperm evolution of the *SPT* lineage, we also incorporated into our investigation non-flowering vascular plants including gymnosperms and lycophytes.

The above studies have led to a manuscript which is a “work-in-progress” that can hopefully be submitted for publication on the completion of a few more experiments. Its main provisional conclusion is that *SPT* does not form part of the basic GRN of the carpel in all angiosperms, but was recruited to the development of CMM tissues much later, in a common ancestor of Brassicaceae. In Brassicaceae, *SPT* appears to contribute to a very specific type of tissue organization, involving the growth from the CMM of a false septum which divides the syncarpous gynoecium into two chambers. However, we have also obtained data relating to an apparent *spt* double-mutant phenotype associated specifically with the stigma in *Petunia* (asterids). The exact list of mutant genes that contribute to this phenotype, however, remains to be clarified, and this awaited clarification could lead to a reappraisal of our current view that *SPT* was recruited to the development of CMM tissues in an ancestor of Brassicaceae. The alternative possibility is that *SPT* may have acquired a role the formation of tissues

at the carpel apex or margins much earlier in the evolution of angiosperms, but this role became more pronounced and specific during the evolution of the novel and atypical gynoecium structure in Brassicaceae.

We show in our manuscript that *SPT* orthologs are very highly expressed in the abaxial domain of the carpel, and in the placenta and ovules, in a very wide spectrum of angiosperms, including some of the most angiosperm basal lineages. However, no mutant phenotypes are known in model angiosperms to be associated with these peaks of *SPT* expression, which may indicate the present of a complex set of genetic redundancies involving genes from outside the *SPT* gene-clade.

In addition to the above questions of *SPT* function and evolution, in which further clarification would be useful, our manuscript provides some clearer insights in a number of areas. These more successfully answered questions include: (1) the phylogenetic origin of the *SPT* lineage in early land plants, (2) evolution of the domain structure of *SPT* proteins in land plants, (3) conservation of biochemical properties and downstream pathways associated with widely-diverged *SPT* orthologs, (4) evolutionary changes to the *SPT*-promoter which may have led to its current role in Brassicaceae-specific gynoecium morphology and (5), the origin of lineages labelled as “*ALCATRAZ*” (*ALC*) in Solanaceae and Brassicaceae.

We envisage submitting a revised version of this manuscript after the completion of our ongoing analyses of *spt* mutants of *Petunia*, which we expect to clear up some of the remaining important questions left partially unanswered in the current manuscript version.



**Evolution of the transcription factor *SPATULA* and its role in *Arabidopsis* gynoecium development.**

Ana C. Rivarola-Sena<sup>1</sup>, Aurélie C. Vialette<sup>1</sup>, Amélie Andres-Robin<sup>1</sup>, Pierre Chambrier<sup>1</sup>, Loïc Bideau<sup>1</sup>, Patrice Morel<sup>1</sup>, Michiel Vandebussche<sup>1</sup>, Roberto Solano<sup>2</sup>, Jose Manuel Franco-Zorrilla<sup>2</sup> and Charlie P. Scutt<sup>1\*</sup>

1. Laboratoire Reproduction et Développement des Plantes, Univ. Lyon, ENS de Lyon, UCB Lyon-1, CNRS, INRAe, F-69342 Lyon, France
2. Centro Nacional de Biotecnología-Consejo Superior de Investigaciones Científicas, Campus Universidad Autónoma, 28049 Madrid, Spain

\*For correspondence: [charlie.scutt@ens-lyon.fr](mailto:charlie.scutt@ens-lyon.fr),

Tel +33 4 72 72 86 03, Fax +33 4 72 72 86 00

Keywords: *SPATULA*, *ALCATRAZ*, *PHYTOCHROME INTERACTING FACTOR5*, *PHYTOCHROME INTERACTING FACTOR7*, basic Helix-Loop-Helix, carpel, gynoecium, carpel margin meristem, flower, angiosperm, evolution, *Arabidopsis*, *Amborella*, *Nymphaea*, *Petunia*, *Picea*, *Selaginella*

## Abstract

*SPATULA (SPT)* encodes a bHLH transcription factor in *Arabidopsis thaliana* that plays important roles in the development of the style, stigma and replum tissues of the gynoecium. Though the regulation, expression pattern, interactions and functions of *SPT* have been well-described in *Arabidopsis*, comparable data are largely lacking for *SPT* orthologs in other angiosperm groups. The main objective of this work is to provide evidence for the timing of *SPT*'s recruitment to the development of apical and marginal gynoecium tissues, and to identify molecular changes which might have been responsible for this novel role. For this, we have chosen species that occupy key phylogenetic positions to reconstruct the evolution of the *SPT* lineage from early vascular plants to present-day *Arabidopsis*. We situate an early structural change in the *SPT* protein, which may have been a prerequisite for its current developmental roles, to a common ancestor of euphyllophytes (ferns+seed plants). We also demonstrate widespread conservation in the biochemical capacities of angiosperm *SPT* proteins, and in the identities of the target-genes they regulate. We report expression differences between *SPT* orthologs in diverse angiosperm taxa, and suggest elements of *SPT*-promoter structures which may be linked to these. We conclude the *Arabidopsis SPT* lineage to have acquired its current role in the style and septum in a common ancestor of Brassicaceae, mainly through changes to its expression pattern, rather than to its coding sequence or to the identities of its direct transcriptional targets. We furthermore attempt to clarify questions of gene orthology between *SPT*-like sequences from diverse core-eudicot families.

## 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 undergo cell division to generate the placenta, ovules and vertical septum or adaxial replum (Ferrandiz et al., 2010). The carpel margins then later differentiate to generate a dehiscence zone and separation layer, both of which contribute to fruit dehiscence.

Many genes are known to function in tissues that develop from the CMM, including the basic Helix-Loop-Helix (bHLH) transcription factor *SPATULA (SPT)*; Heisler et al., 2001). In *spt* loss-of-function mutants, the gynoecium remains unfused at the apex and the septum fails to develop properly (Alvarez & Smyth, 1999). Fertility is reduced in *spt* mutants due to a reduction in the size and extent of the

pollen transmitting tissue within the style and septum, in which cells fails to properly elongate. Outside the flower, *SPT* also functions in seed dormancy (Penfield et al., 2005) and in leaf and cotyledon expansion (Josse et al., 2011). In gynoecium and fruit tissues, *SPT* acts partially redundantly with its close paralog *ALCATRAZ* (*ALC*), though the latter plays the major role in the development of fruit dehiscence zones and shows no single mutant phenotype at earlier stages in the developing gynoecium (Groszmann et al., 2011).

*SPT* is capable of homodimerization, and of heterodimerization with both *ALC* (Groszmann et al., 2011) and with the bHLH factors *HECATE1-3* (*HEC1-3*; Gremski et al., 2007), which are redundantly necessary for normal style and septum development. *SPT* can also form heterodimers with *INDEHISCENT* (*IND*), a close relative of *HEC1-3* (Girin et al., 2011). Reymond et al. (2012) identified 24 putatively direct targets of *SPT*, including a high proportion of genes that are known to be regulated also by phytochrome-interacting transcription factors (*PIFs*) which, like *SPT* and *ALC*, belong to Group 24 of the very large bHLH family (Carretero-Paulet et al., 2010). These light-regulated target genes are known to act downstream of *PIFs* to promote cell-elongation, and thereby bring about shade avoidance responses in stem internodes and the hypocotyl. They are hypothesized also to act downstream of *SPT* to promote cell-elongation in the style and transmitting tissue of the gynoecium. Indeed, Reymond et al. (2012) proposed that *SPT* evolved from a *PIF* by the loss of an active phytochrome binding (*APB*) domain, thereby possibly escaping from light-regulation acting through phytochrome photoreceptors.

In addition to light-regulated targets, *SPT* has been shown to regulate several genes related to auxin signaling, one of which encodes the serine/threonine protein kinase *PINOID* (*PID*), which positively regulates cellular auxin efflux (Friml et al., 2004). *SPT* has been shown to interact with *IND* to downregulate *PID* in the style and carpel margins (Girin et al., 2011). A further direct target of *SPT*, *ARABIDOPSIS RESPONSE REGULATORI* (*ARR1*), promotes cytokinin-dependent cell division at the carpel margins (Irepan Reyes-Olalde et al., 2017).

The domain structure of *SPT*-like proteins has been analyzed in a wide variety of angiosperms, including magnoliids, monocots, basal eudicots, asterids and rosids (Groszmann et al., 2008; Ines Ortiz-Ramirez et al., 2018). *SPT* expression patterns have furthermore been studied in the basal eudicot *Boccinia frutescens* (Papaveraceae; Zumajo-Cardona et al., 2017), and comparison of these data with *SPT* expression in *Arabidopsis* (Heisler et al., 2001; Groszmann et al., 2010) suggested the role of *SPT* in the development of tissues from the carpel margins to be widely conserved in eudicots. However, functional-genetic studies using virus-induced gene silencing (*VIGS*) in three Solanaceae species recently found no *SPT*-knockdown phenotypes associated with early gynoecium development (Ines

Ortiz-Ramirez et al., 2019). Instead, these studies concluded *SPT*-like genes in Solanaceae to control petal expansion, leaf pigmentation and lignification in the fruit. Ortiz-Ramirez et al. (2019) speculate that other genetic systems may be acting redundantly with *SPT*-like genes to control gynoecium development in Solanaceae, or else that VIGS methods may not have produced a sufficient reduction in *SPT* mRNA levels to reveal a loss-of-function phenotype in Solanaceae models.

Here, we attempt to reconstruct the evolutionary history of the *SPT* lineage and use a series of *in-vitro*, *in-vivo* and *in-silico* approaches to try to determine the stage at which *SPT* acquired its current role in the development of tissues that arise from the carpel margins of the *Arabidopsis* gynoecium. We furthermore attempt to identify molecular changes that might have been responsible for this recruitment. We conclude that *Arabidopsis SPT* probably acquired major aspects of its role in the gynoecium in a common ancestor of Brassicaceae, rather than at a much earlier evolutionary stage within angiosperms. We present evidence to suggest that the acquisition of this role did not involve a major change in the biochemical capacities of *SPT* with regard to protein dimerization or protein-DNA interactions. We also find no evidence of any major evolutionary change in the direct targets regulated by *SPT* during angiosperm evolution. Rather, we conclude that *SPT*'s novel role in the Brassicaceae gynoecium resulted from a change to its expression pattern, and we present preliminary data suggesting that major rearrangements in the *SPT* promoter region may have correlated with this change.

## Results

The *SPT* lineage is conserved throughout euphyllophytes and shows a complex history of gene duplication, gene loss and structural modification.

To provide a framework to understand the evolutionary history of the *SPT* clade and its role in gynoecium development, we first performed a phylogenetic analysis of *SPT*-like sequences, together with selected sequences from Group 24 of the bHLH family (Carretero-Paulet et al., 2010). These sequences were sampled from a wide range of vascular plants, some of which have sequenced genomes (SI Tab. 1). The results of the analysis, which used *HEC* sequences from Group 31 of the bHLH family as a close out-group (Carretero-Paulet et al., 2010), are shown in Fig. 1.

Phylogenetic analysis shows both *SPT* and *ALC* from *Arabidopsis* in a clade of *SPT*-like sequences that lies in sister position to the PIF7 clade of phytochrome-interacting transcription factors. Genes from the monilophyte (fern) species *Azolla filiculoides* and *Salvinia cucullata* occupy relatively basal and well-supported positions within the *SPT* clade, while a gene from the lycophyte *Selaginella moellendorffii* forms a basal polytomy, together with the PIF7 and *SPT* clades. Further sampling of basal vascular plant taxa will be necessary to reach firmer conclusions, but these preliminary data

suggest that the *SPT* lineage may have originated in a common ancestor of euphyllophytes (ferns+seed plants), perhaps by duplication of an *PIF7*-like gene.

Our phylogenetic analysis provides evidence of two important duplication events in the *SPT* lineage that took place within angiosperms. The first of these was found at the base of the core-eudicot clade, and apparently gave rise to the *SPT* and *ALC* lineages present in Solanaceae and other asterids, as identified by Pabon-Mora et al. (2014). This gene duplication corresponds in timing with the well-documented *At-γ* genome triplication event (Jaillon et al., 2007). It appears, however, from our data that only one of the two sister lineages generated at this time was retained in rosids.

The second and more recent major duplication observed in the present study occurred somewhere around the base of Brassicaceae and appears to have generated the Brassicaceae *SPT* and *ALC* lineages, as defined by *SPT* and *ALC* in *Arabidopsis*. It therefore appears that genes labelled “*ALC*” in Brassicaceae (Rajani & Sundaresan, 2001; Groszmann et al., 2011) and Solanaceae (Pabon-Mora et al., 2014; Ines Ortiz-Ramirez et al., 2019) may not be orthologous between these families. Rather, the Solanaceae *SPT* lineage may be a pro-ortholog of *SPT* and *ALC* in Brassicaceae, while “*ALC*” genes from Solanaceae and other asterids may have no true orthologs in Brassicaceae, this lineage having been lost early in the evolution of the rosid clade.

We subjected a range of sequences from the *SPT+PIF7* clade in vascular plants, and also *Arabidopsis PIF5* for comparative purposes, to structural analyses using MEME software (Fig. 2A). *PIF5* was chosen as this gene, together with its paralog *PIF4*, was shown to be redundantly necessary for the rescue of *Arabidopsis spt* mutants under specific light conditions (Reymond et al., 2012). All the bHLH proteins we analyzed showed a well-conserved bHLH domain, which is known to be responsible for DNA-binding, homodimerisation, and heterodimerisation with other bHLH proteins.

Of the proteins analyzed here, only *Arabidopsis PIF5* and the putative *PIF7/SPT* pro-ortholog from *S. moellendorffii* possess an N-terminal APB domain (Figs 2B-C). Our findings suggest, therefore, that the *SPT* lineage lost an APB domain, presumably through a genetic truncation event, in a common ancestor of euphyllophytes, shortly after its origin through duplication (Fig. 2D). The evolution of *SPT* through the loss of the APB domain was already proposed by Reymond et al. (2012), though these authors could only conclude this event to have taken place in or before the MRCA of living seed plants. As noted by Reymond et al. (2012), the *SPT* lineage clearly lost its APB domain, and potentially therefore escaped from light-regulation acting through phytochrome, long before the evolutionary origin of the carpel in a common ancestor of angiosperms.

Groszmann et al. (2008) reported SPT orthologs from eudicots to possess two functionally important domains, in addition to the bHLH domain, that were not found to be conserved in monocots. These domains correspond to an amphipathic helix that was determined by deletion experiments to enhance *SPT* function in the gynoecium, and an acidic domain that proved essential to this same gynoecium function. Multiple alignment of bHLH sequences indeed shows sparse amino-acid similarity, outside eudicots, over the amphipathic domain identified by Groszmann et al. (2008), as shown in Fig 2E. However, it is possible that functionally equivalent regions containing similarly charged residues may be present in SPT orthologs from outside the eudicots, but are difficult to identify by multiple alignment. The acidic domain reported by Groszmann et al. (2008) is, by contrast, clearly present in SPT orthologs from basal taxa including *Amborella* and *Picea* (Fig 2F). Amino-acid similarities in this domain are also conserved to a significant extent in the putative PIF7/SPT pro-ortholog from *S. moellendorffii* and in PIF proteins including *Arabidopsis* PIF5. The essential acidic domain of eudicot SPT orthologs appears, therefore, to be widely conserved in angiosperms, and in Group 24 bHLH proteins as a whole, even if it has become lost or indistinct in the *SPT* lineage in monocots.

*In-vitro* assays suggest SPT proteins to have conserved both important cofactors and target genes, at least since the MRCA of living angiosperms.

The phylogeny and domain structure of SPT-like proteins provide interesting insights into the origin of this lineage, but fail to indicate any molecular changes that may have been responsible for *SPT*'s recruitment to gynoecium development, or suggest when this recruitment occurred. To throw some light on these questions, we attempted to determine whether the interactors and target genes of SPT proteins have changed over evolutionary time, particularly during the evolution of angiosperms.

We first used yeast-2-hybrid analyses to determine whether the SPT protein from the likely basal-most living angiosperm, *Amborella trichopoda*, was capable of homodimerization, and/or of heterodimerization with the *HEC*-like genes from the same species, as does its ortholog in *Arabidopsis*. *Amborella* contains two *HEC*-like genes, one of which, *AtrHEC1/2*, is the likely pro-ortholog of *Arabidopsis HEC1* and *HEC2*, while the other, *Atr\_bHLH87*, is the likely pro-ortholog of *Arabidopsis HEC3* and *IND* (Pabon-Mora et al., 2014). The results of this analysis (Fig. 3) show that *Amborella* SPT is able to homodimerize, and to heterodimerize with both *Amborella* *HEC* proteins, strongly suggesting that both of these types of physical interactions involving SPT proteins have been conserved in the *Arabidopsis* and *Amborella* lineages since the MRCA of living angiosperms.

We then attempted to compare the likely direct targets of *SPT*-clade transcription factors, this time extending our analysis to cover the gymnosperm *Picea abies*. *SPT* coding sequences from

*Arabidopsis*, *Amborella* and *Picea* were overexpressed in *E. coli* as protein fusions to a Maltose Binding Protein domain, and analyzed using Protein Binding Microarrays (Godoy et al., 2011) to determine their *in-vitro* DNA-binding preferences. In each case, three high-scoring position weight matrices (PWMs) were derived from the first three top-scoring single 8-mers identified for each factor (SI Fig 2 and Fig 4).

