

A BROADER VIEW FOR PLANT EVODEVO: NOVEL APPROACHES FOR DIVERSE MODEL SYSTEMS

EDITED BY : Verónica S. Di Stilio, Rainer Melzer and Jocelyn C. Hall
PUBLISHED IN : Frontiers in Plant Science





frontiers

Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-153-1

DOI 10.3389/978-2-88945-153-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

A BROADER VIEW FOR PLANT EVODEVO: NOVEL APPROACHES FOR DIVERSE MODEL SYSTEMS

Topic Editors:

Verónica S. Di Stilio, University of Wahsington, USA

Rainer Melzer, University College Dublin, Ireland

Jocelyn C. Hall, University of Alberta, Canada



Mature reproductive sporophyte of the emerging model fern *Ceratopteris richardii*
(photograph by Andrew A. Plackett)

This collection attempts to integrate work pertaining to a fundamental question in plant evolution: What are the molecular underpinnings for the origin of different plant forms? Among the many facets this question touches are the transition to land, the emergence of vascular plants, the origin of the seed and the origin and diversification of floral form. We aim to bring to the forefront the most salient and original plant systems and approaches within an inclusive phylogenetic context that encompasses representatives of the major lineages of land plants.

Citation: Di Stilio, V. S., Melzer, R., Hall, J. C., eds. (2017). A Broader View for Plant EvoDevo: Novel Approaches for Diverse Model Systems. Lausanne: Frontiers Media.
doi: 10.3389/978-2-88945-153-1

Table of Contents

05 Editorial: A Broader View for Plant EvoDevo: Novel Approaches for Diverse Model Systems

Verónica S. Di Stilio, Rainer Melzer and Jocelyn C. Hall

Section 1: Novel experimental approaches

07 Interspecies Gene Transfer as a Method for Understanding the Genetic Basis for Evolutionary Change: Progress, Pitfalls, and Prospects

Lachezar A. Nikolov and Miltos Tsiantis

15 Evolutionary Transgenomics: Prospects and Challenges

Raul Correa and David A. Baum

26 Evolution of the Plant Reproduction Master Regulators LFY and the MADS Transcription Factors: The Role of Protein Structure in the Evolutionary Development of the Flower

Catarina S. Silva, Sriharsha Puranik, Adam Round, Martha Brennich, Agnès Jourdain, François Parcy, Veronique Hugouvieux and Chloe Zubieta

Section 2: Emerging model systems

44 Ferns: The Missing Link in Shoot Evolution and Development

Andrew R. G. Plackett, Verónica S. Di Stilio and Jane A. Langdale

63 Sequenced Genomes and Rapidly Emerging Technologies Pave the Way for Conifer Evolutionary Developmental Biology

Daniel Uddenberg, Shirin Akhter, Prashanth Ramachandran, Jens F. Sundström and Annelie Carlsbecker

71 Petunia, Your Next Supermodel?

Michiel Vandenbussche, Pierre Chambrier, Suzanne Rodrigues Bento and Patrice Morel

82 Evolution of Catkins: Inflorescence Morphology of Selected Salicaceae in an Evolutionary and Developmental Context

Quentin C. B. Cronk, Isabelle Needham and Paula J. Rudall

95 The Phenotypic and Genetic Underpinnings of Flower Size in Polemoniaceae

Jacob B. Landis, Rebecca D. O'Toole, Kayla L. Ventura, Matthew A. Gitzendanner, David G. Oppenheimer, Douglas E. Soltis and Pamela S. Soltis

115 Loss of YABBY2-Like Gene Expression May Underlie the Evolution of the Lamina Style in Canna and Contribute to Floral Morphological Diversity in the Zingiberales

Kelsie Morioka, Roxana Yockteng, Ana M. R. Almeida and Chelsea D. Specht

127 Flower Development and Perianth Identity Candidate Genes in the Basal Angiosperm Aristolochia fimbriata (Piperales: Aristolochiaceae)

Natalia Pabón-Mora, Harold Suárez-Baron, Barbara A. Ambrose and Favio González

Section 3: Conservation vs. divergence of developmental mechanisms

147 *A Role of TDIF Peptide Signaling in Vascular Cell Differentiation is Conserved Among Euphyllophytes*

Yuki Hirakawa and John L. Bowman

157 *A Conserved Role for the NAM/miR164 Developmental Module Reveals a Common Mechanism Underlying Carpel Margin Fusion in Monocarpous and Syncarpous Eurosids*

Aurélie C. M. Vialette-Guiraud, Aurélie Chauvet, Juliana Gutierrez-Mazariegos, Alexis Eschstruth, Pascal Ratet and Charles P. Scutt

168 *Prevalent Exon-Intron Structural Changes in the APETALA1/FRUITFULL, SEPALLATA, AGAMOUS-LIKE6, and FLOWERING LOCUS C MADS-Box Gene Subfamilies Provide New Insights into Their Evolution*

Xianxian Yu, Xiaoshan Duan, Rui Zhang, Xuehao Fu, Lingling Ye, Hongzhi Kong, Guixia Xu and Hongyan Shan



Editorial: A Broader View for Plant EvoDevo: Novel Approaches for Diverse Model Systems

Verónica S. Di Stilio^{1*}, Rainer Melzer² and Jocelyn C. Hall³

¹ Department of Biology, University of Washington, Seattle, WA, USA, ² School of Biology and Environmental Science, University College Dublin, Dublin, Ireland, ³ Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada

Keywords: evolution of development, emerging model systems, land plants, plant morphology, evolutionary transitions, developmental genetics, phylogenetics, molecular toolkit

Editorial on the Research Topic

A Broader View for Plant EvoDevo: Novel Approaches for Diverse Model Systems

For many years, a main focus of plant evolutionary developmental biology was studying the expression and phylogenetic history of genes implicated in developmental pathways. This approach has been enormously successful in identifying potentially conserved gene regulatory circuits that underlie major pattern formation processes in plants. Importantly, hypotheses were often generated on how changes in these gene regulatory-circuits led to the evolution of different plant forms. However, for quite some time, experimental testing of many of these hypotheses proved difficult, simply because the adequate molecular biology toolkit was not available across many plant lineages. This situation has changed dramatically in recent years. The advent of next generation sequencing considerably facilitated sequencing genomes and transcriptomes of plants throughout the phylogeny. Virus-induced gene silencing and the establishment of transformation methods for non-model plants enabled direct testing of gene functions on a wide phylogenetic spectrum, and elaborate biophysical techniques are increasingly applied to analyze changes in protein function during evolution. Furthermore, bioinformatics as well as systems biology are used to integrate the available data into a more coherent understanding of a fundamental question in plant evolution: What are the molecular underpinnings for the origin of different plant forms? Among the many facets this question touches are the transition to land, the emergence of vascular plants, the origin of the seed and the origin and diversification of floral form. In this research topic we highlight emerging model systems across the land plant phylogeny as well as exciting current approaches, including genomics, biophysics, gene networks and transgenesis of plants from diverse lineages. We aim to bring to the forefront the most salient and original plant systems and approaches within an inclusive phylogenetic context that encompasses representatives of the major lineages of land plants.

Among the novel experimental approaches that can be applied in a variety of systems, three articles offer perspectives on methodologies for the study of the diversification of form and the divergence of species. In interspecies gene transfer (IGT), Nikolov and Tsiantis describe how candidate genes from a donor species are added to the wildtype genome of a recipient species to test their causality in the divergence of the two species. In evolutionary transgenomics, Correa and Baum describe the transfer of whole genomic fragments between species to identify novel genes of large effect without prior commitment to candidate genes. Silva et al. discuss the importance of biophysical studies for understanding morphological evolution. They show that small changes in the amino acid sequence of floral developmental regulators can lead to drastically altered protein–protein interaction patterns that may in turn have contributed to the evolution of the flower. This result illustrates that a more integrated approach—using genetics, biophysics and phylogenetics—is necessary to understand the evolution of development.

OPEN ACCESS

Edited by:

Neelima Roy Sinha,
University of California, Davis, USA

Reviewed by:

Jessica M. Budke,
University of Tennessee, USA

*Correspondence:

Verónica S. Di Stilio
distilio@u.washington.edu

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 10 December 2016

Accepted: 11 January 2017

Published: 26 January 2017

Citation:

Di Stilio VS, Melzer R and Hall JC
(2017) Editorial: A Broader View for
Plant EvoDevo: Novel Approaches for
Diverse Model Systems.
Front. Plant Sci. 8:61.
doi: 10.3389/fpls.2017.00061

Several articles present emerging model systems for the study of plant evo–devo, from shoot evolution to changes in inflorescence and floral traits. Plackett et al. highlight the use of the emerging model fern *Ceratopteris richardii*, in the sister lineage to seed plants, for the study of the evolution and development of shoots, and of the genetic regulation of shoot apical meristems. Among the non-flowering seed plants, conifers represent a close extant relative of flowering plants, making them especially interesting from an evolutionary developmental perspective. Uddenberg et al. highlight the importance of studying conifers and suggest that next generation sequencing and improved transformation protocols will make them more accessible to evo–devo studies. Within angiosperms, Vandenbussche et al. argue that new “supermodels” are required to comprehensively study the evolution of gene function that would ideally be as amenable to genetic analyses as *Arabidopsis*. The authors suggest that petunia could be one of those supermodels and provide an extensive overview of the genetic resources available for this system. Cronk et al. describe the evolution of catkin inflorescences in Salicaceae (poplars and willow) illustrating how the morphological richness of the Salicaceae coupled with the rapidly expanding genomic resources make this, of all woody plant families, particularly promising for genome-enabled evolutionary developmental biology. Landis et al. present *Saltugilia* (Polemoniaceae) as a model for the study of flower size (corolla tube length), a trait central to pollination syndrome. They find two independent evolutionary transitions to long corollas, and a correlation of long corollas with an increase in jigsaw cell size and number and with the up regulation of genes associated with cell wall formation and organization. Morioka et al. examine floral diversity in Zingiberales, where members of Cannaceae have a laminar style that plays an important role in pollination interactions. Expression and evolution of genes involved in adaxial/abaxial polarity reveal a complex evolutionary history and suggest that loss of expression lead to this novel feature in *Canna*. Pabón-Mora et al. investigate the genetic basis of the highly derived and fused morphology of *Aristolochia fimbriata*. Developmental and comparative gene expression data support that the fused perianth is derived from sepals, not petals. Their data also reveal that A-class genes in the classic ABCE model do not contribute to perianth identity in this system. This finding provided further evidence that *Arabidopsis* A-class orthologs rarely contribute to perianth identity in other taxa.

Three articles present evidence for opposing forces in the evolution of developmental mechanisms: conservation and divergence of gene and protein function. On the one hand, Hirakawa and Bowman show evidence for the conservation of protein function of the CLE family peptide hormone Tracheary

element Differentiation Inhibitory Factor (TDIF) in regulating procambial cell fate, an important aspect in the evolutionary transition to vascular plants. The study performed evolutionary and functional comparative analyses, using protein assays, among representatives of major lineages in vascular plants and concluded that TDIF was integrated into shoot xylem differentiation in the euphyllophyte lineage (ferns and seed plants), after the split from lycophytes. Vialette-Guiraud et al. also present evidence for a conserved gene regulatory circuit, this time during flower development: They show that a genetic module consisting of microRNA164 and NAM transcription factors is responsible for the fusion of carpel margins in eurosids. The authors further suggest that the same gene regulatory circuit could have contributed to the emergence of the closed carpel very early during angiosperm evolution, and might thus have been involved in the origin of one of the most important evolutionary novelties in angiosperms. On the other hand, Yu et al. show that the exon-intron structure of genes is labile and influences the evolution of flower development gene lineages, frequently following gene duplication and speciation events. Their work suggests, for example, that an unstable gene structure in the *AGL6* lineage may have contributed to its functional diversification in the flowering plants and to its divergence from the *SEP* lineage, which went on to become the major mediator of angiosperm floral quartets.

We hope that the publication of this research topic will promote renewed interest in the need for a larger inventory of model systems and approaches to facilitate an increasingly broader and more meaningful view of the evolution of plant development.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

VD was funded by NSF-IOS 1121669 and RM acknowledges support from University College Dublin.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Di Stilio, Melzer and Hall. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Interspecies Gene Transfer as a Method for Understanding the Genetic Basis for Evolutionary Change: Progress, Pitfalls, and Prospects

Lachezar A. Nikolov and Miltos Tsiantis *

Department of Comparative Development and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington, USA

Reviewed by:

Ana Maria Rocha De Almeida,
University of California at Berkeley,
USA

Raul Correa,
Baylor College of Medicine, USA

*Correspondence:

Miltos Tsiantis
tsiantis@mpipz.mpg.de

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 06 October 2015

Accepted: 30 November 2015

Published: 22 December 2015

Citation:

Nikolov LA and Tsiantis M (2015)
Interspecies Gene Transfer as a
Method for Understanding the
Genetic Basis for Evolutionary
Change: Progress, Pitfalls, and
Prospects. *Front. Plant Sci.* 6:1135.
doi: 10.3389/fpls.2015.01135

The recent revolution in high throughput sequencing and associated applications provides excellent opportunities to catalog variation in DNA sequences and gene expression between species. However, understanding the astonishing diversity of the Tree of Life requires understanding the phenotypic consequences of such variation and identification of those rare genetic changes that are causal to diversity. One way to study the genetic basis for trait diversity is to apply a transgenic approach and introduce genes of interest from a donor into a recipient species. Such interspecies gene transfer (IGT) is based on the premise that if a gene is causal to the morphological divergence of the two species, the transfer will endow the recipient with properties of the donor. Extensions of this approach further allow identifying novel loci for the diversification of form and investigating *cis*- and *trans*-contributions to morphological evolution. Here we review recent examples from both plant and animal systems that have employed IGT to provide insight into the genetic basis of evolutionary change. We outline the practice of IGT, its methodological strengths and weaknesses, and consider guidelines for its application, emphasizing the importance of phylogenetic distance, character polarity, and life history. We also discuss future perspectives for exploiting IGT in the context of expanding genomic resources in emerging experimental systems and advances in genome editing.

Keywords: *Cardamine hirsuta*, evolution of morphology, leaf development, regulatory evolution

As species diverge, so do their genomes and morphologies. Regulatory evolution and consequent modifications of transcriptional networks of a broadly conserved repertoire of developmental genes are believed to be at the heart of evolutionary change in morphology (Carroll, 2008; Peter and Davidson, 2011). Regulatory changes involve modification of *cis*-regulatory elements and the *trans*-environment, and understanding these processes is critical for understanding how traits diversify. However, to pinpoint the precise genetic changes that underlie morphological diversity at different evolutionary scales remains a fundamental challenge. One way forward is to follow a transgenic approach, and transfer a gene suspected to contribute to the divergence between two species with contrasting morphologies. Such interspecies gene transfer (IGT) is based on the premise that if

a gene is causal to divergence, the transfer will endow the recipient with properties of the donor (Figure 1A).

THE PREMISE

IGT is a functional test for sufficiency to study the underlying genetic basis of trait divergence and can be used between species that do not hybridize. As such it complements the classical approach based on genetic crosses. Examining the contribution of the transgene on the trait under study can indirectly provide information about its underlying genetic architecture (Figures 1B–D). The transfer of the entire interrogated locus, including its non-coding regulatory elements, allows studying evolutionary events concerning both coding and regulatory sequences and these have distinct outcomes in the context of IGT (Figure 1E). Protein divergence underlies the trait divergence in two species when the coding sequence of one species is able to elicit a phenotypic change in the other species, whereas the endogenous copy under the same promoter does not (see Kramer, 2015 for details). To further characterize biochemical divergence, which can manifest as a metabolic difference or as differences in the expression of downstream genes, the amino acid differences between the two proteins can be interrogated, for example in *in vitro* assays (e.g., Hoekstra et al., 2006). When two species diverged morphologically but the protein function did not change during evolution, expression difference underlying the divergence is suspected. In this case, coding sequences from both species under the same promoter may be able to elicit phenotypic change in the recipient but if a *cis*-regulatory change is causal, only the entire locus from the donor will be similarly potent. Alternatively, if transfer of the entire locus from the donor has no detectable effect on the recipient's morphology, *trans*-regulatory change, a combination of *cis*- and *trans*-changes, and downstream gene divergence are plausible explanations. In all cases, the experiment should be interpreted in the context of other critical data, such as loss-of-function phenotypes, expression analyses, and the phylogenetic distribution of character states.

Transfer of a heterologous locus into the recipient genetic background will result in a phenotypic change when three criteria are satisfied. First, the encoded protein can perform biochemically; second, it is expressed in the correct (or at least developmentally meaningful) time and place; and third, enough of the gene regulatory network for the trait is intact in the recipient. On the other hand, insufficient dominance resulting for example from absence of synergistic activities, and substantial divergence owing to co-evolution between *cis*-elements and *trans*-factors will render the heterologous locus non-functional. Thus, in its current use, IGT is a one-by-one locus approach that is not well suited for assessing the degree of functional and causal interdependence between endogenous genes in the donor. One of the first transgenic studies to understand the genetic basis of morphological evolution in animals examined the wing pigmentation of fruit flies in the *melanogaster* group (Gompel et al., 2005). To understand the origin of a novel wing spot, *yellow* 5' regulatory sequence of *Drosophila biarmipes*, which

features a spot, was fused to a reporter and introduced into the spot-free *D. melanogaster*. The reporter displayed expression pattern similar but not identical to the one observed in the donor *D. biarmipes*, which suggests that divergence at the *yellow* regulatory region contributed to the novel wing pigmentation pattern. It also revealed additional *trans*-factors that confer the precise spatio-temporal expression of the spot (Gompel et al., 2005). Furthermore, introducing a partial *yellow* locus of *D. biarmipes* was not sufficient to generate a spot in *D. melanogaster*, indicating that additional loci are involved.

OTHER NOTABLE APPLICATIONS

IGT is a powerful test for the contribution of candidate loci known to affect a given trait in other species. Extensive use of the method in an evolutionary context has been made in studies of the evolution of angiosperm leaf shape (Figure 2A; Hay and Tsiantis, 2006; Barkoulas et al., 2008; Vlad et al., 2014; Rast-Somssich et al., 2015). These experiments showed that two apparently independent developmental modules contribute to leaf complexity in the family Brassicaceae. One involves class I *KNOTTED1*-like homeobox (*KNOX*) genes where *cis*-regulatory changes underlie the divergence between the simple-leaved *A. thaliana* and its compound-leaved relative *Cardamine hirsuta* (Hay and Tsiantis, 2006; Barkoulas et al., 2008; Rast-Somssich et al., 2015). Transferring *KNOX* gene paralogs from *C. hirsuta* into *A. thaliana* provides evidence for an inverse relationship between pleiotropy of a gene and its potential to evolve variants able to alter morphology in an IGT experiment (Rast-Somssich et al., 2015). The other involves the *REDUCED COMPLEXITY* (*RCO*) homeobox gene, which is a member of a tandem three-gene cluster in many mustard species (Vlad et al., 2014). Having lost *RCO* from its genome, which likely contributed to leaf shape simplification, *A. thaliana* retains only one member of this cluster, *LATE MERISTEM IDENTITY 1* (*LMII*). In *C. hirsuta*, *RCO* and *LMII* are expressed in near complementary domains, the former at the base of developing leaflets, and the latter along leaflet margins and in the stipules, respectively. Importantly, when expressed from the *RCO* promoter both *RCO* and *LMII* coding sequences can complement the *C. hirsuta rco* mutant, which exhibit simplified leaves, supporting the idea that regulatory rather than coding divergence underlies the functional differences of these two paralogs (Vlad et al., 2014). Using transformation to move the entire genomic *RCO* locus from *C. hirsuta* into *A. thaliana*, which is in principle a functional equivalent of the *C. hirsuta rco* mutant, produces deep lobes in the otherwise nearly smooth leaf margin of *A. thaliana*. Similar results were obtained with the *Capsella* homolog of *RCO* (Sicard et al., 2014). Thus, an introduction of a single gene is capable to reverse-engineer a character lost in *A. thaliana*. The *C. hirsuta RCO* locus is able to modify the simple leaf of *A. thaliana* likely because it represents the derived state due to *RCO* loss but retains the ancestral regulatory landscape that promotes leaf complexity through *RCO* activation. This renders *RCO* a major effect locus that may account for much of the variation in leaf shape in Brassicaceae (Sicard et al., 2014; Vlad et al., 2014).

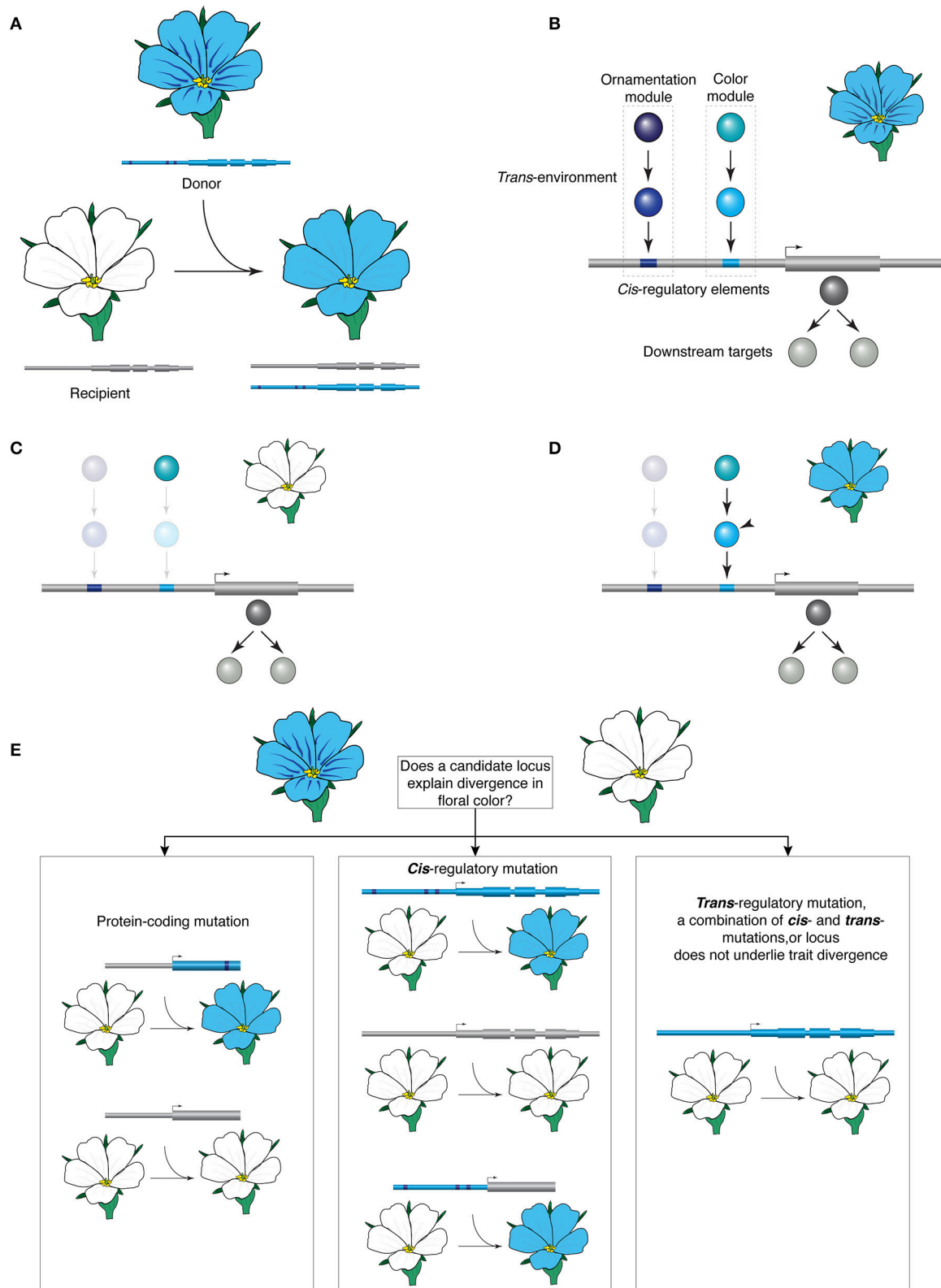


FIGURE 1 | The premise of interspecies gene transfer. (A) Transferring a locus from the blue-flowered donor species is able to transform the color of the white flowers of the recipient; note that the transfer does not recapitulate the ornamentation (the nectar guides) of the donor in the recipient. **(B–D)** Hypothetical network (Continued)

FIGURE 1 | Continued

controlling color and ornamentation expression. **(B)** Light and dark blue transcription factors (TFs) interact directly with the *cis*-regulatory elements of a key blue pigment synthesis gene, and control the color and the ornamentation of the blue flower, respectively. These TFs are in turn regulated by upstream regulators. **(C)** In the white-flowered species, pigment synthesis is abolished via inactivation mutations in TFs but the blue pigment synthesis genes (gray) and the upstream regulator for blue color (light blue) remain intact. **(D)** Introducing the light blue TF (arrowhead) restores blue pigment synthesis in a white-flowered recipient but does not transfer the ornamentation pattern. **(E)** Possible outcomes of an IGT experiment designed to test the contribution of a candidate locus to the divergence in color. In protein divergence, expression of the coding sequence of the donor under the recipient's promoter will result in change of phenotype. If *cis*-regulatory evolution underlies phenotype divergence, the coding sequence of the recipient expressed under the donor's promoter and the entire locus of the donor may be sufficient for phenotypic change. *Trans*-regulatory mutation, a combination of *cis*- and *trans*-mutations, or lack of involvement of the locus are possible when transferring the entire locus of the donor does not reconstitute the phenotype in the recipient.

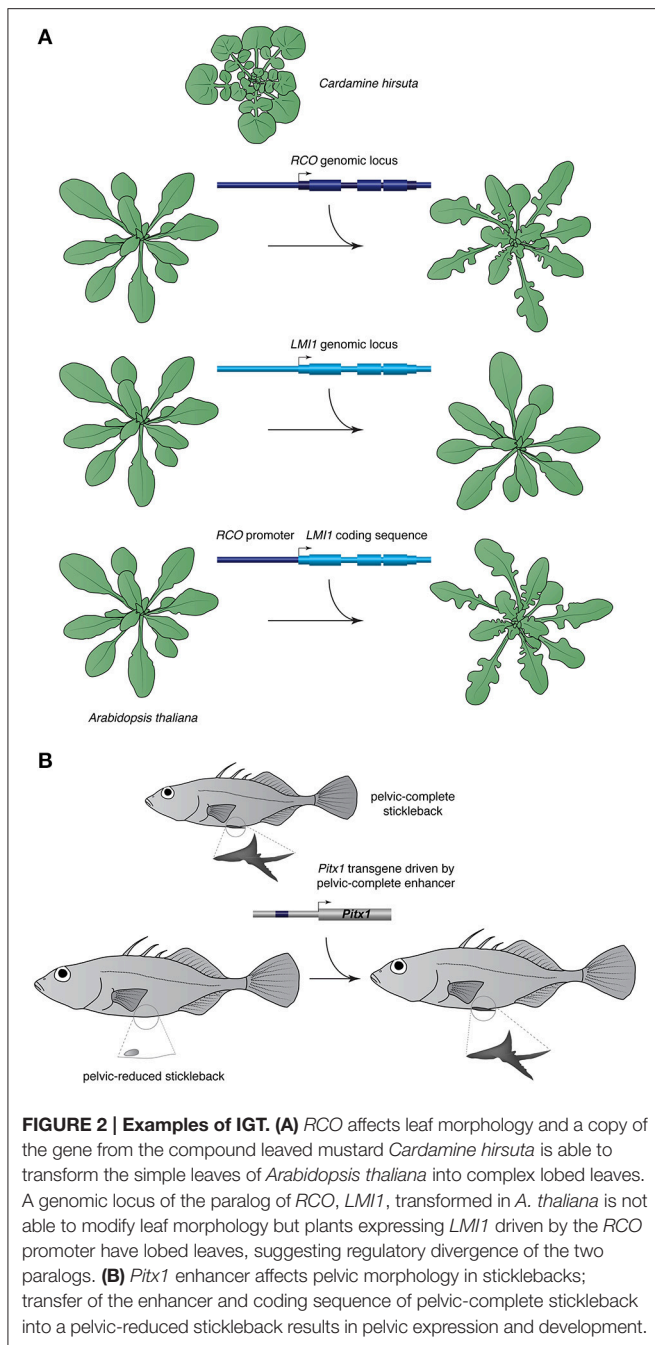
That reintroduction of a single gene can restore a morphological state was also demonstrated in threespine stickleback fish, where reductions or loss of the pelvic girdle and spines, which feature prominently in marine populations, have been lost several times independently after freshwater transition (**Figure 2B**; Chan et al., 2010). Linkage mapping has identified a region containing the *Pituitary homeobox 1* (*Pitx1*) gene to account for much of the variance in pelvic size, and although *Pitx1* protein sequence in pelvic-reduced sticklebacks is identical to their marine ancestors', its expression is abolished in the pelvic region, suggesting a causal regulatory mutation (Shapiro et al., 2004). The mutation was mapped to a deletion in the upstream noncoding region of *Pitx1* in pelvic-reduced sticklebacks that contains a tissue-specific enhancer (Chan et al., 2010). Introducing the *Pitx1* enhancer and coding sequence into fertilized eggs of pelvic-reduced fish resulted in enlarged pelvic girdle and external pelvic spine in transgenic fish, demonstrating the functional significance of *Pitx1* in pelvic development.