Twenty six target genes have been proposed as direct or immediate targets of *Arabidopsis* SPT (Girin et al., 2011; Reymond et al., 2012; Irepan Reyes-Olalde et al., 2017), all but one of which show G-boxes (CACGTG) in their putative promoter regions. In agreement with this observation, all three top-scoring 8-mers for each of the three factors compared were also found to contain G-boxes (Fig. 4). Previous studies of SPT targets in *Arabidopsis* have considered entire intergenic regions as potential promoter regions. However, this approach is not feasible in *Amborella* or *Picea* as average intergenic distances are much greater in these species than in *Arabidopsis*. Also, in *Picea*, the current genome assembly is far from complete and would not, in most cases, be expected to yield full intergenic regions.

To make a direct comparison between the three species under analysis, we therefore defined “promoters” as the 3 kb of sequence immediately upstream of the initiation codon in each gene of interest, regardless of the length (known or unknown) of the intergenic region concerned. Orthologs of putative *Arabidopsis* SPT-targets from *Amborella* and *Picea* were identified using BLAST searches, and their putative orthology was confirmed using Maximum Likelihood phylogenetic analyses in SeaView (Gouy et al., 2010; results available on request). Promoter sequences were obtained for these orthologs from Phytozome, or from Congenie in the case of *Picea*. The required 3-kb of upstream sequence could be obtained for all but three of the *Picea* genes of interest. All sequences were scanned in RSAT (Nguyen et al., 2018) using the MotifScan program, using an arbitrary cut-off value of 6.0 to define hits. Overlapping sites, sites identified on both DNA strands, and sites identified using more than one of the three top-scoring matrices for each factor were reduced to a single site in the scan reports provided in SI Tab. 2 and summarized further to indicate only the number of high-scoring sites detected in each promoter in the summary given in Tab. 1. Interestingly, almost one-third (30/95) of the sites detected (SI Tab. 2) do not contain a full consensus G-box, despite giving a score of 6.0 or more in the analysis. Some *Arabidopsis* promoters of interest were not confirmed as targets in this analysis, in some cases because their only G-box-containing motifs were to be found at >3 kb upstream of the initiation codon, and in others because these motifs yielded scores of <6.0. Accordingly, 18 of 24 putative *Arabidopsis* target promoters analyzed were found to contain at least one high-affinity

potential binding site for SPT, while the equivalent figures were 13/22 for *Amborella* and 10/13 for *Picea* (not including genes for which truncated sequences only were available).

To determine whether the above figures represented a significant enrichment in sites predicted to bind SPT with high affinity, we further obtained datasets of 3-kb upstream sequences from the entire sequenced genomes of *Arabidopsis* and *Amborella*. It was not, however, possible to obtain equivalent data from the relatively incomplete genome assembly of *Picea*. We scanned the *Arabidopsis* and *Amborella* promoter datasets with the corresponding top-scoring SPT PWM, and then used the hypergeometric test to evaluate the significance of the results obtained, compared to the equivalent scans of *Arabidopsis* or *Amborella* putative SPT-target promoters. This procedure identified significantly more promoters containing high-scoring sites among putative targets, at p-values of  $6.00 \times 10^{-7}$  and 0.0150 for *Arabidopsis* and *Amborella*, respectively (SI Tab. 3, Analysis 1). High scoring sites were over-represented in putative SPT promoters by a factors of 4.83 and 2.66 in *Arabidopsis* and *Amborella*, respectively (SI Tab. 3, Analysis 2). As the proportion of putative target promoters containing at least one high-scoring site was higher in *Picea* than *Amborella*, we suppose that an equal or lower p-value might have been obtained for *Picea*, had the necessary sequence data been available to make a full analysis.

These data suggest that direct target genes of SPT have been largely conserved, at least since the MRCA of living angiosperms, and probably since that of living seed plants. These targets include genes identified by Reymond et al. (2012), which encode proteins significantly over-represented in shade-avoidance- and auxin-related processes, among others. Two further 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*, though this confirmation was not possible in *Picea* due to a lack of sequence data from the upstream regions of these genes.

It seems, from the above *in-vitro* and *in-silico* studies, that most of the major direct molecular interactions known to involve *Arabidopsis* SPT, both with its protein cofactors and its direct target genes, have likely been conserved, at least since the MRCA of living angiosperms.

*SPT* orthologs from a wide range of vascular plants can complement strong loss-of-function *spt* mutations in transgenic *Arabidopsis*.

The conservation of the *in-vitro* biochemical properties of *Arabidopsis* SPT with its orthologs from basal angiosperms and gymnosperms led us to ask whether the *in-vivo* biological properties of SPT-like proteins might be similarly conserved between very distant taxa. Indeed, Reymond et al. (2012)

had found that PIF4 and 5, two paralogous bHLH proteins from outside the SPT clade, contributed to the rescue of *spt* loss-of-function mutants grown under low red/far-red light ratios, which suggested that even these evolutionarily distant molecules might be able to replace the biochemical functions of SPT. Furthermore, Groszmann et al. (2008) and Groszmann et al. (2011) had already demonstrated the rescue of *spt-2* mutants using both *Arabidopsis ALC* and an *SPT* ortholog from *Solanum lycopersicum* (tomato).

To extend the above type of analysis over a greater phylogenetic range, we transformed strong loss-of-function *Arabidopsis spt-11* mutant plants using the coding sequences from a range of *SPT* orthologs, pro-orthologs and other bHLH Group 24 members, from taxa as far removed as lycophytes, all driven using the fully functional 6.3-kb promoter sequence of *Arabidopsis SPT* defined by Groszmann et al. (2010). Two versions each of a putative *PIF7/SPT* pro-ortholog from *S. moellendorffii* and of *Arabidopsis PIF5* were included in these studies, one of which had been artificially truncated to remove its N-terminal APB domain while leaving intact its initiation codon and reading frame. The untransformed *Arabidopsis spt-11* mutant typically 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, due to poor pollen-tube transmission prior to fertilization (Fig. 5A-C). At least 16 independent transformants were observed for each transgene construct tested, and untransformed mutants were grown for comparison in each batch of plants observed.

All the constructs tested were able, in at least some of the transgenic lines analyzed, to fully restore fusion of the gynoecium apex (Fig 5D-N, Tab. 2). There was no significant difference (at the  $p < 0.05$  level) in the efficiency of the constructs tested, compared to the positive control *AthSPT::AthSPT*, with the exceptions of: *pSPT::PIF5*, *pSPT::SmoPIF7/SPT* and *pSPT::Smo $\Delta$ PIF7/SPT*, which proved less efficient than the others. Interestingly, the proportion of lines restored was higher in the N-terminally deleted versions of both *AthSPT::Smo PIF7/SPT* and *AthSPT::AthPIF5*. The constructs tested varied more markedly in their capacity to fully restore fruit-size/fertility (Fig. 6, Tab. 2). Accordingly, *AthSPT::AthALC*, *AthSPT::NcaSPT* and *AthSPT::PaxALC* all complemented the fruit-size phenotype with no significant difference in efficiency from the positive control construct (at  $p < 0.05$ ). *AthSPT::AtrSPT*, *AthSPT::AthPIF5* and *AthSPT:: $\Delta$ AthPIF5* complemented the fruit-size phenotype with intermediate efficiency ( $p$ -value between 0.001 and 0.05), while the remaining constructs tested showed very low efficiency ( $p < 0.001$ ), with no complementation in any transformed line generated using *AthSPT::PabSPT*. Again, the N-terminally deleted versions of *AthSPT::AthPIF5* and

*AthSPT::SmoPIF7/SPT* were more efficient at complementing the fruit mutant phenotype than the unmodified versions, though the difference was not significant in the latter case.

These results indicate the *in-vivo* biological activity of the *SPT* coding sequence in the *Arabidopsis* gynoecium to be broadly conserved with widely diverged proteins from Group 24 of the bHLH family. Though some differences in complementation efficiency were observed between constructs, these did not appear to correlate with any given level of phylogenetic distance from *Arabidopsis SPT*. The results obtained using *AthPIF5* are interesting and support the conclusion of Reymond et al. (2012) that the rescue of *Arabidopsis spt* mutants grown under low red/far-red light ratios may depend in part on this gene.

In general, the results of these complementation experiments do not indicate a level of phylogenetic distance within bHLH Group 24 genes in angiosperms at which the rescue of *Arabidopsis spt* mutants is no longer possible, and therefore do not help to establish a stage during land plant evolution at which the *SPT* lineage might have acquired the biochemical properties needed to play its current developmental role in the *Arabidopsis* gynoecium. Indeed, these results are consistent with the hypothesis that almost any coding sequence from the *PIF7/SPT* clade in vascular plants, or even perhaps from the entire Group 24 of bHLH proteins, might be able to replace the function of *AthSPT* in transgenic *Arabidopsis*, providing it is appropriately expressed from the *AthSPT* promoter.

*SPATULA* orthologs show similarities in gynoecium expression patterns between widely diverged angiosperms, though strong expression in the style and replum appears specific to Brassicaceae.

The results of the above *in-vitro* and *in-vivo* analyses indicate that the biochemical properties of *SPT* proteins have been largely conserved in vascular plants for over >400 MY, and therefore suggest that evolutionary changes to the *SPT* coding region may not have been involved in the recruitment of *SPT* to gynoecium development. These considerations led us to ask whether the role of *SPT* in the gynoecium might have originated from a change in its expression pattern. We therefore tested the expression of *SPT* genes in the asterid (eudicots) model *Petunia axillaris*, and in the ANA-grade (basal) angiosperms *Amborella trichopoda* and *Nymphaea thermarum*.

*Petunia axillaris* contains two *SPT*-like lineages which we refer to here as *PaxSPT* (orthologous to previously identified Solanaceae *SPT* genes; Ines Ortiz-Ramirez et al., 2019) and *PaxALC* (orthologous to previously identified Solanaceae *ALC* genes; Ines Ortiz-Ramirez et al., 2019). *PaxSPT* is expressed in the L1 cell-layer of both the floral meristem and stamen primordia (Fig. 7A-B), with a peak of expression also at the centre of the early gynoecium in the region where the placenta will form

(Fig. 7B-C). *PaxSPT* is also 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. 7A). At later stages (Fig. 7D-F), very 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 stylar transmitting tissue. *PaxALC* is less strongly and widely expressed than *PaxSPT* (Fig. 7G-L), showing clearest 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.

*AtrSPT*, the *SPT* ortholog from the dioecious basal angiosperm *Amborella trichopoda*, is expressed in the stamens of male flower buds, particularly in the zones that will form the anther loculi (Fig. 8A-B). In female buds, *AtrSPT* is expressed very strongly in carpel primordia and moderately in young tepals (Fig. 8C). Strong *AtrSPT* expression can be observed at later stages (Fig. 8D-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*, the *SPT* ortholog from the basal angiosperm *Nymphaea thermarum*, is expressed during the formation of the anther loculi (Fig. 8F) and in the L1 cell-layer of all floral organs (Fig. 8F-G). *NthSPT* is also strongly expressed adaxially in the gynoecium and in the placenta and developing ovules (Fig. 8F-G).

Important aspects of *SPT* expression are thus conserved between the three widely-diverged angiosperms studied here, and with the well-documented expression of *SPT* in *Arabidopsis* (Heisler et al., 2001; Groszmann et al., 2010). In particular, *SPT* orthologs appear highly expressed in anther loculi, ovules and adaxial tissues of the carpel/gynoecium. Interestingly, the very strong expression of *SPT* in the transmitting tissue in *Arabidopsis* is not matched by similar levels of expression of the two *SPT*-like genes in *Petunia axillaris*. Both of the basal angiosperms studied in the present work lack specialized transmitting tissue, so no direct comparison with core eudicots is possible on this point.

#### Promoters of *SPT* orthologs show diverse activities in transgenic *Arabidopsis*.

The *in-situ* hybridization studies described above, combined with published data, revealed both similarities and differences between the expression patterns of *SPT* orthologs in widely diverged angiosperms. We were curious to know whether apparent differences, such as the much higher expression of *SPT* in the stylar tissues in *Arabidopsis* compared with other species, were due to differences between *SPT*-promoter sequences, or to some other factor. To investigate this question, we decided to test the activity of *SPT* promoters from a range of species in transgenic *Arabidopsis*. An important caveat of this work is that observed differences in expression between promoters in

heterogeneous systems may be complex to interpret with regard to promoter evolution: the promoter sequences may have diverged between taxa, but so too may trans-acting components that interact with them. By contrast, similar expression patterns generated by orthologous promoters in a heterologous system may provide useful information, pointing in particular to a likely conservation of activity in the promoters under comparison.

Promoter activities of *SPT* orthologs from three species representing diverse seed plants were tested using GUS reporter-gene expression in transgenic *Arabidopsis*, together with a positive-control construct containing the 6.3-kb upstream sequence of *AthSPT* (Groszmann et al., 2010). The control construct generated very high reporter activity in the style and extending down into the replum within the ovary (Fig. 9A-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* also generated strong gynoecium expression in transgenic *Arabidopsis* (Fig. 9C-D, Tab. 3), but the precise pattern generated by this construct was in contrast to that of the control, with generalized expression in the valve tissues and no visible expression in the style. The upstream sequence of *PabSPT*, the likely *SPT* ortholog from the gymnosperm *Picea abies*, generated very high reporter-gene activity in sepals and at the base of pedicels, but no detectable reporter expression in the gynoecium (Fig. 9E-F, Tab. 3). Finally, the orthologous promoter from the *NcaSPT* gene of the basal angiosperm *Nymphaea careulea*, a very close relative of *N. thermarum* from the same subgroup of African waterlilies, generated a low level of reporter expression in anthers, but no detectable expression in the gynoecium (Fig. 9G-H, Tab. 3).

Of the three species tested here in heterogeneous combinations, none showed *SPT*-promoter activity that was very similar to that of the native *Arabidopsis* *SPT* promoter. Arguably, the reporter-gene expression generated by *PaxSPT* was most similar to that of *Arabidopsis*, as strong expression in the gynoecium was noted in both species, though the precise tissues-specific patterns observed were clearly distinct. The results of these experiments are not easy to interpret, but seem broadly to indicate that *SPT* promoters have likely undergone evolutionary changes of significance to gene expression patterns during angiosperm evolution, both before and after the origin of the core eudicots. We considered from this tentative conclusion that it would be interesting to compare *SPT*-promoter sequences in detail to identify elements that may account for the similarities and differences observed in *SPT*-expression between taxa.

### Comparison of *SPT* promoters in core eudicots reveals novel motifs and structural rearrangements of potential functional significance.

To attempt to identify conserved, functional elements of *SPT* promoters, we compared promoters from 20 species of core eudicots, including seven species of Solanaceae, five species of Brassicaceae, and eight species of other rosids. The use of MEME software identified ten motifs shared between two or more of the species under comparison (Fig. 10). The Brassicaceae species analyzed showed a particularly striking similarity in the spatial arrangement of conserved motifs (Fig. 11). Groszmann et al. (2010) had already 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 3. However, Motifs 4, 9 and 10 are newly identified motifs that occur in a well-conserved cluster between the Regions 5 and 6 of Groszmann et al. (2010).

We found Solanaceae *SPT* promoters to share numerous conserved motifs with Brassicaceae. For example, the *P. axillaris* *SPT* promoter shares seven of the ten motifs identified in *Arabidopsis*, lacking only Motifs 2, 7 and 9 (Fig. 11). 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 suppressers 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 with 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 two hybrid motifs identified by Groszmann et al. (2010) were identified in the present study as Motifs 1 and 3 (Fig. 10). 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. 11). 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 Solanaceae cluster is also typically augmented by the presence of Motif 4 near its proximal end. Motif 4 is universally present too in the Brassicaceae promoters analyzed, but in much more variable positions, and in most cases far upstream of the well-conserved cluster of Motifs 1-6-7-3 near the transcriptional start site. Motif 7, present in Brassicaceae but absent in Solanaceae *SPT* promoters, contains a CAAG box, already identified by Groszmann et al. (2010). It is worth noting, however, that despite the absence of this element, the

*Pax\_SPT* promoter from *Petunia axillaris* is capable of directing strong reporter expression in the gynoecium of transgenic *Arabidopsis*, albeit with a different tissue-specific pattern to the native *Arabidopsis SPT* promoter (Fig. 9C-D).

Mutant analyses provide inconclusive data on the biological role of *SPT*-like genes in *Petunia*.

To investigate the biological role of *SPT*-like genes in *Petunia hybrida* (a hybrid species from a cross including *P. axillaris*), we identified lines carrying transposon insertions in both the *Pax\_SPT* and *Pax\_ALC* genes. Homozygous single *Pax\_spt-1* or *Pax\_alc-1* mutants showed no detectable phenotypic differences to wild-type plants (data not shown). However, a double homozygous mutant containing the same two alleles in the Mitchell genetic background showed a gynoecium phenotype in which the stigmatic surface lacked the normal papillae present in wild-type plants (Fig 12). This phenotype is distinct from the gynoecium phenotype of *Arabidopsis spt* or *spt alc* mutants, in which stigmatic tissues may be reduced, but some stigmatic papillae are still typically present. It is, however, in good agreement with the expression patterns of *PaxSPT* and *PaxALC*, which overlap in the L1 cell-layer in the stigma (Fig 6L).

The same two alleles *Pax-spt-1* and *Pax-alc-1*, combined in the W138 (transposon line) genetic background, showed a more subtle stigmatic phenotype in which stigmatic papillae were present, though with some areas of the stigmatic surface devoid of papillae. However, a similar phenotype was observed also in wild-type W138 plants, so this phenotype cannot be ascribed to mutations in *PaxSPT* and *PaxALC*. A second mutant allele of each gene was then obtained, *Pax\_spt-2* and *Pax\_alc-2*, and transferred into both the Mitchell and W115 genetic backgrounds by genetic crossing. Homozygous double mutants carrying these alleles, however, showed no mutant phenotype.

At present, it is unclear whether a third gene may have been mutated, and contributed to the clear stigma phenotype originally observed in the *Pax\_spt-1 Pax\_alc-1* double mutant in the Mitchell background. It is also not sure that all of the alleles investigated produced null mutant phenotypes. It is furthermore possible that environmental conditions (e.g. ambient light conditions) may have had an impact on allelic penetrance and have varied subtly between experiments. These questions will require further careful analyses in several mutant lines and strictly controlled environmental conditions to reach a definitive answer.

## Discussion

### The origin of the *SPATULA* lineage in vascular plants.

We have traced the likely origin of the *SPT* lineage to a common ancestor of euphyllophytes (ferns+seed plants; Fig 1). We confirm the findings of Reymond et al. (2012) that a defining characteristic of SPT-clade proteins appears to be the loss of an N-terminal active phytochrome-binding (APB) domain, present in their close relatives the PIF proteins, which is responsible in the latter for interactions with the active form of phytochrome photoreceptors (Fig. 2A-C). In the more distantly related lycophytes, a putative pro-ortholog of *SPT* is equally closely related to the *PIF7* lineage, and encodes an N-terminal APB-like domain (Fig. 2D). We provide evidence that the biochemical characteristics of *Arabidopsis SPT* can be replaced by a high proportion of *SPT* coding sequences from diverse land plant taxa (Figs 5 and 6), suggesting that the SPT protein in *Arabidopsis* has not acquired and retained significant novel biochemical capacities since its distant ancestor in the MRCA of euphyllophytes. This conclusion is supported by the observation that SPT proteins from the basal angiosperm *Amborella* and the gymnosperm *Picea* both bind preferentially to G-box-containing motifs (Fig. 4), as does SPT in *Arabidopsis*. Furthermore, the SPT protein from *Amborella* is capable of equivalent protein-protein interactions to those shown by *Arabidopsis SPT* (Fig. 3). The downstream pathways regulated by *Arabidopsis SPT*, as assessed by the presence of high-affinity SPT-binding sites in putative target promoters, appear also to have been largely conserved, at least since the MRCA of living seed plants (Tab. 1, SI. Tab 2).

Groszmann et al. (2008) identified two domains outside the bHLH domain that supported SPT functions in transgenic *Arabidopsis*: an amphipathic helix and an acidic domain. These domains were identified as specific to SPT/ALC-like proteins from eudicots, being absent in SPT orthologs from monocots. The acidic domain identified by Groszmann et al. (2008) is, however, encoded in most *SPT* orthologs/pro-orthologs including those from basal angiosperms, gymnosperms, and the lycophyte *Selaginella moellendorffii*, as well as in other PIF genes such as *AthPIF5* (Fig. 2F). The amphipathic helix of eudicot *SPT* orthologs is less easy to recognize in other taxa (Fig. 2G), but it is possible that a structurally similar region, or several regions split over different parts of the protein molecule, might functionally replace this amphipathic helix in non-eudicot SPT and related proteins and thus contribute to the *in-vivo* functional equivalence of these with *Arabidopsis SPT*, as observed in the present work (Figs 5 and 6).

### When and how did the role of *SPATULA* arise in the development of carpel margin tissues?

In *Arabidopsis*, *SPT* controls the development of tissues arising from the carpel margin meristem (CMM), including the pollen-transmitting tissue of the style and adaxial replum. Solanaceae species also contain a specialized transmitting tissue in the style. However, *SPT*-like genes are much less strongly expressed in these tissues in Solanaceae, compared to *Arabidopsis*, and studies using VIGS (Ines Ortiz-Ramirez et al., 2019) and mutant approaches (Fig. 12) have failed to show any clear phenotype associated with loss of *SPT* function in these tissues. It therefore seems likely that the *SPT* lineage was recruited to the development of CMM tissues in a common ancestor of Brassicaceae, and that it does not share this role in widely diverged angiosperms such as asterids or basal angiosperms. The replum of *Arabidopsis* is a false septum that grows from the carpel margins to divide the ovary into two loculi. The ovary of *Petunia* (Solanaceae) also contains two loculi, but these are separated by congenitally fused ovary-wall tissues, rather than a false septum. Thus, the anatomy of the typical Brassicaceae gynoecium is highly specialized, and it seems highly plausible that modifications to the role of *SPT* could have been involved in the evolution of this family-specific anatomical arrangement. Our study suggests *Arabidopsis* *SPT* to possess few if any unique biochemical properties, compared to its orthologs from diverse vascular plants, and also that its downstream genetic pathways and protein-protein interactions are largely conserved, at least within angiosperms. We therefore propose that a change to *SPT*'s expression pattern was likely responsible for its recruitment to CMM development in Brassicaceae. Solanaceae *SPT* promoters in transgenic *Arabidopsis* fail to induce high levels of reporter-gene expression in the style and replum, though are highly active in other floral tissues, suggesting that changes to the *SPT* promoter in a common ancestor of Brassicaceae may have been responsible for the evolutionary acquisition of novel functions in the style and replum. Groszmann et al. (2010) previously showed that a conserved, proximal region of the *Arabidopsis* *SPT* promoter, containing closely spaced G-box and AuxRE motifs, was essential to stylar and replum expression. Interestingly, such hybrid motifs are typically present also in the *SPT* promoters of Solanaceae species, and also occur in clusters, though with a different order and position relative to the proximal end of the promoter. Therefore, we tentatively suggest that structural rearrangements to the *SPT* promoter, combined with the introduction of a CAAT box, may have been involved in the acquisition by *SPT* of novel gynoecium expression patterns in a common ancestor of Brassicaceae. This hypothesis might be tested by making subtly modified versions of a Solanaceae *SPT* promoter to determine which modifications might produce high levels of reporter expression in the style and replum of transgenic *Arabidopsis*.

We show *SPT* to be very strongly expressed in the placentae, ovules and abaxial regions of the gynoecium in *Petunia* and in the basal angiosperms *Amborella* and *Nymphaea*. However, the functional significance of this expression is unclear as mutant and VIGS approaches have not identified any clear phenotypes associated with the loss of *SPT* function in these tissues in Brassicaceae or Solanaceae models (Alvarez & Smyth, 1999; Groszmann et al., 2011; Ines Ortiz-Ramirez et al., 2019). Given our findings (Figs 5 and 6) and those of Reymond et al. (2012) relating to the biological equivalence between *SPT* and certain PIF proteins, it could be that *SPT*-like genes are acting redundantly with PIFs in these tissues of high expression which have so-far revealed no mutant phenotypes.

#### The origin of *ALCATRAZ* lineages in Brassicaceae and asterids.

There is currently in the literature a debate over the relationship between genes named “*ALCATRAZ*” (*ALC*) in Brassicaceae, compared to those in distantly-related core-eudicot taxa such as Solanaceae. Groszmann et al. (2011) considered the *Arabidopsis* *ALC* lineage to have arisen by a duplication of the pre-existing *SPT* lineage within Brassicales (more precisely in a common ancestor of Brassicaceae, Cleomaceae and Capparaceae), and this conclusion is in agreement with the results of recent Bayesian phylogenetic analyses (Pfannebecker et al., 2017). Considerations of synteny suggested (Groszmann et al., 2011; Bowers et al., 2003), in agreement with the above findings, the *SPT/ALC* duplication to have corresponded to the At- $\beta$  whole genome duplication (Bowers et al., 2003) within Brassicales. However, more recent studies (Pabon-Mora et al., 2014; Ines Ortiz-Ramirez et al., 2019) have identified putative *ALC* orthologs from Solanaceae and other asterids and propose the *ALC* lineage to have separated from that of *SPT* in a common ancestor of core eudicots, probably at around the time of the At- $\gamma$  genome triplication event. According to this latter hypothesis, both the *SPT* and *ALC* lineages would have been conserved in asterids (including Solanaceae) and rosids (including Brassicaceae).

It has been suggested that long-branch attraction (LBA) in Bayesian phylogenies may have distorted the phylogenetic position of *ALC* genes, specifically in Brassicaceae (Ines Ortiz-Ramirez et al., 2019). However, the Maximum Likelihood phylogeny presented here (Fig. 1) is in agreement with the conclusions of Groszmann et al. (2011), and with the Bayesian analyses of Pfannebecker et al. (2017). We do not consider it likely that LBA could have pushed the Brassicaceae *ALC* lineage to an erroneously terminal position in molecular phylogenies, as LBA tends to have the opposite effect of

pushing fast-evolving lineages to erroneously basal positions. Synteny between duplicated blocks of *Arabidopsis* chromosomes has already been used to support the origin of the *AthSPT* and *AthALC* lineages within Brassicales (Groszmann et al., 2011). It may soon be possible to use a more extensive combined phylogeny+synteny approach (Murat et al., 2017) to clear up any remaining questions relating to the origin of *SPT/ALC*-like genes in diverse core eudicot groups.

## Materials and Methods

### Database-searching, phylogenetic reconstructions and DNA-sequence analyses.

Sequences of interest were identified by TBLASTn (Altschul et al., 1997) from databases given in SI Tab. 1. Amino-acid alignments were performed with MUSCLE in SeaView (Gouy et al., 2010), using default parameters. Alignments were used to generate maximum likelihood phylogenies in PhyML, employing the LG evolutionary model. Branch support was provided by the aLRT method (Anisimova & Gascuel, 2006). Conserved motifs in promoters and coding sequences were identified using MEME Suite 5.1.1 (<http://meme-suite.org/tools/meme>).

### Plant material and nucleic acid extraction.

Mutant and wild-type *Arabidopsis thaliana* 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. *Pinus taeda* needle material was a kind gift from Charles Dana Nelson, University of Kentucky, Lexington KY, USA. *Amborella* plants were obtained in 2010 in collaboration with Gildas Gâteblé and Bruno Fogliani (IAC, St Michel, New Caledonia) and then maintained in Lyon in a greenhouse under conditions of ~70% relative humidity, 18-30°C, and natural daylight attenuated by artificial shade-covering. 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 transformations.

*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 by standard “floral dip” procedures (Clough & Bent, 1998) and transformants selected on kanamycin (250 µg/L)-containing media.

### Transposon-mutant screening.

*Petunia hybrida* plants were grown under 18h/6h light/dark cycles at 22°C. A transposon mutant collection of *P. hybrida* W138 was screened for insertions in two *SPT/ALC* homologues as described by Vandebussche et al. (2008). Primers used for genotyping, by PCR-amplification across transposon borders, are given in SI Tab. 4. Mutant alleles of interest were backcrossed into Mitchell and WII5 genetic backgrounds and combined by crossing and PCR-screening of F2 populations to detect double mutant homozygotes. A double-mutant homozygote in the W138 (transposon) line was also constructed.

### 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 through SDS-PAGE analysis of cell lysates. The resulting recombinant proteins were analyzed on PBM11 protein binding microarrays (PBMs), as described by Godoy et al. (2011) and the resulting data analyzed as described by Berger & Bulyk (2009) to produce position weight matrices (PWMs) describing DNA-binding preferences, which were converted to diagrams using ENOLOGOS (Workman et al., 2005). PWMs were also used to scan promoter sequences in RSAT MotifScan (Nguyen et al., 2018). 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 described in SI Tab. 3.

### 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 *Atr\_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 into the *pGAD24-GW* expression vector (Clontech) as a translational fusion to the *GAL4* activation domain (AD). The resulting *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 AD and DBD of *GAL4*.

### GUS reporter gene analysis

Promoter sequences of *SPT* orthologs were PCR-amplified using high-fidelity thermostable polymerases and primers given in SI Tab. 5. The resulting molecules were inserted by recombination in the *pENTR/D/TOPO* (Invitrogen) vector and sequenced using automated Sanger DNA sequencing reactions to ensure the absence of mutations. 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, as described above. 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- $\beta$ -glucuronide 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. 6. The resulting DNA molecules were inserted by recombination into the *pENTR/D/TOPO* vector (Invitrogen) and sequenced in automated Sanger sequencing reactions to verify the absence of mutations. Two different 5'-primers were used to amplify each of *AthPIF5* and *SmoPIF7/SPT*, one of which caused a 5'-truncation of the APB domain. 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 required coding sequence were then transferred to *pK7m24GW* plasmids (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 using the resulting plasmids, 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 using a stage-temperature of -10°C to -20°C and a tube-voltage of 10 Kv.

### 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).

## **Author contributions**

ACRS performed the phylogenetic and bioinformatics analyses, most of the *in-situ* hybridizations, the molecular cloning and genetic transformations, all phenotypic characterizations and microscopy, helped prepare recombinant proteins for *in vitro* studies, and wrote the manuscript. ACV cloned and prepared recombinant proteins and tutored ACRS and LB. AAR performed Y2H analyses and *in situ* hybridizations on *Amborella* tissues. PC prepared recombinant proteins, tutored ACRS and cared for growing plant material. LB performed phylogenies of SPT-target genes and an earlier version of the RSAT analysis. PM and MV screened for *Petunia spt* mutants and supervised their phenotypic characterization by ACRS. RS and JMFZ performed PBM analyses. CPS coordinated the work and corrected the manuscript.

## **Acknowledgements**

This work was supported by ANR grant “ORANGe” (2013-18) and through core funding from CNRS, INRAe, ENS-Lyon and Université Claude-Bernard Lyon-1. ACRS was supported by a BECAL studentship from the Government of Paraguay. We thank Gildas Gâteblé, Bruno Fogliani, Paula Rudall, Carlos Magdalena and Charles Dana Nelson for generously supplying plant material.

## References

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Alvarez, J., & Smyth, D. R. (1999). CRABS CLAW and SPATULA, two Arabidopsis genes that control carpel development in parallel with AGAMOUS. *Development*, *126*(11), 2377–2386.
- Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, *55*(4), 539–552. <https://doi.org/10.1080/10635150600755453>
- Berger, M. F., & Bulyk, M. L. (2009). Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols*, *4*(3), 393–411. <https://doi.org/10.1038/nprot.2008.195>
- Bowers, J. E., Chapman, B. A., Rong, J. K., & Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, *422*(6930), 433–438. <https://doi.org/10.1038/nature01521>
- Carretero-Paulet, L., Galstyan, A., Roig-Villanova, I., Martinez-Garcia, J. F., Bilbao-Castro, J. R., & Robertson, D. L. (2010). Genome-Wide Classification and Evolutionary Analysis of the bHLH Family of Transcription Factors in Arabidopsis, Poplar, Rice, Moss, and Algae. *Plant Physiology*, *153*(3), 1398–1412. <https://doi.org/10.1104/pp.110.153593>
- Clough, S. J., & Bent, A. F. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal*, *16*(6), 735–743. <https://doi.org/10.1046/j.1365-313x.1998.00343.x>
- de Bossoreille, S., Morel, P., Trehin, C., & Negrutiu, I. (2018). REBELOTE, a regulator of floral determinacy in Arabidopsis thaliana, interacts with both nucleolar and nucleoplasmic proteins. *Febs Open Bio*, *8*(10), 1636–1648. <https://doi.org/10.1002/2211-5463.12504>
- Ferrandiz, C., Fourquin, C., Prunet, N., Scutt, C. P., Sundberg, E., Trehin, C., & Vialette-Guiraud, A. C. M. (2010). Carpel Development. In J. C. Kader & M. Delseny (Eds.), *Advances in Botanical Research*, Vol 55 (Vol. 55, pp. 1–73).
- Franco-Zorrilla, J. M., Lopez-Vidriero, I., Carrasco, J. L., 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(6), 2367–2372. <https://doi.org/10.1073/pnas.1316278111>
- Friml, J., Yang, X., Michniewicz, M., Weijers, D., Quint, A., Tietz, O., Benjamins, R., Ouwerkerk, P. B. F., Ljung, K., Sandberg, G., Hooykaas, P. J. J., Palme, K., & Offringa, R. (2004). A PINOID-dependent binary switch in apical-basal PIN polar targeting directs auxin efflux. *Science*, 306(5697), 862–865. <https://doi.org/10.1126/science.1100618>
- Girin, T., Paicu, T., Stephenson, P., Fuentes, S., Koerner, E., O'Brien, M., Sorefan, K., Wood, T. A., Balanza, V., Ferrandiz, C., Smyth, D. R., & Ostergaard, L. (2011). INDEHISCENT and SPATULA Interact to Specify Carpel and Valve Margin Tissue and Thus Promote Seed Dispersal in Arabidopsis. *Plant Cell*, 23(10), 3641–3653. <https://doi.org/10.1105/tpc.111.090944>
- Godoy, M., Franco-Zorrilla, J. M., Perez-Perez, J., Oliveros, J. C., Lorenzo, O., & Solano, R. (2011). Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors. *Plant Journal*, 66(4), 700–711. <https://doi.org/10.1111/j.1365-313X.2011.04519.x>
- 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(2), 221–224. <https://doi.org/10.1093/molbev/msp259>
- Gremski, K., Ditta, G., & Yanofsky, M. F. (2007). The HECATE genes regulate female reproductive tract development in Arabidopsis thaliana. *Development*, 134(20), 3593–3601. <https://doi.org/10.1242/dev.011510>
- Groszmann, M., Bylstra, Y., Lampugnani, E. R., & Smyth, D. R. (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(5), 1495–1508. <https://doi.org/10.1093/jxb/erq015>
- Groszmann, M., Paicu, T., Alvarez, J. P., Swain, S. M., & Smyth, D. R. (2011). SPATULA and ALCATRAZ, are partially redundant, functionally diverging bHLH genes required for Arabidopsis gynoecium and fruit development. *The Plant Journal*, 68(5), 816–829. <https://doi.org/10.1111/j.1365-313X.2011.04732.x>