The *Pitx1* and *RCO* examples highlight the advantage of recipients with loss-of-function phenotypes in transgenic rescue experiments. Derived gain-of-function phenotypes can also be transferred to provide evidence for sufficiency. For example the trait of four abdominal bristles from *Drosophila quadrilineata* can be transferred to the two-bristled *D. melanogaster* via the *scute* enhancer from *D. quadrilineata* but not via transferring homologous enhancers from species with only two abdominal bristles (Marcellini and Simpson, 2006). In another example, transferring the promoter and coding sequence of the plasma membrane ATPase *HMA4* from *Arabidopsis halleri*, which exhibits heavy metal hyperaccumulation to the non-accumulator *A. thaliana* resulted in increased *HMA4* transcript levels (Hanikenne et al., 2008). The transgenic *A. thaliana* plants also showed zinc distribution in the root comparable to *A. halleri* suggestive of zinc partitioning and tolerance, but toxic shoot zinc hypersensitivity characteristic of wild type *A. thaliana*. This finding indicates that additional genes are necessary to reconstitute all facets of the hyperaccumulator syndrome in plants.

In a study designed to investigate a possible contribution of the *LEAFY* (*LFY*) transcription factor to the divergence of plant architecture in Brassicaceae, the entire *LFY* locus from the rosette flowering crucifers *Ionopsidium acaule* (*IacLFY*), *Idaho scapigera* (*IscLFY1*), and *Leavenworthia crassa* (*LcrLFY*) was independently introduced into *A. thaliana lfy-6* mutant background, which shows defects in floral meristem identity

(Yoon and Baum, 2004). The *IacLFY* locus was able to rescue the *lfy* phenotype as expected for regulatory and protein conservation and thus cannot explain rosette flowering in *I. acaule*. In contrast, *IscLFY1* rescued some aspects of the *lfy* floral phenotype in *A. thaliana*, but generated developmental defects, such as bracteate flowers (bracts normally abort in Brassicaceae), shortened internodes, and occasionally aerial rosettes resembling the phenotype of the donor, suggesting that *IscLFY1* may contribute to rosette flowering (Yoon and Baum, 2004). Similarly, *LcrLFY* partially rescued the floral *lfy* phenotype but some transgenic lines produced terminal flowers as in wild *L. crassa* plants. These observations imply different mechanisms for rosette flowering in the studied species, but the complex transgenic phenotypes make interpretation difficult, likely because *LFY* affects many downstream processes beyond plant architecture (Winter et al., 2011). Pleiotropic effects may hinder donor phenotype reconstitution using developmental master regulators even when a complex phenotype is reduced to well-defined principle components. Despite rigorous phenotypic analysis, a study assessing the contribution of the transcription factors *doublesex* and *fruitless*, which coordinate sex-specific functions, to species-specific male courtship dance revealed that although transgenes from four *Drosophila* species were able to rescue *D. melanogaster* courtship behavior, no elements of the ritualized species-specific dance were transferred (Cande et al., 2014).

The test for sufficiency can be extended to a forward screen to find novel genes contributing to morphological divergence between species, as initially proposed under the term transgenomics (Baum, 2002; Correa and Baum, 2015). A proof-of-concept study reported the introduction of ca. 4% of the genome of *Leavenworthia alabamica*, a relative of *A. thaliana* that differs in a number of traits, into *A. thaliana* to screen for changes in morphology consistent with the presence of a transgene (Correa et al., 2012). The technique holds much promise when larger portions of the genome are introduced into the donor and more primary transformants are screened. A transgenomic screen of a large insert library from the salt tolerant mustard *Eutrema salsuginea* into *A. thaliana*, which represent two divergent lineages in the mustard family, revealed a stress tolerant candidate locus (Wang et al., 2010). A similar study to identify factors for drought and alkaline tolerance of the resurrection plant *Boea hygrometrica* (Gesneriaceae), an asterid, as a donor and the rosoid *A. thaliana* as a recipient revealed a retro-element fragment conferring improved photochemical



efficiency and membrane integrity under osmotic and alkaline stress (Zhao et al., 2014). Although, the precise mechanisms by which these loci confer tolerance are currently unknown, this approach has a potential to identify putative causative variants that can be examined further.

PRACTICAL CONSIDERATIONS

IGT is a versatile tool and can be applied in both forward (i.e., transgenomics) and reverse genetics context to obtain functional

information for genes identified in forward genetic screens (e.g., EMS mutagenesis screens), as well as candidates from comparative gene expression studies and targeted transcriptome profiles (e.g., from laser capture microdissection- and INTACT-derived cell specific transcriptomes; Nelson et al., 2006; Deal and Henikoff, 2011). In that respect, the technique can be used successfully to study both homologs of known morphologically important genes, as well as non-obvious candidates identified through high-throughput genomic and transcriptomic approaches. IGT is most revealing in combination with data on gene expression and the biochemical conservation of the protein. The native expression of the locus in the donor and its reconstituted expression in the recipient can be characterized by promoter reporters to determine whether the transgene will be expressed in a functionally relevant position (i.e., corresponding to the donor's) and to further infer *cis*-regulatory changes in the promoter or *trans*-changes in factors upstream of the candidate locus in the genetic hierarchy. To test the biochemical potential of the protein to alter form, the coding sequence of the donor species can be expressed under a broadly active promoter (e.g., *CaMV 35S*, *pRPS5a*, and *Ubi* promoters) in the recipient species. Many eukaryotic genes are alternatively spliced, and to reduce cloning efforts and avoid bias in calling splice variants, constructs containing the entire exon-intron structure between the start and the stop codon can be transferred into the recipient species to allow processing by the endogenous splicing machinery. This approach may also allow identification of control elements, such as intronic enhancers. Since the resulting phenotype may be difficult to interpret due to pleiotropic defects that reflect expression that is too broad in space or time, or has a very high level, expression in a narrower domain known for its strong morphogenetic properties, such as leaf margin (Hagemann and Gleissberg, 1996), vertebrate limb bud, and insect imaginal disk may be more suitable to assess biochemical function. However, implicit assumption of when and where the protein is expressed may not be fully congruent with its native pattern of expression, and may prevent the protein from eliciting function. Transferring the entire locus, including its coding sequence and regulatory elements will reconstitute the spatiotemporal context where the protein operates in the donor if it is functional in the recipient's *trans*-environment.

In plants, the IGT constructs are generally introduced into the donor genome using *Agrobacterium*-mediated transformation. A transformation event often results in introducing multiple copies of the transgene so studying single T-insertion lines in detail is preferable. Since dosage alone can account for much of the observed phenotypic change, it is critical to confirm that such effects are not causal by independently transforming the recipient's endogenous copy as a control and comparing the phenotypic distribution. Another useful control is introducing the transgene into recipient's null mutant background in a complementation test; however, few species outside of the established models permit such experiment. To avoid positional effects and circumvent transgene silencing, multiple independent T-insertions are to be analyzed, and rare phenotypes should be treated with caution.

CRITICAL APPRAISAL

IGT is particularly valuable for determining the genetic basis for morphological variation between reproductively isolated species when classical genetics methods including QTL analysis are not feasible. Genetic transformation is a prerequisite for IGT. Since the introduced transgene is in hemizygous state and resides in the recipient genome along with the endogenous copy, typically only gain-of-function phenotypes are accessible. As such, IGT is a test for sufficiency, which can determine whether the introduced copy alone has the potential to recreate and is thus likely causative to the donor phenotype. Based on these assumptions, the IGT strategy is particularly powerful if the donor represents the ancestral state of the studied character, and the recipient exhibits a loss-of-function derived state. Reconstructing the ancestral phenotype then also suggests that the rest of the transcriptional network underlying the trait (or a network that is functionally equivalent) is intact in the recipient. Alternatively, a donor with derived character state may elicit phenotypic response in a recipient that lacks the trait (ancestral state) if the introduced locus can function alone (e.g., many metabolic enzymes), or is capable of coopting appropriate downstream targets. The second scenario is most likely if the gene regulatory hierarchy of the trait is not particularly complex and the number of loci involved is not too large. Thus, knowledge of character polarity distribution is useful in the experimental design and interpretation of IGT. As directionality of evolutionary change is often difficult to infer from incompletely sampled phylogenies, extending the common garden experiment in ecology to genetics (i.e., swap of promoters or entire loci between a donor and a recipient in an equivalent of a reciprocal transplant) circumvents the need to understand the phylogenetic distribution of morphological states (Hay and Tsiantis, 2006; Kellogg, 2006a,b; Gordon and Ruvinsky, 2012). Because we are often constrained to a particular focal species as a donor, there is more flexibility in selecting the recipient species—the more advanced experimental model in a given phylogenetic proximity is a reasonable choice. While it is impossible to select a recipient that differs from the donor only by the character under study or to introduce the transgene in an ancestor prior to the acquisition of the character state, related species with similar life history and growth habit can be used to obtain an interpretable phenotypic readout. In plants, the model *A. thaliana* can serve as a reference recipient species in many IGT studies due to practical considerations, such as reliable transformation, rapid life cycle, and lack of prolonged seed dormancy. However, the IGT outcomes are particularly sensitive to the choice of donor and recipient species (Ruvinsky and Ruvkin, 2003; Barriere and Ruvinsky, 2014; Gordon et al., 2015).

A major issue with the IGT approach is the interpretation of the resulting phenotypes. The approach attempts to atomize a trait by interrogating an individual locus and its contribution, which is an advantage for interpreting gene transfers of major phenotypic and minimal pleiotropic effects. Traits with simpler underlying genetic architecture, such as ones determining certain physiological traits are readily amenable to study by IGT (Hanikenne et al., 2008; Wang et al., 2010). These traits appear less sensitive to the phylogenetic distance between

the donor and recipient genomes, and phenotypic changes that are more straightforward to quantify. In complex traits evolved by accumulation of many changes, sequence divergence and the co-evolution of *cis*- and *trans*-elements becomes more substantial and the contribution of any individual locus diminishes, which confounds interpretation. Similarity in body plan, life history, and growth habit, and the ability to make clear homology statements improve interpretability. With increase of phylogenetic distance, homology inference for at least some characters becomes more difficult. This is a reason why there are limitations to the conclusions that can be drawn from transfer of genes between distantly related species with widely divergent morphologies, for example when attempting to understand the origin of flowering by transferring orthologs of seed plant floral identity genes from mosses into *A. thaliana*, though such experiments can inform on the biochemical potential of the proteins in questions in divergent lineages (Maizel et al., 2005). A measure of functional conservation of the gene regulatory machinery at a phylogenetic scale might be obtained from reporter gene analyses, either through reciprocal swap of 5' regulatory sequences, or via expression of heterologous regulatory sequence in a single reference species (Kalay and Wittkopp, 2010). Reporter expression demonstrated limited conservation of regulatory sequence and/or *trans*-factors between *Drosophila melanogaster* enhancers that were moved into *Caenorhabditis elegans* by transformation, and extensive conservation in swaps between *C. elegans* and *C. briggsae*, with several specific instances of functional divergence (Ruvinsky and Ruvkin, 2003). Similar experiments with the regulatory sequences of eight genes from four *Caenorhabditis* nematodes in *C. elegans* revealed overall conservation of expression pattern with broader expression in other cell types, which was interpreted as a sign of functional divergence (Barriere and Ruvinsky, 2014). Similar trend with at least partial conservation of gene expression patterns not reflected in sequence similarity persists at a phylum level for nematodes that diverged more than 400 million years ago (Gordon et al., 2015). Therefore, depending on the complexity of the trait under study the approach can be applicable at various scales and informed judgment that considers lineage specific evolutionary patterns and rates is needed to decide on the most appropriate experimental setup.

SUMMARY AND FUTURE PROSPECTS

Traditionally, identifying the loci underlying trait divergence is based on crosses between two populations or species with contrasting morphologies and examining the phenotypic distribution of traits in the progeny. The precise proportion of gene activity that accounts for morphological divergence of reproductively isolated species, however, is difficult to conceptualize. IGT provides a platform to address this question transgenically by interrogating candidate loci in their entirety and via their regulatory and protein-coding components. A weakness of the approach is the underlying assumption that a certain aspect of the trait can be reducible to a single gene (see Baum, 2013, for a discussion on the causation between individuated traits and developmental-causal genes,

and Orgogozo et al., 2015), which is only an operational approximation that will have limited validity when multigene interactions underlie the diversification of the studied trait. To offset this shortcoming, an extension of the method to introduce several transgenes into the recipient, which collectively may reconstruct a functional module underlying the trait under study, can be applied. Some of the characteristics of the transgenic approach, such as dosage and integration effects, which obstruct interpretability, can be overcome by gene replacement of an endogenous gene with a transgene placed in comparable genomic position via homologous recombination (Puchta, 2002). Not all species are amenable to gene replacement, which makes emerging technologies in genome editing (e.g., programmable nucleases and the CRISPR-Cas system, which allow direct and precise manipulation of gene function) particularly promising to specifically target loci in their endogenous genomic context. The explosion of sequencing information from a wide range of organisms should greatly facilitate the broad application of IGT (Rowan et al., 2011). Sequencing information in combination with improved tools for genome editing will advance the versatility of the platform through introducing judiciously distributed species pairs at key phylogenetic positions (Jenner and Wills, 2007; Abzhinov et al., 2008). The progress in editing technologies also allows for a more straightforward application of other techniques used to assess gene contribution to phenotypic change, such as the reciprocal hemizygosity test, which compares

the phenotypes of reciprocal hybrids that are genetically identical throughout the genome except at the test locus (Stern, 2014). In addition to opening new avenues for comparative research and contributing to the shift from studying the pattern of variation to providing a mechanistic insight into the genetic basis of evolutionary change, IGT also offers the conceptual background for the reverse engineering of traits of practical interest through synthetic biology.

AUTHOR CONTRIBUTIONS

LN and MT wrote and approved the manuscript.

ACKNOWLEDGMENTS

This manuscript is supported by a Deutsche Forschungsgemeinschaft (DFG) 'Adaptomics' grant TS 229/1-1, a DFG Collaborative Research Center SFB 680 grant on Evolutionary Innovations, and a core grant from the Max Planck Society to MT, and a Humboldt Research Fellowship to LN. MT also acknowledges support from the Gatsby Charitable Foundation, the Cluster of Excellence on Plant Sciences and BBSRC grant BB/H011455/1 on interspecific gene transfer. We thank Elena Kramer for critically reading the manuscript and members of Tsiantis laboratory for discussions and for sharing unpublished data.

REFERENCES

- Abzhinov, A., Extavour, C. G., Groover, A., Hodges, S. A., Hoekstra, H. E., Kramer, E. M., et al. (2008). Are we there yet? Tracking the development of new model systems. *Trends Genet.* 24, 353–360. doi: 10.1016/j.tig.2008.04.002
- Barkoulas, M., Hay, A., Kougiumoutzi, E., and Tsiantis, M. (2008). A developmental framework for dissected leaf formation in the *Arabidopsis* relative *Cardamine hirsuta*. *Nat. Genet.* 40, 1136–1141. doi: 10.1038/ng.189
- Barrière, A., and Ruvinsky, I. (2014). Pervasive divergence of transcriptional gene regulation in *Caenorhabditis* nematodes. *PLoS Genet.* 10:e1004435. doi: 10.1371/journal.pgen.1004435
- Baum, D. A. (2002). "Identifying the genetic causes of phenotypic evolution: a review of experimental strategies," in *Developmental genetics and evolution*, eds Q. C. B. Cronk, R. M. Bateman, and J. A. Hawkins (London: Taylor & Francis), 493–507.
- Baum, D. A. (2013). Developmental causation and the problem of homology. *Philos. Theory Biol.* 5:e403. doi: 10.3998/ptb.6959004.0005.003
- Cande, J., Stern, D. L., Morita, T., Prud'homme, B., and Gompel, N. (2014). Looking under the lamp post: neither *fruitless* nor *doublesex* has evolved to generate divergent male courtship in *Drosophila*. *Cell Rep.* 8, 363–370. doi: 10.1016/j.celrep.2014.06.023
- Carroll, S. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36. doi: 10.1016/j.cell.2008.06.030
- Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Shapiro, M. D., Brady, S. D., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 327, 302–305. doi: 10.1126/science.1182213
- Correa, R., and Baum, D. (2015). Evolutionary transgenomics: prospects and challenges. *Front. Plant. Sci.* 6:858. doi: 10.3389/fpls.2015.00858
- Correa, R., Stanga, J., Larget, B., Roznowski, A., Shu, G., Dilkes, B., et al. (2012). An assessment of transgenomics as a tool for identifying genes involved in the evolutionary differentiation of closely related plant species. *New Phytol.* 193, 494–503. doi: 10.1111/j.1469-8137.2011.03949
- Deal, R. B., and Henikoff, S. (2011). The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*. *Nat. Prot.* 6, 56–68. doi: 10.1038/nprot.2010.175
- Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., and Carroll, S. B. (2005). Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* 433, 481–487. doi: 10.1038/nature03235
- Gordon, K. L., Arthur, R. K., and Ruvinsky, I. (2015). Phylum-level conservation of regulatory information in nematodes despite extensive non-coding sequence divergence. *PLoS Genet.* 11:e1005268. doi: 10.1371/journal.pgen.1005268
- Gordon, K. L., and Ruvinsky, I. (2012). Tempo and mode in evolution of transcriptional regulation. *PLoS Genet.* 8:e1002432. doi: 10.1371/journal.pgen.1002432
- Hagemann, W., and Gleissberg, S. (1996). Organogenetic capacity of leaves: the significance of marginal blastozones in angiosperms. *Plant Syst. Evol.* 199, 121–152. doi: 10.1007/BF00984901
- Hanikenne, M., Talke, I. N., Haydon, M. J., Lanz, C., Nolte, A., Motte, P., et al. (2008). Evolution of metal hyperaccumulation required *cis*-regulatory changes and triplication of *HMA4*. *Nature* 453, 391–396. doi: 10.1038/nature06877
- Hay, A., and Tsiantis, M. (2006). The genetic basis for differences in leaf form between *Arabidopsis thaliana* and its wild relative *Cardamine hirsuta*. *Nat. Genet.* 38, 942–947. doi: 10.1038/ng1835
- Hoekstra, H. E., Hirschmann, R. J., Bunday, R. A., Insel, P. A., and Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313, 101–104. doi: 10.1126/science.1126121
- Jenner, R. A., and Wills, M. A. (2007). The choice of model organisms in evo-devo. *Nat. Rev. Genet.* 8, 311–314. doi: 10.1038/nrg2062
- Kalay, G., and Wittkopp, P. J. (2010). Nomadic enhancers: tissue specific *cis*-regulatory elements of yellow have divergent genomic positions among *Drosophila* species. *PLoS Genet.* 6:e1001222. doi: 10.1371/journal.pgen.1001222
- Kellogg, E. A. (2006a). Progress and challenges in studies of the evolution of development. *J. Exp. Bot.* 57, 3505–3516. doi: 10.1093/jxb/erl132

- Kellogg, E. A. (2006b). The difference between simple and compound leaves. *Nat. Genet.* 38, 865–866. doi: 10.1038/ng0806-865
- Kramer, E. M. (2015). A stranger in a strange land: the utility and interpretation of heterologous expression. *Front. Plant. Sci.* 6:734. doi: 10.3389/fpls.2015.00734
- Maizel, A., Busch, M. A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M., et al. (2005). The floral regulator LEAFY evolves by substitution in the DNA binding domain. *Science* 308, 260–263. doi: 10.1126/science.1108229
- Marcellini, S., and Simpson, P. (2006). Two or four bristles: functional evolution of an enhancer of *scute* in Drosophilidae. *PLoS Biol.* 4:e386. doi: 10.1371/journal.pbio.0040386
- Nelson, T., Tausta, S. L., Gandotra, N., and Liu, T. (2006). Laser microdissection of plant tissue: what you see is what you get. *Ann. Rev. Plant Biol.* 57, 181–201. doi: 10.1146/annurev.arplant.56.032604.144138
- Orgogozo, V., Morizot, B., and Martin, A. (2015). The differential view of genotype-phenotype relationships. *Front. Genet.* 6:179. doi: 10.3389/fgene.2015.00179
- Peter, I. S., and Davidson, E. H. (2011). Evolution of gene regulatory networks controlling body plan development. *Cell* 144, 970–985. doi: 10.1016/j.cell.2011.02.017
- Puchta, H. (2002). Gene replacement by homologous recombination in plants. *Plant Mol. Biol.* 48, 173–182. doi: 10.1023/A:1013761821763
- Rast-Somssich, M. I., Broholm, S., Jenkins, H., Canales, C., Vlad, D., Kwantes, M., et al. (2015). Alternate wiring of a KNOXI genetic network underlies differences in leaf development of *A. thaliana* and *C. hirsuta*. *Genes Dev.* 29, 2391–2404. doi: 10.1101/gad.269050.115
- Rowan, B. A., Weigel, D., and Koenig, D. (2011). Developmental genetics and new sequencing technologies: the rise of nonmodel organisms. *Dev. Cell.* 21, 65–76. doi: 10.1016/j.devcel.2011.05.021
- Ruvinsky, I., and Ruvkin, G. (2003). Functional tests of enhancer conservation between distantly related species. *Development* 130, 5133–5142. doi: 10.1242/dev.00711
- Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., et al. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428, 717–723. doi: 10.1038/nature02415
- Sicard, A., Thamm, A., Marona, C., Lee, Y. W., Wahl, V., Stinchcombe, J. R., et al. (2014). Repeated evolutionary changes of leaf morphology caused by mutations to a homeobox gene. *Curr. Biol.* 24, 1880–1886. doi: 10.1016/j.cub.2014.06.061
- Stern, D. L. (2014). Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. *Trends Genet.* 30, 547–554. doi: 10.1016/j.tig.2014.09.006
- Vlad, D., Kierzkowski, D., Rast, M. I., Vuolo, F., Dello Ioio, R., Galinha, C., et al. (2014). Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. *Science* 343, 780–783. doi: 10.1126/science.1248384
- Wang, W., Wu, Y., Li, Y., Xie, J., Zhang, Z., Deng, Z., et al. (2010). A large insert *Thellungiella halophila* BIBAC library for genomics and identification of stress tolerance genes. *Plant Mol. Biol.* 72, 91–99. doi: 10.1007/s11103-009-9553-3
- Winter, C. M., Austin, R. S., Blanvillain-Baufumé, S., Reback, M. A., Monniaux, M., Wu, M. F., et al. (2011). Leafy target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. *Dev. Cell.* 20, 430–443. doi: 10.1016/j.devcel.2011.03.019
- Yoon, H. S., and Baum, D. A. (2004). Transgenic study of parallelism in plant molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6524–6529. doi: 10.1073/pnas.0401824101
- Zhao, Y., Xu, T., Shen, C.-Y., Xu, G.-H., Chen, S.-X., Song, L.-Z., et al. (2014). Identification of a retroelement from the resurrection plant *Boea hygrometrica* that confers osmotic and alkaline tolerance in *Arabidopsis thaliana*. *PLoS ONE* 9:e98098. doi: 10.1371/journal.pone.0098098

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Nikolov and Tsiantis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evolutionary transgenomics: prospects and challenges

Raul Correa¹ and David A. Baum^{2*}

¹ Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA, ² Department of Botany, University of Wisconsin-Madison, Madison, WI, USA

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington, USA

Reviewed by:

Marcelo Camier Domelas,
Universidade Estadual de Campinas,
Brazil
William Oki Wong,
Institute of Botany, Chinese Academy
of Sciences, China

*Correspondence:

David A. Baum
dbaum@wisc.edu

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 01 August 2015

Accepted: 28 September 2015

Published: 20 October 2015

Citation:

Correa R and Baum DA (2015)
Evolutionary transgenomics:
prospects and challenges.
Front. Plant Sci. 6:858.
doi: 10.3389/fpls.2015.00858

Many advances in our understanding of the genetic basis of species differences have arisen from transformation experiments, which allow us to study the effect of genes from one species (the donor) when placed in the genetic background of another species (the recipient). Such interspecies transformation experiments are usually focused on candidate genes – genes that, based on work in model systems, are suspected to be responsible for certain phenotypic differences between the donor and recipient species. We suggest that the high efficiency of transformation in a few plant species, most notably *Arabidopsis thaliana*, combined with the small size of typical plant genes and their *cis*-regulatory regions allow implementation of a screening strategy that does not depend upon *a priori* candidate gene identification. This approach, transgenomics, entails moving many large genomic inserts of a donor species into the wild type background of a recipient species and then screening for dominant phenotypic effects. As a proof of concept, we recently conducted a transgenomic screen that analyzed more than 1100 random, large genomic inserts of the Alabama gladeless *Leavenworthia alabamica* for dominant phenotypic effects in the *A. thaliana* background. This screen identified one insert that shortens fruit and decreases *A. thaliana* fertility. In this paper we discuss the principles of transgenomic screens and suggest methods to help minimize the frequencies of false positive and false negative results. We argue that, because transgenomics avoids committing in advance to candidate genes it has the potential to help us identify truly novel genes or cryptic functions of known genes. Given the valuable knowledge that is likely to be gained, we believe the time is ripe for the plant evolutionary community to invest in transgenomic screens, at least in the mustard family Brassicaceae where many species are amenable to efficient transformation.

Keywords: developmental system drift, evolution, evo-devo, genetic screens, speciation genes, transgenomics, transformation

INTRODUCTION

At its most general, evolutionary developmental biology, “evo-devo,” seeks to understand how development, the translation of a genotype to a phenotype in a given environment, constrains, or enables phenotypic evolution (Stern, 2000; Wagner et al., 2000; Arthur, 2002; Carroll et al., 2004; Lee et al., 2014). The core data that are needed to achieve such understanding are genetic and developmental changes that have been shown experimentally to cause particular evolutionary transitions from ancestral to derived phenotypes. While much evo-devo research can focus on phenotypic variation within living populations, there has long been an interest in also studying

characters that differ between living species, where one species manifests the ancestral character state and the other manifests the derived. How, then, can we experimentally determine the genetic and developmental basis of traits that differ between species?

Until now, the search for genes responsible for species differences has mainly exploited either candidate gene or quantitative trait locus (QTL) approaches. Candidate gene methods use information from genetic model systems to hypothesize that a certain genetic change caused the transition from the ancestral to the derived character state and then set about to test that hypothesis using comparative studies of two (or more) species that differ for the trait. The test usually involves comparative expression studies combined with various functional studies, which might include knocking down gene expression and/or moving the candidate gene among species either by crossing or transgenic methods. However, candidate gene approaches are limited to phenotypes whose developmental basis is well understood in model genetic systems. While much of what we know about evo-devo comes from candidate gene studies, they have limitations. In particular, when a phenotype is caused by unpredictable genes, as can happen due to neofunctionalization [e.g., (Vlad et al., 2014)], candidate gene approaches will come-up empty. Indeed, using only a candidate gene approach would make it very hard to answer one of the key questions in evo-devo: how many different genetic pathways are available for the evolution of a new trait?

Currently, the main alternative to candidate gene approaches is QTL analysis. This involves crossing two species with contrasting traits and then looking for cosegregation of the trait with genetic markers in the F2 or later generations. The goal is to positionally clone genes causing phenotypic differences and, eventually, home in on the sequence difference causing the trait difference. However, QTL analysis is limited to cases where species are capable of being crossed. Further, it is notoriously difficult to clone the gene underlying a QTL in non-model species.

In this paper we will argue that evolutionary transgenomics (Baum, 2002) represents a third method, complementary to the other two, that could be used to identify the genes responsible for species' differences. An evolutionary transgenic screen involves introducing genomic fragments of a donor species into the genome of a recipient species (or perhaps a divergent ecotype) and screening the resulting transgenic lines for phenotypic effects. Such screens have the advantage of not being limited to crossable species and yet being able to find genes that would not have been predicted *a priori*. We will suggest that while transgenomics poses practical challenges, it has great potential value for plant evo-devo research in those taxa that are readily transformed and may help us identify many genes of evolutionary and developmental interest that would otherwise be difficult to discover.