- Groszmann, M., Paicu, T., & Smyth, D. R. (2008). Functional domains of SPATULA, a bHLH transcription factor involved in carpel and fruit development in Arabidopsis. *The Plant Journal*, *55*(1), 40–52. <https://doi.org/10.1111/j.1365-313X.2008.03469.x>
- Heisler, M. G. B., Atkinson, A., Bylstra, Y. H., Walsh, R., & Smyth, D. R. (2001). SPATULA, a gene that controls development of carpel margin tissues in Arabidopsis, encodes a bHLH protein. *Development*, *128*(7), 1089–1098.
- 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*(2), 252–261. <https://doi.org/10.1093/pcp/pcp184>
- Ines Ortiz-Ramirez, C., Giraldo, M. A., Ferrandiz, C., & Pabon-Mora, N. (2019). Expression and function of the bHLH genes ALCATRAZ and SPATULA in selected Solanaceae species. *Plant Journal*, *99*(4), 686–702. <https://doi.org/10.1111/tpj.14352>
- Ines Ortiz-Ramirez, C., Plata-Arboleda, S., & Pabon-Mora, N. (2018). Evolution of genes associated with gynoeceum patterning and fruit development in Solanaceae. *Annals of Botany*, *121*(6), 1211–1230. <https://doi.org/10.1093/aob/mcy007>
- Irepan Reyes-Olalde, J., Zuniga-Mayo, V. M., Serwatowska, J., Chavez Montes, R. A., Lozano-Sotomayor, P., Herrera-Ubaldo, H., Gonzalez-Aguilera, K. L., Ballester, P., Ripoll, J. J., Ezquer, I., Paolo, D., Heyl, A., Colombo, L., Yanofsky, M. F., Ferrandiz, C., Marsch-Martinez, N., & de Folter, S. (2017). The bHLH transcription factor SPATULA enables cytokinin signaling, and both activate auxin biosynthesis and transport genes at the medial domain of the gynoeceum. *Plos Genetics*, *13*(4), e1006726. <https://doi.org/10.1371/journal.pgen.1006726>
- Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Huguency, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., ... Wincker, P. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, *449*(7161), 463–U5. <https://doi.org/10.1038/nature06148>
- Josse, E.-M., Gan, Y., Bou-Torrent, J., Stewart, K. L., Gilday, A. D., Jeffree, C. E., Vaistij, F. E., Martinez-Garcia, J. F., Nagy, F., Graham, I. A., & Halliday, K. J. (2011). A DELLA in Disguise: SPATULA Restrains the Growth of the Developing Arabidopsis Seedling. *Plant Cell*, *23*(4), 1337–1351. <https://doi.org/10.1105/tpc.110.082594>

- Karimi, M., Bleys, A., Vanderhaeghen, R., & Hilson, P. (2007). Building blocks for plant gene assembly. *Plant Physiology*, *145*(4), 1183–1191. <https://doi.org/10.1104/pp.107.110411>
- Morel, P., Heijmans, K., Ament, K., Chopy, M., Trehin, C., Chambrier, P., Bento, S. R., Bimbo, A., & Vandenbussche, M. (2018). The Floral C-Lineage Genes Trigger Nectary Development in *Petunia* and *Arabidopsis*. *Plant Cell*, *30*(9), 2020–2037. <https://doi.org/10.1105/tpc.18.00425>
- Murat, F., Armero, A., Pont, C., Klopp, C., & Salse, J. (2017). Reconstructing the genome of the most recent common ancestor of flowering plants. *Nature Genetics*, *49*(4), 490+. <https://doi.org/10.1038/ng.3813>
- Nguyen, N. T. T., Contreras-Moreira, B., Castro-Mondragon, J. A., Santana-Garcia, W., Ossio, R., Robles-Espinoza, C. D., Bahin, M., Collombet, S., Vincens, P., Thieffry, D., van Helden, J., Medina-Rivera, A., & Thomas-Chollier, M. (2018). RSAT 2018: Regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Research*, *46*(W1), W209–W214. <https://doi.org/10.1093/nar/gky317>
- Pabon-Mora, N., Wong, G. K.-S., & Ambrose, B. A. (2014). Evolution of fruit development genes in flowering plants. *Frontiers in Plant Science*, *5*, 300. <https://doi.org/10.3389/fpls.2014.00300>
- Penfield, S., Josse, E. M., Kannangara, R., Gilday, A. D., Halliday, K. J., & Graham, I. A. (2005). Cold and light control seed germination through the bHLH transcription factor SPATULA. *Current Biology*, *15*(22), 1998–2006. <https://doi.org/10.1016/j.cub.2005.11.010>
- Pfannebecker, K. C., Lange, M., Rupp, O., & Becker, A. (2017). Seed Plant-Specific Gene Lineages Involved in Carpel Development. *Molecular Biology and Evolution*, *34*(4), 925–942. <https://doi.org/10.1093/molbev/msw297>
- Rajani, S., & Sundaresan, V. (2001). The *Arabidopsis* myc/bHLH gene ALCATRAZ enables cell separation in fruit dehiscence. *Current Biology*, *11*(24), 1914–1922. [https://doi.org/10.1016/S0960-9822\(01\)00593-0](https://doi.org/10.1016/S0960-9822(01)00593-0)
- Reymond, M. C., Brunoud, G., Chauvet, A., Martinez-Garcia, J. F., Martin-Magniette, M.-L., Moneger, F., & Scutt, C. P. (2012). A Light-Regulated Genetic Module Was Recruited to Carpel Development in *Arabidopsis* following a Structural Change to SPATULA. *Plant Cell*, *24*(7), 2812–2825. <https://doi.org/10.1105/tpc.112.097915>
- Vandenbussche, M., Janssen, A., Zethof, J., van Orsouw, N., Peters, J., van Eijk, M. J. T., Rijpkema, A. S., Schneiders, H., Santhanam, P., de Been, M., van Tunen, A., & Gerats, T. (2008).

Generation of a 3D indexed *Petunia* insertion database for reverse genetics. *Plant Journal*, 54(6), 1105–1114. <https://doi.org/10.1111/j.1365-313X.2008.03482.x>

Vialette-Guiraud, A. C. M., Alaux, M., Legeai, F., Finet, C., Chambrier, P., Brown, S. C., Chauvet, A., Magdalena, C., Rudall, P. J., & Scutt, C. P. (2011). *Cabomba* as a model for studies of early angiosperm evolution. *Annals of Botany*, 108(4), 589–598. <https://doi.org/10.1093/aob/mcr088>

Workman, C. T., Yin, Y. T., Corcoran, D. L., Ideker, T., Stormo, G. D., & Benos, P. V. (2005). EnoLOGOS: a versatile web tool for energy normalized sequence logos. *Nucleic Acids Research*, 33, W389–W392. <https://doi.org/10.1093/nar/gki439>

Zumajo-Cardona, C., Ambrose, B. A., & 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. <https://doi.org/10.1186/s13227-017-0068-8>

Table 1. Summary of high-scoring SPT-binding sites in putative target promoters.

		<i>Arabidopsis thaliana</i>		<i>Amborella trichopoda</i>		<i>Picea abies</i>	
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
		AT2G44910.1	4				
4	HDZip3	AT3G60390.1	1	AmTrV1scaffold00111.49	0	MA_57689g0010	0
		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
				AmTrV1scaffold17.24	1		
7	SAUR14	AT4G38840.1	2	AmTrV1scaffold00245.2	0	MA_10431311g0020	0
				AmTrV1scaffold00047.151	0		
8	BR6OX2	AT3G30180.1	0	AmTrV1scaffold00047.152	0	MA_31668g0010	2
				AmTrV1scaffold00013.244	12		
9	AFP3/4	AT3G02140.1	3	AmTrV1scaffold00154.36	1	MA_2575g0010	8
				AmTrV1scaffold00021.115	2		
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
				AmTrV1scaffold00021.257	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
				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 (p-values <0.05 underlined).

		Gynoecium fusion	%	<i>p</i> -value (comparison to <i>pSPT::SPT</i> )	Fruit size	%	<i>p</i> -value (comparison to <i>pSPT::SPT</i> )
<i>pSPT::SPT</i>	21	21	100	N/A	18	85.7	N/A
<i>pSPT::AtrSPT</i>	25	20	80	0.0536	11	44	<b><u>0.0054</u></b>
<i>pSPT::NcaSPT</i>	23	21	91.3	0.4894	18	78.3	0.701
<i>pSPT::PabSPT</i>	25	21	84	0.1142	0	0	<b><u>&lt;0.0001</u></b>
<i>pSPT::PaxSPT</i>	22	19	86.4	0.2326	4	18.2	<b><u>&lt;0.0001</u></b>
<i>pSPT::PIF5</i>	25	19	76	<b><u>0.0247</u></b>	13	52	<b><u>0.026</u></b>
<i>pSPT::ΔPIF5</i>	22	20	90.9	0.4884	13	59.1	0.0883
<i>pSPT::SmoPIF7/SPT</i>	25	12	48	<b><u>&lt;0.0001</u></b>	8	32	<b><u>0.0003</u></b>
<i>pSPT::SmoΔPIF7/SPT</i>	19	15	78.9	<b><u>0.0424</u></b>	6	31.6	<b><u>0.0009</u></b>
<i>pSPT::AthALC</i>	20	19	95	0.4878	16	80	0.6965
<i>pSPT::PaxALC</i>	16	16	87.5	1	13	81.3	1

Table 3. Gus reporter assays in transgenic Arabidopsis

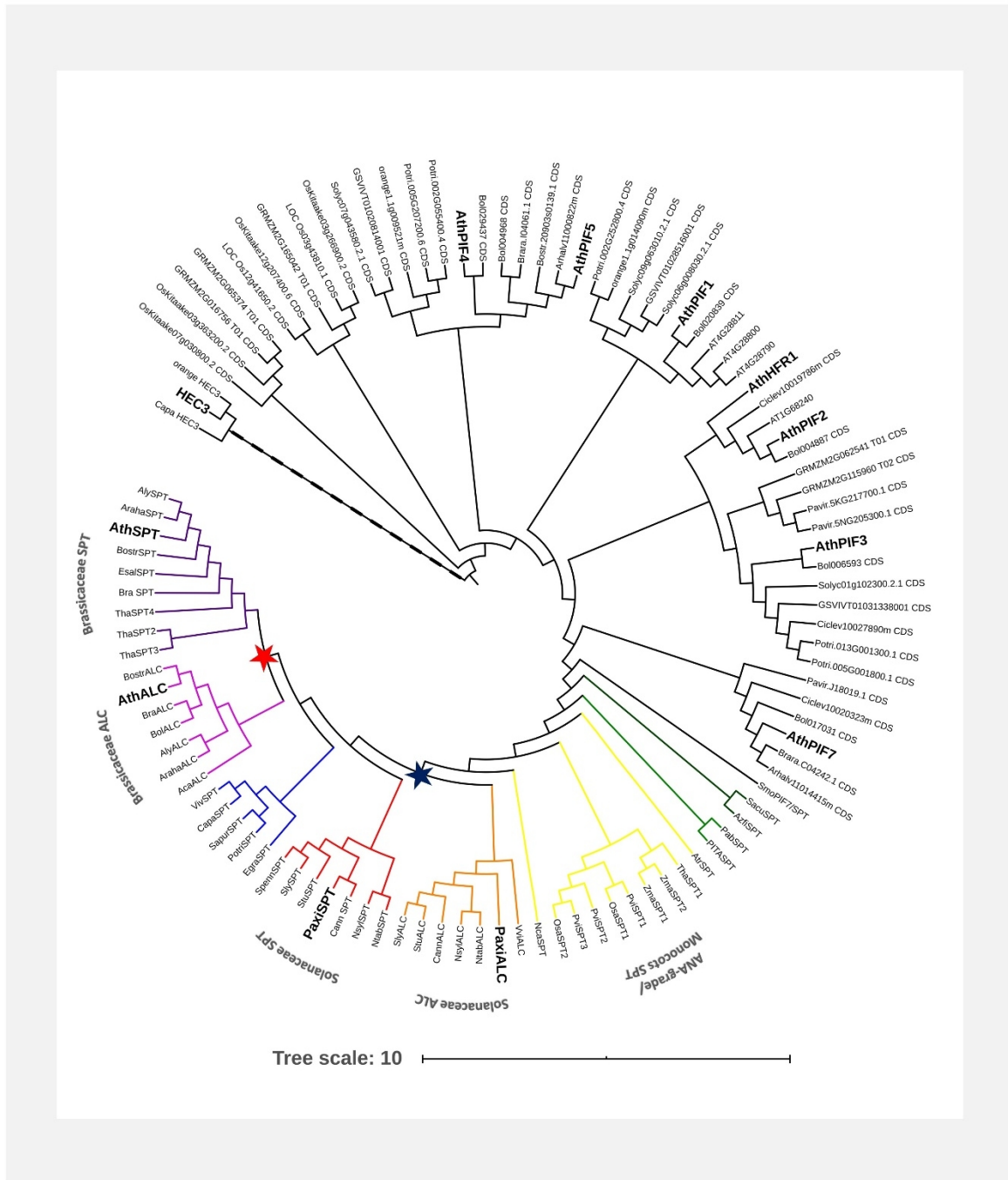
<b>Promoter fragment</b>	<b>No. of transformants</b>	<b>GUS +</b>	<b>Tissues showing GUS staining</b>
Ath pSPT_6253 bp (positive control)	17	14	sepals, petals, carpel margins, ovary wall, style, stigma
Nca pSPT_6277 bp	25	7	young floral bud, anthers
Pax pSPT_5006 bp	19	11	sepals, ovary wall
Pta pSPT_1659 bp	25	18	sepals, petals, base of pedicel

Table 4. Petunia transposon-mutant analyses.

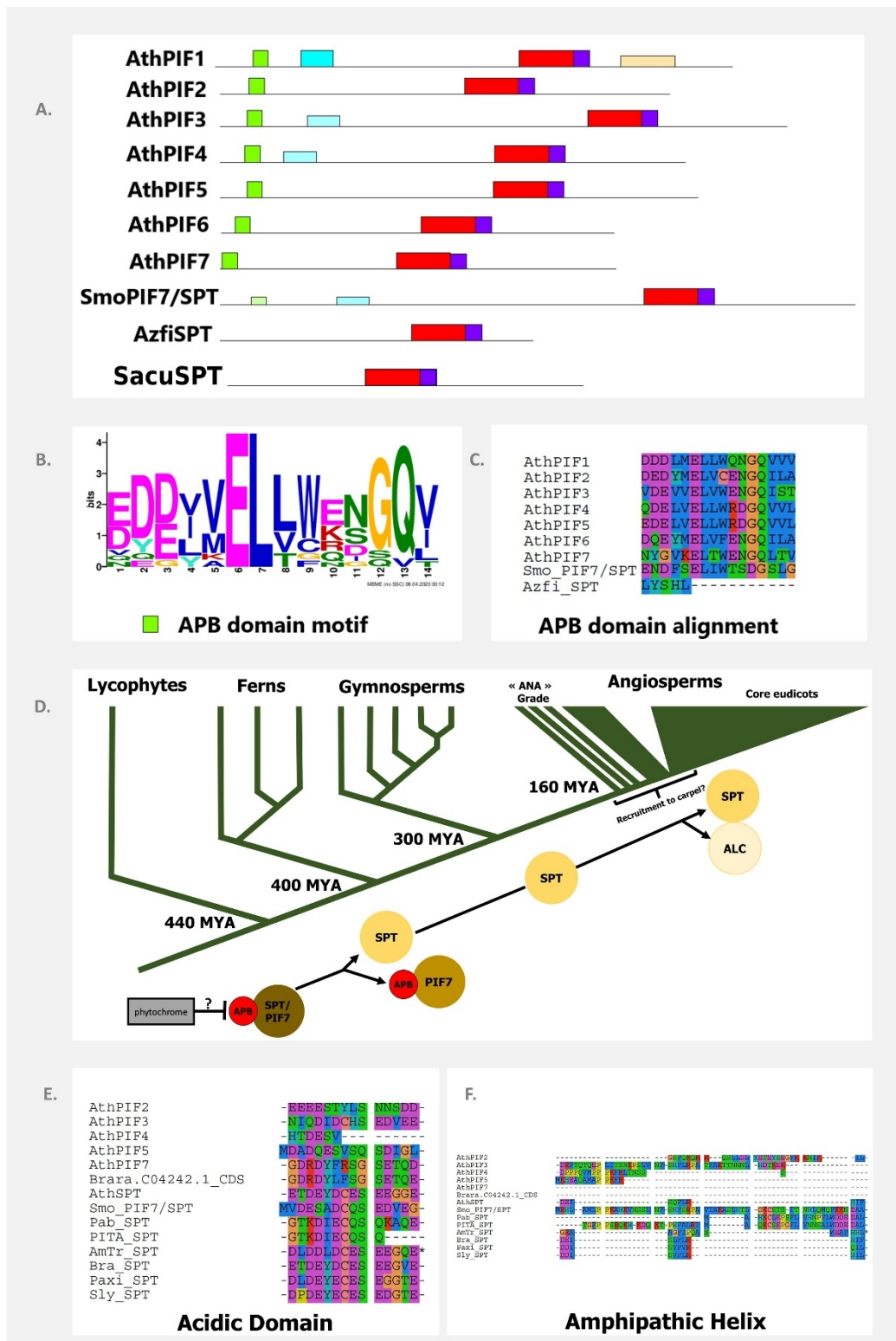
<b>Gene</b>	<b>Allele</b>	<b>Position of insertion in coding sequence.</b>
<i>PaxSPT</i>	<i>Pax_spt-1</i>	437 bp
<i>PaxSPT</i>	<i>Pax_spt-2</i>	18 bp
<i>PaxALC</i>	<i>Pax_alc-1</i>	252 bp
<i>PaxALC</i>	<i>Pax_alc-2</i>	267 bp