Systematic screens in yeast, including screens of plant cDNA libraries, have successfully been used to identify genes controlling cell-level phenotypes [e.g., (Papayan and Kochian, 2004; Liu et al., 2007)]. However, there have been few attempts to screen the genome of one multicellular eukaryote in that of another. Because there are many plants that are closely related, show

many phenotypic differences, have compact genomes, and are readily transformed, plants are better suited to transgenomics than are most animals. But are the benefits to be gained likely to outweigh the work entailed in conducting evolutionary transgenic experiments? In this paper we use evolutionary theory and data from prior interspecies transformation experiments and our published, pilot transgenic screen (Correa et al., 2012) to assess the approach and how it could best be implemented. We conclude that the time is ripe to develop transgenomics resources at least in Brassicaceae Burnett, a clade of flowering plants, many of whose species can now be transformed with high efficiency.

PRINCIPLES OF TRANSGENOMICS

In an evolutionary transgenic screen, fragments of genomic DNA from a donor species are added to the wildtype genome of a recipient species. We then screen primary transformants (T1s) to look for phenotypic effects that might be due to the inserted DNA. Since plant transformation usually entails inserting an extra piece of DNA rather than homologous replacement of endogenous sequences, inserts will only cause phenotypes in T1s if they act in a *transdominant* manner. That is to say, one copy of the foreign gene must manifest a phenotypic effect even in the presence of two functional copies of that gene (if any exist in the recipient genome). Genetic theory suggests two primary causes of a *transdominant* phenotype.

The first potential cause of a *transdominant* phenotypic effect is the addition of a supernumerary gene copy to the genome, a gene dosage effect. It is well documented that changes in gene dosage can have phenotypic effects [reviewed by (Birchler and Veitia, 2007)]. This is most obvious when aneuploids (e.g., trisomics) yield distinct phenotypes, including lethality or sterility. Dosage effects presumably result from additional copies altering the balance of expression of genes in regulatory pathways. Addition of a single additional gene (as would occur in a hemizygous T1 transgenic plant) might be expected to increase expression level by approximately 50%, or higher if multiple insertions of the transgene occur. However, the actual effect on expression will vary subject to position effects and whether or not transgene silencing is triggered.

It is not known how often dosage alone will yield a visible phenotype in a transgenic line, and this might vary depending on the phylogenetic distance between donor and recipient species. However, based on prior transgenic data it seems likely that many dosage effects will primarily be quantitative. For example, it has been found that adding extra *Arabidopsis thaliana* (L.) Heynh. *LFY* transgenes to a wildtype *A. thaliana* background has a dosage dependent effect on flowering time (Blazquez et al., 1997).

Dosage effects do not depend on sequence divergence between donor and recipient species. Quite the contrary – a dosage effect depends upon conservation of molecular function between the endogenous and exogenous gene copies. This leads to a powerful test for discriminating dosage effects from other mechanisms: if the phenotypic effect can be replicated by introducing the homologous fragment of the recipient species back into the

recipient species, then dosage is likely to be the cause of the observed phenotype. If, there is no homologous region, or if the homologous region fails to cause the phenotype, then dosage is unlikely to be responsible.

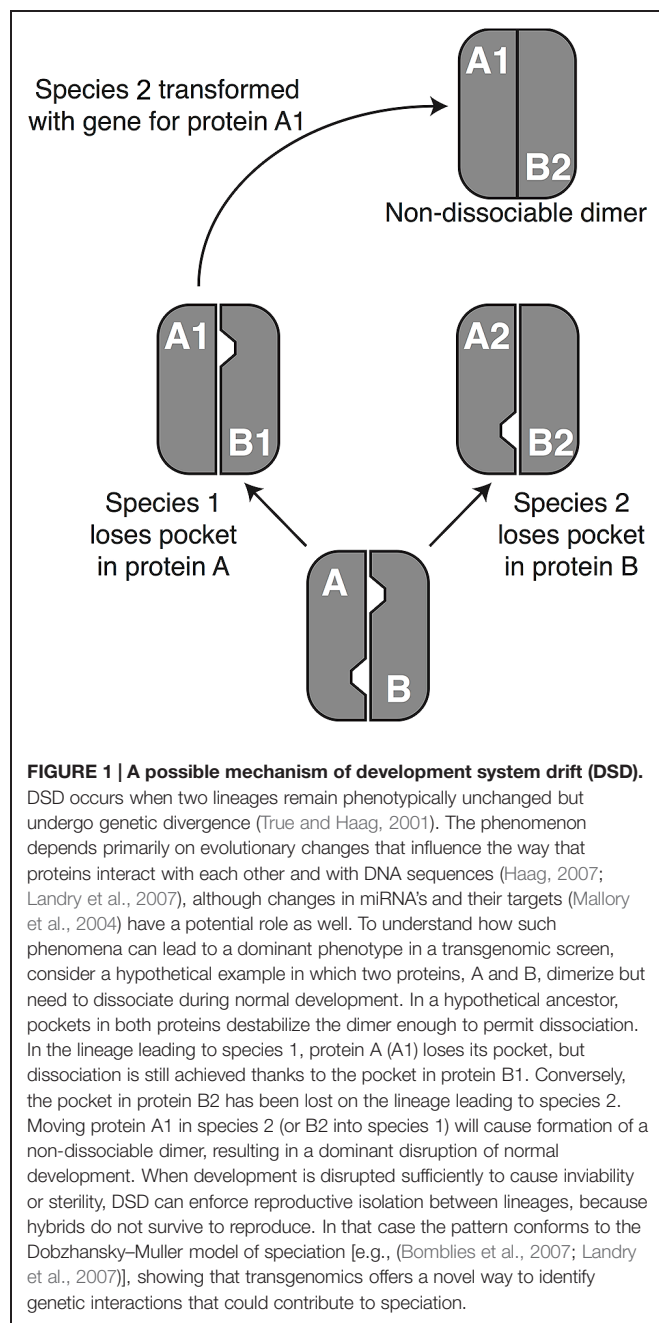
The second potential cause of a *transdominant* phenotype is evolutionary divergence between the donor and recipient genes. This could arise through one of two mechanisms, developmental system drift (DSD) or phenotypic divergence. DSD occurs when proteins and/or regulatory DNA/RNA sequences coevolve without altering visible phenotypes (True and Haag, 2001). The “drift” in underlying molecular mechanisms can cause a gene from the donor species to malfunction in the recipient genome in a such a way that a phenotype is seen, analogous to transgressive segregation, which is often seen in QTL studies (Rieseberg et al., 1999). To make the concept more concrete, **Figure 1** shows a hypothetical example involving a protein with two subunits that must be disassembled for proper development, with disassembly requiring at least one “pocket” of low attraction between the two subunits. Reciprocal loss of the pocket in the two subunits could result in a case in which subunit A from species 1 yields a dominant-negative phenotypic effect when placed in the genome of species 2. It is worth noting the similarities between this DSD model and Dobzhansky–Muller interactions (Kondrashov et al., 2002; Bomblies et al., 2007). Indeed, one exciting aspect of transgenomics is its potential to identify potential hybrid inviability genes.

The alternative explanation of a *transdominant* phenotype is that phenotypic evolution has been driven by sequence evolution (whether in coding or regulatory regions) at genes of large effect. In this case, a gene can carry the donor species’ phenotype into the recipient species’ genome (**Figure 2**). This is expected to happen when a difference in phenotype is due to a fully or partially dominant mutation on the lineage leading to the donor species or to a recessive mutation on the lineage leading to the recipient species (**Figure 3**). Assuming that the relative frequency of dominant and recessive mutations is about equal on the two evolutionary lineages, first principles would suggest that a complete transgenomic screen would be able to detect about 50% of the major genes explaining phenotypic differences between the donor and recipient species.

If we knew how many of the phenotypic differences between species were due to genes of large effect we could predict the frequency with which transgenomic lines will manifest a phenotype that resembles the donor species. Unfortunately, we are not aware of any relevant quantitative data. Indeed, one of the most compelling reasons to conduct transgenomic screens is because they will help quantify the frequency of evolution via genes of large effect, something that has long been a source of controversy (Gottlieb, 1984; Doebley and Lukens, 1998; Hoekstra and Coyne, 2007).

ALTERNATIVE TRANSGENOMIC STRATEGIES

In considering transgenomics two main experimental approaches suggest themselves. One approach is a shotgun strategy, where



we generate a genomic library of a donor species in a suitable bacterium, e.g., *Agrobacterium tumefaciens* (Smith et Town.) Conn, and then introduce this *en masse* (or perhaps in pools) into a population of recipient plants (**Figure 4**). T1s would be screened for phenotypes of interest and, when a phenotype is observed, we would determine *post hoc* what insert had been introduced into the recipient's genome. Alternatively, a clone-by-clone strategy can be followed in which we isolate individual clones from a genomic library and use each clone for multiple transformations of the same recipient species to identify repeatable phenotypic effects (**Figure 4**).

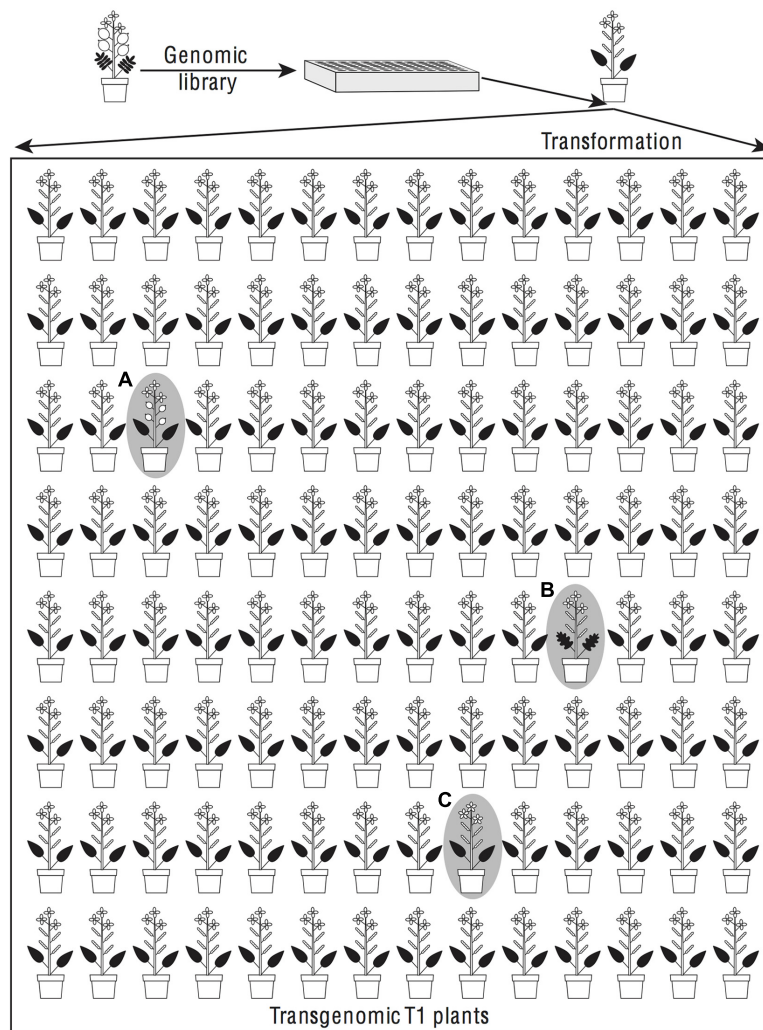


FIGURE 2 | Dominant phenotypic effects found in a transgenic screen could reflect phenotypic divergence or developmental system drift between donor and recipient species. In a transgenic screen random genomic fragments from a donor species are introduced by transformation into a recipient species. The transformants are screened for phenotypes that differ from the recipient species. When an insert causes a phenotype of the donor species to be found, for example round, flattened fruits (**plant A**), it is likely that the transgene contains a major gene contributing to the evolution of that phenotype. If the phenotype is intermediate between the donor and recipient species, for example pinnately lobed (**plant B**) rather than either entire or pinnately compound leaves, then the transgene is a candidate for being one of several genes that changed during phenotypic divergence. If the phenotype resembles neither the donor nor the recipient species, as for example in having five rather than four petals (**plant C**), then developmental system drift is suggested.

A shotgun strategy has the advantage of quickly generating large populations of transformant plants to screen for phenotypes of interest. Nonetheless, it does have some serious drawbacks. (1) When interesting phenotypes are seen it might not be trivial to isolate the responsible genomic fragment, especially for large inserts or for inserts that cause sterility meaning that one could not obtain much transgenic plant tissue. (2) Because each clone will be introduced into only one recipient, a phenotypic effect could be a false positive due to genetic or microenvironmental differences among transformed plants. (3) Clones that cause dominant early lethality will not be identified so their frequency in the genome could not be assessed. (4) A shotgun-transformed

pool could not readily interface with existing genomic information to yield a durable resource for other researchers to utilize.

A clone-by-clone strategy has the virtue that the identity of inserts is easily determined (by sequencing the clone) and one can obtain multiple T1 plants per clone, reducing the false positive problem. Furthermore, inserts causing early lethality can, at least theoretically, be identified by their inability to generate mature transgenic plants and, once a population of transgenic lines has been made, it represents a durable resource that could be screened repeatedly for different phenotypes in a diversity of growth conditions. On the other hand, a clone-by-clone strategy requires more work to separate and bulk-up individual

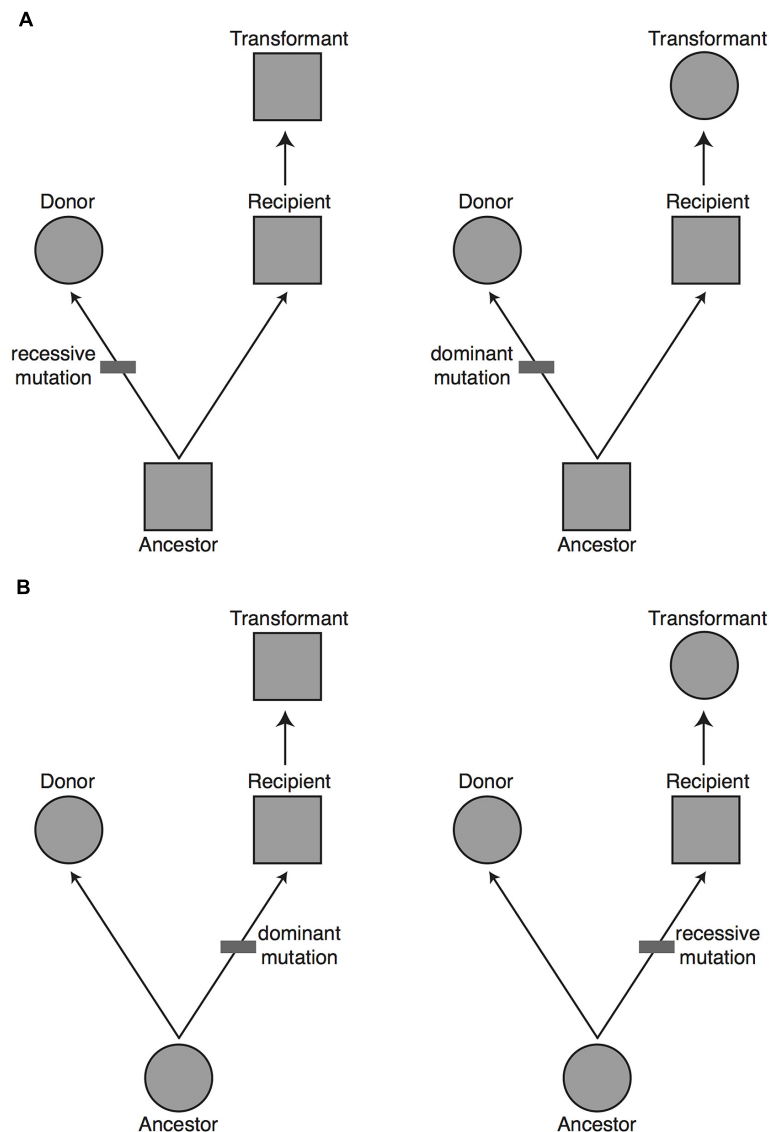


FIGURE 3 | The proportion of major genes expected to be uncovered in a transgenomic screen. Based on evolutionary principles we would predict that 50% of the major genes responsible for a phenotypic difference between the donor and recipient species will be found in a unidirectional transgenomic screen. To see why this is so, imagine a potential donor and recipient species for a transgenomic screen that differ in a phenotype (circle vs. square, respectively). Without further information it is equally likely that: **(A)** the donor species has the derived phenotype, with a change having occurred on the lineage from the common ancestor to the donor species (upper panels), or **(B)** the donor has the ancestral phenotype, with a change having occurred on the lineage from the common ancestor to the recipient species (lower panels). The mutation that gave rise to the derived phenotype could have been fully recessive or at least partially dominant. If the donor has a derived phenotype that is dominant (top right), or it has an ancestral phenotype that is dominant (bottom right), then moving the causal gene into the recipient will yield a dominant phenotype. In approximately 50% of cases (left panels) the causal gene will not yield a dominant phenotype when moved from the donor to the recipient. The major genes missed in a unidirectional transgenomic screen could theoretically be found with a reciprocal screen in which the genome of the former recipient species is screened in the background of the former donor species.

clones and requires one to do many more individual plant transformations.

Whether one uses a shotgun or clone-by-clone approach, once transgenomics lines are found to have phenotypic effects a number of different downstream experiments can be undertaken. Before, or in parallel with, standard methods for studying gene-function (e.g., isolation of T-DNA insert lines, double mutant assays, expression studies) experiments should be

conducted to assess the role of sequence divergence between donor and recipient species in explaining the phenotype. Some critical experiments will include: (1) repeating transformation to confirm that the phenotype is caused by the insert; (2) subcloning the insert to identify the causal gene region; (3) introducing the homologous gene from the recipient species as an extra copy to assess if the result is due to gene dosage; (4) generating chimeric constructs between the donor and recipient genes to

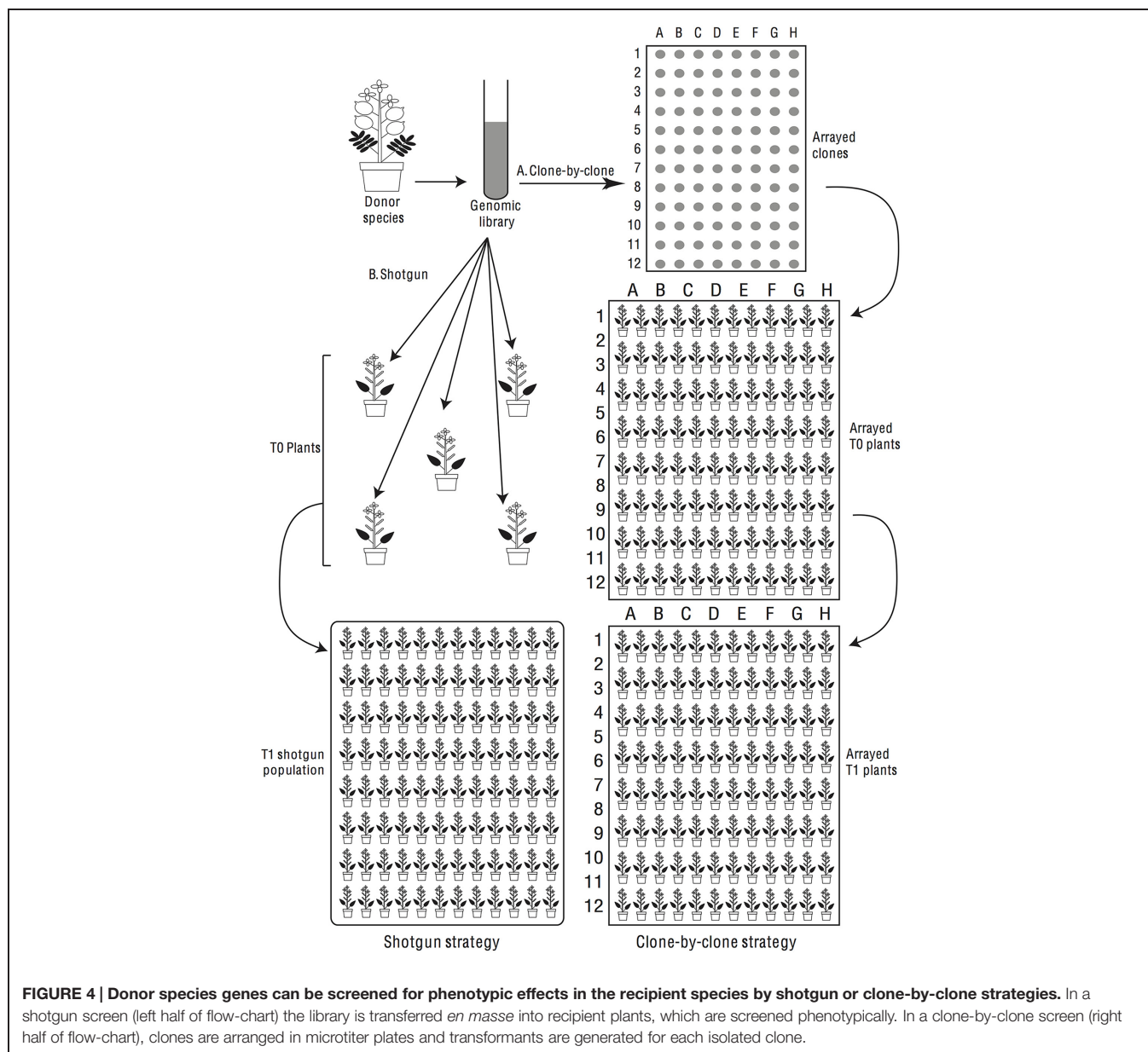


FIGURE 4 | Donor species genes can be screened for phenotypic effects in the recipient species by shotgun or clone-by-clone strategies. In a shotgun screen (left half of flow-chart) the library is transferred *en masse* into recipient plants, which are screened phenotypically. In a clone-by-clone screen (right half of flow-chart), clones are arranged in microtiter plates and transformants are generated for each isolated clone.

locate the causal differences, and; (5) isolating the homologous gene from additional species to assess the correlation between the phenotype of the donor species and the ability of the gene to confer an effect in transgenic lines of the recipient species. Through such experiments, there is every reason to hope that we could eventually arrive at a clear understanding of the molecular and developmental basis of the transgene's effect on phenotype and, in some cases, shed light on the evolution of phenotypic differences between species.

A PILOT CLONE-BY-CLONE SCREEN

Correa et al. (2012) piloted a clone-by-clone transgenomic screen using genomic clones with ~20 kb inserts from the gladecress

Leavenworthia alabamica Rollins introduced into *A. thaliana*. *Leavenworthia alabamica*, and *A. thaliana* differ in almost all visible morphologies and yet they are both members of Brassicaceae Lineage I (Beilstein et al., 2008, 2010; Franzke et al., 2011). While this was a large experiment, high-throughput “drip” transformation allowed one graduate student and 2–3 undergraduate assistants to screen T1s for as many as 750 clones per month.

Of the 1134 *L. alabamica* clones screened, 84 produced an initial T1 that deviated from “normal”. However, in only eight cases was the initial “abnormal” phenotype repeated in additional independent T1s. Correa et al. (2012) focused on one clone that was shown to cause stunted fruit and increased seed abortion in a *transdominant* manner. Sequencing of the clone insert suggests that this effect is most likely explained by a gene region that

shows homology to the *A. thaliana* SLOW-WALKER2 (SWA2) gene. SWA2 encodes a protein that is important for ribosome biogenesis, and *A. thaliana* *swa2* mutants manifest shortened fruit and seed abortion (Li et al., 2009). Follow-on experiments have confirmed that a subclone containing the region of SWA2 homology is sufficient to cause these phenotypes and have detected *L. alabamica* SWA2-like mRNA in developing fruit (Wu and Baum, unpublished data).

What lessons were learned thanks to this pilot transgenomic screen? On the plus side, we established that a clone-by-clone strategy is feasible, at least when *A. thaliana* is the recipient, and showed that it is possible to use this strategy to identify a gene that can alter the phenotype of a recipient in a *transdominant* manner. Back-of-the-envelope calculation would suggest that if 750 clones can be screened per month by a small team piloting the approach for the first time, a larger and more experienced team could realistically screen more than 10,000 clones per year making it plausible that one could screen a moderate size genome to saturation.

COULD TRANSGENOMICS IDENTIFY MAJOR GENES EXPLAINING SPECIES DIFFERENCES?

The single gene identified by Correa et al. (2012) is likely to reflect DSD rather than phenotypic divergence. The question this raises is whether this result indicates that the method has limited utility for scientists whose goal is to identify genes that explain species differences.

The first fact to emphasize is that the screen conducted by Correa et al. (2012) covered less than 5% of the donor species' genome. Furthermore, considering that with a false negative rate of at least 50%, only 2.5% of the donor genome was effectively screened. This means that it would be grossly premature to take the results of Correa et al. (2012) as indicating that transgenomics cannot find species-differentiating genes. Thus, the best we can do is to look at prior experiments in which single genes have been moved between closely species to evaluate how often we might expect to find a dominant phenotypic effect in a transgenomic screen that is indicative of some functional role in explaining species differences (Correa et al., 2012).

A majority of experiments in which a gene (coding region or cDNA) from one species is moved into a different species use either a broadly active promoter such as 35S or the homologous promoter from the recipient species. The most common result from such studies is that the exogenous gene functions equivalently to the endogenous gene [e.g., (Whipple et al., 2004; Maizel et al., 2005; Dornelas and Rodriguez, 2006; Busch and Zachgo, 2007)]. Sometimes, especially with distantly related donors, the exogenous gene shows reduced functionality resembling a partial loss-of-function allele [e.g., (Tzeng and Yang, 2001; Maizel et al., 2005)]. Some studies using 35S have combined transformation data with evidence on comparative gene expression to show that altered expression of a single functionally conserved gene likely contributed to the evolution of plant phenotypes (Doczi et al., 2005; He and Saedler, 2005; Lee

et al., 2005; Maizel et al., 2005; Zahn et al., 2005; Hay and Tsiantis, 2006; Busch and Zachgo, 2007; Hovav et al., 2007). In one case, the dominant effect of a full-length transgene was demonstrated using introgression, rather than transformation (Hovav et al., 2007). Likewise, some experiments using 35S have yielded novel phenotypes, with examples including pathogen resistance (Lee and Lee, 2005; Lee et al., 2005; Zahn et al., 2005; Carlson et al., 2006), stress-tolerance (Hsieh et al., 2002; Liu et al., 2007; Wang et al., 2007), the production of novel secondary metabolites (de Majnik et al., 2000; Mietkiewska et al., 2004; Mori et al., 2004), and even donor species-like morphology (Fourquin et al., 2013) that hint at evolutionary divergence in protein function sufficient to cause a dominant phenotype in a transgenomics screen.

Another class of interspecies transformation experiments move homologous *cis*-regulatory ("promoter") regions with reporters from one species to another to see if expression patterns are conserved [e.g., (Lin et al., 1993; Burger et al., 2006)] or divergent [e.g., (Doczi et al., 2005; Hay and Tsiantis, 2006)]. When a donor species' promoter drives expression in developmental stages or tissue types where the recipient species' promoter is inactive [e.g., (Zahn et al., 2005; de Martino et al., 2006)] there is a potential for a dominant phenotypic effect to be found in a transgenomics screen.

More direct evidence comes from those few studies that have moved genomic fragments with both *cis*-regulatory and coding regions between species. One example comes from work on the evolution of self-compatibility in *A. thaliana*. Genomic fragments from the S-locus of the self-incompatible *A. lyrata* (L.) O'Kane et Al-Shehbaz, when used to transform wildtype plants of the self-compatible *A. thaliana*, converted the latter to self-incompatibility – the effect varying with ecotype of the recipient plants (Nasrallah et al., 2002, 2004). This effect appears to be due to divergence in the coding region (loss of gene function in *A. thaliana*) rather than the evolution of regulatory regions. Similarly, Vlad et al. (2014) used transformation to show that the presence of the gene *REDUCED COMPLEXITY* (*RCO*) in the genomes of *A. lyrata* and *Cardamine hirsuta* L., but its loss in *A. thaliana*, largely explains the dissected leaf shape of the former species and the simple leaves of the latter. Another example involves the introduction of *LFY* and *TFL1* genes from different Brassicaceae species into *A. thaliana*. In all three cases involving *LFY* (Yoon and Baum, 2004; Sliwinski et al., 2007) and the one case involving *TFL1* (Liu et al., 2011), the transgene resulted in a novel phenotype and these were shown to also occur in a wildtype background showing transdominance (Sliwinski et al., 2006, 2007; Liu et al., 2011). Similarly, studies of *REPLUMLESS* (*RPL*), a gene that promotes fruit dehiscence, showed that introducing the *Arabidopsis* *RPL* gene into *Brassica* is sufficient to induce *Arabidopsis*-like fruit dehiscence in the recipient (Arnaud et al., 2011). As was also found in the *LFY* and *TFL1* experiments, the transgene effect is not replicated when using the endogenous gene copy, showing that the effect is due to sequence divergence (specifically in *cis*-regulatory regions) rather than to a gene dosage effect.

Taken together, evolutionary theory and prior candidate gene interspecies transformation experiments suggest that a transgenomics screen has a high potential to uncover

transdominant phenotypic effects. Given this we should assess how best to design future transgenomic screens to maximize their efficiency.

PRACTICALITIES OF SCREENING

One of the striking findings of Correa et al. (2012) was the high false positive rate. Specifically, out of 84 cases where an abnormal phenotype was seen in an initial T1, only eight recurred in further T1s from the same clone, and only one was definitively shown to be due to a *L. alabamica* insert. It is not surprising that false positives arise since there is likely to be genetic variation among the plants used for transformation and even wildtype plants occasionally manifest phenotypic abnormalities (Hempel and Feldman, 1995). Furthermore, transformation and *Agrobacterium* infection are both potentially mutagenic, so some abnormalities likely reflect *de novo* mutation. However, although some false positives are inevitable, it is worth considering strategies for reducing their frequency so as to avoid wasting time and effort following up phenotypes that are not caused by the transgene.

One strategy that we have explored is to immediately grow-up and screen ~5 T1s per clone instead of just one. Once one is screening seedlings on plates for a selective resistance trait, it is not much more difficult to transplant and retain five rather than one T1. Requiring that phenotypes recur in at least 2–3 of the independent T1s from the same clone would go a long way toward excluding phenotypic effects that are due to a position effect or an insertional mutation in an endogenous gene, since independent T1s are not expected to have their inserts integrated at the same *A. thaliana* locus. Furthermore, if the locations of plants used to test a particular clone are randomized within growth rooms, there is little chance that multiple T1s would have the same phenotype because of a shared microenvironment.

An additional benefit of screening several T1s per clone is that it should also reduce the number of false negatives: overlooking a causally important clone due to a lack of a visible phenotype in the initial T1 screened. Despite clear evidence that the clone containing the SWA2-like region causes increased seed abortion and reduced fruit size, Correa et al. (2012) noted that about one half of the T1s containing the clone showed a wild type morphology. This matches other experiments that have reported that many clones fail to manifest a phenotype in *A. thaliana* due to transgene silencing (Morel et al., 2000; Schubert et al., 2004). However, while there are benefits to screening several rather than just one T1, such an approach significantly increases the space needed to conduct the initial screen. So, if space limitations rather than labor limitations are paramount, this strategy might not be worth deploying.

We have also tried a further embellishment to reduce the impact of genetic differences among the transformed (T0) plants. This involved growing ~5 independent T1 plants germinated as usual on selective agar plates and in parallel growing the same number of plants from the same seed stock but isolated from non-selective plates. Based on *Arabidopsis* transformation efficiency, non-selectively grown plants are much (ca. two orders

of magnitude) more likely to lack donor species DNA than to contain it. This means that phenotypic effects that recur in the T1 plants but are absent in non-selectively grown siblings are much more likely to be due to a *transdominant* donor species gene. However, this represents a further doubling of the space required and additional work for making plates, transplanting seedlings, and scoring plants for abnormalities. Based on our informal experimentation, we are doubtful that this additional work would yield sufficient benefits to make it worthwhile.

CHOICE OF DONOR AND RECIPIENT SPECIES

Whichever transgenomic strategy is used, the choice of donor and recipient species is important. As the preeminent plant genetic model system, blessed with efficient transformation methods, *A. thaliana* is the ideal recipient species with which to initially implement and assess a transgenomics approach. While other readily transformed plant species such as rice, tobacco, and petunia are worth considering, we believe that the efficiency of dip-transformation makes Brassicaceae the clade of choice in which to first try transgenomic research.

The strengths of transgenomics would be greatest when working with a donor species that is closely related to the recipient so that many of the core developmental process and genes are shared, but distant enough that many visible phenotypes differ. Alternatively, if one is committed to a particular trait (e.g., salt tolerance, metal hyperaccumulation, leaf shape, etc.), one can pick a donor species that differs from *A. thaliana* in at least that phenotype. What we do not know is how far from the phylogenetic neighborhood of *A. thaliana* you can go before most phenotypic effects are hard to make sense of: *Brassica* L., *Cleome* L., papaya, cotton, tobacco, rice, moss?

An additional consideration is the availability of genetic tools in the donor species. If the donor genome has been sequenced it will be that much easier to home in on causal regions once an interesting phenotype has been found. Also, if the donor species is one that can be transformed then it will be possible to introduce *A. thaliana* genes, which may be useful for exploring gene function. Furthermore, if both species can be transformed with high-throughput methods, it becomes possible to undertake a full, bidirectional transgenomics screen, which should allow one to identify almost all genes of large effect that have contributed to the phenotypic divergence of the two species (Figure 3).

CLONING STRATEGY

Once the donor and recipient species have been identified, a number of detailed practical issues will need to be addressed, many of which could significantly influence the efficacy of the screen. Among these is the strategy used to assemble a genomic library, most notably insert size, choice of vector, selectable marker, and possible target-enrichment strategy.

Large inserts will presumably allow for more rapid screening of a genome to completion. A further advantage of long inserts

is they are less likely to include truncated genes, which can yield false positive results if, for example, a truncated donor species protein binds to and prevents a partner protein from interacting normally with the homologous recipient species protein (Villagarcia et al., 2012) or if the transgene region is missing a negative regulatory protein domain or a *cis*-repressor element. On the other hand, longer clone inserts are more difficult to work with in the lab and typically show lower transformation efficiencies. Furthermore, inserts that are too large to allow amplification from T-DNA primers will make it markedly more difficult to isolate the insert sequence from transformant plants, as might be necessary for a shotgun transgenomic screen.

Vectors should be ones that can achieve sufficient transformation efficiency for inserts of the target size, yet should tend to yield only one insert per transgenic line so as to minimize confounding dosage effects. Additionally, all things being equal, it might be helpful to engineer a vector that results in inserts being flanked by insulator sequences (She et al., 2010; Singer and Cox, 2013). By reducing the extent to which gene expression is affected by where in the genome an insert lands, insulator sequences might reduce line-to-line variability and, thus, lower the false negative rate.

The choice of selectable marker will be guided by speed of screening as well as the desirability of being able to identify transgenic lines at a very early developmental stage so that genes causing early lethality can be found. For example, it might be possible to use a selectable marker that causes embryos to fluoresce (Ali et al., 2012), allowing transformants to be visually identified as seeds and then grown up, non-selectively, on soil.

The final factor to consider in designing a transgenomic screen is whether it might be possible to manipulate the genomic library to increase the proportion of clones that includes potentially causal genes. While transgenomics is premised on the idea that we want to focus on genes with their native *cis*-regulatory machinery, meaning that we must include abundant non-coding content in our inserts, there is certainly much of the genome that is, *a priori*, less likely to cause informative *transdominant* phenotypes in a foreign genome. Our initial focus might be to enrich for the gene-space or, if we are focusing on a particular phenotype, we might be specifically interested in enriching for genes that are expressed in a particular organ. There is a diversity of methods available for enriching genomic libraries (Cronn et al., 2012), though most are optimized for short rather than long inserts. Possibilities include enriching for low-copy number genes based on melting kinetics (Yuan et al., 2003) or methylation (Palmer et al., 2003) or using hybridization against expressed genes (Lovett et al., 1991; Fu et al., 2010). While further method development would be needed, there is abundant scope for generating transgenomic libraries where a far higher proportion of clones contain causally important genes, thereby making screens dramatically more efficient.

PROSPECTS FOR TRANSGENOMICS

Transgenomics has great potential for contributing to the study of gene function in genetic model species and for identifying the molecular changes that underlie phenotypic evolution. While transgenomic screens will require a significant investment of effort and money, evolutionary theory and data from candidate gene transformation experiments show that, at least in Brassicaceae, the costs may well be outweighed by the fundamental data that will be obtained.

The initial ventures in transgenomics should aim to answer some important outstanding questions. In particular it is critical that an effort be made to quantify the proportion of a genome that causes different kinds of dominant phenotypes (morphological changes, sterility, lethality, etc.) as a function of phylogenetic distance between donor and recipient species. Further, there would be great value in studying two closely related species and conducting a complete, reciprocal transgenomics screen in order to determine the total number of major genes responsible for their different phenotypes. Lastly, we believe it would be beneficial for at least the *Arabidopsis* community to invest in an ordered transgenomics resource deposited in stock centers. If sets of sequenced clones from different donor species were each associated with transgenic seed, a researcher interested in a particular gene family could order up seed containing exogenous versions of that gene to look for phenotypes that may reveal aspects of gene function. Similarly, a scientist seeking new genes involved in the development of a phenotype could identify a donor and recipient species pair that differ in the phenotype and then could screen the corresponding transgenomic lines.

Looking to the future, it seems likely that transgenomics will emerge as an important tool for basic plant research. It is also possible that transgenomics could have applied importance. As a complement to traditional mass selection and candidate gene genetic modification, breeders may find it useful to introduce genetic variation in bulk from foreign species followed by screening and selection based on desirable traits that emerge. For all these reasons, we hope that this article will stimulate efforts to develop transgenomics as a tool for plant genetic research.

ACKNOWLEDGMENTS

We are very grateful to the following for thoughtful discussion and helpful suggestions over the years we have been exploring transgenomics: S. Cutler, B. Deakin, B. Dilkes, J. Doebley, B. Glover, E. Kramer, P. Krysan, N. Liu, P. Masson, T. C. Pires, S. D. Smith, K. Shimizu, M. K. Sliwinski, J. Stanga, and D. Weigel. DB gratefully acknowledges funding from the NSF (IOS-0641428), the UW-Madison Vilas fund, and the Guggenheim Foundation. We thank Melody Wu for her work on *L. alabamica* SWA2. Artwork was prepared with assistance from Kandis Elliot.

REFERENCES

- Ali, M. A., Shah, K. H., and Bohlmann, H. (2012). pMAA-Red: a new pPZP-derived vector for fast visual screening of transgenic *Arabidopsis* plants at the seed stage. *BMC Biotechnol.* 12:37. doi: 10.1186/1472-6750-12-37
- Arnaud, N., Lawrenson, T., Østergaard, L., and Sablowski, R. (2011). The same regulatory point mutation changed seed-dispersal structures in evolution and domestication. *Curr. Biol.* 21, 5–5. doi: 10.1016/j.cub.2011.06.008
- Arthur, W. (2002). The emerging conceptual framework of evolutionary developmental biology. *Nature* 415, 757–764. doi: 10.1038/415757a
- Baum, D. A. (2002). "Identifying the genetic causes of phenotypic evolution: a review of experimental strategies," in *Developmental Genetics and Plant Evolution*, eds Q. C. B. Cronk, R. M. Bateman, and J. A. Hawkins (London: Taylor & Francis), 493–507.
- Beilstein, M. A., Al-Shehbaz, I. A., Mathews, S., and Kellogg, E. A. (2008). Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: tribes and trichomes revisited. *Am. J. Bot.* 95, 1307–1327. doi: 10.3732/ajb.0800065
- Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R., and Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18724–18728. doi: 10.1073/pnas.0909766107
- Birchler, J. A., and Veitia, R. A. (2007). The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* 19, 395–402. doi: 10.1105/tpc.106.049338
- Blazquez, M. A., Soowal, L. N., Lee, I., and Weigel, D. (1997). LEAFY expression and flower initiation in *Arabidopsis*. *Development* 124, 3835–3844.
- Bombliès, K., Lempe, J., Epple, P., Warthmann, N., Lanz, C., Dangel, J. L., et al. (2007). Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol.* 5:e236. doi: 10.1371/journal.pbio.0050236
- Burger, A. L., Watts, L., and Botha, F. C. (2006). Grapevine promoter directs gene expression in the nectaries of transgenic tobacco. *Physiol. Plant.* 126, 418–434. doi: 10.1111/j.1399-3054.2006.00598.x
- Busch, A., and Zachgo, S. (2007). Control of corolla monosymmetry in the Brassicaceae *Iberis amara*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16714–16719. doi: 10.1073/pnas.0705338104
- Carlson, A., Skadsen, R., and Kaeppler, H. F. (2006). Barley hordothionin accumulates in transgenic oat seeds and purified protein retains antifungal properties in vitro. *In Vitro Cell. Dev. Biol. Plant* 42, 318–323. doi: 10.1079/IVP2006780
- Carroll, S. B., Grenier, J. K., and Weatherbee, S. D. (2004). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Oxford: Wiley-Blackwell.
- Correa, R., Stanga, J., Larget, B., Roznowski, A., Shu, G., Dilkes, B., et al. (2012). An assessment of transgenomics as a tool for identifying genes involved in the evolutionary differentiation of closely related plant species. *New Phytol.* 193, 494–503. doi: 10.1111/j.1469-8137.2011.03949.x
- Cronn, R., Knaus, B. J., Liston, A., Maughan, P. J., Parks, M., Syring, J. V., et al. (2012). Targeted enrichment strategies for next-generation plant biology. *Am. J. Bot.* 99, 291–311. doi: 10.3732/ajb.1100356
- de Majnik, J., Weinman, J. J., Djordjevic, M. A., Rolfe, B. G., Tanner, G. J., Joseph, R. G., et al. (2000). Anthocyanin regulatory gene expression in transgenic white clover can result in an altered pattern of pigmentation. *Austral. J. Plant Physiol.* 27, 659–667.
- de Martino, G., Pan, I., Emmanuel, E., Levy, A., and Irish, V. F. (2006). Functional analyses of two tomato APETALA3 genes demonstrate diversification in their roles in regulating floral development. *Plant Cell* 18, 1833–1845. doi: 10.1105/tpc.106.042978
- Doczi, R., Kondrak, M., Kovacs, G., Beczner, F., and Banfalvi, Z. (2005). Conservation of the drought-inducible DS2 genes and divergences from their ASR paralogues in solanaceous species. *Plant Physiol. Biochem.* 43, 269–276.
- Doebley, J., and Lukens, L. (1998). Transcriptional regulators and the evolution of plant form. *Plant Cell* 10, 1075–1082. doi: 10.2307/3870712
- Dornelas, M. C., and Rodriguez, A. P. M. (2006). The tropical cedar tree (*Cedrela fissilis* Vell., Meliaceae) homolog of the *Arabidopsis* LEAFY gene is expressed in reproductive tissues and can complement *Arabidopsis* leafy mutants. *Planta* 223, 306–314. doi: 10.1007/s00425-005-0086-y
- Fourquin, C., Del Cerro, C., Victoria, F. C., Viallette-Guiraud, A., De Oliveira, A. C., and Ferrandiz, C. (2013). A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in *Medicago* genus. *Plant Physiol.* 162, 907–917. doi: 10.1104/pp.113.217570
- Franzke, A., Lysak, M. A., Al-Shehbaz, I. A., Koch, M. A., and Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci.* 16, 9–9. doi: 10.1016/j.tplants.2010.11.005
- Fu, Y., Springer, N. M., Gerhardt, D. J., Ying, K., Yeh, C. T., Wu, W., et al. (2010). Repeat subtraction-mediated sequence capture from a complex genome. *Plant J.* 62, 898–909. doi: 10.1111/j.1365-313X.2010.04196.x
- Gottlieb, L. D. (1984). Genetics and morphological evolution in plants. *Am. Nat.* 123, 681–709. doi: 10.1086/284231
- Haag, E. S. (2007). Compensatory vs. pseudocompensatory evolution in molecular and developmental interactions. *Genetica* 129, 45–55. doi: 10.1007/s10709-006-0032-3
- Hay, A., and Tsiantis, M. (2006). The genetic basis for differences in leaf form between *Arabidopsis thaliana* and its wild relative *Cardamine hirsuta*. *Nat. Genet.* 38, 942–947. doi: 10.1038/ng1835
- He, C. Y., and Saedler, H. (2005). Heterotopic expression of MPF2 is the key to the evolution of the Chinese lantern of *Physalis*, a morphological novelty in Solanaceae. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5779–5784. doi: 10.1073/pnas.0501877102
- Hempel, F. D., and Feldman, L. J. (1995). Specification of chimeric flowering shoots in wild-type *Arabidopsis*. *Plant J.* 8, 725–731. doi: 10.1046/j.1365-313X.1995.08050725.x
- Hoekstra, H. E., and Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61, 995–1016. doi: 10.1111/j.1558-5646.2007.00105.x
- Hovav, R., Chehanovsky, N., Moy, M., Jetter, R., and Schaffer, A. A. (2007). The identification of a gene (Cwp1), silenced during Solanum evolution, which causes cuticle microfissuring and dehydration when expressed in tomato fruit. *Plant J.* 52, 627–639. doi: 10.1111/j.1365-313X.2007.03265.x
- Hsieh, T. H., Lee, J. T., Charng, Y. Y., and Chan, M. T. (2002). Tomato plants ectopically expressing *Arabidopsis* CBF1 show enhanced resistance to water deficit stress. *Plant Physiol.* 130, 618–626. doi: 10.1104/pp.006783
- Kondrashov, A. S., Sunyaev, S., and Kondrashov, F. A. (2002). Dobzhansky-Muller incompatibilities in protein evolution. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14878–14883. doi: 10.1073/pnas.232565499
- Landry, C. R., Hartl, D. L., and Ranz, J. M. (2007). Genome clashes in hybrids: insights from gene expression. *Heredity* 99, 483–493. doi: 10.1038/sj.hdy.6801045
- Lee, J.-H., Cho, Y.-S., Yoon, H.-S., Suh, M. C., Moon, J., Lee, I., et al. (2005). Conservation and divergence of FCA function between *Arabidopsis* and rice. *Plant Mol. Biol.* 58, 823–838. doi: 10.1007/s11103-005-8105-8
- Lee, S. Y., and Lee, D. H. (2005). Expression of MBR4, a TIR-NBS type of apple R gene, confers resistance to bacterial spot disease in *Arabidopsis*. *J. Plant Biol.* 48, 220–228. doi: 10.1007/BF03030411
- Lee, Y. W., Gould, B. A., and Stinchcombe, J. R. (2014). Identifying the genes underlying quantitative traits: a rationale for the QTN programme. *AoB Plants* 6:plu004. doi: 10.1093/aobpla/plu004
- Li, N., Yuan, L., Liu, N., Shi, D., Li, X., Tang, Z., et al. (2009). SLOW WALKER2, a NOCI/MAK21 homologue, is essential for coordinated cell cycle progression during female gametophyte development in *Arabidopsis*. *Plant Physiol.* 151, 1486–1497. doi: 10.1104/pp.109.142414
- Lin, E., Burns, D. J. W., and Gardner, R. C. (1993). Fruit developmental regulation of the kiwifruit actinidin promoter is conserved in transgenic petunia plants. *Plant Mol. Biol.* 23, 489–499. doi: 10.1007/BF00019297
- Liu, N., Chen, A. P., Zhong, N. Q., Wang, F., Wang, H. Y., and Xia, G. X. (2007). Functional screening of salt stress-related genes from *Thellungiella halophila* using fission yeast system. *Physiol. Plant.* 129, 671–678. doi: 10.1111/j.1399-3054.2007.00857.x
- Liu, N., Sliwinski, M. K., Correa, R., and Baum, D. A. (2011). Possible contributions of TERMINAL FLOWER 1 to the evolution of rosette flowering

- in *Leavenworthia* (Brassicaceae). *New Phytol.* 189, 616–628. doi: 10.1111/j.1469-8137.2010.03511.x
- Lovett, M., Kere, J., and Hinton, L. M. (1991). Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9628–9632. doi: 10.1073/pnas.88.21.9628
- Maizel, A., Busch, M. A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M., et al. (2005). The floral regulator *LEAFY* evolves by substitutions in the DNA binding domain. *Science* 308, 260–263. doi: 10.1126/science.1108229
- Mallory, A. C., Reinhart, B. J., Jones-Rhoades, M. W., Tang, G., Zamore, P. D., Barton, M. K., et al. (2004). MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region. *EMBO J.* 23, 3356–3364. doi: 10.1038/sj.emboj.7600340
- Mietkiewska, E., Giblin, E. M., Wang, S., Barton, D. L., Dirpaul, J., Brost, J. M., et al. (2004). Seed-specific heterologous expression of a nasturtium FAE gene in *Arabidopsis* results in a dramatic increase in the proportion of erucic acid. *Plant Physiol.* 136, 2665–2675. doi: 10.1104/pp.104.046839
- Morel, J. B., Mourrain, P., Beclin, C., and Vaucheret, H. (2000). DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in *Arabidopsis*. *Curr. Biol.* 10, 1591–1594. doi: 10.1016/S0960-9822(00)00862-9
- Mori, S., Kobayashi, H., Hoshi, Y., Kondo, M., and Nakano, M. (2004). Heterologous expression of the flavonoid 3',5'-hydroxylase gene of *Vinca* major alters flower color in transgenic *Petunia hybrida*. *Plant Cell Rep.* 22, 415–421. doi: 10.1007/s00299-003-0709-3
- Nasrallah, M. E., Liu, P., and Nasrallah, J. B. (2002). Generation of self-incompatible *Arabidopsis thaliana* by transfer of two S locus genes from *A. lyrata*. *Science* 297, 247–249. doi: 10.1126/science.1072205
- Nasrallah, M. E., Liu, P., Sherman-Broyles, S., Boggs, N. A., and Nasrallah, J. B. (2004). Natural variation in expression of self-incompatibility in *Arabidopsis thaliana*: Implications for the evolution of selfing. *Proc. Natl. Acad. Sci. U.S.A.* 101, 16070–16074. doi: 10.1073/pnas.0406970101
- Palmer, L. E., Rabinowicz, P. D., O'Shaughnessy, A. L., Balija, V. S., Nascimento, L. U., Dike, S., et al. (2003). Maize genome sequencing by methylation filtration. *Science* 302, 2115–2117. doi: 10.1126/science.1091265
- Papayan, A., and Kochian, L. V. (2004). Identification of *Thlaspi caerulescens* genes that may be involved in heavy metal hyperaccumulation and tolerance. Characterization of a novel heavy metal transporting ATPase. *Plant Physiol.* 136, 3814–3823.
- Rieseberg, L. H., Archer, M. A., and Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity* 83(Pt 4), 363–372. doi: 10.1038/sj.hdy.6886170
- Schubert, D., Lechtenberg, B., Forsbach, A., Gils, M., Bahadur, S., and Schmidt, R. (2004). Silencing in *Arabidopsis* T-DNA transformants: the predominant role of a gene-specific RNA sensing mechanism versus position effects. *Plant Cell* 16, 2561–2572. doi: 10.1105/tpc.104.024547
- She, W., Lin, W., Zhu, Y., Chen, Y., Jin, W., Yang, Y., et al. (2010). The gypsy insulator of *Drosophila melanogaster*, together with its binding protein suppressor of Hairy-wing, facilitate high and precise expression of transgenes in *Arabidopsis thaliana*. *Genetics* 185, 1141–1150. doi: 10.1534/genetics.110.117960
- Singer, S., and Cox, K. (2013). A gypsy-like sequence from *Arabidopsis thaliana* exhibits enhancer-blocking activity in transgenic plants. *J. Plant Biochem. Biotechnol.* 22, 35–42. doi: 10.1007/s13562-012-0108-3
- Sliwinski, M. K., Bosch, J. A., Yoon, H. S., Von Balthazar, M., and Baum, D. A. (2007). The role of two *LEAFY* paralogs from *Idahoia scapigera* (Brassicaceae) in the evolution of a derived plant architecture. *Plant J.* 51, 211–219. doi: 10.1111/j.1365-3113.2007.03148.x
- Sliwinski, M. K., White, M. A., Maizel, A., Weigel, D., and Baum, D. A. (2006). Evolutionary divergence of *LFY* function in the mustards *Arabidopsis thaliana* and *Leavenworthia crassa*. *Plant Mol. Biol.* 62, 279–289. doi: 10.1007/s11103-006-9020-3
- Stern, D. L. (2000). Evolutionary developmental biology and the problem of variation. *Evolution* 54, 1079–1091. doi: 10.1111/j.0014-3820.2000.tb00544.x
- True, J. R., and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* 3, 109–119. doi: 10.1046/j.1525-142x.2001.003002109.x
- Tzeng, T. Y., and Yang, C. H. (2001). A MADS box gene from lily (*Lilium longiflorum*) is sufficient to generate dominant negative mutation by interacting with PISTILLATA (PI) in *Arabidopsis thaliana*. *Plant Cell Physiol.* 42, 1156–1168. doi: 10.1093/pcp/pce151
- Villagarcia, H., Morin, A. C., Shpak, E. D., and Khodakovskaya, M. V. (2012). Modification of tomato growth by expression of truncated ERECTA protein from *Arabidopsis thaliana*. *J. Exp. Bot.* 63, 6493–6504. doi: 10.1093/jxb/ers305
- Vlad, D., Kierzkowski, D., Rast, M. I., Vuolo, F., Dello Ioio, R., Galinha, C., et al. (2014). Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. *Science* 343, 780–783. doi: 10.1126/science.1248384
- Wagner, G. P., Chiu, C.-H., and Laubichler, M. (2000). Developmental evolution as a mechanistic science: the inference from developmental mechanisms to evolutionary processes. *Am. Zool.* 40, 819–831. doi: 10.1093/icb/40.5.819
- Wang, M. Y., Gu, D., Liu, T. S., Wang, Z. Q., Guo, X. Y., Hou, W., et al. (2007). Overexpression of a putative maize calcineurin B-like protein in *Arabidopsis* confers salt tolerance. *Plant Mol. Biol.* 65, 733–746. doi: 10.1007/s11103-007-9238-8
- Whipple, C. J., Ciceri, P., Padilla, C. M., Ambrose, B. A., Bandong, S. L., and Schmidt, R. J. (2004). Conservation of B-class floral homeotic gene function between maize and *Arabidopsis*. *Development* 131, 6083–6091. doi: 10.1242/dev.01523
- Yoon, H. S., and Baum, D. A. (2004). Transgenic study of parallelism in plant morphological evolution. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6524–6529. doi: 10.1073/pnas.0401824101
- Yuan, Y., Sanmiguel, P. J., and Bennetzen, J. L. (2003). High-Cot sequence analysis of the maize genome. *Plant J.* 34, 249–255. doi: 10.1046/j.1365-3113.2003.01716.x
- Zahn, L. M., Leebens-Mack, J., DePamphilis, C. W., Ma, H., and Theissen, G. (2005). To B or not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. *J. Heredity* 96, 225–240. doi: 10.1093/jhered/esi033

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Correa and Baum. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evolution of the Plant Reproduction Master Regulators LFY and the MADS Transcription Factors: The Role of Protein Structure in the Evolutionary Development of the Flower

OPEN ACCESS

Edited by:

Rainer Melzer,
University College Dublin, Ireland

Reviewed by:

Marcelo Camier Domelas,
Universidade Estadual de Campinas,
Brazil
William Oki Wong,
Institute of Botany – Chinese
Academy of Sciences, China
Florian Ruempler,
Friedrich Schiller University Jena,
Germany

*Correspondence:

Chloe Zubieta
chloe.zubieta@cea.fr

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 28 August 2015

Accepted: 11 December 2015

Published: 06 January 2016

Citation:

Silva CS, Puranik S, Round A,
Brennich M, Jourdain A, Parcy F,
Hugouvieux V and Zubieta C (2016)
Evolution of the Plant Reproduction
Master Regulators LFY and the MADS
Transcription Factors: The Role
of Protein Structure in the Evolutionary
Development of the Flower.
Front. Plant Sci. 6:1193.
doi: 10.3389/fpls.2015.01193

Catarina S. Silva^{1,2,3,4†}, Sriharsha Puranik^{5†}, Adam Round^{6,7,8}, Martha Brennich⁵,
Agnès Jourdain^{1,2,3,4}, François Parcy^{1,2,3,4}, Veronique Hugouvieux^{1,2,3,4} and
Chloe Zubieta^{1,2,3,4*}

¹ CNRS, Laboratoire de Physiologie Cellulaire & Végétale, UMR 5168, Grenoble, France, ² Laboratoire de Physiologie Cellulaire & Végétale, University of Grenoble Alpes, Grenoble, France, ³ Commissariat à l'Energie Atomique et aux Energies Alternatives, Direction des Sciences du Vivant, Laboratoire de Physiologie Cellulaire & Végétale, Institut de Recherches en Technologies et Sciences pour le Vivant, Grenoble, France, ⁴ Laboratoire de Physiologie Cellulaire & Végétale, Institut National de la Recherche Agronomique, Grenoble, France, ⁵ European Synchrotron Radiation Facility, Structural Biology Group, Grenoble, France, ⁶ European Molecular Biology Laboratory, Grenoble Outstation, Grenoble, France, ⁷ Unit for Virus Host-Cell Interactions, University of Grenoble Alpes-EMBL-CNRS, Grenoble, France, ⁸ Faculty of Natural Sciences, Keele University, Keele, UK

Understanding the evolutionary leap from non-flowering (gymnosperms) to flowering (angiosperms) plants and the origin and vast diversification of the floral form has been one of the focuses of plant evolutionary developmental biology. The evolving diversity and increasing complexity of organisms is often due to relatively small changes in genes that direct development. These “developmental control genes” and the transcription factors (TFs) they encode, are at the origin of most morphological changes. TFs such as LEAFY (LFY) and the MADS-domain TFs act as central regulators in key developmental processes of plant reproduction including the floral transition in angiosperms and the specification of the male and female organs in both gymnosperms and angiosperms. In addition to advances in genome wide profiling and forward and reverse genetic screening, structural techniques are becoming important tools in unraveling TF function by providing atomic and molecular level information that was lacking in purely genetic approaches. Here, we summarize previous structural work and present additional biophysical and biochemical studies of the key master regulators of plant reproduction – LEAFY and the MADS-domain TFs SEPALLATA3 and AGAMOUS. We discuss the impact of structural biology on our understanding of the complex evolutionary process leading to the development of the bisexual flower.