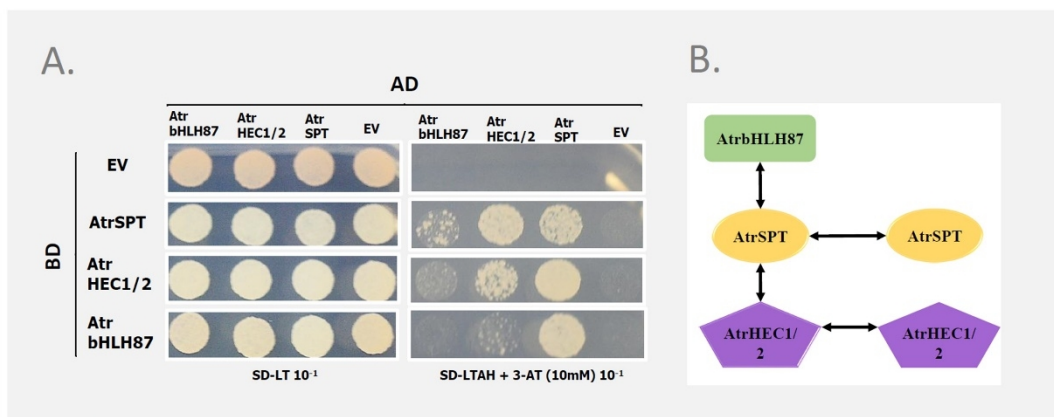
<b>Background</b>	<b>Allele combination</b>	<b>Mutant phenotype</b>
Mitchell	<i>Pax_spt-1/alc-1</i>	Stigmatic papillae absent
W138	<i>Pax_spt-1/alc-1</i>	None
Mitchell	<i>Pax_spt-2/alc-2</i>	None
W115	<i>Pax_spt-2/alc-2</i>	None



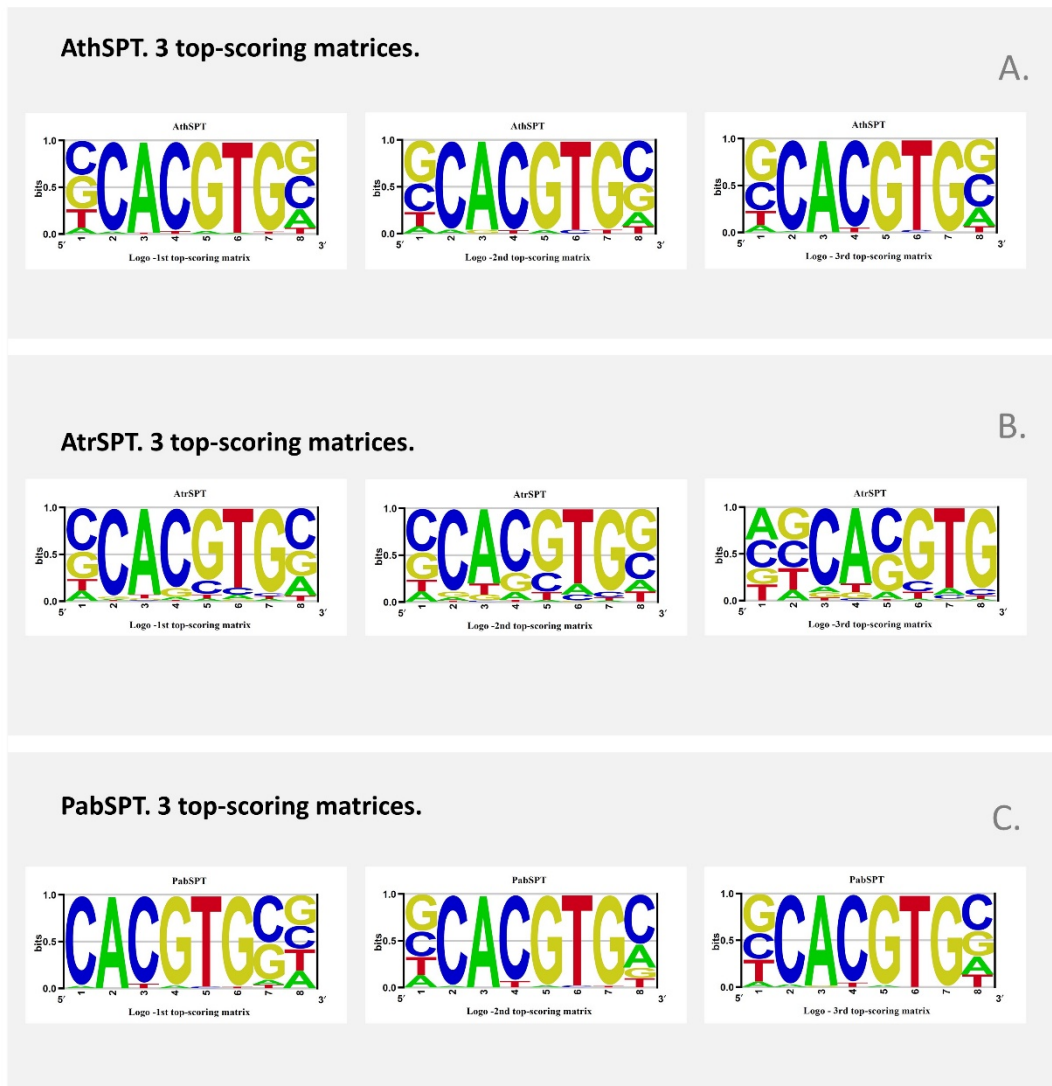
**Figure 1.** PhyML phylogeny of SPATULA/ALCATRAZ-like proteins within bHLH Subfamily 24. *Arabidopsis thaliana* HECATE3 (AthHEC3) and orthologues from bHLH Subfamily 31 were used as an outgroup. Two inferred duplications, referred to in the text, are shown at the base of core-eudicots (blue star) and within Brassicales (red star).



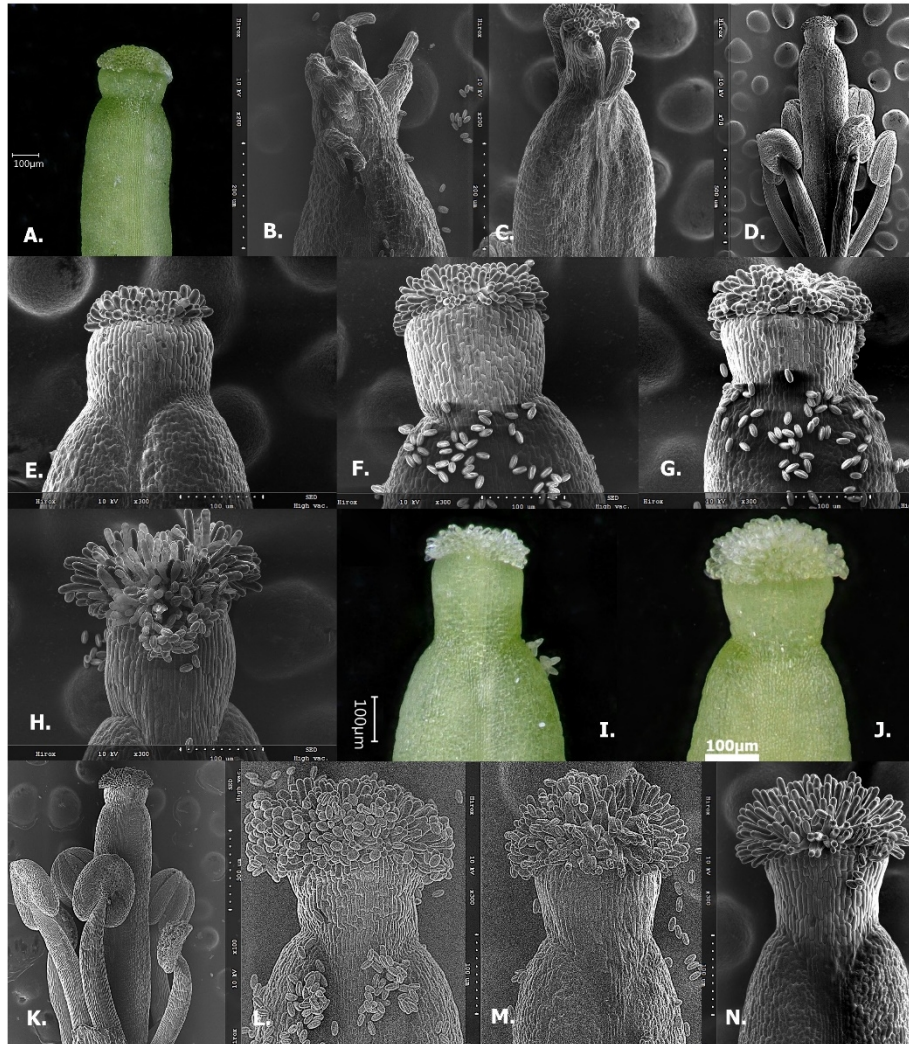
**Figure 2.** A. Conservation of the APB domain in PIF proteins and in SmoPIF7/SPT (red bars = BHLH domain, dark blue bars = beta strand, green bars = APB domain, light blue and beige bars = unspecified conserved motifs detected by MEME). The height and color intensity of the bars indicates the degree of conservation. B. The APB motif from a MEME analysis. C. Alignment of APB domains using MUSCLE. D. SPT/ALC lineage-evolution in land plants. E. and F. MUSCLE alignments of SPT domains based on E. Alignment of acidic domains previously identified core eudicot SPT/ALC by Groszmann et al. (2011). F. Alignment of amphipathic helices previously identified core eudicot SPT/ALC by Groszmann et al. (2011).



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



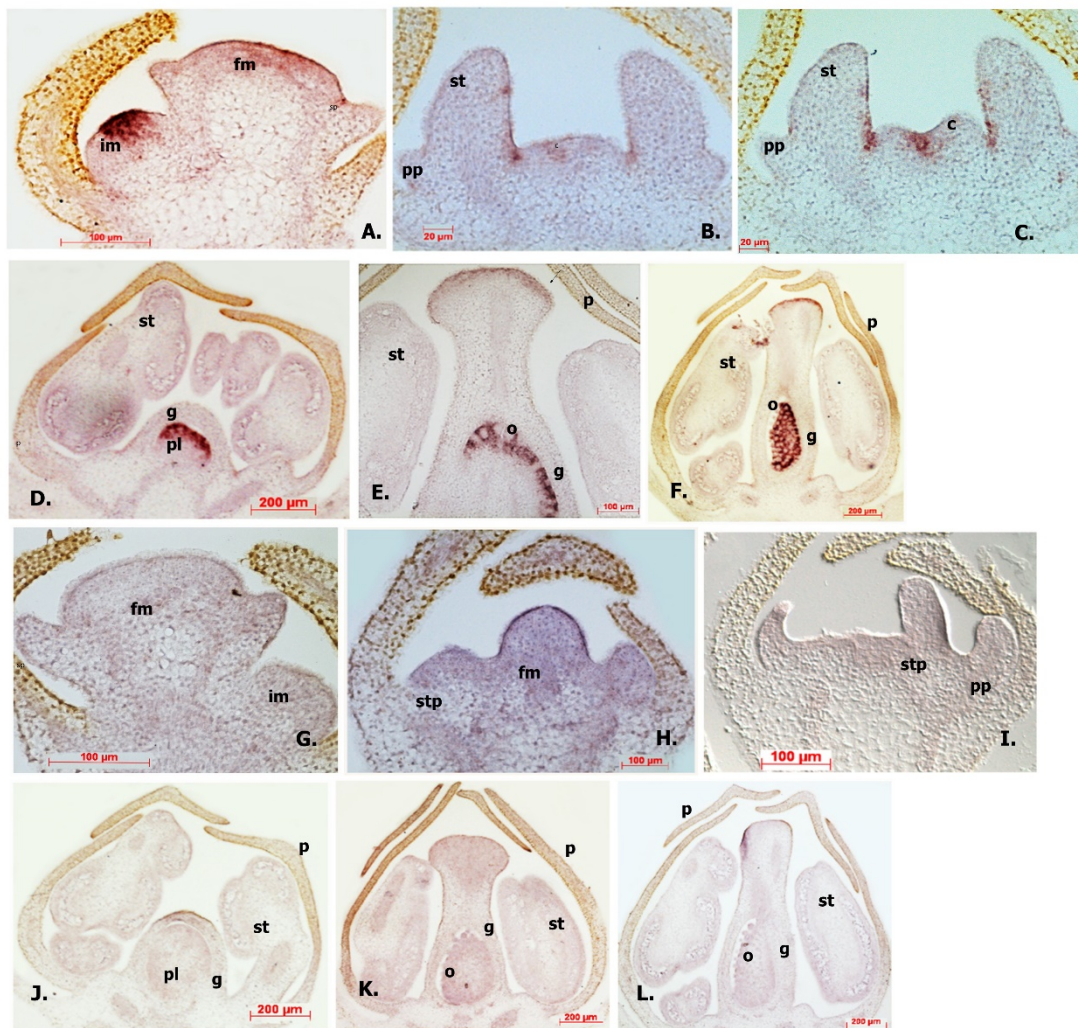
**Figure 4.** DNA-binding logos of SPT orthologues from *Arabidopsis thaliana*, *Amborella trichopoda* and *Picea abies*.



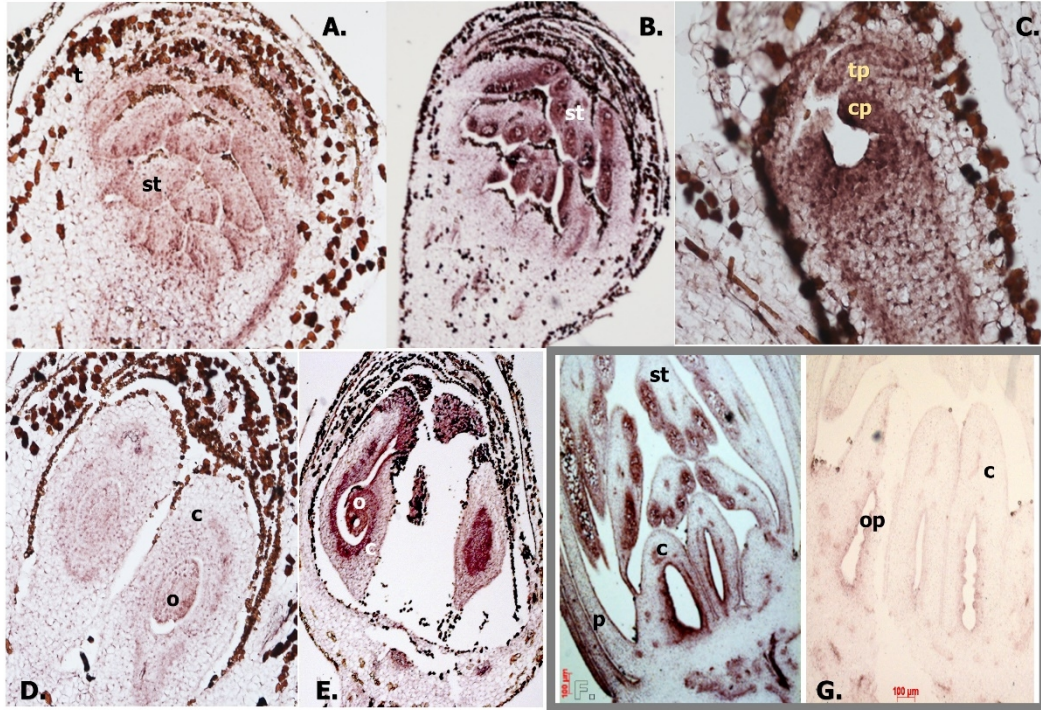
**Figure 5.** Complementation of *Arabidopsis thaliana spt-11* carpel-apex mutant phenotype using various bHLH coding sequences. **A** WT carpel apex. **B-C** *spt-11* style and stigma. **D** *spt-11* mutants transformed with *AthSPT* (positive control). **E – L** *spt-11* mutants showing fully restored carpel development following transformation with: **E** *pPST::AtrSPT*, **F** *pPST::NcaSPT*, **G** *pPST::PabSPT*, **H** *pPST::PaxSPT*, **I** *pPST::PIF5*, **J** *pPST::ΔPIF5*, **K** *pPST::SmoPIF7/SPT*, **L** *pPST::SmoΔPIF7/SPT*. **M** *pPST::AthALC*, **N** *pPST::PaxALC*