Keywords: evolution, SEPALLATA3, AGAMOUS, LEAFY, protein crystallography, small angle X-ray scattering, homology modeling

INTRODUCTION

The evolution of streptophytes (green plants), chronicled by the fossil record, follows a trajectory from simple green algae, to the earliest land plants (mosses, hornworts, liverworts), to free-sporing vascular plants (lycopsids including extant clubmosses, quillworts and spike mosses and monilophytes such as ferns and horsetails) and finally culminating with more complex seed plants (**Figure 1**). As the climate changed and became less favorable to spore-forming lycophtyes and monilophytes, spermatophytes (seed plants) were able to supplant these spore-forming vascular plants to become the majority of land plant species. The radiation of seed plants was due in large part to their ability to reproduce without the necessity of water for the dispersal of pollen or successful fertilization, as in the case of mosses and ferns. The reproductive adaptations in seed plants acted as a driver for terrestrial colonization and played a key role in their radiation across a wide range of habitats.

Extant seed plants are further divided into two sister groups, the gymnosperms and the angiosperms. Gymnosperms have naked seeds unprotected by a carpel and generally develop as the result of a single fertilization event. Exceptions exist as is the case of the genus *Ephedra* and *Gnetum* (Friedman, 1990; Friedman and Carmichael, 1996). In contrast, angiosperm seeds are enclosed and protected by the carpel and result from a double fertilization event that ensures the simultaneous development of the zygote and nutritive tissues, the endosperm (Lord and Russell, 2002). In addition to these variations in fertilization and seed development, the most striking difference between gymnosperms and angiosperms is the evolutionary innovation of the angiosperm flower. This novel arrangement joins the male and female organs into one reproductively competent structure. While the evolution of green plants from algae to seed plants follows a relatively smooth path in the fossil record, the evolution of the flower in angiosperms represents an evolutionary leap lacking an extensive step-wise fossil record. Since the time of Charles Darwin, the “abominable mystery” of flower origins and the unprecedented explosive radiation of angiosperm species have been the subject of extensive study and speculation (Burkhardt et al., 1985; Friedman, 2009).

In contrast to gymnosperm cones, which are unisexual and lack an enveloping perianth (sterile outer organs), angiosperm flowers have both male and female reproductive organs on a single axis surrounded by sepals and petals. A typical angiosperm flower is composed of four organs arranged in four concentric whorls. The outermost whorl contains the green protective sepals, followed by a whorl of petals involved in flower opening and pollinator attraction, the next whorl contains the stamens that produce pollen and constitute the male gametophyte, and finally the inner most whorl comprising the pistil, composed of one or more carpels, that contain the ovules. This basic floral architecture can vary across angiosperms. For example, basal angiosperms may contain tepals, sterile outer organs that cannot be differentiated into distinct sepals and petals. In addition, the number of flower parts and their arrangement around the central axis of the flower may vary as in orchids where the male and female organs are fused. However, the essential characteristic of

the flower, co-localized male and female organs, is retained across all angiosperm species and acts as a defining trait.

Angiosperm and Gymnosperm Evolution

One of the central questions in plant evolutionary developmental biology is how the flower, a bisexual compacted reproductive structure, evolved and what were the underlying molecular mechanisms for this dramatic morphological change. Extant gymnosperms and angiosperms separated ~300 Mya (Zhang et al., 2004), with angiosperms quickly achieving an unprecedented level of species dominance, with over 350,000 extant species, in a dramatically short evolutionary timescale. However, simple morphological comparisons between gymnosperm cones and angiosperm flowers offer limited insight into flower evolution (Bateman et al., 2006; Frohlich and Chase, 2007). An understanding of the abrupt appearance of the flower from gymnosperm cones requires not only a fossil record to probe the changing morphologies of plant reproductive structures, but also a molecular basis derived from genome sequencing, molecular biology and structural biology. Impressive progress has been made in understanding the gene networks that regulate plant reproduction in angiosperms and, albeit to a lesser extent, also in gymnosperms. Due to extensive forward and reverse genetic studies (Coen and Meyerowitz, 1991; Saedler et al., 2001; Theissen and Saedler, 2001; Krizek and Fletcher, 2005) and whole genome sequencing in model plants such as thale cress (*Arabidopsis thaliana*), snapdragon (*Antirrhinum majus*) and petunia (*Petunia x hybrida*), as well as the large scale gene sequencing initiatives such as the 1000 plant genomes project (<https://sites.google.com/a/ualberta.ca/onekp/home>) and the complete sequencing and annotation of the first gymnosperm genome from Norway spruce (Nystedt et al., 2013), many of the genes which regulate the transition from vegetative to reproductive growth in angiosperms and gymnosperms have been identified.

Gene Regulatory Networks Controlling Plant Reproductive Development

Despite the morphological difference between angiosperm and gymnosperm reproductive structures, a comparison of the genes responsible for male and female organ development demonstrates a high degree of conservation. Based on studies in angiosperm model plants such as *Arabidopsis*, development is switched from a vegetative to a reproductive program based on exogenous environmental and endogenous developmental signals such as plant age. This switch is orchestrated by the high level regulator of reproductive development, *LEAFY* (*LFY*), a gene that is conserved in gymnosperms and angiosperms (Vazquez-Lobo et al., 2007; Moyroud et al., 2010) and which has recently been identified in green algae, suggesting ancestral functions predating land plants (Sayou et al., 2014). Interestingly, while existing primarily as a single copy gene in most angiosperms, gymnosperms have two paralogous *LFY*-like genes-*LFY* and *NEEDLY* (*NLY*; Frohlich and Meyerowitz, 1997; Vazquez-Lobo et al., 2007), the only known exception being the gymnosperm genus *Gnetum* where *NLY* is absent (Frohlich and

MADS-box genes have undergone multiple duplication events, leading to a more extensive gene network in angiosperms versus their sister gymnosperms (**Figure 1**). *LFY*, *NLY* and the MADS-box genes all encode transcription factors (TFs). These TFs act as master regulators and are able to direct extensive downstream gene networks. Recent work examining the function of *LFY*, *NLY*

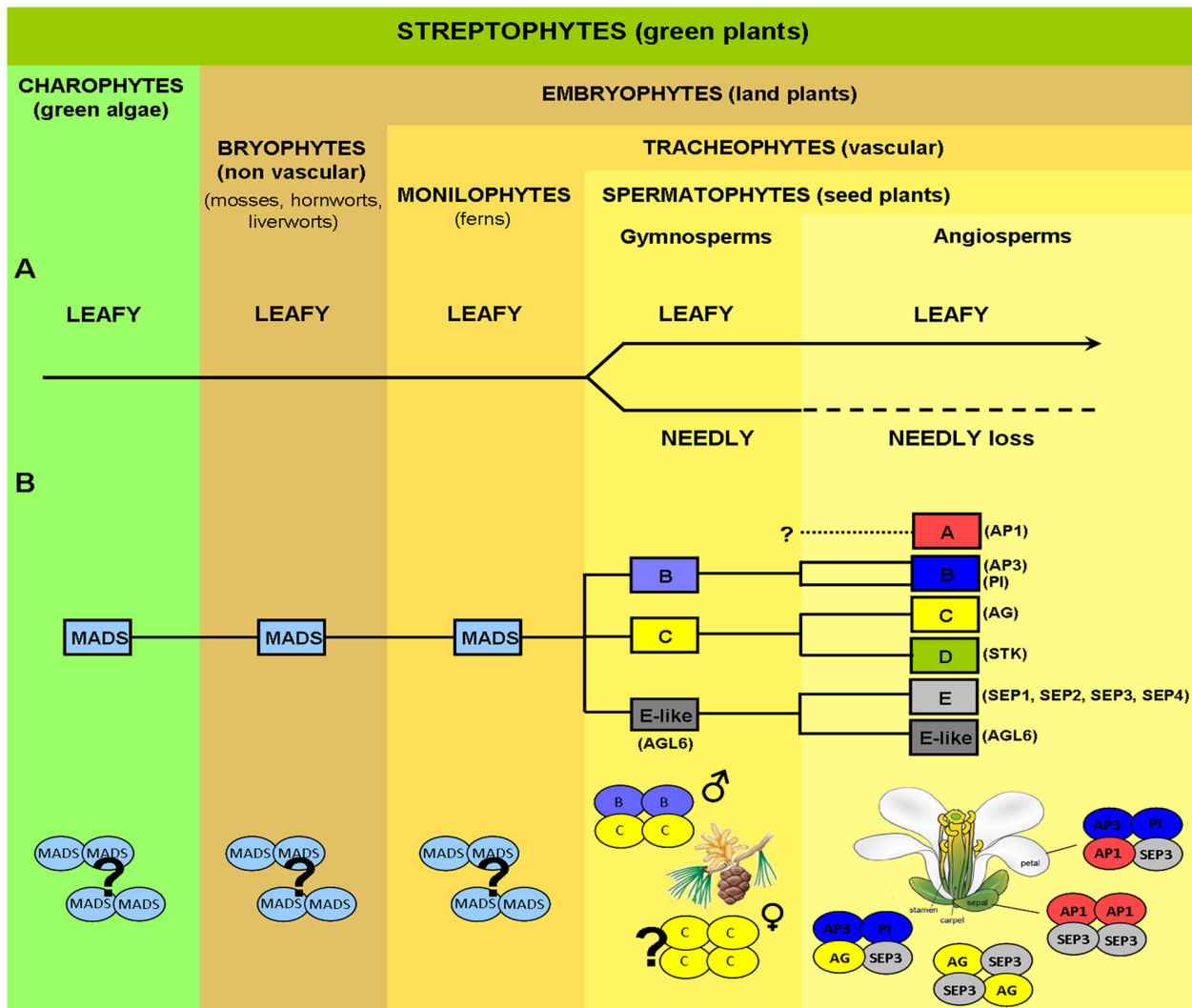


FIGURE 1 | Evolution of key genes controlling plant reproductive development. (A) Evolution of *LEAFY* (*LFY*) from green algae to angiosperms. *LFY* exists mostly as a single-copy gene in all streptophytes (green plants), with the exception of gymnosperms where a *LFY*-like paralog, *NEEDLY* (*NLY*), originated after a major duplication event (the only possible exception being the genus *Gnetum*). In gymnosperms, *LFY* and *NLY* are consistently expressed in both male (pollen-bearing) and female (seed-bearing) cones, in a spatiotemporal coordinated manner. In the angiosperm lineage, *NLY* was subsequently lost, with *LFY* now regulating the expression of genes responsible for both the male and female organs in the unified bisexual flower. **(B)** MADS-box homeotic gene family. MADS-box genes are present in the most simple green algae and, as plants became more complex, the MADS-box gene family expanded via multiple duplication and specification events. Putative orthologs of class B, C, and E-like (*AGL6*) floral homeotic genes have been isolated from different gymnosperms (conifers, gnetophytes, ginkgophytes, and cycads) as shown schematically by yellow and blue colored ovals. In contrast, *SEP*-like genes, the second subfamily conferring E-class function, as well as A-class genes, seem to be absent in extant gymnosperms but are present in all angiosperms. In gymnosperms, expression patterns of putative B and C-class gene orthologs resemble those of B and C-class genes in angiosperms, with B-class genes being expressed on male reproductive organs, whereas C-class genes are expressed in both male and female organs. In gymnosperms C-class proteins alone or C and B-class proteins together seem capable of forming tetrameric complexes (without any additional partners), which define, respectively, the female and male organs in these organisms as indicated. In angiosperms tetramer formation is dependent on the *SEPALATTA* (E-class) TFs which act as hubs by mediating interactions among proteins from different floral homeotic classes, strictly determining floral organ identity. Question marks indicate uncertainty as to physiological oligomerisation state, AP1, *APETALA1*; AP3, *APETALA3*; PI, *PISTILLATA*; AG, *AGAMOUS*; STK, *SEEDSTICK*; SEP, *SEPALATTA*; AGL6, *AGAMOUS LIKE 6*.

and the MADS TFs at the protein level has greatly advanced our understanding of how relatively small changes in a few key regulatory TFs can result in large differences at the morphological level of the organism. Current hypotheses point to changes in a few key genes, and the TFs they encode, as determining factors in the evolution of plant reproduction and the formation of the flower (Theissen, 2000, 2005; Zahn et al., 2005a,b; Theissen and Melzer, 2007; Melzer et al., 2010).

The Role of LFY and LFY/NLY in Angiosperms and Gymnosperms

In angiosperms such as *Arabidopsis* or *Antirrhinum*, the switch to reproductive growth involves the conversion of the shoot apical meristem (SAM) to an inflorescence meristem (IM). The IM will in turn generate the floral meristem (FM) on its flanks. The development of a FM can be divided into two main steps (1) the formation of a specific zone within the IM, called the anlage, from which the FM will arise and (2) the growth of the FM primordia and subsequent differentiation into the floral organs. It is a balance between inflorescence identity genes such as *TERMINAL FLOWER 1* (*TFL1*) and FM identity genes such as *LFY* that determines the acquisition of flower identity. *TFL1* is predominantly expressed in the IM and acts as a repressor, preventing *LFY* and the MADS-box gene, *APETALA1* (*AP1*), expression (Liljegren et al., 1999). Increasing levels of *LFY* act as a committing step in FM identity, with *LFY* repressing expression of *TFL1* and inducing the expression of FM identity genes such as *AP1* (Parcy et al., 1998; Wagner et al., 1999; Kaufmann et al., 2010; Moyroud et al., 2010; Winter et al., 2011).

In gymnosperms, *LFY* and *NLY* expression patterns overlap in male and female cones early in development with expression patterns diverging later into mutually exclusive but complementary domains, resulting in higher *LFY* expression levels in male cones and higher *NLY* expression in female cones (Shindo et al., 2001; Dornelas and Rodriguez, 2005; Vazquez-Lobo et al., 2007). Originally, the *NLY* gene was thought to exclusively specify gymnosperm female reproductive structures (seed-bearing cone) in *Pinus radiata* (Mouradov et al., 1998), whereas its paralogous gene *LFY* appeared restricted to the male pollen-carrying cones (Mellerowicz et al., 1998). However, subsequent findings of *LFY* orthologs being expressed in female cones of gnetophytes and congeneric conifers (Carlsbecker et al., 2004; Dornelas and Rodriguez, 2005), demonstrated concurrent expression of both genes in male and female reproductive structures. Thus, *LFY* and *NLY* from gymnosperms are both necessary to act as regulators of male and female cone development, likely fulfilling a similar critical role in plant reproduction as the single copy angiosperm *LFY*.

The Roles of the MADS-Box Genes and MADS TFs in Organ Identity

Once the FM is specified, *LFY* activates additional floral organ identity genes including the MADS-box genes *AP3*, *AG*, and *SEP3* (Weigel and Meyerowitz, 1993; Busch et al., 1999; Wagner et al., 1999; Lamb et al., 2002; Lohmann and Weigel, 2002; Winter

et al., 2011). To date there is no direct evidence that gymnosperm *LFY* or *NLY* directly regulate MADS-box genes in gymnosperms as *LFY* does in angiosperms, although this is possible and warrants study. Once expressed, the overlapping patterns of the MADS-box genes will specify floral organ identity as outlined in the ABC(D)E model (Schwarz-Sommer et al., 1990; Coen and Meyerowitz, 1991) and for review see (Sablowski, 2010). In essence, the MADS-box genes can be divided into classes A–E with A+E genes necessary for sepal development, A+B+E genes specifying petals, B+C+E genes specifying stamen, C+E genes specifying carpels and D+E specifying ovules (Theissen and Saedler, 1995; Theissen, 2000; Honma and Goto, 2001; Ng and Yanofsky, 2001; Theissen and Saedler, 2001; Favaro et al., 2003). In *Arabidopsis* the class A genes are *APETALA1* (*AP1*) and *APETALA2* (*AP2*), class B genes are *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), class C is *AGAMOUS* (*AG*) and class E are *SEPALLATA1,2,3,4* (*SEP1,2,3,4*). Except for *AP2*, all the floral homeotic genes in the ABC(D)E model encode MADS-domain TFs. The molecular mechanism of action of these proteins is explained by the floral quartet model, in which the A-E class genes encode TFs which are able to homo and heterotetramerise in specific combinations, resulting in the activation or repression of distinct downstream target genes and thus specifying floral organ identity (Honma and Goto, 2001; Theissen, 2001).

Gymnosperms possess B- and C-like MADS-box genes with their expression patterns resembling B- and C- class genes in angiosperms (Tandre et al., 1995; Sundstrom et al., 1999; Becker et al., 2002, 2003; Jager et al., 2003; Melzer et al., 2010; Wang et al., 2010; Gramzow et al., 2014). Indeed, several studies have described the expression of C-like genes in both male and female cones, while B-like gene expression appeared to be restricted to male cones (Sundstrom and Engstrom, 2002; Wang et al., 2010). Complementation studies have demonstrated that B and C homologs are well-conserved between gymnosperms and angiosperms as B and C genes from gymnosperms can nearly fully restore a wild type flower phenotype (Winter et al., 2002; Zhang et al., 2004). In addition, the gymnosperm MADS-domain TFs from the B and C class appear competent to form homo and heterotetramers, similarly to their angiosperm orthologs (Figure 1; Wang et al., 2010). Interestingly, the *SEP* subfamily members are absent in gymnosperms but are present in all major lineages of extant angiosperms (Zahn et al., 2005a). Based on phylogenetic analysis, the closest relative of the *SEP* subfamily is the *AGL6* subfamily, which is found in both angiosperms and gymnosperms (Becker and Theissen, 2003; De Bodt et al., 2003; Martinez-Castilla and Alvarez-Buylla, 2003; Nam et al., 2003; Zahn et al., 2005a). Similarly to class E *SEP* genes in angiosperms, *AGL6*-like genes are predominantly expressed in reproductive tissues in gymnosperms (for review see, Melzer et al., 2010) and represent the closest homologs to the *SEP* subfamily. Changes in the regulation of B and C class genes during evolution coupled with the appearance of the *SEP*-like genes and the dependence on the *SEP* TFs to form tetrameric MADS protein complexes, have been proposed to be crucial for the appearance of the bisexual flower. By requiring the *SEP* TFs to form transcriptionally active complexes

with other homeotic MADS TFs, male and female organ identity may have become more easily co-regulated due to the multiple roles of the SEPs in specifying all reproductive organs.

The gene regulatory networks directing plant reproduction in gymnosperms and angiosperms are becoming more well-defined and the changes in key genes in gymnosperms and angiosperms which may be at the nexus of flower origins have been identified based on genetics studies in angiosperms and large scale sequencing initiatives in most plant lineages. However, only recently has the structure-function relationship of the proteins encoded by these key genes been determined. Here, we summarize available structural studies and provide new data to show how changes at the protein level in the key regulators LFY, NLY, and MADS-domain TFs potentially result in new functionality. Using biophysical data as a foundation, we probe the molecular mechanisms underlying the emergence and evolution of the novel reproductive architecture of the angiosperm flower and discuss how biochemistry and structural biology can provide new insights into evolutionary developmental biology.

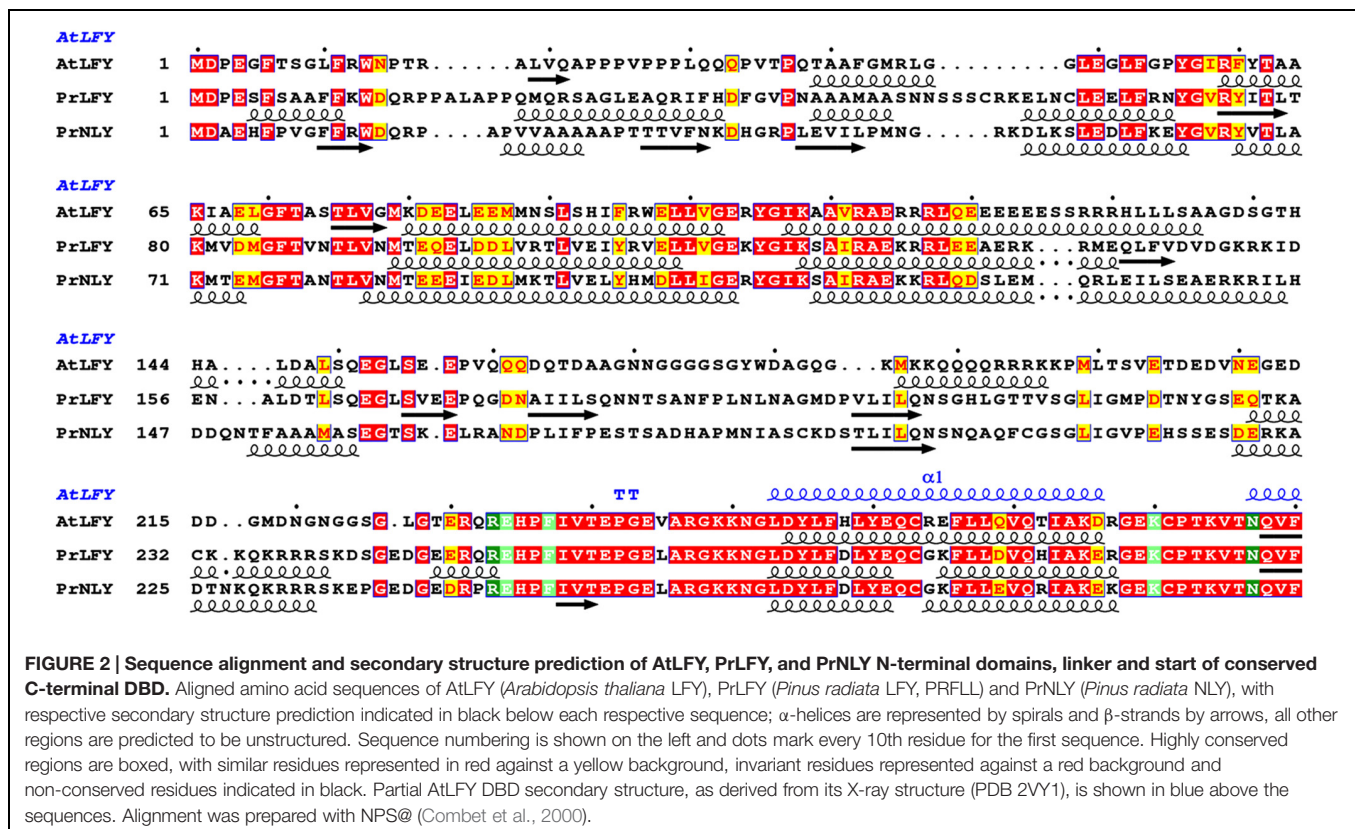
MATERIALS AND METHODS

Sequence Alignments

Sequence alignments were performed using the server NPS@ (Network Protein Sequence Analysis; Combet et al., 2000).

Sequences were aligned with ClustalW (Thompson et al., 1994) using the default parameters for both pairwise alignment and multiple alignment sections. Where appropriate, secondary structure predictions were carried out with PREDATOR (DSSP) using the NPS@ server. Protein sequences used were obtained from GenBank and the 1000 Plants (1KP) initiative (<http://www.onekp.com>). Resulting alignments and secondary structure predictions were rendered with ESPrnt (Robert and Gouet, 2014).

For the LFY/NLY sequence alignments (Figures 2 and 3) the sequences used are as follows: AtLFY (*A. thaliana* LFY, AED97525.1), OsLFY (*Oryza sativa japonica* LFY, RFL, AHX83808.1), AmtLFY (*Amborella trichopoda* LFY, AmboLFY, AGV98899.1), PrLFY (*Pinus radiata* LFY, PRFL, AAB51587.1), GbLFY (*Ginkgo biloba* LFY, ADD64700.1), WmLFY (*Welwitschia mirabilis* LFY, AAF23870.1), PrNLY (*P. radiata* NLY, AAB68601.1), PaNLY (*Pinus armandii* NLY, ADO33969.1), GbNLY (*G. biloba* NLY, AAF77074.1), and WmNLY (*W. mirabilis* NLY, AAD38872.1). For MADS-domain TFs sequence alignments (Figures 4 and 5) the sequences used are: *A. thaliana* SEP3 (AEE30503.1), SEP1 (AED92208.1), SEP2 (AEE73791.1), AP3 (AEE79216.1), PI (AED92817.1), AP1 (AEE34887.1), AG (AEE84111.1), AGL6 (AEC10582.1), SOC1 (AEC10583.1), SVP (AEC07320.1), and FLC (AED91498.1); *Gnetum gnemon* GGM2 (CAB44448.1), GGM3 (CAB44449.1), GGM15 (CAC13991.1), GGM9 (CAB44455.1), and GGM11 (CAB44457.1); *Picea abies* DAL11 (AAF18373.1), DAL12 (AAF18375.1), DAL13 (AAF18377.1), DAL2 (CAA55867.1),



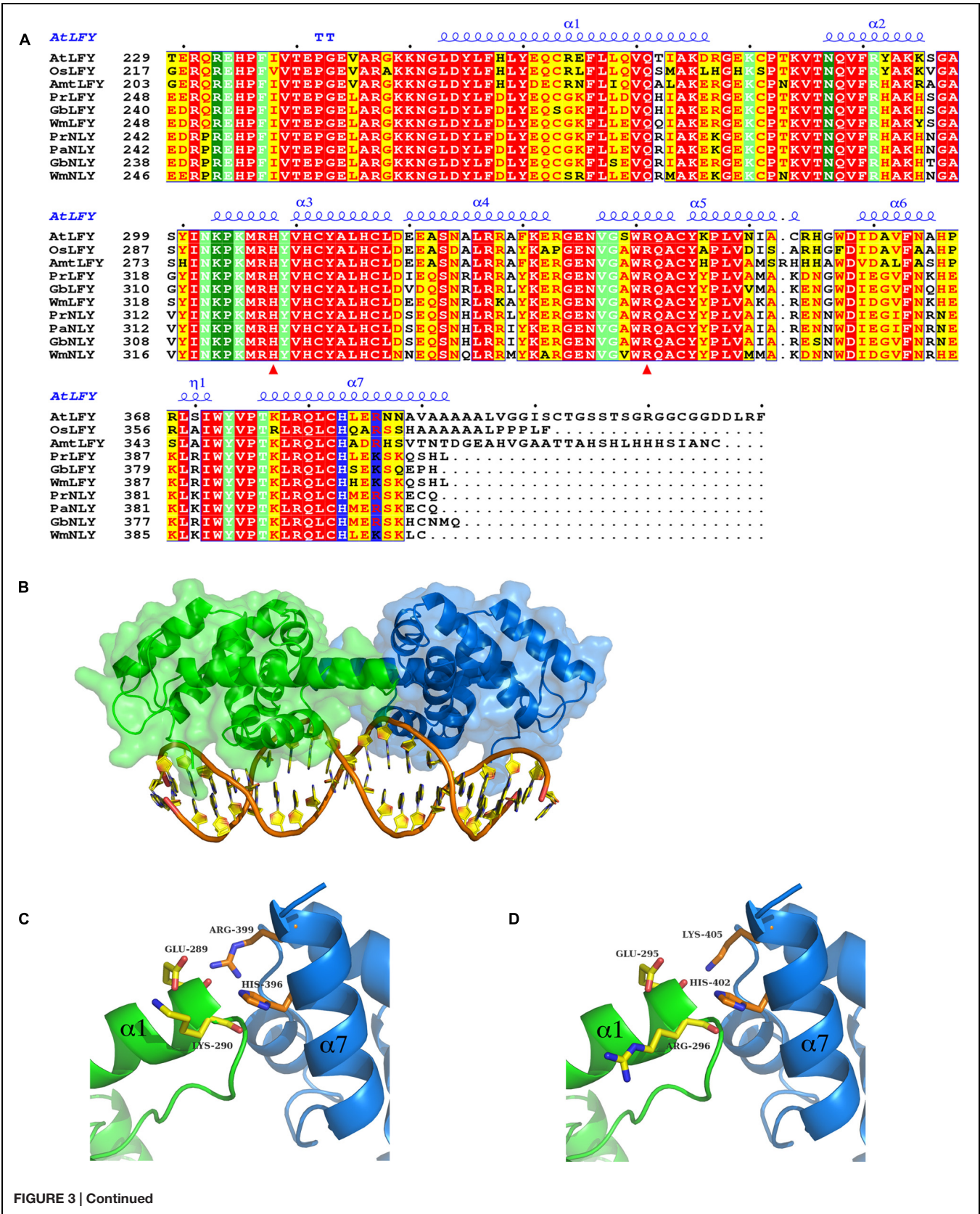


FIGURE 3 | Continued**Sequence alignment and homology models of the DNA binding domain (DBD) of LFY and NLY.**

(A) Sequence alignment of LEAFY (LFY) and NEEDLY (NLY) DBDs. Aligned C-terminal DBD amino acid sequences of AtLFY (*A. thaliana* LFY, GenBank AED97525.1), OsLFY (*Oryza sativa japonica* LFY, RFL), AmlLFY (*Amborella trichopoda* LFY, AmboLFY), PrLFY (*Pinus radiata* LFY, PRFL), GbLFY (*Ginkgo biloba* LFY), WmLFY (*Welwitschia mirabilis* LFY), PrNLY (*Pinus radiata* NLY), PaNLY (*Pinus armandii* NLY), GbNLY (*G. biloba* NLY), and WmNLY (*W. mirabilis* NLY). All sequences are numbered and dots mark every tenth residue above the sequences. Highly conserved regions are boxed, with similar residues represented in red against a yellow background, invariant residues represented against a red background and non-conserved residues indicated in black. The secondary structure annotation of AtLFY DBD, as derived from its three-dimensional X-ray structure (PDB 2VY1), is depicted in blue on top of the aligned sequences [α -helices (α); strict β -turn (TT); 3_{10} -helix (η)]. Residues involved in interactions with the DNA are highlighted in dark-green (direct contact with DNA bases) and light-green (sugar phosphate backbone contacts); residues involved in dimerisation are depicted in blue. Red triangles indicate residues important for determining DNA half-site specificity. The AtLFY protein sequence (AED97525.1) differs from the AtLFY sequences in Hames et al. (2008) and (Sayou et al., 2014; AAA32826) by a four residue deletion after residue 152 resulting in a -4 sequence shift. **(B)** Homology model of *Pinus radiata* NEEDLY (PrNLY) DBD based on AtLFY DBD X-ray structure (PDB 2VY1). Monomers are represented in green and blue as cartoons with a partial transparent surface; bound DNA is represented in orange and gold. PrNLY DBD adopts the same seven α -helix fold, contacting the DNA through both the minor and major grooves with complete conservation of all DNA-binding amino acid residues determined for AtLFY DBD. **(C)** Close-up view of the dimerisation interface of PrNLY. Monomers are colored as per **(B)** and side chain residues involved in putative hydrogen bonding interactions are shown and labeled. **(D)** Close-up view of the dimerisation interface of PrLFY. Colors and residues as per **(C)**.