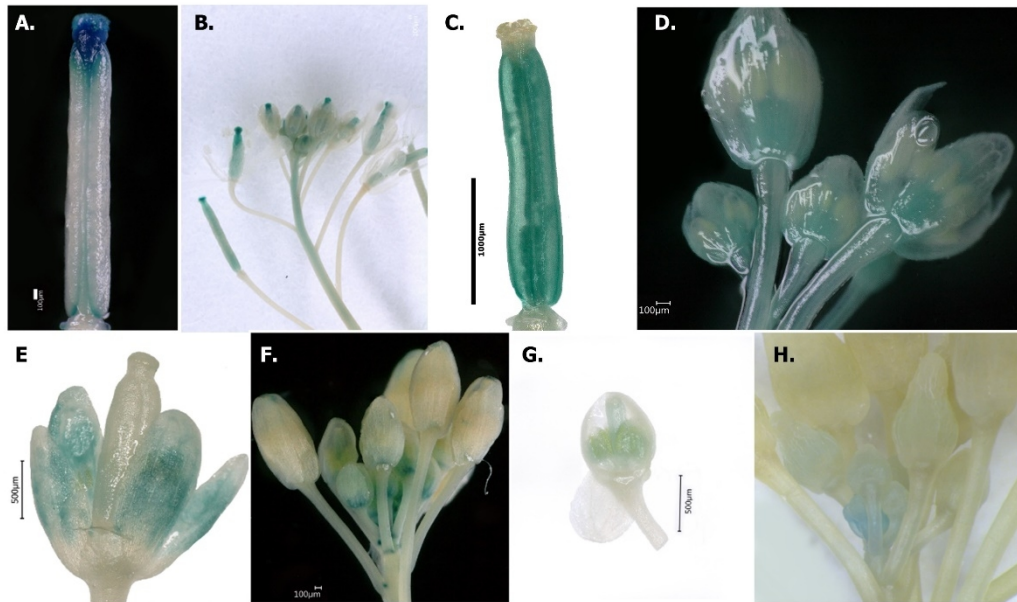
**Figure 6.** Complementation of *Arabidopsis thaliana spt-11* silique mutant phenotypes using various bHLH coding sequences: **A** AthSPT (positive control). **B – J** Completely restored phenotypes after transformation with: **B** *pPST::AtrSPT*, **C** *pSPT::NcaSPT*, **D** *pSPT::PaxSPT*. **E** full-length *pSPT::PIF5* **F** truncated *pSPT::ΔPIF5*. **G** full-length *pSPT::SmoSPT/PIF7*, **H** truncated *pSPT::SmoΔSPT/PIF7*. **I** *pSPT::AthALC*, **J** *pSPT::PaxALC*. **K** partially restored phenotype after transformation with *pSPT::PabSPT*. Untransformed siliques are shown at top-left. Scale bars: WT, A-E, I and J = 1000 μm; G, H and K = 200 μm; *spt-11* and F = 100 μm.



**Figure 7.** Expression of *PaxSPT* (A-F) and *PaxALC* (G-L) in longitudinal sections of *Petunia axilaris* floral buds. 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.

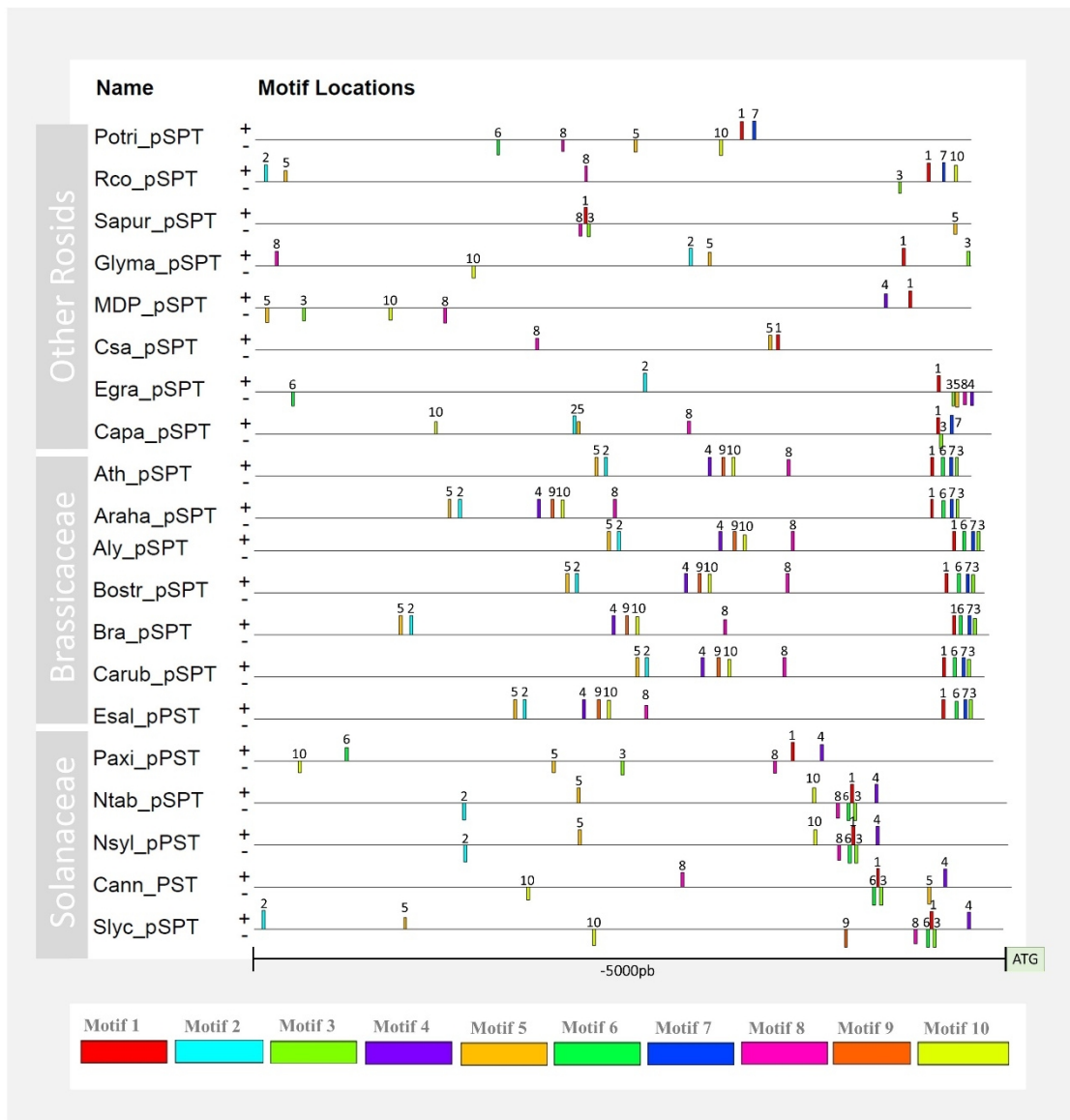


**Figure 8.** Expression of *SPT* orthologs from ANA-grade angiosperms. **A-B** *Amborella trichopoda* male flower buds, **C-E** *Amborella trichopoda* female flower buds, **F-G** *Nymphaea thermarum* flower buds. Key: c = carpel, cp = carpel primordium, o = ovule, op = ovule primordia, p = petal, st = stamen, sc = stigmatic crest t = tepal, tp = tepal primordium.

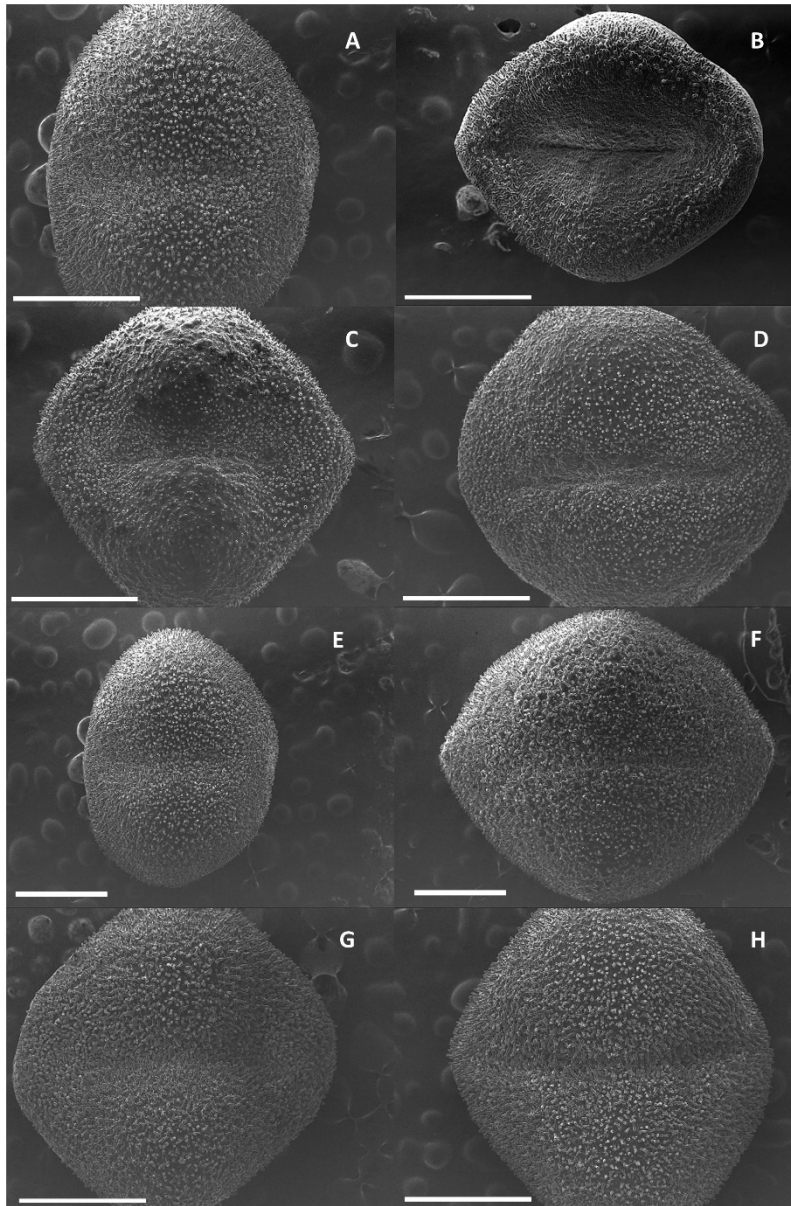


**Figure 9.** *Arabidopsis* transformed using GUS reporter constructs containing promoter sequences from *SPT* orthologs. **A and B** *Ath\_pSPT* from *A. thaliana* (positive control). **C and D** *Pax\_pSPT* from *Petunia axilaris*, **E and F** *Pta\_pSPT* from *Pinus taeda*, **G and H** *Nca\_pSPT* from *Nymphaea caerulea*.





**Figure 11. MEME analysis of upstream sequences from 20 core-eudicot SPT orthologs.** Of the 10 motifs searched, Motif 1 and Motif 8 were conserved on 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 axilaris*, **Ntab** = *Nicotiana tabacum*, **Nsyl** = *Nicotiana sylvestris*, **Cann** = *Capsicum annuum*, **Slyc** = *Solanum lycopersicum*.



**Figure 12.** Stigmatic surfaces of *Petunia hybrida* transposon-insertion mutants in genetic backgrounds Mitchell, W138 and W115 (WT on left, mutant on right). **A/B** Mitchell *spt-1 alc-1*. **C/D**: W138 *spt-2 alc-2*. **E/F**: Mitchell *spt-2 alc-2*. **G/H**: W115 *spt-1 alc-1*. Bars = 500 $\mu$ m.

## Supplementary information

### SI Tables

SI Tab 1. SPT orthologues, PIF-like and out-group references

SI Tab 2. High-scoring SPT binding sites revealed using RSAT MotifScan

SI Tab 3. Statistical analysis of SPT targets and target sites.

SI Tab 4. Petunia mutant screening primers

Gene	Allele	Primer Name	Sequence
PaxSPT	<i>Pax_spt-2</i>	MLY1423	TTAAtcctgtaatatattaaactgtcttc
		MLY0697	catcaggttcatgatcaggttcag
PaxALC	<i>Pax_alc-1/ Pax_alc-2</i>	MLY0699	CATACTCATCTAGATCATTGTCCATAG
		MLY0766	CTCGATTTTATCAACCGGAAGGTGT
PaxSPT	<i>Pax_spt-1</i>	MLY0767	CGAAATTATCAACCTTATCATCATCTAG
		MLY0768	GTACAGAGGATTTTGTAGTGGGAAGC
		MLY0769	TTCAGACAAATTGTGGACTTCAGC

SI Tab 5. Promoter amplification primers

Specie Name	Primer Name	Sequence
<i>Arabidopsis thaliana</i>	pSPT6253_1	CACCACTATTAAAAATTGGGAACA
	pSPT6253_2	TACACCAACAACAAAAAAAAAAGC
<i>Amborella trichopoda</i>	AtrpSPT6253 F	CACCCTCTCACCCCTACAAAACC
	AtrpSPT6253 R	GGTTTTATGGTGGCTATTGTTG
<i>Nymphaea caerulea</i>	SP/pSPTNymp_1401/F	CACCGGTGGCGGATTCCGGCAA
	ASP/pSPTNymp_1401/R	GACGACTCCGCATCGGAA
	SP/pSPTNymp_6277/F	CACCTTATTCGATGAACTATGA
<i>Petunia axillaris</i>	ASP/pSPTNymp_6277/R	GACGACTCCGCATCGGAAA
	pPax450_5006/F	CACCCAAGATAACTGTTGGAAA
	Paxi450_5006_2	AATATTTTTCACAAAAACAAGAC
<i>Pinus taeda</i>	1.2pSPTPita_for	CACCAAGCGCTATATGATTTATATTATG
	ASP/pPita_6275/R	TTATTGGTAACCGTCTGTG

SI Tab 6. CDS amplification primers

Specie Name	Gene	Primer Name	Sequence	
<i>Arabidopsis thaliana</i>	SPT	AthSPT_F1	CACCATGATATCACAGAGAGAAGA	
		AthSPT_R1	TCAAGTAATTCGATCTTTTAG	
	ALC	TOPO SP/ALC/F	CACCATGGGTGATTCTGACGTC	
		TOPO ASP/ALC/R	TCAAAGCAGAGTGGCTGT	
	PIF5	AthPI5_F	CACCATGGAACAAGTGTGCTGATTGGA	
		AthPI5_R	TCAGCCTATTTTACCCATATGAAGAC	
	$\Delta$ PIF5	AthPI5_F_delta	CACCATGGATGGTCAAGTGGTTTTACAAAGCCA	
		AthPI5_R	TCAGCCTATTTTACCCATATGAAGAC	
	<i>Amborella trichopoda</i>	SPT	AtrSPT_F1	CACCATGCCCTTCTCCCCCTCTTCTTCCT
			AtrSPT_R1	CTACGTAATTTGCGACCTAGCAGCAGCTTG
<i>Nymphaea caerulea</i>	SPT	NewNceSPT_For	CACCATGGCGGAGATCTACGCCTCCCCGC	
		NewNceSPT_Rev	TTAGGCATCCCTGTTGGTTGAATAGAAGG	
<i>Petunia axillaris</i>	SPT	TOPO SP/PaxSPT/F	CACCATGGCTAATAACAATATG	
		TOPO ASP/PaxSPT/R	ATACAAATGTGGGGCAAG	
	ALC	TOPO SP/Pax_ALC/F	CACCATGGCGGATCCATACAAATCAAC	
		TOPO ASP/Pax_ALC/R	CTAAGAAGACACCCCAGA	
<i>Picea abies</i>	SPT	PabSPT_F1	CACCATGGCCCATAAATGTTTA	
		PabSPT_R1	TTATTTCTCTTGCAAGGA	
		SmoSPT/PIF7_F	CACCATGAATCATTGTGTTCCCGACTGCGA	
<i>Selaginella moellendorffii</i>	PIF7/SPT	SmoSPT/PIF7_R	CTATTGCATAACAACAACCTCCACCCGT	
		SmoSPT/PIF7_F_delta	CACCATGAATTCGCAGTGGGTCTCTCAGGAATC	
	$\Delta$ PIF7/SPT	SmoSPT/PIF7_R	CTATTGCATAACAACAACCTCCACCCGT	

## **SI Figures**

SI Fig 1. amino acid alignment used in phylogeny

SI Fig 2. Nine PWMs as a .txt file

Conserved regions defined by Groszmann et al., (2010) underlined

Region 6

TTATCTCGACACAACCTCTAATTCAGTTTGACCGGACTGAACACTTTTATCAATATTAATGTTGAGTTGTT  
 GACCTATATATTATTGTAATAATGAATAAACCAACGCCATGATTGCGTACTATCTATCAGACTTTATAGAG  
 TAACTACTTTGATTAACCATTACATTTTGTTTTATGAGAATGTTTTTCAGTTATATATCTTTATCAACAC  
 AGTTGAAACCTGTTCTTCTATGTTAAAGCCTTTGAATTTGAGTAATTTTACTTTTGATAGCTAGCCTTTT  
 TAAAATGAATGATTATATTGTTAGGTTCCAAAGTGGGGAGTTGCTTGGGGATCTCATTGACGTTGATC  
 TTGAATTTTGATAGAGATCCTAATGATCTCGTGGTCCCTCTCGCAGAGTTTATGTAACCATCTAATATA  
 TAACCTTCATGCTCGGATATGGAGAAATCCAAAACCTTAAGAAATGCACGTATTAGCATTGCTAGTTGTATAT  
 ATGTTATGATTTTTGCGTCAAATGCAGAACAAAATAAAGCAAGAAGAGACCAACAATCTTCATATGGAGA  
 ATTGTAATTTGCCAAATCAGCAATATATAGATGCTTGAACAATAAAAAGATAACCGAGTTTAACATTA  
 GACCAAAAAAAGAAAGATAACCGCTACAGAATTTGTGACGAAAAGAAAAGCTTTGGAGAAAAAAC  
 AGAGTAAGAACTGTTTTTGTGTGTGAGAAAGATTTGTCCTTAGGTCGCATAATCTTTTAGGATGTAGTT  
 ATATGTGTTAAGTGTAACACATGCATATCATTACTTCTGCAATATATCTCTTCGAGTTTCGTAAGGT  
 ACTTATGTTTGGTTGTCTTCGAACTACCAACGAACAACCTTAAAAAACATTGATCTTTTAAATAGCGTG  
 AGTTTGATTTATGTTATCCTTGTATGTTGATGATGATTACGGAAACTTTGGTCTGTTTTAGATGGGTA  
 AATAGTTGTATATGAGTAAATATGATATTGCCAGTGTTTATATAAAGAAAAGGGAAACGATGATAAAA  
 GGTGAAAAAGAAGTTTAGGATCATCTTTGTTTTTATTTTGTTCGACTTTTCAATCAGACAAGACAA  
 CGAGGTATGGGTCTTTAACATACAGAGACTCAACTTGAAAAATCTATCAAACCACTTAAAAAACCA  
 GAGAAAACATCAACTATCTTTTGAATCTAGACAAAATAAAAAGTGTATCATGCTGATTTTGAACAAA  
 TTATTAAGAGATCTCAACCATCGTTTCATTAATATCTTTGACAATCTAGAACCGGTAGTAAATCTTGT  
 CTTTTTATGTGTATATCTTGTGAAATTTGATAAGGATATCAAATAATGAGATATGATGTTGAGTAAT  
 TGAGTAAATCAGATCCCTTGTTTAATTTGACGATTAAACGCTTTTTTCATTATTTTTTTTCGATTTTG  
 TTGTGTATGTTCTCAATTACAAATCCACTATAAGCATTGGACTCTACAATGTCATTTCTAATTTCTGA  
 TATTTAATAAACAATTTTGTATATGATAAATAGTACCAGAAAAACGAACCTAATAAATAAATTTGA  
 ATCAAAAGCAGATGCTGAACAAATAGCTGCAGCTTTTAGCAATTTGCTCCCAATATGCTTTTCTTTTGT  
 CAAAAATGTTATATACGAGGTTAAGTGAATTTGCAATGAAATAAATGATGCAGTTTGTGCATTTCATCAA  
 GCGAACAAAGTACGATTGATGTTTTTCAGTGCAAAACTAAAAATAAATACTACTAACAACATATATAGTGA  
 AATATATATAACTTTAACTTCTTTTTTTTTAAGTCTTCATACGCTGCGAACGATCAAAATATATTTACAA  
 ATTTACGACCAATATTTAAAAATACTTCATTAAGAATTTGATGCTGCTGTCTGCACAGTTGAGAAGCA  
 ATTTGAAACTTTTCGTGCATTCATTTGTTCCGCGCAAATGGATATGTGAACGTTGTCATCATTATTAT  
 AGAATTTCTATAAAAACCTTACTATAAAAACCTAGATATAAAGCTGGATCATATAAAAATGAATTTACTAA  
 AAATTCGATAGTTAATTTATAGCAGATGTTATGTTCCAATTTGAAAGCATAACGACGAGGTATATAGTTGA  
 AAACACACACCGTAAAGTTAATAATTTTCAACAACACAAAGCAAAATCAAAGTCCGCAAGTAATTTAAGCA  
 TGTCAGTAGCATGGGCTTCTTATATACTAGTCAATAGAATAAACTAGCAAAACAAGCAACTAATGTATTCT  
 TGTTTATCAGCCACGGGTTACAATATCATACAAAATTTGAATACTAATTAATGGTAACAAGTAAAAAAA  
 CAAATTAAGTAAATGGAATACTTTTGTCAAACAACCAAGACGTATAACTTTTCGTTTTCTATAGATTAA  
 TGGACTTCTTAAAAATCTCCTCATCAGATTAACCTTTGAGATATACAAATACAGTTTTTGTTCCTTCTA  
 AATGATATGAATATTAACCTTTATCGATTTTCATCCGTAGCAGATTTCCATTTAAATAATAAACTATGAGA  
 AAAACAGATAAAGGTTGTATATTATTGTTACCCCAAAAAAATAAATAAATAAATAAATAAATAAATAAATA  
 TGTGTTCTCACGTCGACAGGAGGAAAAGTTTTGGGAGAGTAAAAACATTTAATATTACGACTAGTTTGA  
 AACCGTGAGCTGACACAAGCTCATTGCTAATGCTACAGTAACAGCTACCTTCACTTTTAACTAAATGACA  
 GACCAATCAATTTAACCTCTGTTTTCTTAGCTGGCGGCTGACAGACACTCTCCCTCTCCATGCCATA  
 AAATCTCAAAGACTGTTTTAAAAAATAA\*ATGTTTTAGCTTTAACTGCTTTTTTTTTGTTGTTGGTGA

Motifs identified in this work, shown in red

Motif 5

Motif 2

Motif 4

Motif 9

Motif 10

Region 5

Motif 8

Region 4

Region 3

Region 2

Motif 1

Region 1

Motif 6

Motif 7

Motif 3

↑  
Alternative translational start (+1)

↑  
Translational start (+40)

SI Fig. 3. Conserved regions and motifs in core-eudicot *SPT* promoters, from Groszmann et al. (2010) and the present study



# **GENERAL DISCUSSION AND PERSPECTIVES**

---

## Discussion and perspectives

### Identification of genetic modules conserved between two divergent species for evo-devo studies of angiosperms.

The main goal of this thesis has been to study the molecular basis of carpel evolution. In the first of the manuscripts included in the thesis, I identified and compared genetic modules from two key species for the study of angiosperm evolution: the basal angiosperm *Amborella trichopoda* and the model angiosperm *Arabidopsis thaliana*. We showed that it was possible to compare the contents and expression patterns of genetic modules between these very distant species and identify modules that have been conserved since the evolutionary separation of the two lineages concerned. The methodology that we developed in this pilot study shows promise, and could perhaps be developed further to make a valuable contribution to macroevolutionary studies of the origin of the carpel and other evo-devo topics.

We identified in this study the importance of maximizing the compatibility of datasets to be compared, and we also highlighted difficulties in selecting equivalent samples from very diverse species due both to uncertain homology between organ systems and to heterochrony. Though this type of analysis can be carried out, as in the present work, using a mixture of data obtained from diverse studies, it may frequently be better to use custom data, specifically obtained for the purpose, by applying equivalent sampling and bioinformatics procedures to all species to be compared, as was done in the study of Kivivirta et al. (2019). To extend our present pilot study further, it may be useful to include in future analyses a wider range of flower developmental stages for each species, and perhaps also to include more species in the analysis, conditionally on the availability of the appropriate sequenced genomes.

Using such an approach, we envisage that it may be possible to track changes in GRNs across angiosperm evolution and determine the composition of the carpel GRN at any node of interest in angiosperm phylogeny. It would also be useful to extend this type of analysis to gymnosperms. Finally, as the molecular regulation of flower development in distantly related species will show differences related to different types of anatomical organization, it will be important to include histological analyses of flowers from the chosen species to gain full insight into the evolution of the genetic modules which regulate floral organ development.

### The role of *SPATULA* in the gynoecium and its evolution in angiosperms

*SPATULA* may have played a role in the female reproductive tissues of the MRCA of living angiosperms, as suggested by the conservation of strong *SPT* expression in the placentae and ovule primordia of the ANA-grade species *Amborella tricopoda* and *Nymphaea thermarum*, as well as in numerous core-eudicot models such as *Arabidopsis thaliana* and *Petunia* spp. However, none of the core-eudicot models so far investigated have revealed a function associated with this very clear and highly conserved peak of gene expression in ovules and placentae. At present, we lack functional approaches to perform the equivalent experiments in ANA-grade species. The probability is not high of developing in the near future functional genetics approaches in *Amborella*. By contrast, *N. thermarum* offers many advantages as a potential functional-genetics model. The recent publication of the *N. thermarum* genome (Povilus et al., 2020) should facilitate the task of genetically modifying this species, maybe using CRISPR-Cas9 technologies (Mao et al., 2019).

The current role of *SPT* in tissues such as the style, stigma and transmitting tract of *A. thaliana* seems to be a recent acquisition, probably near the base of Brassicaceae. Based on the results obtained in the present work, we can suggest that the acquisition of this role by *SPT* occurred through changes to its expression pattern, probably at least in part brought about by changes to the *SPT* promoter itself. In our promoter analyses, we identified ten motifs grouped in highly conserved regions within Brassicaceae *SPT* promoters, most of which coincide with regions of high sequence conservation previously reported by Grozsmann et al. (2010) (SI Figure 3). Interestingly, seven of these ten motifs are present also in *SPT* promoters of Solanaceae species. However, in Solanaceae species, these conserved motifs are located at very variable distances from the start codon, compared to the arrangement in Brassicaceae, and several of them do not occur in the same order or close juxtaposition as they do in Brassicaceae spp. Similar distinctions in *SPT*-promoter structure are noted when comparing other rosids with Brassicaceae.

To test the functional significance of the above observations, it would be very interesting to make native and rearranged versions of the *SPT* promoters from a range of angiosperm species such as *Nymphaea thermarum* (ANA-grade, Nymphaeales, Nymphaeaceae), *Petunia axillaris* (core-eudicots, asterids, Solanaceae) and *Carica papaya* (core-eudicots, rosids, Caricaceae), which show varying levels of similarity in *SPT*-promoter structure with *Arabidopsis*. Such experiments could identify any precise changes in promoter structure which were linked to the acquisition by *Arabidopsis* *SPT* of its current role in style, stigma and replum tissues.

Although VIGS approaches have not identified any clear phenotypes associated with the loss of *SPT* function in Solanaceae species (Ortiz-Ramírez et al., 2019), it would be interesting to continue with

the analysis of mutants in *Petunia*, begun in the present study, perhaps by adding other genetic backgrounds and testing all allelic combinations in each of them, in order to reliably confirm the absence of a carpel phenotype linked to *SPT* and *ALC* functions.

## **The position of *SPATULA* orthologs in gene-expression modules identified using WGCNA.**

The second manuscript included in this thesis focused on *SPT*, and so it is logical to ask where *SPT* orthologs were to be found in the results obtained from the analysis of gene-expression modules performed in the first manuscript of this thesis. For reasons that we do not know, *Arabidopsis thaliana* *SPT*, its paralog *AthALC*, and their mutual pro-ortholog *AtrSPT* (AM\_TR\_V1\_scaffold00046.26) from *Amborella trichopoda*, were not placed in an orthogroup in the global analysis of gene orthology performed in PHYTOZOME (SI Tab 1 of Manuscript/Part 1 of this thesis). However, these three genes were present in the *Arabidopsis* and *Amborella* transcriptome data analyzed in our WGCNA gene-clustering procedures. According to these analyses, *AthSPT* (AT4G36390) and *AthALC* (AT5G67110) were placed by WGCNA in the Ath\_red and Ath\_blue modules, respectively, while *AtrSPT* was placed in the Atr\_turquoise module (Manuscript 1, SI Tab 3). Subsequent statistical analyses using the hypergeometric test gave upper-tail (for positive correlations) p-values of p=0.92 and p=0.63 for comparisons of Atr\_turquoise with Ath\_red and Ath\_blue, respectively (Manuscript 1, SI Tab 6). This means there was no significant positive correlation between either of the modules in *Arabidopsis* that contained *AthSPT* and *AthALC* with the *Amborella* module that contained *AtrSPT*. This observation is perhaps unsurprising as the expression data presented in the second manuscript of this thesis (Manuscript/Part 2, Fig. 8) shows that *AtrSPT*'s expression in *Amborella* flowers differs in many respects from the published expression profiles of both of its orthologs in *Arabidopsis* (Groszmann et al., 2010, 2011; Heisler et al., 2001). This result from gene-clustering (Manuscript 1) is thus consistent with the conclusion of Manuscript 2 that the *SPT* lineage adopted novel expression profiles and functions in a common ancestor of Brassicaceae, long after the divergence of the *Arabidopsis* and *Amborella* lineages in the MRCA of living angiosperms.

## The future of carpel evolution studies

To continue evolutionary studies of the carpel and floral organs in general, it will be interesting to concentrate efforts in species that occupy key phylogenetic positions such as *Nymphaea thermarum*, which additionally shows potential as a possible functional-genetic model. If stable transform of *N. thermarum* becomes possible, an interesting approach to test would be Interspecies Gene Transfer (IGT) (Nikolov & Tsiantis, 2015). In this method, entire genetic loci of interest, including non-coding regulatory elements, are transferred between species. In this way, both regulatory and coding sequences can be analyzed to study molecular changes that lead to evolutionary events. Beyond the genetic transformation of new model species, we could also use transcriptomic approaches to study the regulatory network of flower development in a range of both model and non-model species. The laser-capture microdissection method used in Manuscript 1 of the present work, followed by RNA-seq, and/or single-cell RNA sequencing techniques, could also be used in *N. thermarum* and other species occupying key phylogenetic positions.

Studies of individual genes (rather than transcriptomics approaches) will also be very important to further elucidate the molecular basis of floral-organ evolution. In the case of the carpel, there are many further candidate genes to be analyzed, such as the YABBY gene *CRABS CLAW (CRC)*, which has a carpel-specific expression profile in *A. thaliana* (J. Alvarez & Smyth, 1999; John Alvarez & Smyth, 2002). YABBY members are expressed in the abaxial side of lateral organs (Siegfried et al., 1999). In *A. thaliana*, *CRC* is expressed specifically in the abaxial domain of the carpel wall (John Alvarez & Smyth, 2002), and previous work shows that *CRC* conserves this expression pattern in ANA-grade angiosperms (Chloe Fourquin et al., 2007; Chloé Fourquin et al., 2005), and was most probably part of the carpel gene regulatory network of the MRCA of flowering plants (Pfannebecker et al., 2017b). It furthermore seems that the origin of the *CRC* lineage coincides with the origin of angiosperms, after the Epsilon whole genome duplication event (Becker, 2020).

However, there are still unanswered questions surrounding *CRC* and its contribution to carpel origin and evolution. For instance: which other YABBY lineage is paralogous to *CRC*, what confers on *CRC* its specificities in carpel development, and how were these acquired? Further pertinent questions include: what are *CRC*'s direct target genes, and how were these recruited? Similar questions can be asked of many further genes of importance to the carpel GRN. Clearly, as we continue in the evolutionary study of flowering plants, integrative approaches will be necessary to better comprehend the contribution of each candidate gene to carpel origin and evolution. These approaches will need to

integrate many levels of study, from morphological analyses and structural biology to the epigenetic mechanisms that operate on gene expression to bring about developmental processes.