DAL1 (CAA56864.1), and DAL14 (AGR53802.1). Numbers indicated correspond to GenBank accession numbers.

Homology Modeling

The homology model of the DNA-binding domain (DBD) of *Pinus radiata* NLY (PrNLY) and LFY (PrLFY) proteins

were built using the SWISS-MODEL server (Arnold et al., 2006; Biasini et al., 2014, swissmodel.expasy.org). Based on the sequence alignment between PrNLY and AtLFY DBDs the PrNLY partial sequence [E242-Q404] was fed to the server, as well as the AtLFY DBD PDB structure (PDB 2VY1, GenBank accession AAA32826). The homology model of PrNLY DBD comprises residues [R246-K401]. The same procedure was applied to PrLFY for which the partial sequence [Q251-H410] was fed to the server; the PrLFY homology model comprises residues [R252-K407]. Each of the models was superimposed on the AtLFY DBD structure (Hames et al., 2008) using COOT (Emsley et al., 2010); the DNA coordinates added to the composite homology models were taken from the AtLFY structure. The cartoon model representation was made using the program Pymol (The PyMOL Molecular Graphics System, 2010).

SEP3^(75–178) Mutagenesis, Expression, and Purification

SEP3^(75–178) construct (wild type) was cloned into the expression vector pESPRIT002 (Hart and Tarendeau, 2006; Guilligay et al., 2008) using the AatII and NotI restriction sites. The plasmid contains an N-terminal 6x-His tag followed by a TEV protease cleavage site. All mutants produced were generated using the SEP3^(75–178) construct as the template and using Phusion polymerase (NEB) according to the manufacturer's protocol. The oligonucleotides used for mutagenesis are provided in Table 1.

SEP3^(75–178) and all the tetramerisation mutant constructs were overproduced in *Escherichia coli* BL21 (DE3) CodonPlusRIL (Agilent Technologies; Puranik et al., 2014); all dimerisation mutant constructs were overproduced in *E. coli* Rosetta2 (DE3) pLysS cells (this study). Cells were grown at 37°C in Luria-Bertani (LB) culture medium supplemented with kanamycin (50 mg/mL) and chloramphenicol (37 mg/mL), until an OD₆₀₀ of 0.7–0.8

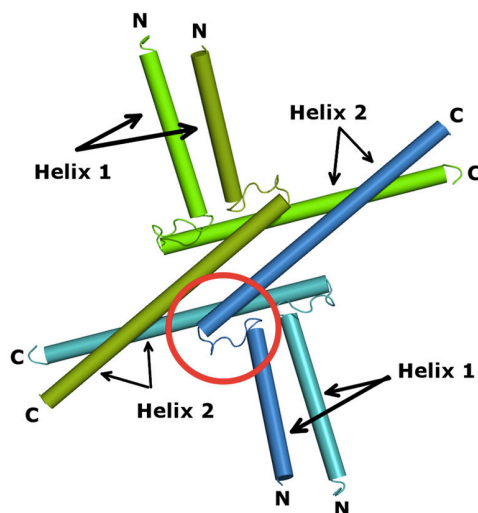
PaDAL11	1MGRGKLEIKMIESNRKVTTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
PaDAL13	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
PaDAL12	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
AtPI	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
GgGGM15	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
AtSEP1	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
AtSEP2	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
AtSEP3	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
GgGGM9	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
PaDAL1	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
GgGGM11	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
PaDAL14	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
AtAGL6	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
AtAP1	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
AtFLC	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
AtSOC1	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
AtSVP	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
GgGGM3	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
PaDAL2	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
AtAG	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
GgGGM2	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...
AtAP3	1MGRGKLEIKMIENPTNRQVTSKRRGGTFKKAR	ELSVLCBAEVALIVFSTGNHQEAS...

FIGURE 4 | Sequence alignment of MADS TFs M-domain. Aligned M-domain amino acid sequences of *A. thaliana* SEP3, SEP1, SEP2, AP3, PI, AP1, AG, AGL6, SOC1, SVP, and FLC; the gymnosperm *Gnetum gnemon* GGM2, GGM15 (AP3/PI-like), GGM3 (AG-like), and GGM9, GGM11 (AGL6-like); and the gymnosperm *Picea abies* DAL11, DAL12, DAL13 (AP3/PI-like), DAL2 (AG-like), and DAL1, DAL14 (AGL6-like) proteins. Sequence numbering is indicated on the left, with every tenth residue marked by black dots above the sequences. Highly conserved regions are boxed, with similar residues represented in red against a yellow background, invariant residues represented against a red background and non-conserved residues indicated in black.

A

AtSEP3		$\alpha 1$		TT	$\alpha 2$		$\eta 1$
AtSEP3	80	PNVPSREALAVELSSQEEYILKLRERYDALQRTQNNLL	EDL	GPLSTKELESEROLDSSLKQIRALRTQFM	LDOLNDLQSKERMLTETNKTLLRLRL		
AtSEP1	80	VNN . KPAKELEN . SYREYILKGRYENLQQRQNNLL	EDL	GPLNSKELEQLEROLDGSLKQVRSIKTQYMLD	LDLONKEQMLLETNRLAMKL		
AtSEP2	80	VNN . KPAKELEN . SYREYILKGRYENLQQRQNNLL	EDL	GPLNSKELEQLEROLDGSLKQVRCIKTQYMLD	LDLQGGHEILLDANRALSMKL		
AtAGL6	78	SNN . KPETTQS . WCQEVTKLKSYESLVRTNRRNLL	EDL	GEMVKEIQALERQLEAALATATQRKTQVMMEEM	EDLQKKERQLGDINKQLKIKF		
AtAP1	80	LIA PESDVNTNWSMEYNRLKAKTELLERNQRHLL	EDL	QAMSKEIQNLQQLDTALKHIRTTRKNQOLMYES	INELQKKEKATQEQNSMLSKQI		
AtFLC	80	ALD HQSKALNYGSHYELLELVDSKLV	SNVKNVSDALVQLEEHLELTALSVTRAKKTEL	MLKLVENLKEKEKMLKEENQVLASQM			
AtSOC1	79	TKP VSEENMHLKYEAANMMKIEQLKASKRKL	LGEGITC	SIIEELQQLIEQLKSVKCIARAKTQVFKEQIE	QLKQKEKALAAENKLS . . .		
AtSVP	79	LDQ PSLELQLVENS DHARMSKETADKSHRL	QMRGEE	LGDLIEELQQLIEKALETGLTRVIETKSDKIM	SEISELQKKEKQMLMDENKRLRQQG		
AtAG	95	TGS VAEINAQYYQESAKLRQQTISI	QNSNRQLMGETI	GSMSEKELRNLEGRLEERSITRIRSKKNEL	LFSEIDYMQKREVLDHNDNQILRAKI		
GGM3	79	GGA IAESNAQYQQEAVKLRQQTIDVLNN	QIRHYMGEC	LQSMTIKELKQLE	GKLEKGLGRVRSKRNEKLLEDITLQRR	EDNLIRENEYIRNKI	
GGM2	79	DIE LOIMGQELIKERRENEKLRSKLRY	MMGEDI	GELKIAQLEKLEHDLLESALRLVRRKKD	HAWDYQRTILKKVKLOYALHEQMSRQL		
AtAP3	81	ATQ YERMQETKRKLLETNRNLR	TQIKQRL	GECLDEL	DIQELRRLEDEMENTFKLVRRKFKSLGN	QIETTKKKNKSQDQIQNLIHEL	
AtPI	81	DAK HENLSNEIDRTIKKENDSLQLEL	RHLK	GEDIQ	SINLKNLMAVEHATEHGLDKVRDHQME	LISKRRNEKMMAEQRQLTFQLQQQE	

B



C

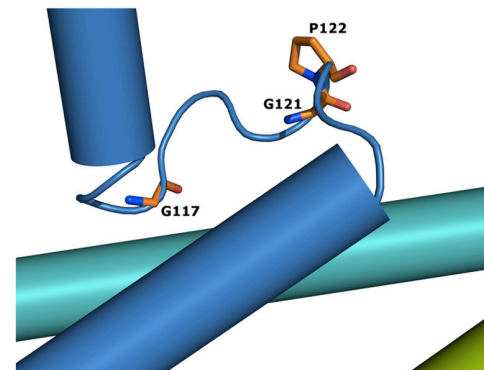


FIGURE 5 | MADS TFs oligomerisation domain. (A) Sequence alignment spanning the SEP3 crystallographic structure (PDB 4OX0). All sequences are from *A. thaliana* SEP3, SEP1, SEP2, AP3, PI, AP1, AG, AGL6, SOC1, SVP, and FLC and the gymnosperm *G. gnemon* GGM2 and GGM3 proteins. Numbering is indicated at the start of the sequences and every tenth residue indicated by a black dot above the SEP3 sequence. Highly conserved regions are boxed in blue with a white background; strictly conserved residues are depicted by a red square below the sequence. The secondary structure elements of AtSEP3 K-domain (PDB 4OX0), are shown in blue above the sequences (α ; TT; η). Residues involved in dimerisation and tetramerisation in SEP3 K-domain structure are highlighted in violet and cyan, respectively. The kink region between helices 1 and 2 is framed in red; Gly and Pro residues present within the kink regions are highlighted in green; Gly residues in the N-terminal region of helix 2 are highlighted in yellow. **(B)** Structure of SEP3 K-domain (PDB 4OX0). The oligomerisation domains of SEP3 are represented as cylinders; each monomer, composed of two distinct helices (helices 1 and 2), is colored uniquely in green, dark green, blue, and light blue. N and C-terminal regions are indicated. **(C)** Close-up of the SEP3 kink between helices. Glycine and proline residues are depicted as sticks colored by atom with carbons in orange.

was reached. At this point, protein expression was induced by addition of 1 mM isopropyl- β -D-galactopyranoside (IPTG) and the temperature reduced to 20°C; expression was continued for 16 h (overnight). Cells were harvested by centrifugation at 6000 rpm for 30 min at 4°C and then resuspended in Buffer A [30 mM Tris pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 2 mM TCEP] to which benzonase (Sigma) and protease inhibitors (Roche EDTA-free) were added. Cells were disrupted by sonication, followed by centrifugation at 25000 rpm for 40 min at 4°C, to remove cell debris. The cell lysate was then passed onto a column containing 1 mL of Ni-Sepharose High-Performance resin (GE-Healthcare), previously equilibrated with Buffer A. Bound protein was washed in two steps: high salt (30 mM Tris pH 8.0, 1 M NaCl, 5% glycerol, 2 mM TCEP) and low imidazole concentration (buffer A + 20 mM Imidazole); and subsequently eluted with Buffer B (30 mM Tris pH 8.0,

300 mM NaCl, 5% glycerol, 250 mM Imidazole, 2 mM TCEP). Fractions of interest were pooled and dialysed overnight at 4°C, against Buffer A and in the presence of 2% (w/w) TEV protease, in order to cleave the 6xHis tag. The protein sample was passed over the same 1 mL Ni-Sepharose column, in order to deplete the His-tagged TEV protease and remove uncut protein from the cleaved protein sample. The purified protein was then concentrated and applied onto a size exclusion Superdex 200 10/300 GL column (GE Healthcare), pre-equilibrated with Buffer A. SEP3^(75–178) and all mutants were purified following this same protocol.

EMSA Experiments

AG, SEP3 full length wild type, and SEP3 mutants (L171A, L115R, SEP3^{ΔC}) were cloned into a pSPUTK plasmid and used for *in vitro* transcription translation (Promega SP6 High Yield

Expression System). SEP3^{ΔC} contains residues 1–160 with a –LADG-stop terminating sequence corresponding to a complete truncation of the C-terminal domain. Protein expression was performed as per the manufacturer's protocol and used without further purification. *SOC1* promoter DNA (121 bp *SOC1* specific DNA) comprising two CArG boxes was used as per (Kaufmann et al., 2009). Two *SOC1* promoter DNA fragments were generated with either the first or second CArG box mutated. Mutations were generated using a 1 kb *SOC1* promoter DNA as the template (inserted into pCR blunt vector) and using Phusion polymerase (NEB) according to the manufacturer's protocol. CArG-box 1 was mutated with the forward primer 5'-CGTGTCTAAAGAGGCATTGACATATGACGTCCTCG GATTACTAAAG-3' and the reverse primer 5'-CTTTAG TAATCCGAGGGACGTCATATGTCAAATGCCTCTTTAGACACG (CArG-box 1 mutation is underlined); and CArG-box 2 mutated with the forward primer 5'-GTGGCA CCAAAAAATATACATATGACGAGATAAAATTGTTAATC G-3' and the reverse primer 5'-CGATTAACAATTTT ATCTCGTCATATGTATATTTTTTGGTGCCAC-3' (CArG-box 2 mutation is underlined). Final 145 bp mutated *SOC1*

DNA fragments were then PCR amplified using the primers 5'-CTAAAGAGGCATTTGACATATGACGTCCTCG (fwd) and 5'-GATTAACAATT TTATCTCCAAAAAAGGATATTTTTTTTG G (rev) for CArG-box 1, and 5'-CTAAAGAGGCATTTG CTATTTTTTGGTCCCTCG (fwd) and 5'-GATTAACAATTTA TCTCGTCATATGTATATTTTTTTTGG (rev) for CArG-box 2 mutated DNAs, respectively. *SOC1* DNA labeled with DY-682 (Dyomics GmbH, wild type) or Cy5 (Eurofins, mutated CArG-boxes) was used at a concentration of approximately 5–10 nM for all reactions in a protein binding buffer containing 7 mM HEPES, pH 7.0, 1 mM BSA, 1 mM EDTA, 1 mM DTT, 2.5% CHAPS, 6% glycerol, 0.06 mg/ml salmon sperm DNA, 1.3 mM spermidine. 4 μl of TnT protein mix was added directly without purification to the binding buffer to a final volume of 20 μl.

AG Expression and Purification

AG^(74–173) was cloned into a pESPRIT002 vector using NotI and AatII restriction sites. The construct contained an N-terminal TEV protease cleavable poly-histidine tag (Hart and Tarendeau, 2006; Guilligay et al., 2008). The protein was overexpressed in *E. coli* BL21 Star (DE3)pLysS cells (Life Technologies). Cells were grown in Luria Bertani medium in the presence of 50 mg/ml kanamycin and 35 mg/ml chloramphenicol at 37°C and 180 rpm to an optical density $A_{600} = 0.8$ after which time the temperature was lowered to 20°C and 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added for induction. After 16 h, the cells were harvested by centrifugation at 6000 rpm and 4°C for 15 min and resuspended in lysis buffer containing 30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% (v/v) glycerol, 20% (w/v) sucrose and 1x protease inhibitors (Roche EDTA-free). Cells were lysed by sonication and the insoluble fraction pelleted by centrifugation at 25000 rpm and 4°C for 30 min. The pellet was resuspended in denaturation buffer [30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% (v/v) glycerol, 8 M Urea] and incubated for 1 h at room temperature. The solubilized fraction was applied to a 5 ml Ni-NTA column pre-equilibrated with denaturation buffer, followed by a wash with 10 CV of wash buffer (30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol, 8 M Urea, 30 mM imidazole) and eluted with 3 CV of elution buffer (30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol, 8 M Urea, 300 mM Imidazole). The eluted fraction was dialysed step-wise against 6, 4, and 2 M urea plus 30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol. After the final dialysis step, the protein was applied to a size exclusion chromatography column (Superdex 75 10/300 GL, GE Healthcare) pre-equilibrated with gel filtration buffer [30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% (v/v) glycerol]. The purity of the final fractions was assessed using SDS-PAGE. Fractions of interest were pooled and incubated overnight with TEV protease to remove the poly-histidine tag. After depletion of TEV and uncleaved protein over a 5 ml Ni-NTA column, the cleaved AG^(74–173) was loaded onto a Superdex S75 10/300 GL column as a final purification step and the fractions of interest pooled and concentrated to approximately 4 mg/ml for SAXS studies.

TABLE 1 | Oligonucleotides used for SEP3^(75–178) mutagenesis.

Mutant	Oligonucleotides
M150A	5'-CTCTCAGGACACAGTTTGCCTTGACCAGCTCAAC-3' 5'-GTTGAGCTGGTCAAGCGCAAACTGTGCTCTGAGAG-3'
L154A	5'-CAGTTTATGCTTGACCGAGCGAACGATCTTCAGAGTAAGG-3' 5'-CCTTACTCTGAAGATCGTTGCGCTGGTCAAGCATAACTG-3'
L171A	5'-CTGACTGAGACAAATAAACTGCAAGACTAAGGTTAGCTGATGG-3' 5'-CCATCAGCTAACCTTAGTCTTGACAGTTTATTTGTCTCAGTCAG-3'
Splice ^{Δ161–174}	5'-CCAGCTCAACGATCTTCAGAGTAAGCTAGCTGATGGATGAGAGACAAATAAACTCTAAGACTAAGG-3' (forward) 5'-CCTTAGTCTTAGAGTTTTATTTGTCTCTCATCCATCAGCTAGCTTACTCTGAAGATCGTTGAGCTGG-3' (reverse)
Y98K	5'-CTTAGTAGCCAGCAGGAGAGCTCAAGCTTAAGGAGCG-3' 5'-CGTCTCTTAAGCTTGAGCTTCTCTGCTGGCTACTAG-3'
Y105N	5'-CTCAAGCTTAAGGAGCGTAATGACGCTTACAAAGAACCC-3' 5'-GGGTTCTTTGTAAGGCGTCATTACGCTCCTTAAGCTTGAG-3'
L115R	5'-CAAAGAACCCAAAGGAATAGGTTGGGAGAAGATCTTGGACCTC-3' 5'-GAGGTCCAAGATCTTCTCCCAACCTATTCCTTTGGGTTCTTTG-3'
L131V	5'-CTAAGTACAAAGGAGCTTGAGTCAGTTGAGAGACAGCTTGATTC-3' 5'-GAATCAAGCTGTCTCTCAACTGACTCAAGCTCCTTTGTACTTAG-3'
L135M	5'-CTTGAGTCACTTGAGAGACAGATGGATTCTTCCTTGAAGC-3' 5'-GCTTCAAGGAAGAATCCATCTGTCTCTCAAGTGAAGTCAAG-3'
L135A	5'-GTCACCTTGAGAGACAGGCTGATTCTTCTTGAAGC-3' 5'-GCTTCAAGGAAGAATCAGCTGTCTCTCAAGTGAC-3'

SAXS Data Collection

An on-line hplc system (Viscotek, Malvern Instruments) was attached directly to the sample inlet valve of the BM29 sample changer (European Synchrotron Radiation Facility, bioSAXS bending magnet beamline 29; Pernot et al., 2013; Round et al., 2015). The protein sample (50 μ l) was injected onto the column (Superdex 75 3.2/300 PC, GE Healthcare) after column equilibration. Buffers were degassed prior to the run and a flow rate of 0.1 ml/min at room temperature was used. Buffers used were as described above. All data from the run was collected using a sample to detector (Pilatus 1 M Dectris) distance of 2.86 m corresponding to an s range of 0.04–4.9 nm^{-1} . Approximately 1800 frames (1 frame/sec) per hplc run were collected. Initial data processing was performed automatically using the EDNA pipeline (Incardona et al., 2009), generating radially integrated, calibrated, and normalized 1-D profiles for each frame. All frames were compared to the initial frame and matching frames were merged to create the reference buffer. Any subsequent frames which differed from the reference buffer were subtracted and then processed within the EDNA pipeline using tools from the EMBL-HH ATSAS suite (Petoukhov and Svergun, 2007). The invariants calculated by the ATSAS autoRg tool were used to select a subset of frames from the peak scattering intensity. The 49 frames corresponding to the highest protein concentration were merged manually and used for all further data processing and model fitting. Molecular weight for the protein was estimated based on the correlated volume (Rambo and Tainer, 2013). The approximate molecular weight was 21 kDa, corresponding to a dimer. The volume of 36 nm^3 was calculated using the GNOM interface of the cross platform version of PRIMUS for the ATSAS software suite.

AG Model Fitting

Homology models for AG^(74–173) were generated based on the SEP3 structure (PDB 4OX0; Puranik et al., 2014). For the elongated conformation, the kink between helices 1 and 2 was removed, the helices superposed and residues corresponding to the flexible region between the helices built in manually using COOT with idealized geometry and no secondary structure restraints. The model for the bent conformation was generated by threading the sequence of AG^(74–173) directly onto the SEP3 dimer (4OX0). Structures corresponding to two different dimer conformations (bent and elongated) were used to calculate theoretical scattering curves. These curves were compared with the experimental data using CRY SOL (Svergun et al., 1995).

RESULTS AND DISCUSSION

LEAFY and NEEDLY Structure and Function-Homology Modeling of the DBDs

The angiosperm *LFY* gene is most often found as a single copy (Brunkard et al., 2015), however, gymnosperms possess two paralogous genes- *LFY* and *NLY*, born from an ancient duplication which occurred before the divergence of the

angiosperm and extant gymnosperm lineages. Examination of the genomes of gymnosperms available through the 1000 plant genomes project as well as all partial deposited sequences reveals that *LFY* and *NLY* are present in all gymnosperm genomes characterized to date, with the exception of the genus *Gnetum* where *NLY* is absent. The proteins the *LFY* and *NLY* genes encode comprise two distinct domains, a partially conserved N-terminal domain (**Figure 2**) important for complex formation and a highly conserved C-terminal DBD (70% sequence identity between AtLFY and WmNLY, for example; **Figure 3A**), with connecting regions presenting a higher degree of variability. In order to probe the function of these proteins, we first aligned the DBDs of LFY and NLY using ClustalW in order to assess conservation of DNA-binding specificity (**Figure 3A**). We observed that the DBDs of LFY and NLY are highly conserved in all seed plants based on sequence alignment. To investigate any potential changes in quaternary structure or putative alterations in the DNA-binding interface, the crystal structure of the DBD of LFY from *A. thaliana* (Hames et al., 2008) was used as a homology model to generate 3D models of the DBDs of gymnosperm LFY and NLY (**Figure 3B**) using SWISS-MODEL with default parameters. Comparison of the primary sequences with secondary, tertiary and quaternary structure derived from the crystallographic data revealed that the DBDs are structurally identical and all amino acids involved in direct contacts with DNA are completely conserved between angiosperm LFY (aLFY), gymnosperm LFY (gLIFY), and NLY. In addition, the dimerisation interface recently described as a key component in DNA binding specificity (Sayou et al., 2014) is also highly conserved between gymnosperms and angiosperms as shown in **Figure 3**. However, while AtLFY His383 is almost completely conserved in both angiosperms and gymnosperms, based on all available sequence data, the residue at position AtLFY 386 varies as either an arginine in aLFY and NLY (Arg399 in PrNLY, **Figure 3C**) or by substitution as a lysine in gLIFY (Lys405 in PrLFY, **Figure 3D**). Arginine and lysine fulfill similar structural roles and can substitute for one another due to the conserved positive charge and hydrogen bonding ability of the primary ϵ -amine and guanidine group for lysine and arginine, respectively (Sokalingam et al., 2012). However the higher pKa and longer size of the arginine side chain may affect the hydrogen bonding interaction with the carbonyl oxygen of residue 276 (AtLFY; residue 289, PrNLY, **Figure 3C**) and cannot be ruled out as affecting dimer stability, and possibly conformation (relative positioning of the monomers). Overall, the high degree of sequence identity between the DBD of aLFY, gLIFY, and NLY implies a likely conserved recognition of cognate DNA sequences. Recent studies by Sayou et al. (2014) have demonstrated the evolutionary trajectory of LFY from green algae to moss to angiosperms based on structural and biochemical studies of several DBDs including those of *Klebsormidium subtile* LFY (algae), *Physcomitrella patens* LFY (moss), and *Arabidopsis* LFY (Sayou et al., 2014). The distantly related LFY from algae, moss and angiosperms were shown to bind different DNA motifs due to small changes in the LFY dimerisation interface, as well as in two other key amino acids (AtLFY His308 and Arg341) that determine the DNA half-site sequence recognized (**Figure 3A**), as previously determined through a combination of structural and

SELEX experiments (Sayou et al., 2014). However, the SELEX motif for the DBD of gymnosperm *G. biloba* LFY (GbLFY) is almost identical to the SELEX motif for the DBD of AtLFY (Sayou et al., 2014). As the dimerisation interface and all residues directly contacting the DNA are highly conserved in angiosperms and gymnosperms for aLFY, gLFY and NLY, this suggests the proteins are able to bind the same or very similar DNA motifs. Thus, it is probable that the DBD of LFY/NLY in higher plants became fixed, with conservation of DNA binding and dimerisation motifs. While bound DNA sequences and DNA binding matrices are available for LFY from *A. thaliana* based on multiple ChIP-seq and SELEX studies, no such data is available for gLFY or NLY with the exception of the GbLFY motif. Additional data would be important to confirm that there are no subtle allosteric effects that may tune the DNA binding specificity of these different paralogs, a possibility that cannot be excluded based on available data.

Functional Implications of Complex Formation- the Role of the N-terminal Domain in LFY and NLY Function

Interestingly, functional studies do not show full complementation of a *lfy* mutant in *A. thaliana* by either *gLFY* (from *P. radiata*) or *NLY* (from *W. mirabilis*; Maizel et al., 2005). If the DBDs are able to recognize the same DNA sequences, why do gLFY and NLY less efficiently complement the *Arabidopsis lfy* mutant? One explanation relies on complex formation with ternary factors that may tune DNA binding specificity, for example through multi-site binding of different adjacent *cis*-elements. This suggests that differences in target gene regulation for aLFY, gLFY, and NLY likely rests on the structure and function of the N-terminal non-conserved regions of the LFY and NLY proteins. While the DBDs are virtually identical, the sequence conservation in the N-terminal regions of aLFY, gLFY, and NLY is much lower (**Figures 2 and 3A**).