# REFERENCES

---

## References

- Alvarez, J., & Smyth, D. R. (1999). CRABS CLAW and SPATULA, two Arabidopsis genes that control carpel development in parallel with AGAMOUS. *Development*, 126(11), 2377–2386.
- Alvarez, John, & Smyth, D. R. (2002). CRABS CLAW and SPATULA Genes Regulate Growth and Pattern Formation during Gynoecium Development in Arabidopsis thaliana. *International Journal of Plant Sciences*, 163(1), 17–41. <https://doi.org/10.1086/324178>
- Andres-Robin, A., Reymond, M. C., Dupire, A., Battu, V., Dubrulle, N., Mouille, G., Lefebvre, V., Pelloux, J., Boudaoud, A., Traas, J., Scutt, C. P., & Monéger, F. (2018). Evidence for the Regulation of Gynoecium Morphogenesis by ETTIN via Cell Wall Dynamics. *Plant Physiology*, 178(3), 1222–1232. <https://doi.org/10.1104/pp.18.00745>
- Becker, A. (2020). A molecular update on the origin of the carpel. *Current Opinion in Plant Biology*, 53, 15–22. <https://doi.org/10.1016/j.pbi.2019.08.009>
- Becker, A., & Ehlers, K. (2016). Arabidopsis flower development—Of protein complexes, targets, and transport. *Protoplasma*, 253(2), 219–230. <https://doi.org/10.1007/s00709-015-0812-7>
- Bernhardt, P., Sage, T., Weston, P., Azuma, H., Lam, M., Thien, L. B., & Bruhl, J. (2003). The pollination of *Trimenia moorei* (Trimeniaceae): Floral volatiles, insect/wind pollen vectors and stigmatic self-incompatibility in a basal angiosperm. *Annals of Botany*, 92(3), 445–458. <https://doi.org/10.1093/aob/mcg157>
- Carvalho, M. de A., Lana, C. C., Bengtson, P., & Sa, N. de P. (2017). Late Aptian (Cretaceous) climate changes in northeastern Brazil: A reconstruction based on indicator species analysis (IndVal). *Palaeogeography Palaeoclimatology Palaeoecology*, 485, 543–560. <https://doi.org/10.1016/j.palaeo.2017.07.011>
- COEN, E., & MEYEROWITZ, E. (1991). THE WAR OF THE WHORLS - GENETIC INTERACTIONS CONTROLLING FLOWER DEVELOPMENT. *Nature*, 353(6339), 31–37. <https://doi.org/10.1038/353031a0>

- Coiro, M., Doyle, J. A., & Hilton, J. (2019). How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *New Phytologist*, *223*(1), 83–99.  
<https://doi.org/10.1111/nph.15708>
- COLOMBO, L., FRANKEN, J., KOETJE, E., VANWENT, J., DONS, H., ANGENENT, G., & VANTUNEN, A. (1995). THE PETUNIA MADS BOX GENE FBP11 DETERMINES OVULE IDENTITY. *Plant Cell*, *7*(11), 1859–1868. <https://doi.org/10.1105/tpc.7.11.1859>
- de Boer, H. J., Eppinga, M. B., Wassen, M. J., & Dekker, S. C. (2012). A critical transition in leaf evolution facilitated the Cretaceous angiosperm revolution. *Nature Communications*, *3*(1), 1221. <https://doi.org/10.1038/ncomms2217>
- De-Paula, O. C., Assis, L. C. S., & Craene, L. P. R. de. (2018). Unbuttoning the Ancestral Flower of Angiosperms. *Trends in Plant Science*, *23*(7), 551–554.  
<https://doi.org/10.1016/j.tplants.2018.05.006>
- Doyle, J. A. (1999). The rise of angiosperms as seen in the African Cretaceous pollen record. In K. Heine, L. Scott, A. Cadman, & R. Verhoeven (Eds.), *Palaeoecology of Africa and the Surrounding Islands, Vol. 26* (Vol. 26, pp. 3–29).
- Doyle, James A. (2008). Integrating Molecular Phylogenetic and Paleobotanical Evidence on Origin of the Flower. *International Journal of Plant Sciences*, *169*(7), 816–843.  
<https://doi.org/10.1086/589887>
- Doyle, James A. (2012). Molecular and Fossil Evidence on the Origin of Angiosperms. *Annual Review of Earth and Planetary Sciences*, *40*(1), 301–326. <https://doi.org/10.1146/annurev-earth-042711-105313>
- Doyle, James A., Jardiné, S., & Doerenkamp, A. (n.d.). Afropollis, a new genus of early angiosperm pollen, with notes of the Cretaceous palynostratigraphy and paleoenvironments of Northern Gondwana. *Bulletin Des Centres de Recherches Exploration-Production Elf-Aquitaine*, *6*, 39–117.
- Endress, P. K. (2011). Evolutionary diversification of the flowers in angiosperms. *American Journal of Botany*, *98*(3), 370–396. <https://doi.org/10.3732/ajb.1000299>

- Endress, P. K., & Doyle, J. A. (2007). Floral phyllotaxis in basal angiosperms: Development and evolution. *Current Opinion in Plant Biology*, *10*(1), 52–57.  
<https://doi.org/10.1016/j.pbi.2006.11.007>
- Endress, P. K., & Doyle, J. A. (2009). Reconstructing the ancestral angiosperm flower and its initial specializations. *American Journal of Botany*, *96*(1), 22–66.  
<https://doi.org/10.3732/ajb.0800047>
- Endress, P. K., & Doyle, J. A. (2015). Ancestral traits and specializations in the flowers of the basal grade of living angiosperms. *TAXON*, *64*(6), 1093–1116. <https://doi.org/10.12705/646.1>
- Fourquin, Chloe, Vinauger-Douard, M., Chambrier, P., Berne-Dedieu, A., & Scutt, C. P. (2007). Functional Conservation between CRABS CLAW Orthologues from Widely Diverged Angiosperms. *Annals of Botany*, *100*(3), 651–657. <https://doi.org/10.1093/aob/mcm136>
- Fourquin, Chloé, Vinauger-Douard, M., Fogliani, B., Dumas, C., & Scutt, C. P. (2005). Evidence that CRABS CLAW and TOUSLED have conserved their roles in carpel development since the ancestor of the extant angiosperms. *Proceedings of the National Academy of Sciences*, *102*(12), 4649–4654. <https://doi.org/10.1073/pnas.0409577102>
- Frohlich, M. W. (2003). An evolutionary scenario for the origin of flowers. *Nature Reviews Genetics*, *4*(7), 559–566. <https://doi.org/10.1038/nrg1114>
- Frohlich, Michael W., & Chase, M. W. (2007). After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature*, *450*(7173), 1184–1189.  
<https://doi.org/10.1038/nature06393>
- Groszmann, M., Bylstra, Y., Lampugnani, E. R., & Smyth, D. R. (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*(5), 1495–1508. <https://doi.org/10.1093/jxb/erq015>
- Groszmann, M., Paicu, T., Alvarez, J. P., Swain, S. M., & Smyth, D. R. (2011). SPATULA and ALCATRAZ, are partially redundant, functionally diverging bHLH genes required for Arabidopsis gynoecium and fruit development. *Plant Journal*, *68*(5), 816–829.  
<https://doi.org/10.1111/j.1365-313X.2011.04732.x>

- Group, T. A. P. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, *161*(2), 105–121. <https://doi.org/10.1111/j.1095-8339.2009.00996.x>
- Group, T. A. P., Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., Mabberley, D. J., Sennikov, A. N., Soltis, P. S., & Stevens, P. F. (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), 1–20. <https://doi.org/10.1111/boj.12385>
- Heisler, M. G. B., Atkinson, A., Bylstra, Y. H., Walsh, R., & Smyth, D. R. (2001). SPATULA, a gene that controls development of carpel margin tissues in Arabidopsis, encodes a bHLH protein. *Development*, *128*(7), 1089–1098.
- Herendeen, P. S., Friis, E. M., Pedersen, K. R., & Crane, P. R. (2017). Palaeobotanical redux: Revisiting the age of the angiosperms. *Nature Plants*, *3*(3), 1–8. <https://doi.org/10.1038/nplants.2017.15>
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., & dePamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, *473*(7345), 97–U113. <https://doi.org/10.1038/nature09916>
- Kivivirta, K., Herbert, D., Lange, M., Beuerlein, K., Altmüller, J., & Becker, A. (2019). A protocol for laser microdissection (LMD) followed by transcriptome analysis of plant reproductive tissue in phylogenetically distant angiosperms. *Plant Methods*, *15*(1), 1–11. <https://doi.org/10.1186/s13007-019-0536-3>
- Li, H.-T., Yi, T.-S., Gao, L.-M., Ma, P.-F., Zhang, T., Yang, J.-B., Gitzendanner, M. A., Fritsch, P. W., Cai, J., Luo, Y., Wang, H., van der Bank, M., Zhang, S.-D., Wang, Q.-F., Wang, J., Zhang, Z.-R., Fu, C.-N., Yang, J., Hollingsworth, P. M., ... Li, D.-Z. (2019). Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants*, *5*(5), 461–470. <https://doi.org/10.1038/s41477-019-0421-0>

- Litt, A., & Kramer, E. M. (2010). The ABC model and the diversification of floral organ identity. *Seminars in Cell & Developmental Biology*, 21(1), 129–137.  
<https://doi.org/10.1016/j.semcdb.2009.11.019>
- Mandel, J. R. (2019). A Jurassic leap for flowering plants. *Nature Plants*, 5(5), 455–456.  
<https://doi.org/10.1038/s41477-019-0423-y>
- Moyroud, E., Monniaux, M., Thévenon, E., Dumas, R., Scutt, C. P., Frohlich, M. W., & Parcy, F. (2017). A link between LEAFY and B-gene homologues in *Welwitschia mirabilis* sheds light on ancestral mechanisms prefiguring floral development. *New Phytologist*, 216(2), 469–481.  
<https://doi.org/10.1111/nph.14483>
- Nikolov, L. A., & Tsiantis, M. (2015). Interspecies Gene Transfer as a Method for Understanding the Genetic Basis for Evolutionary Change: Progress, Pitfalls, and Prospects. *Frontiers in Plant Science*, 6. <https://doi.org/10.3389/fpls.2015.01135>
- Okamoto, J. K., Caster, B., Villaruel, R., VanMontagu, M., & Jofuku, K. D. (1997). The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 94(13), 7076–7081. <https://doi.org/10.1073/pnas.94.13.7076>
- Ortiz-Ramírez, C. I., Giraldo, M. A., Ferrándiz, C., & Pabón-Mora, N. (2019). Expression and function of the bHLH genes ALCATRAZ and SPATULA in selected Solanaceae species. *The Plant Journal*, 99(4), 686–702. <https://doi.org/10.1111/tpj.14352>
- Ortiz-Ramírez, C. I., Plata-Arboleda, S., & Pabón-Mora, N. (2018). Evolution of genes associated with gynoecium patterning and fruit development in Solanaceae. *Annals of Botany*, 121(6), 1211–1230. <https://doi.org/10.1093/aob/mcy007>
- Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., & Yanofsky, M. F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, 405(6783), 200–203.  
<https://doi.org/10.1038/35012103>
- Pfannebecker, K. C., Lange, M., Rupp, O., & Becker, A. (2017a). An Evolutionary Framework for Carpel Developmental Control Genes. *Molecular Biology and Evolution*, 34(2), 330–348.  
<https://doi.org/10.1093/molbev/msw229>

- Pfannebecker, K. C., Lange, M., Rupp, O., & Becker, A. (2017b). Seed Plant-Specific Gene Lineages Involved in Carpel Development. *Molecular Biology and Evolution*, *34*(4), 925–942. <https://doi.org/10.1093/molbev/msw297>
- Povilus, R. A., DaCosta, J. M., Grassa, C., Satyaki, P. R. V., Moeglein, M., Jaenisch, J., Xi, Z., Mathews, S., Gehring, M., Davis, C. C., & Friedman, W. E. (2020). Water lily (*Nymphaea thermarum*) genome reveals variable genomic signatures of ancient vascular cambium losses. *Proceedings of the National Academy of Sciences*, *117*(15), 8649–8656. <https://doi.org/10.1073/pnas.1922873117>
- Project, A. G. (2013). The Amborella Genome and the Evolution of Flowering Plants. *Science*, *342*(6165). <https://doi.org/10.1126/science.1241089>
- Ramírez-Barahona, S., Sauquet, H., & Magallón, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology & Evolution*, 1–7. <https://doi.org/10.1038/s41559-020-1241-3>
- Reyes, E., Nadot, S., von Balthazar, M., Schönenberger, J., & Sauquet, H. (2018). Testing the impact of morphological rate heterogeneity on ancestral state reconstruction of five floral traits in angiosperms. *Scientific Reports*, *8*(1), 9473. <https://doi.org/10.1038/s41598-018-27750-1>
- Reyes-Olalde, J. I., Zúñiga-Mayo, V. M., Marsch-Martínez, N., & Folter, S. de. (2017). Synergistic relationship between auxin and cytokinin in the ovary and the participation of the transcription factor SPATULA. *Plant Signaling & Behavior*, *12*(10), e1376158. <https://doi.org/10.1080/15592324.2017.1376158>
- Reyes-Olalde, J. I., Zúñiga-Mayo, V. M., Montes, R. A. C., Marsch-Martínez, N., & Folter, S. de. (2013). Inside the gynoecium: At the carpel margin. *Trends in Plant Science*, *18*(11), 644–655. <https://doi.org/10.1016/j.tplants.2013.08.002>
- Rümpler, F., & Theißen, G. (2019). Reconstructing the ancestral flower of extant angiosperms: The ‘war of the whorls’ is heating up. *Journal of Experimental Botany*, *70*(10), 2615–2622. <https://doi.org/10.1093/jxb/erz106>

- Salomo, K., Smith, J. F., Feild, T. S., Samain, M.-S., Bond, L., Davidson, C., Zimmers, J., Neinhuis, C., & Wanke, S. (2017). The Emergence of Earliest Angiosperms may be Earlier than Fossil Evidence Indicates. *Systematic Botany*, *42*(4), 607–619.  
<https://doi.org/10.1600/036364417X696438>
- Sauquet, H., Balthazar, M. von, Magallón, S., Doyle, J. A., Endress, P. K., Bales, E. J., Morais, E. B. de, Bull-Hereñu, K., Carrive, L., Chartier, M., Chomicki, G., Coiro, M., Cornette, R., Ottra, J. H. L. E., Epicoco, C., Foster, C. S. P., Jabbour, F., Haevermans, A., Haevermans, T., ... Schönenberger, J. (2017). The ancestral flower of angiosperms and its early diversification. *Nature Communications*, *8*(1), 1–10. <https://doi.org/10.1038/ncomms16047>
- Schönenberger, J. (2005). Rise from the ashes—The reconstruction of charcoal fossil flowers. *Trends in Plant Science*, *10*(9), 436–443. <https://doi.org/10.1016/j.tplants.2005.07.006>
- Scutt, Charles P. (2018). The Origin of Angiosperms. In L. Nuno de la Rosa & G. Müller (Eds.), *Evolutionary Developmental Biology: A Reference Guide* (pp. 1–20). Springer International Publishing. [https://doi.org/10.1007/978-3-319-33038-9\\_60-1](https://doi.org/10.1007/978-3-319-33038-9_60-1)
- Scutt, Charlie P., Vinauger-Douard, M., Fourquin, C., Finet, C., & Dumas, C. (2006). An evolutionary perspective on the regulation of carpel development. *Journal of Experimental Botany*, *57*(10), 2143–2152. <https://doi.org/10.1093/jxb/erj188>
- Siegfried, K. R., Eshed, Y., Baum, S. F., Otsuga, D., Drews, G. N., & Bowman, J. L. (1999). Members of the YABBY gene family specify abaxial cell fate in Arabidopsis. *Development*, *126*(18), 4117–4128.
- Smith, S. A., Beaulieu, J. M., & Donoghue, M. J. (2010). An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences*, *107*(13), 5897–5902. <https://doi.org/10.1073/pnas.1001225107>
- SMYTH, D., BOWMAN, J., & MEYEROWITZ, E. (1990). EARLY FLOWER DEVELOPMENT IN ARABIDOPSIS. *Plant Cell*, *2*(8), 755–767.
- Sokoloff, D. D., Remizowa, M. V., Bateman, R. M., & Rudall, P. J. (2018). Was the ancestral angiosperm flower whorled throughout? *American Journal of Botany*, *105*(1), 5–15.  
<https://doi.org/10.1002/ajb2.1003>

- Soltis, D. E., Chanderbali, A. S., Kim, S., Buzgo, M., & Soltis, P. S. (2007). The ABC Model and its Applicability to Basal Angiosperms. *Annals of Botany*, *100*(2), 155–163.  
<https://doi.org/10.1093/aob/mcm117>
- Taylor, D. W., Li, H. Q., Dahl, J., Fago, F. J., Zinniker, D., & Moldowan, J. M. (2006). Biogeochemical evidence for the presence of the angiosperm molecular fossil oleanane in Paleozoic and Mesozoic non-angiospermous fossils. *Paleobiology*, *32*(2), 179–190.  
[https://doi.org/10.1666/0094-8373\(2006\)32\[179:BEFTPO\]2.0.CO;2](https://doi.org/10.1666/0094-8373(2006)32[179:BEFTPO]2.0.CO;2)
- Theißen, G., Melzer, R., & Rümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: Linking plant development and evolution. *Development*, *143*(18), 3259–3271. <https://doi.org/10.1242/dev.134080>
- Theissen, G., & Saedler, H. (2001). Plant biology—Floral quartets. *Nature*, *409*(6819), 469–471.  
<https://doi.org/10.1038/35054172>
- THIEN, L., WHITE, D., & YATSU, L. (1983). THE REPRODUCTIVE-BIOLOGY OF A RELICT ILLICIUM-FLORIDANUM ELLIS. *American Journal of Botany*, *70*(5), 719–727.  
<https://doi.org/10.2307/2443126>
- Thomson, B., & Wellmer, F. (2019). Chapter Eight—Molecular regulation of flower development. In U. Grossniklaus (Ed.), *Current Topics in Developmental Biology* (Vol. 131, pp. 185–210). Academic Press. <https://doi.org/10.1016/bs.ctdb.2018.11.007>
- Vea, I. M., & Grimaldi, D. A. (2016). Putting scales into evolutionary time: The divergence of major scale insect lineages (Hemiptera) predates the radiation of modern angiosperm hosts. *Scientific Reports*, *6*(1), 23487. <https://doi.org/10.1038/srep23487>
- von Balthazar, M., Pedersen, K. R., Crane, P. R., Stampanoni, M., & Friis, E. M. (2007). *Potomacanthus lobatus* gen. Et sp nov., a new flower of probable Lauraceae from the Early Cretaceous (Early to Middle Albian) of eastern North America. *American Journal of Botany*, *94*(12), 2041–2053. <https://doi.org/10.3732/ajb.94.12.2041>
- Wang, Y.-Q., 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.' *The Plant Journal*, 64(2), 177–190. <https://doi.org/10.1111/j.1365-313X.2010.04325.x>

Winter, K.-U., Saedler, H., & Theißen, G. (2002). On the origin of class B floral homeotic genes: Functional substitution and dominant inhibition in *Arabidopsis* by expression of an orthologue from the gymnosperm *Gnetum*. *The Plant Journal*, 31(4), 457–475. <https://doi.org/10.1046/j.1365-313X.2002.01375.x>

Zeng, L., Zhang, Q., Sun, R., Kong, H., Zhang, N., & Ma, H. (2014). Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications*, 5(1), 4956. <https://doi.org/10.1038/ncomms5956>

Zúñiga-Mayo, V. M., Gómez-Felipe, A., Herrera-Ubaldo, H., & de Folter, S. (2019). Gynoecium development: Networks in *Arabidopsis* and beyond. *Journal of Experimental Botany*, 70(5), 1447–1460. <https://doi.org/10.1093/jxb/erz026>

---