The ability to interact with specific partners and form different ternary complexes changes the ability of a TF to regulate downstream genes. By retaining the core DBD and the essential DNA-binding functionality, the N-terminal region of the protein could vary, thus leading to relatively smooth changes in gene regulation over the course of evolution by simply tuning the interactions with ternary partners and thus modulating interactions with cognate DNA without requiring altering the DBD itself. The N-terminal ~200 residues of LFY have been shown to be important for dimerisation (Siriwardana and Lamb, 2012) and can possibly play a role in the formation of higher order complexes with chromatin remodelers and other TFs (Wu et al., 2012). Indeed, unfolded, flexible loops, and low-complexity regions exhibit greater variability and tolerance for mutations, as they do not affect the overall fold of the macromolecule. In addition, these regions often have important functions and act as protein-protein interaction surfaces (Dyson and Wright, 2005). While alpha-helices are relatively disfavored as protein-protein interaction interfaces, exposed beta strands, hydrophobic patches and long loops are more likely to play a role in complex formation (Jones and Thornton, 1996; Neuvirth et al., 2004). These structural motifs are able to create relatively planar surfaces

which are often correlated with protein-protein interactions (Hoskins et al., 2006). Few mapping studies of LFY have been performed and only a small number of interaction surfaces with partner proteins have been determined (Chae et al., 2008; Souer et al., 2008; Pastore et al., 2011; Siriwardana and Lamb, 2012; Wu et al., 2012). From the limited data available, however, it seems that several partners interact with the N-terminal region of the protein (Souer et al., 2008; Siriwardana and Lamb, 2012). Structural characterization of the N-terminal domain of LFY would allow determining whether its properties might have changed during evolution.

Due to the loss of NLY in the angiosperm lineage, aLFY likely assumed additional functions, fusing the functionality of NLY, a key regulator of female organ development, and gLFY, an important primary regulator of male cone development, into one fully competent regulator of plant reproduction. As has been recently shown for several conifers (*Picea abies*, *Podocarpus reichei*, and *Taxus globosa*), LFY and NLY have overlapping expression patterns (Vazquez-Lobo et al., 2007; Carlsbecker et al., 2013). This would mitigate any deleterious effects of NLY loss during the gymnosperm/angiosperm split by allowing more facile compensation for NLY function by LFY, as LFY was already present in the same tissues, possessed the same DBD, and likely recognized very similar cognate DNA sequences. Thus, aLFY compensation for NLY/gLFY during reproductive development would not necessitate extensive reprogramming of LFY expression patterns nor require any changes to the gene coding sequence of the DBD, important factors in the successful compensation due to gene loss of NLY in the angiosperm lineage.

MADS-Domain TFs and Their Role in Floral Organ Development

The homeotic class A-E MADS-box genes direct the specification of all the floral organs and as such are central players in flower evolution and development. In gymnosperms, orthologs to the B and C class MADS-box genes (*AP3/PI* and *AG* in *Arabidopsis*) are also present and play important roles in male and female organ development. While the MADS-box gene family has expanded in all land plants, this is most striking in angiosperms due to extensive duplication events giving rise to the class E *SEPALLATA* genes, which are not present in extant gymnosperms (Zahn et al., 2005a). The *SEPALLATA* (*SEP*) proteins have acquired new functionality and act as mediators of interactions between class A, B, and C MADS-domain TFs as shown by yeast two and three hybrid studies, EMSA experiments and *in vivo* studies (Pelaz et al., 2000; Honma and Goto, 2001; Kaufmann et al., 2005; Malcomber and Kellogg, 2005; Theissen and Melzer, 2007; Immink et al., 2009; Mendes et al., 2013). The *SEP* proteins form heteromeric complexes with other MADS TFs and all putative floral organ-specifying tetrameric MADS complexes contain at least one *SEP* protein leading to the specification of the different floral organs (Theissen, 2001; Theissen and Saedler, 2001). Indeed, *sep123* mutants are sterile and unable to produce male or female organs, with the flower converted to a collection of sepaloid-like structures, illustrating the requirement of the *SEP* proteins for proper reproductive

organ formation (Pelaz et al., 2000). Examination of the B and C class MADS TFs in gymnosperms such as *G. gnemon* suggests that tetramerisation can occur and is necessary for male and female organ development (Figure 1). This tetramerisation takes place without the obligatory mediation of the class E-like AGL6 proteins (Wang et al., 2010). However, angiosperms are dependent on the class E SEPs for tetramer formation, as the B and C class TFs have lost their ability to directly interact. Current hypothesis suggest that the changing interaction patterns of the MADS TFs, in particular the requirement of the SEPs to mediate tetramer formation in angiosperms is at the nexus of flower origins (Melzer et al., 2010; Wang et al., 2010). The evolution of the bisexual flower thus requires an understanding, at the protein level, of the MADS TFs, particularly how the SEPs are able to mediate the formation of tetrameric complexes which are critical to the development of all the floral organs.

Our recent crystallographic data of the oligomerisation domain of SEP3 (Puranik et al., 2014), together with mutagenesis studies, sequence alignments and biophysical characterization of the C-class MADS TF AGAMOUS (this study) help to explain the molecular function of the MADS TFs and contribute to our understanding of flower evolution. All MADS homeotic TFs are characterized by a four domain arrangement consisting of a highly conserved DBD “M” domain (~60 amino acid MADS domain, Figure 4), an “I” domain (linker Intervening domain) important for dimerisation, a “K” domain (alpha helical Keratin-like domain) critical for dimerisation and tetramerisation, and a “C” domain (highly variable C-terminal domain) important for different functions including transactivation and higher order complex formation (Kaufmann et al., 2005). Based on the crystal structure of a portion of the I and the full K domain of SEP3 (Puranik et al., 2014) and extensive mutagenesis studies, the dimerisation and tetramerisation interfaces of the MADS-domain TFs can be mapped at the amino acid level (Figures 5A,B). Different amino acids along the dimer and tetramer interface were targeted for mutagenesis studies in order to probe the mechanisms of oligomerisation and stability (Table 2). Mutation of any residue making a direct contact with its partner along the dimer (Leu115, Leu131, Leu135, Tyr98, Tyr105; this study) or tetramer (Met150, Leu154, Leu171; Puranik et al., 2014) interface in SEP3 had a striking effect on oligomerisation, with even a single point mutation greatly destabilizing the complex as determined by size exclusion chromatography and comparison with the wild type protein. This suggests that subtle differences in the amino acids at the dimerisation and tetramerisation surface will shift the oligomerisation equilibrium to favor certain complexes when multiple MADS TFs are present. Examining structure based sequence alignments for the homeotic MADS-domain TFs demonstrates a conservation of hydrophobic residues at the oligomerisation interface, but the size and shape of these residues varies, which will help mediate protein–protein interactions (Figure 5A).

Based on the structure of the SEP3 homotetramer and mutagenesis studies, we probed the formation of hetero-oligomers using electrophoretic mobility shift assays (Figure 6). EMSA experiments and identification of putative complexes were

TABLE 2 | Effect of point mutations on SEP3 oligomerisation.

	SEP3 ^(75–178) constructs	Oligomerisation state
Tetramerisation interface (Puranik et al., 2014)	Wild type	Tetramer /dimer
	M150A	Dimer
	L154A	Dimer
	L171A	Dimer
	Splice ^{Δ161–174}	Dimer
Dimerisation interface (this study)	Y98K	Dimer /monomer
	Y105N	Tetramer /monomer
	L115R	Unstable complex
	L131V	Unstable complex
	L135M	Dimer /monomer
	L135A	Unstable complex

Point mutations targeting the highly conserved residues involved in the putative dimerisation and tetramerisation interface of SEP3 were chosen for mutational analysis. Oligomeric state was determined by size exclusion chromatography. Where two states exist, the predominant species is marked in bold. “Unstable complex” is used to denote a complex mixture of species between monomeric, dimeric, and tetrameric states with no predominance for a particular oligomerisation state.

performed according to previously published work (Smaczniak et al., 2012). SEP3 dimerisation and tetramerisation mutants were tested for DNA binding with AG, all expressed using *in vitro* transcription translation due to the difficulties in producing folded full length MADS TFs using standard recombinant bacterial expression. Sufficient heterodimers and tetramers were produced and a gel shift assay was performed using DNA corresponding to the *SOC1* promoter containing two CArG-box MADS TF binding sites (Figure 6A) and the *SOC1* promoter sequence with either the first or the second CArG-box mutated (Figure 6B). A SEP3 dimerisation-interface mutant, SEP3^{L115R}, was dramatically impaired in its ability to oligomerise based on studies of the K-domain alone (Table 2), however, it was able to bind DNA as a homodimer and heterotetramer with AG, albeit with less efficiency than the wild type SEP3. The SEP3^{L115R} mutant was designed to mimic the sequence of AtAP3, which is unable to form homodimers but still retains the ability to interact with partners such as AtPI (Riechmann et al., 1996; Winter et al., 2002; Yang et al., 2003). Both AtPI and AtAG have a leucine residue at position 115, which is likely able to accommodate the arginine side chain during hetero-oligomer formation. AGAMOUS alone exhibited poor binding to the *SOC1* DNA due either to lower protein production in the *in vitro* transcription translation reaction or non-optimal sequences of the DNA, however, AG heterodimers with SEP3 were able to bind the *SOC1* sequences, suggesting differences in sequence specificity are important for AG homo and heteromer DNA binding interactions. Tetramerisation interface mutants SEP3^{L171A} and a truncation mutant (SEP3^{ΔC}) showed greatly impaired heterotetramerisation with AG, as expected. Altogether, these data provide strong evidence that the homotetramerisation interface observed in the crystal structure of SEP3 is conserved in the formation of heterotetramers.

Changes in the tetramerisation interface in SEP partner MADS proteins also has an effect on oligomer formation. For example, studies of the C-class genes *PLENA* (*PLE*) and

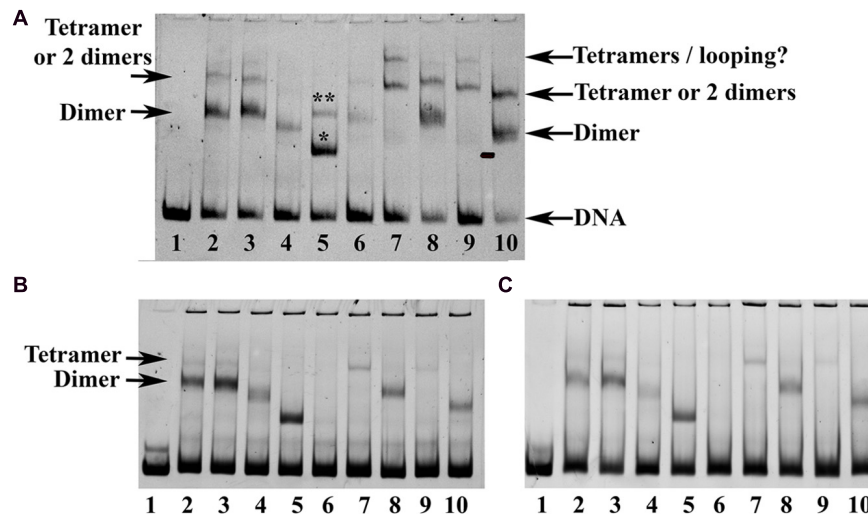


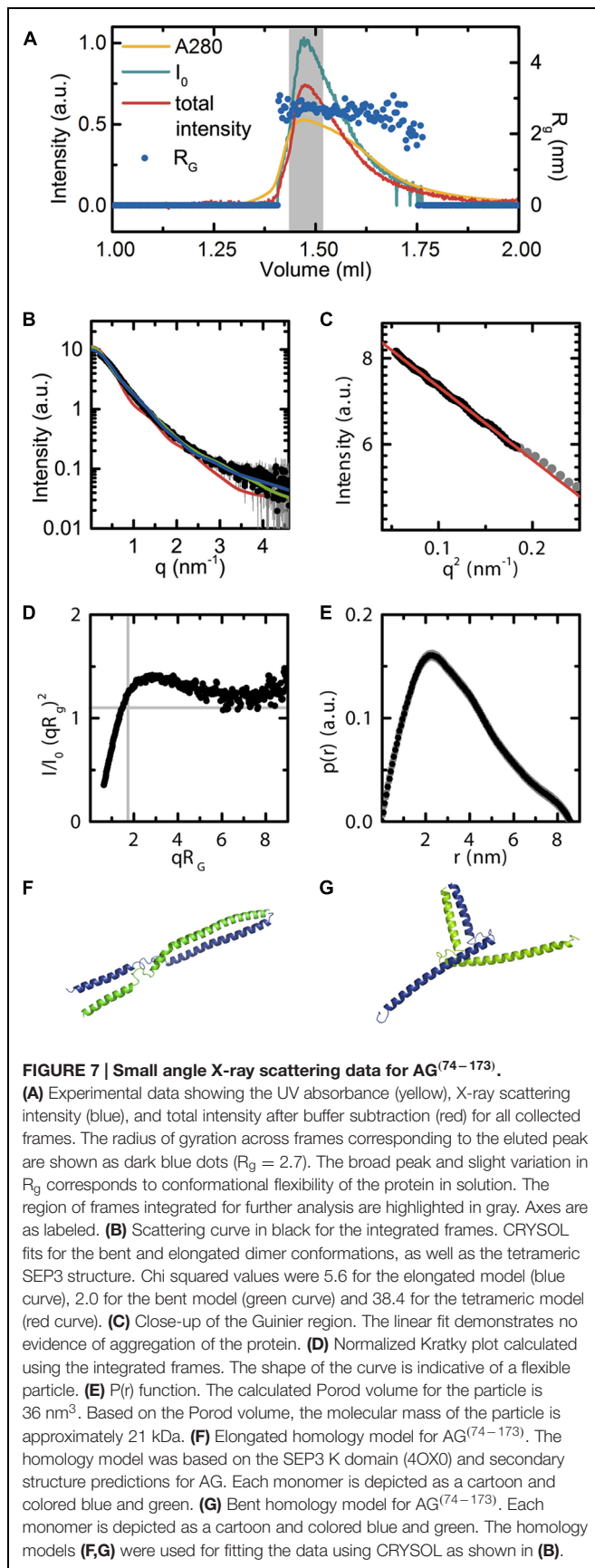
FIGURE 6 | Electrophoretic mobility shift assay (EMSA) for SEP3 and AG. (A) Comparison of the oligomerisation state of SEP3 wild type, dimerisation and tetramerisation mutants with AG using a 121 bp DNA fragment of the SOC1 promoter comprising two CArG boxes. Lane 1 corresponds to DNA alone, lane 2 SEP3 wild type, lane 3 SEP3^{L171A}, lane 4 SEP3^{L115R}, lane 5 SEP3^{ΔC} truncation mutant, lane 6 AG, lane 7 SEP3 + AG, lane 8 SEP3^{L171A} + AG, lane 9 SEP3^{L115R} + AG, lane 10 SEP3^{ΔC} + AG. Putative tetramer and dimer are indicated by arrows. The truncation mutant in lane 5 likely has one or two dimers bound to DNA as indicated above the bands by a single (one dimer bound) or double (two dimers bound) asterisk. A faint highly retarded band corresponding to bound tetramers or tetramer-induced DNA-looping was noted in lanes 7 and 9 and indicated. **(B)** EMSAs run with either CArG box 1 (left) or CArG box 2 (right) mutated (see Materials and Methods). Proteins are as per **(A)**. A faint band for SEP3 wild type and SEP3^{L171A} was noted running as per the tetramer band in **(A)**, suggesting homotetramer formation on a single CArG box site. All proteins were produced via *in vitro* transcription translation using equivalent amounts of template DNA and equivalent volumes of the reaction mixture were added to the final binding reaction. DNA was approximately 5–10 nM and labeled with DY-682 **(A)** or Cy5 **(B)** for imaging.

FARINELLI (*FAR*) from *A. majus* demonstrate that a single amino acid change was responsible for neofunctionality of these duplicated genes with *FAR* able to specify only male organs and *PLE* able to specify both male and female organs in a complementation assay in *Arabidopsis*. This activity was due to a single amino acid insertion in the K domain that altered the oligomerisation capabilities of *PLE* and *FAR* with the *SEPALLATA* proteins (Airoldi et al., 2010). An amino acid insertion shifts the hydrophobic pattern of all amino acids in the leucine zipper tetramerisation interface, thus modulating the hydrophobic protein–protein interface of the putative tetrameric complexes formed by *PLE* and *FAR* with their *SEP* partners.

In addition to the hydrophobic dimer and tetramer interface acting as a driver for oligomerisation, a key component of the MADS TFs oligomerisation propensity is the presence of a kink in between alpha helices 1 and 2 of the K domain (**Figures 5B,C**). Based on sequence alignments of the MADS homeotic TFs, this kink region is highly variable in the family with a tight turn predicted for *SEP1*, *SEP2* and *SEP3* due to the presence of a GlyPro motif (**Figure 5A**). Prolines act as “breakers” in an alpha helix due to their inability to form the appropriate hydrogen bonding interactions between the carbonyl backbone and amide proton due to the presence of the proline side chain. Glycine residues exhibit a high degree of conformational flexibility and have been shown to lead to kinks in alpha helices in soluble and membrane proteins (Wilman et al., 2014). These residues result in the formation of a tight turn and, in the case of *SEP3*, an approximately 90°

bend between alpha helices 1 and 2 (**Figure 5B**). Examination of the sequences of other MADS TFs show scattered glycine and/or proline residues between helices 1 and 2, but not a conservation of the GlyPro motif observed in *SEP1*, *SEP2* and *SEP3*. In order to investigate whether the presence of a GlyPro motif is required for complete opening of helices 1 and 2, we recombinantly overexpressed and purified the K-domain of AG.

The AG^(74–173) construct, spanning the complete AG K domain, was designed based on both secondary structure predictions using PSIPRED (Jones, 1999) and homology modeling with the *SEP3* structure using SWISS-MODEL. This protein was used in small angle X-ray scattering (SAXS) studies to determine oligomerisation state and conformational flexibility of the AG K domain in solution. The AG^(74–173) construct was expressed in *E. coli*, purified from inclusion bodies under denaturing conditions and subsequently refolded. Protein monodispersity and purity were assayed by size exclusion column chromatography (SEC) and SDS-PAGE prior to SAXS experiments. In order to avoid any bias due to protein aggregation or the presence of multiple oligomeric species, the AG^(74–173) construct was purified on-line and the complete elution profile measured directly in the X-ray beam (**Figures 7A–C**). The stable radius of gyration (R_g) across the eluted protein peak corresponding to the highest protein concentration demonstrates that there is one species in solution as the particle size is constant (**Figure 7A**). In contrast to the *SEP3* K-domain, which is predominantly tetrameric in solution (Puranik et al., 2014), the AG K-domain is dimeric. Volume



calculations based on the histogram of interatomic distances for the particle give a volume of 36 nm^3 , corresponding to a molecular mass of approximately 21 kDa, the molecular mass of an AG^(74–173) dimer (**Figure 7E**). AG^(74–173) exhibits a great deal of flexibility based on the Kratky plot (**Figure 7D**), which is characteristic of a highly flexible and/or partially disordered protein in solution. In order to further investigate the possible conformations of the AG^(74–173) dimer, homology models based on the structure of SEP3 (4OX0) were generated in an elongated and bent conformation (**Figures 7F,G**). CRYSOLOG fits (**Figure 7B**) were relatively consistent with either particle shape giving chi-squared values of 5.6 and 2.0 for the elongated and bent conformations, respectively. In contrast, the tetrameric SEP3 structure is inconsistent with the recorded data, giving a chi-squared of 38.4 (**Figure 7B**). The R_g for both dimeric homology models (3.1 nm for the bent and 3.6 nm for the elongated model) was slightly bigger than the calculated R_g of 2.7 nm for the measured data. This variation is attributable to disorder, multiple unmodeled conformations and/or partial unfolding at the termini of the protein. Contamination by a tetramer or soluble aggregates is considered highly unlikely as these species would elute prior to the measured peak and there is no evidence for this in the UV trace or X-ray scattering of the sample.

While possessing glycine residues in the kink region between helices 1 and 2, AG lacks the GlyPro motif seen in SEP1, SEP2 and SEP3. Although it is well-established that AG can form tetrameric complexes, these complexes usually contain a SEP partner. Indeed almost characterized floral organ tetrameric complexes of homeotic MADS TFs from angiosperms to date rely on at least one SEP protein for tetramer formation (Honma and Goto, 2001; Theissen and Saedler, 2001). Thus, the SEPs are able to act as hubs of tetramer formation for other MADS TFs. Because the GlyPro motif forces open helix 2 exposing hydrophobic surfaces, we postulate that the SEP proteins are able to preferentially form tetramers with themselves or other MADS TF proteins and this exposed hydrophobic surface on helix 2 acts as an entropic driving force for oligomerisation.

Some gymnosperm B and C-class MADS TFs are postulated to form tetramers when bound to DNA. *In vitro* studies of GGM2 (*G. gnemon* B-like) and GGM3 (*G. gnemon* C-like) demonstrate that GGM2 can form heterotetramers with GGM3 and that GGM3 is additionally able to homotetramerise when bound to DNA (Wang et al., 2010). Examination of the kink region between helices 1 and 2 as determined from secondary structure predictions and sequence alignments for GGM2 and GGM3 reveals the presence of two glycine residues for GGM2 but no proline. GGM3 has scattered glycines in both the kink region and in the N-terminal portion of helix 2 (**Figure 5A**). We speculate that these glycines will destabilize helix 2 and increase the conformational space the protein is able to sample. Indeed, GGM3 was shown to homotetramerise on DNA with non-optimally spaced binding sites, suggesting additional flexibility of the protein and the tetramerisation interface (Wang et al., 2010). It is likely that the combination

of helix destabilization in GGM3 and the relatively plastic kink region in GGM2 is sufficient to allow the formation of tetrameric complexes when the local concentration of the proteins is relatively high as would be the case when bound to adjacent regions of DNA. Further experiments to probe these interactions and extensive mutagenesis studies would be required to fully determine the rules governing tetramerisation. Nascent tetramerisation capabilities are present in at least some species of gymnosperm MADS TFs, though whether tetramerisation is required for proper gene regulation is less clear. However, in angiosperms, interactions mediated by the SEP class of MADS TFs is required for male and female organ specification and reproductive development. The gene duplication event giving rise to the SEPALLATA class of MADS TFs and their central role in organizing the homeotic MADS TFs into functional tetrameric complexes we hypothesize to be a key component in flower origins and evolution.

Taken together, these data suggest that the interaction surfaces and oligomerisation of the MADS TFs is both variable and highly sensitive to even small alterations in amino acid sequence which would allow for the fast evolution of different interactions within the family. By retaining the core DBD, the essential function of the MADS TFs- DNA binding to specific cognate sequences- would be preserved, but mutations in the auxiliary I, K, and C domains would allow for functional plasticity by changing the identity or altering the affinity of protein interaction partners. This model is very similar to what is observed for aLFY, gLFY, and NLY in which the C-terminal DBD is conserved and the auxiliary N-terminal region involved in protein-protein interactions is allowed to vary, likely changing ternary complex formation and tuning downstream gene regulation.

REFERENCES

- Airoidi, C. A., Bergonzi, S., and Davies, B. (2010). Single amino acid change alters the ability to specify male or female organ identity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18898–18902. doi: 10.1073/pnas.1009050107
- Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195–201. doi: 10.1093/bioinformatics/bti770
- Bateman, R. M., Hilton, J., and Rudall, P. J. (2006). Morphological and molecular phylogenetic context of the angiosperms: contrasting the 'top-down' and 'bottom-up' approaches used to infer the likely characteristics of the first flowers. *J. Exp. Bot.* 57, 3471–3503. doi: 10.1093/jxb/erl128
- Becker, A., Kaufmann, K., Freialdenhoven, A., Vincent, C., Li, M. A., Saedler, H., et al. (2002). A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Mol. Genet. Genomics* 266, 942–950. doi: 10.1007/s00438-001-0615-8
- Becker, A., Saedler, H., and Theissen, G. (2003). Distinct MADS-box gene expression patterns in the reproductive cones of the gymnosperm *Gnetum gnemon*. *Dev. Genes Evol.* 213, 567–572. doi: 10.1007/s00427-003-0358-0
- Becker, A., and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* 29, 464–489. doi: 10.1016/S1055-7903(03)00207-0
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42, W252–W258. doi: 10.1093/nar/gku340

CONCLUSION

Small changes in TFs that do not directly affect the DBD can trigger very striking evolutionary developmental changes in an organism. LFY and the MADS TFs illustrate how small changes at the genetic level lead to dramatic alterations and novel functions at the protein level. While the evolutionary origins of the bisexual angiosperm flower are still unclear, major genetic changes - the loss of NLY and the duplication event resulting in the SEPALLATA genes in angiosperms- likely play key roles. How these genetic changes were able to result in morphological changes requires an integrated study incorporating detailed examination of protein structure and biochemistry. By exploring the protein structure-function relationship, particularly for TFs whose activity impacts entire downstream networks, we can begin to understand the molecular basis for evolution. Structural biology offers an important perspective in probing this relationship for the master regulators, LFY and the MADS TFs, and provides a foundation for understanding how alterations in protein structure lead to the evolution of new functions and new morphologies at the organismal level.

ACKNOWLEDGMENTS

We would like to acknowledge Darren Hart and Philippe Mas for the use of the pESPRIT002 vector, Kerstin Kaufmann and Cezary Smaczniak for EMSA protocols and pSPUTK SEP3 and AG TnT vectors and Renaud Dumas for discussions and critical reading of the manuscript. We are grateful to the ESRF for provision of beamtime and support for the experiments. This work was supported by an ATIP-Avenir (CZ).

- Brunkard, J. O., Runkel, A. M., and Zambryski, P. C. (2015). Evolution. Comment on "A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity." *Science* 347, 621. doi: 10.1126/science.1255437
- Burkhardt, F., Sydney, S., David, K., and William, M. (1985). *The Correspondence of Charles Darwin*. Cambridge: Cambridge University Press.
- Busch, M. A., Bomblies, K., and Weigel, D. (1999). Activation of a floral homeotic gene in *Arabidopsis*. *Science* 285, 585–587. doi: 10.1126/science.285.5427.585
- Carlsbecker, A., Sundstrom, J. F., Englund, M., Uddenberg, D., Izquierdo, L., Kvarnheden, A., et al. (2013). Molecular control of normal and acrocona mutant seed cone development in Norway spruce (*Picea abies*) and the evolution of conifer ovule-bearing organs. *New Phytol.* 200, 261–275. doi: 10.1111/nph.12360
- Carlsbecker, A., Tandre, K., Johanson, U., Englund, M., and Engstrom, P. (2004). The MADS-box gene DAL1 is a potential mediator of the juvenile-to-adult transition in Norway spruce (*Picea abies*). *Plant J.* 40, 546–557. doi: 10.1111/j.1365-313X.2004.02226.x
- Chae, E., Tan, Q. K., Hill, T. A., and Irish, V. F. (2008). An *Arabidopsis* F-box protein acts as a transcriptional co-factor to regulate floral development. *Development* 135, 1235–1245. doi: 10.1242/dev.015842
- Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0
- Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000). NPS@: network protein sequence analysis. *Trends Biochem. Sci.* 25, 147–150. doi: 10.1016/S0968-0004(99)01540-6

- De Bodt, S., Raes, J., Van De Peer, Y., and Theissen, G. (2003). And then there were many: MADS goes genomic. *Trends Plant Sci.* 8, 475–483. doi: 10.1016/j.tplants.2003.09.006
- Dornelas, M. C., and Rodriguez, A. P. (2005). The rubber tree (*Hevea brasiliensis* Muell. Arg.) homologue of the LEAFY/FLORICAULA gene is preferentially expressed in both male and female floral meristems. *J. Exp. Bot.* 56, 1965–1974. doi: 10.1093/jxb/eri194
- Dyson, H. J., and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208. doi: 10.1038/nrm1589
- Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. doi: 10.1107/S0907444910007493
- Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., et al. (2003). MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell* 15, 2603–2611. doi: 10.1105/tpc.015123
- Friedman, W. E. (1990). Double fertilization in ephedra, a nonflowering seed plant: its bearing on the origin of angiosperms. *Science* 247, 951–954. doi: 10.1126/science.247.4945.951
- Friedman, W. E. (2009). The meaning of Darwin's 'abominable mystery.' *Am. J. Bot.* 96, 5–21. doi: 10.3732/ajb.0800150
- Friedman, W. E., and Carmichael, J. S. (1996). Double fertilization in gnetales: implications for understanding reproductive diversification among seed plants. *Int. J. Plant Sci.* 157, S77–S94. doi: 10.1086/297405
- Frohlich, M. W., and Chase, M. W. (2007). After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450, 1184–1189. doi: 10.1038/nature06393
- Frohlich, M. W., and Meyerowitz, E. M. (1997). The search for flower homeotic gene homologs in basal angiosperms and Gnetales: a potential new source of data on the evolutionary origin of flowers. *Int. J. Plant Sci.* 158, S131–S142. doi: 10.1086/297513
- Frohlich, M. W., and Parker, D. S. (2000). The Mostly Male theory of flower evolutionary origins: from genes to fossils. *Syst. Bot.* 25, 155–170. doi: 10.2307/2666635
- Gramzow, L., Ritz, M. S., and Theissen, G. (2010). On the origin of MADS-domain transcription factors. *Trends Genet.* 26, 149–153. doi: 10.1016/j.tig.2010.01.004
- Gramzow, L., Weilandt, L., and Theissen, G. (2014). MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Ann. Bot.* 114, 1407–1429. doi: 10.1093/aob/mcu066
- Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., et al. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.* 15, 500–506. doi: 10.1038/nsmb.1421
- Hames, C., Ptchelkine, D., Grimm, C., Thevenon, E., Moyroud, E., Gerard, F., et al. (2008). Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *EMBO J.* 27, 2628–2637. doi: 10.1038/emboj.2008.184
- Hart, D. J., and Tarendeau, F. (2006). Combinatorial library approaches for improving soluble protein expression in *Escherichia coli*. *Acta Crystallogr. D Biol. Crystallogr.* 62, 19–26. doi: 10.1107/S0907444905036097
- Honma, T., and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525–529. doi: 10.1038/35054083
- Hoskins, J., Lovell, S., and Blundell, T. L. (2006). An algorithm for predicting protein-protein interaction sites: abnormally exposed amino acid residues and secondary structure elements. *Protein Sci.* 15, 1017–1029. doi: 10.1110/ps.051589106
- Immink, R. G., Tonaco, I. A., De Folter, S., Shchennikova, A., Van Dijk, A. D., Busscher-Lange, J., et al. (2009). SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. *Genome Biol.* 10, R24. doi: 10.1186/gb-2009-10-2-r24
- Incardona, M. F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N., and Svensson, O. (2009). EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synchrotron Radiat.* 16, 872–879. doi: 10.1107/S0909049509036681
- Jager, M., Hassanin, A., Manuel, M., Le Guyader, H., and Deutsch, J. (2003). MADS-box genes in *Ginkgo biloba* and the evolution of the AGAMOUS family. *Mol. Biol. Evol.* 20, 842–854. doi: 10.1093/molbev/msg089
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202. doi: 10.1006/jmbi.1999.3091
- Jones, S., and Thornton, J. M. (1996). Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13–20. doi: 10.1073/pnas.93.1.13
- Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347, 183–198. doi: 10.1016/j.gene.2004.12.014
- Kaufmann, K., Muino, J. M., Jauregui, R., Airoidi, C. A., Smaczniak, C., Krajewski, P., et al. (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the *Arabidopsis* flower. *PLoS Biol.* 7:e1000090. doi: 10.1371/journal.pbio.1000090
- Kaufmann, K., Wellmer, F., Muino, J. M., Ferrier, T., Wuest, S. E., Kumar, V., et al. (2010). Orchestration of floral initiation by APETALA1. *Science* 328, 85–89. doi: 10.1126/science.1185244
- Krizek, B. A., and Fletcher, J. C. (2005). Molecular mechanisms of flower development: an armchair guide. *Nat. Rev. Genet.* 6, 688–698. doi: 10.1038/nrg1675
- Lamb, R. S., Hill, T. A., Tan, Q. K., and Irish, V. F. (2002). Regulation of APETALA3 floral homeotic gene expression by meristem identity genes. *Development* 129, 2079–2086.
- Liljgren, S. J., Gustafson-Brown, C., Pinyopich, A., Ditta, G. S., and Yanofsky, M. F. (1999). Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *Plant Cell* 11, 1007–1018. doi: 10.1105/tpc.11.6.1007
- Lohmann, J. U., and Weigel, D. (2002). Building beauty: the genetic control of floral patterning. *Dev. Cell* 2, 135–142. doi: 10.1016/S1534-5807(02)00122-3
- Lord, E. M., and Russell, S. D. (2002). The mechanisms of pollination and fertilization in plants. *Annu. Rev. Cell Dev. Biol.* 18, 81–105. doi: 10.1146/annurev.cellbio.18.012502.083438
- Maizel, A., Busch, M. A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M., et al. (2005). The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science* 308, 260–263. doi: 10.1126/science.1108229
- Malcomber, S. T., and Kellogg, E. A. (2005). SEPALLATA gene diversification: brave new whorls. *Trends Plant Sci.* 10, 427–435. doi: 10.1016/j.tplants.2005.07.008
- Martinez-Castilla, L. P., and Alvarez-Buylla, E. R. (2003). Adaptive evolution in the *Arabidopsis* MADS-box gene family inferred from its complete resolved phylogeny. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13407–13412. doi: 10.1073/pnas.1835864100
- Mellerowicz, E. J., Horgan, K., Walden, A., Coker, A., and Walter, C. (1998). PRFL1—a *Pinus radiata* homologue of FLORICAULA and LEAFY is expressed in buds containing vegetative shoot and undifferentiated male cone primordia. *Planta* 206, 619–629. doi: 10.1007/s004250050440
- Melzer, R., Wang, Y. Q., and Theissen, G. (2010). The naked and the dead: the ABCs of gymnosperm reproduction and the origin of the angiosperm flower. *Semin. Cell Dev. Biol.* 21, 118–128. doi: 10.1016/j.semdb.2009.11.015
- Mendes, M. A., Guerra, R. F., Berns, M. C., Manzo, C., Masiero, S., Finzi, L., et al. (2013). MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in *Arabidopsis*. *Plant Cell* 25, 2560–2572. doi: 10.1105/tpc.112.108688
- Mouradov, A., Glassick, T., Hamdorf, B., Murphy, L., Fowler, B., Marla, S., et al. (1998). NEEDLY, a *Pinus radiata* ortholog of FLORICAULA/LEAFY genes, expressed in both reproductive and vegetative meristems. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6537–6542. doi: 10.1073/pnas.95.11.6537
- Moyroud, E., Kusters, E., Monniaux, M., Koes, R., and Parcy, F. (2010). LEAFY blossoms. *Trends Plant Sci.* 15, 346–352. doi: 10.1016/j.tplants.2010.03.007
- Nam, J., DePamphilis, C. W., Ma, H., and Nei, M. (2003). Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol. Biol. Evol.* 20, 1435–1447. doi: 10.1093/molbev/msg152
- Neuirth, H., Raz, R., and Schreiber, G. (2004). ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J. Mol. Biol.* 338, 181–199. doi: 10.1016/j.jmb.2004.02.040
- Ng, M., and Yanofsky, M. F. (2001). Function and evolution of the plant MADS-box gene family. *Nat. Rev. Genet.* 2, 186–195. doi: 10.1038/35056041
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584. doi: 10.1038/nature12211
- Parcy, F., Nilsson, O., Busch, M. A., Lee, I., and Weigel, D. (1998). A genetic framework for floral patterning. *Nature* 395, 561–566. doi: 10.1038/26903

- Pastore, J. J., Limpuangthip, A., Yamaguchi, N., Wu, M. F., Sang, Y., Han, S. K., et al. (2011). LATE MERISTEM IDENTITY2 acts together with LEAFY to activate APETALA1. *Development* 138, 3189–3198. doi: 10.1242/dev.063073
- Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405, 200–203. doi: 10.1038/35012103
- Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., et al. (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Radiat.* 20, 660–664. doi: 10.1107/S0909049513010431
- Petoukhov, M. V., and Svergun, D. I. (2007). Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr. Opin. Struct. Biol.* 17, 562–571. doi: 10.1016/j.sbi.2007.06.009
- Puranik, S., Acajjaoui, S., Conn, S., Costa, L., Conn, V., Vial, A., et al. (2014). Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in *Arabidopsis*. *Plant Cell* 26, 3603–3615. doi: 10.1105/tpc.114.127910
- Rambo, R. P., and Tainer, J. A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496, 477–481. doi: 10.1038/nature12070
- Riechmann, J. L., Krizek, B. A., and Meyerowitz, E. M. (1996). Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc. Natl. Acad. Sci. U.S.A.* 93, 4793–4798. doi: 10.1073/pnas.93.10.4793
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316
- Round, A., Felisaz, F., Fodinger, L., Gobbo, A., Huet, J., Villard, C., et al. (2015). BioSAXS Sample Changer: a robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments. *Acta Crystallogr. D Biol. Crystallogr.* 71, 67–75. doi: 10.1107/S1399004714026959
- Sablowski, R. (2010). Genes and functions controlled by floral organ identity genes. *Semin. Cell Dev. Biol.* 21, 94–99. doi: 10.1016/j.semcdb.2009.08.008
- Saedler, H., Becker, A., Winter, K. U., Kirchner, C., and Theissen, G. (2001). MADS-box genes are involved in floral development and evolution. *Acta Biochim. Pol.* 48, 351–358.
- Sayou, C., Monniaux, M., Nanao, M. H., Moyroud, E., Brockington, S. F., Thevenon, E., et al. (2014). A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* 343, 645–648. doi: 10.1126/science.1248229
- Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250, 931–936. doi: 10.1126/science.250.4983.931
- Shindo, S., Sakakibara, K., Sano, R., Ueda, K., and Hasebe, M. (2001). Characterization of a FLORICAULA/LEAFY homologue of *Gnetum parvifolium* and its implications for the evolution of reproductive organs in seed plants. *Int. J. Plant Sci.* 162, 1199–1209. doi: 10.1086/323417
- Siriwardana, N. S., and Lamb, R. S. (2012). A conserved domain in the N-terminus is important for LEAFY dimerization and function in *Arabidopsis thaliana*. *Plant J.* 71, 736–749. doi: 10.1111/j.1365-313X.2012.05026.x
- Smaczniak, C., Immink, R. G., Muino, J. M., Blanvillain, R., Busscher, M., Busscher-Lange, J., et al. (2012). Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1560–1565. doi: 10.1073/pnas.1112871109
- Sokalingam, S., Raghunathan, G., Soundarajan, N., and Lee, S. G. (2012). A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS ONE* 7:e40410. doi: 10.1371/journal.pone.0040410
- Souer, E., Rebocho, A. B., Blik, M., Kusters, E., De Bruin, R. A., and Koes, R. (2008). Patterning of inflorescences and flowers by the F-Box protein DOUBLE TOP and the LEAFY homolog ABERRANT LEAF AND FLOWER of petunia. *Plant Cell* 20, 2033–2048. doi: 10.1105/tpc.108.060871
- Sundstrom, J., Carlsbecker, A., Svensson, M. E., Svensson, M., Johanson, U., Theissen, G., et al. (1999). MADS-box genes active in developing pollen cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. *Dev. Genet.* 25, 253–266. doi: 10.1002/(SICI)1520-6408(1999)25:3<253::AID-DVG8>3.0.CO;2-P
- Sundstrom, J., and Engstrom, P. (2002). Conifer reproductive development involves B-type MADS-box genes with distinct and different activities in male organ primordia. *Plant J.* 31, 161–169. doi: 10.1046/j.1365-313X.2002.01343.x
- Svergun, D., Barberato, C., and Koch, M. H. J. (1995). CRYSOLO – A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* 28, 768–773. doi: 10.1107/S002188995007047
- Tandre, K., Albert, V. A., Sundas, A., and Engstrom, P. (1995). Conifer homologues to genes that control floral development in angiosperms. *Plant Mol. Biol.* 27, 69–78. doi: 10.1007/BF00019179
- The PyMOL Molecular Graphics System (2010). *The PyMOL Molecular Graphics System*. New York, NY: Schrödinger, LLC.
- Theissen, G. (2000). Evolutionary developmental genetics of floral symmetry: the revealing power of Linnaeus' monstrous flower. *Bioessays* 22, 209–213. doi: 10.1002/(SICI)1521-1878(200003)22:3<209::AID-BIES1>3.3.CO;2-A
- Theissen, G. (2001). Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* 4, 75–85. doi: 10.1016/S1369-5266(00)00139-4
- Theissen, G. (2005). Birth, life and death of developmental control genes: new challenges for the homology concept. *Theory Biosci.* 124, 199–212. doi: 10.1007/BF02814484
- Theissen, G., and Melzer, R. (2007). Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann. Bot.* 100, 603–619. doi: 10.1093/aob/mcm143
- Theissen, G., and Saedler, H. (1995). MADS-box genes in plant ontogeny and phylogeny: Haeckel's 'biogenetic law' revisited. *Curr. Opin. Genet. Dev.* 5, 628–639. doi: 10.1016/0959-437X(95)80032-8
- Theissen, G., and Saedler, H. (2001). Plant biology. Floral quartets. *Nature* 409, 469–471. doi: 10.1038/35054172
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Vazquez-Lobo, A., Carlsbecker, A., Vergara-Silva, F., Alvarez-Buylla, E. R., Pinero, D., and Engstrom, P. (2007). Characterization of the expression patterns of LEAFY/FLORICAULA and NEEDLY orthologs in female and male cones of the conifer genera *Picea*, *Podocarpus*, and *Taxus*: implications for current evo-devo hypotheses for gymnosperms. *Evol. Dev.* 9, 446–459. doi: 10.1111/j.1525-142X.2007.00182.x
- Wagner, D., Sablowski, R. W., and Meyerowitz, E. M. (1999). Transcriptional activation of APETALA1 by LEAFY. *Science* 285, 582–584. doi: 10.1126/science.285.5427.582
- Wang, Y. Q., Melzer, R., and Theissen, G. (2010). Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm *Gnetum gnemon* provide a clue to the evolutionary origin of 'floral quartets'. *Plant J.* 64, 177–190. doi: 10.1111/j.1365-313X.2010.04325.x
- Weigel, D., and Meyerowitz, E. M. (1993). Activation of floral homeotic genes in *Arabidopsis*. *Science* 261, 1723–1726. doi: 10.1126/science.261.5129.1723
- Wilman, H. R., Shi, J., and Deane, C. M. (2014). Helix kinks are equally prevalent in soluble and membrane proteins. *Proteins* 82, 1960–1970. doi: 10.1002/prot.24550
- Winter, C. M., Austin, R. S., Blanvillain-Baufume, S., Reback, M. A., Monniaux, M., Wu, M. F., et al. (2011). LEAFY target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. *Dev. Cell* 20, 430–443. doi: 10.1016/j.devcel.2011.03.019
- Winter, K. U., Saedler, H., and 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*. *Plant J.* 31, 457–475. doi: 10.1046/j.1365-313X.2002.01375.x
- Wu, M. F., Sang, Y., Bezhani, S., Yamaguchi, N., Han, S. K., Li, Z., et al. (2012). SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the LEAFY and SEPALLATA3 transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3576–3581. doi: 10.1073/pnas.1113409109
- Yang, Y., Fanning, L., and Jack, T. (2003). The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. *Plant J.* 33, 47–59. doi: 10.1046/j.0960-7412.2003.01473.x

- Zahn, L. M., Kong, H., Leebens-Mack, J. H., Kim, S., Soltis, P. S., Landherr, L. L., et al. (2005a). The evolution of the SEPALLATA subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* 169, 2209–2223. doi: 10.1534/genetics.104.037770
- Zahn, L. M., Leebens-Mack, J., DePamphilis, C. W., Ma, H., and Theissen, G. (2005b). To B or Not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. *J. Hered.* 96, 225–240. doi: 10.1093/jhered/esi033
- Zhang, P., Tan, H. T., Pwee, K. H., and Kumar, P. P. (2004). Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. *Plant J.* 37, 566–577. doi: 10.1046/j.1365-313X.2003.01983.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The Guest Associate Editor Rainer Melzer declares that, despite having collaborated with the author Chloe Zubieta, the review process was handled objectively.

Copyright © 2016 Silva, Puranik, Round, Brennich, Jourdain, Parcy, Hugouvieux and Zubieta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Ferns: the missing link in shoot evolution and development

Andrew R. G. Plackett^{1*}, Verónica S. Di Stilio² and Jane A. Langdale¹

¹ Department of Plant Sciences, University of Oxford, Oxford, UK, ² Department of Biology, University of Washington, Seattle, WA, USA

OPEN ACCESS

Edited by:

Catherine Anne Kidner,
University of Edinburgh, UK

Reviewed by:

John Bowman,
Monash University, Australia
Neelima Roy Sinha,
University of California, Davis, USA
Jo Ann Banks,
Purdue University, USA

*Correspondence:

Andrew R. G. Plackett
andrew.plackett@plants.ox.ac.uk

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 29 August 2015

Accepted: 23 October 2015

Published: 06 November 2015

Citation:

Plackett ARG, Di Stilio VS
and Langdale JA (2015) Ferns:
the missing link in shoot evolution
and development.
Front. Plant Sci. 6:972.
doi: 10.3389/fpls.2015.00972

Shoot development in land plants is a remarkably complex process that gives rise to an extreme diversity of forms. Our current understanding of shoot developmental mechanisms comes almost entirely from studies of angiosperms (flowering plants), the most recently diverged plant lineage. Shoot development in angiosperms is based around a layered multicellular apical meristem that produces lateral organs and/or secondary meristems from populations of founder cells at its periphery. In contrast, non-seed plant shoots develop from either single apical initials or from a small population of morphologically distinct apical cells. Although developmental and molecular information is becoming available for non-flowering plants, such as the model moss *Physcomitrella patens*, making valid comparisons between highly divergent lineages is extremely challenging. As sister group to the seed plants, the monilophytes (ferns and relatives) represent an excellent phylogenetic midpoint of comparison for unlocking the evolution of shoot developmental mechanisms, and recent technical advances have finally made transgenic analysis possible in the emerging model fern *Ceratopteris richardii*. This review compares and contrasts our current understanding of shoot development in different land plant lineages with the aim of highlighting the potential role that the fern *C. richardii* could play in shedding light on the evolution of underlying genetic regulatory mechanisms.

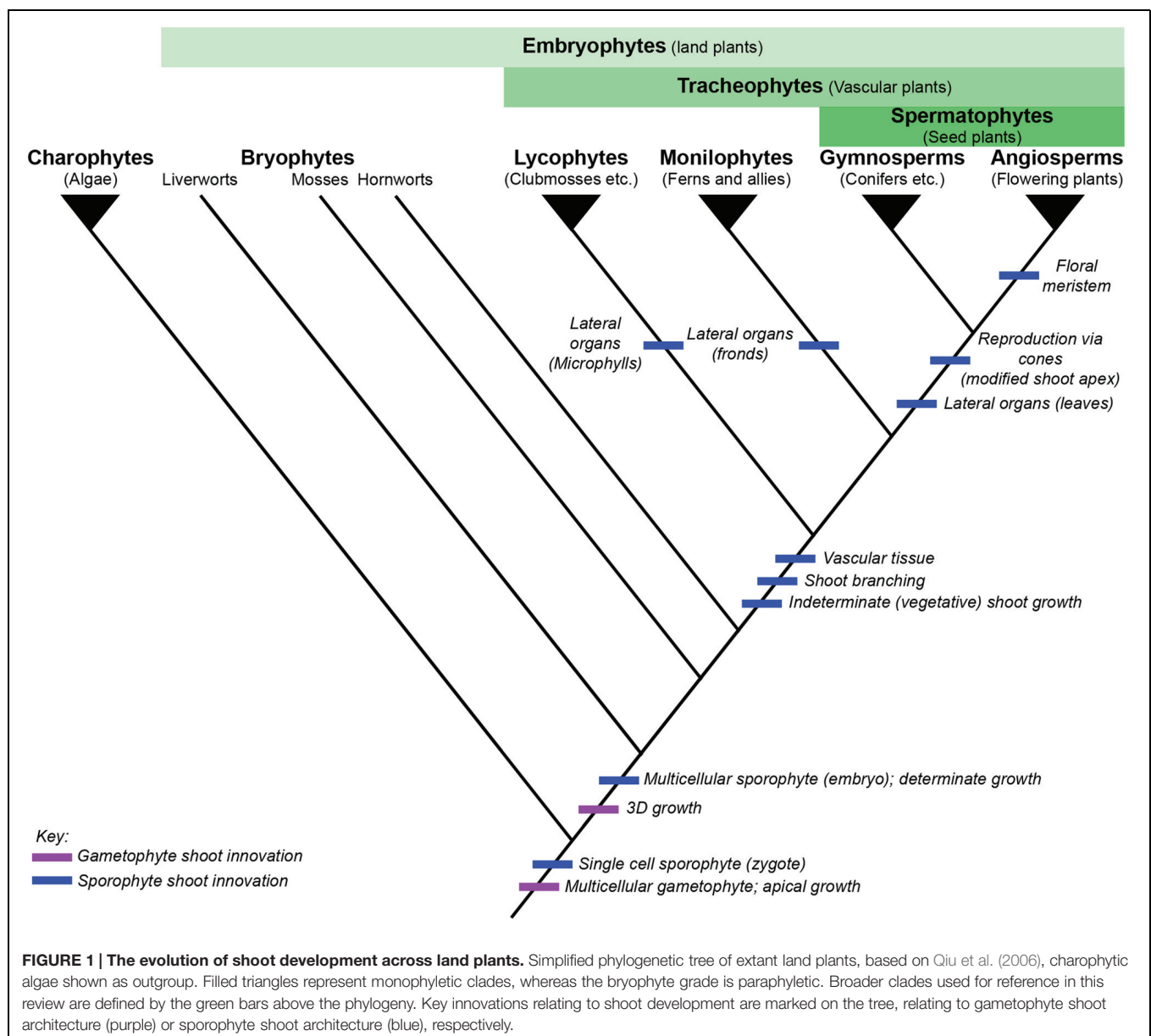
Keywords: plant, evolution, development, shoot, monilophyte, fern, *Ceratopteris*

INTRODUCTION

Land plants (embryophytes) evolved from aquatic green algae ~470 million years ago, with phylogenetic analyses consistently positioning charophytic (streptophyte) algae as the closest extant sister group (Karol et al., 2001; Lewis and McCourt, 2004; Wodniok et al., 2011; Ruhfel et al., 2014). Whilst charophytes exhibit a range of vegetative body plans in the haploid (gametophyte) generation of the lifecycle (reviewed in Niklas and Kutschera, 2010), the diploid (sporophyte) generation of the lifecycle is unicellular; the single-celled product of gamete fusion (zygote) directly undergoes meiosis. By contrast, in all land plants the zygote undergoes intervening mitotic divisions to create a multicellular sporophyte (the embryo), the uppermost part of which has become specialized into a photosynthetic shoot. Although a multicellular sporophyte is the defining characteristic of land plants, the structure has undergone enormous diversification and elaboration during evolution, from simple and transient (as in the stalked sporangium in most bryophytes) to highly complex and long-lived (as in the tree-forms of various vascular plants). In all cases, however, meiosis ultimately generates haploid gametophytes to complete the lifecycle.

Successive land plant lineages have innovated new sporophytic shoot structures, leading to increasing morphological and physiological complexity (**Figure 1**). Understanding the genetic mechanisms underlying the origins and continued modification of the land plant shoot is one of the primary aims of research into plant evolution and development (evo-devo). Although the characterization of evolutionary trajectories is not always straightforward, because many lineages that contained informative intermediate characters are now extinct, reconstruction is possible through comparison of extant species to infer plesiomorphies (ancestral traits) and apomorphies (derived traits). Our understanding of how land plant morphologies evolved is based mostly on comparative developmental studies between representative model species, predominantly the flowering plants *Arabidopsis thaliana* and

Oryza sativa and the moss *Physcomitrella patens*. Further models are increasingly being exploited as experimental systems for molecular analyses, including the liverwort *Marchantia polymorpha* and lycopods in the genus *Selaginella* (the genome of *Selaginella moellendorffii* has been sequenced (Banks et al., 2011) whilst the bulk of developmental data comes from *S. kraussiana*). A substantial amount of detailed developmental data has been accumulated in these and other non-seed plant lineages (reviews include White and Turner, 1995; Banks, 2009; Renzaglia et al., 2009; Ligrone et al., 2012; Vasco et al., 2013, and references therein). Key developmental characteristics relating to shoot development are summarized and compared between the different models discussed in this review in **Figure 2**. However, at present the bulk of gene function data available outside of the angiosperms is from the moss *P. patens*. The large evolutionary



					Spermatophytes (Seed plants)		
					Tracheophytes (Vascular plants)		
Embryophytes (land plants)							
Lineage:		Liverworts	Mosses	Hornworts	Lycophytes	Monilophytes	Angiosperms
Model species:		<i>Marchantia polymorpha</i>	<i>Physcomitrella patens</i>	<i>Anthoceros agrestis</i>	<i>Selaginella kraussiana</i>	<i>Ceratopteris richardii</i>	<i>Arabidopsis thaliana</i>
Gametophyte	Growth Habit	Thalloid ^{1,2,3}	Leafy shoot ^{3,5}	Thalloid ^{3,7}	Endosporic; ^{10,11} Thalloid (female) Determinate (male)	Thalloid ¹⁵	Determinate ^{20,21}
	Shoot Apex	Single AC ^{1,2,3}	Single AC ⁵	Single AC ^{3,7}	None (female) ¹⁰ None (male) ¹¹	Single AC; ¹⁵ Notch meristem (Hermaphrodite)	None ^{20,21}
	Lateral Organs	None ^{1,2}	Phyllidia ⁵	None ⁷	None ^{10,11}	None ¹⁵	None ^{20,21}
	Phyllotaxy	None ^{1,2}	Spiral ⁵	None ⁷	None ^{10,11}	None ¹⁵	None ^{20,21}
	Organ Apex	None ^{1,2}	Single AC ⁵	None ⁷	None ^{10,11}	None ¹⁵	None ^{20,21}
	Shoot Branching	Apical bifurcation ^{1,2,3}	Axillary branches ⁶	Apical bifurcation ^{3,8}	None ^{10,11}	None ¹⁵	None ^{20,21}
Sporophyte	Growth Habit	Determinate ⁴	Determinate ⁴	Determinate ^{4,9}	Leafy shoot ¹²	Leafy shoot ¹⁶	Leafy shoot ²²
	Shoot Apex	None ⁴	Transient AC and ⁴ intercalary meristem	Transient ^{4,9} basal meristem	Paired ACs ¹³	Single AC ¹⁷	Multicellular ²² meristem
	Lateral Organs	None ⁴	None ⁴	None ^{4,9}	Microphylls ^{12,14}	Fronds ^{14,16,17}	Leaves ^{14,22}
	Phyllotaxy	None ⁴	None ⁴	None ^{4,9}	Spiral (juvenile); ¹³ four ranks (adult)	Spiral ¹⁸	Spiral ²²
	Organ Apex	None ⁴	None ⁴	None ^{4,9}	Paired ACs ¹³	Single AC (frond); ¹⁷ Paired ACs (pinna) ¹⁹	Multicellular ²² primordium
	Shoot Branching	None ⁴	None ⁴	None ^{4,9}	Apical bifurcation; ¹³ Angle meristems ¹²	None ¹⁵	Axillary branches ²³
References:		¹ Leitgeb, 1881 ² Kny, 1890 ³ Reviewed in Ligrone <i>et al.</i> , 2012 ⁴ Reviewed in Kato and Akiyama, 2005 ⁵ Harrison <i>et al.</i> , 2009 ⁶ Coudert <i>et al.</i> , 2015	³ Reviewed in Ligrone <i>et al.</i> , 2012 ⁴ Reviewed in Kato and Akiyama, 2005 ⁵ Harrison <i>et al.</i> , 2009 ⁶ Coudert <i>et al.</i> , 2015	³ Reviewed in Ligrone <i>et al.</i> , 2012 ⁴ Reviewed in Kato and Akiyama, 2005 ⁷ Renzaglia, 1978 ⁸ Schuster, 1984b ⁹ Bartlett, 1928	¹⁰ Robert, 1971 ¹¹ Robert, 1973 ¹² Reviewed in Banks, 2009 ¹³ Harrison <i>et al.</i> , 2007 ¹⁴ Reviewed in Tomescu, 2009	¹⁴ Reviewed in Tomescu, 2009 ¹⁵ Reviewed in Banks, 1999 ¹⁶ Reviewed in Hickok <i>et al.</i> , 1995 ¹⁷ Hou and Hill, 2002 ¹⁸ Reviewed in White and Turner, 1995 ¹⁹ Hill, 2001	¹⁴ Reviewed in Tomescu, 2009 ²⁰ Reviewed in McCormick, 2004 ²¹ Reviewed in Yadegari and Drews, 2004 ²² Reviewed in Gaillochot and Lohmann, 2015 ²³ Reviewed in Janssen <i>et al.</i> , 2014

FIGURE 2 | Comparing shoot development across model land plants. Comparison of shoot characteristics found in the gametophyte (pink) and sporophyte (blue) generations of select developmental/genetic model plant species, including examples from most extant land plant (embryophyte) lineages. Broader clade denominations between these lineages are indicated by green bars above (see **Figure 1**). It should be noted that not all model species are entirely representative of development within each lineage, in particular the mosses and liverworts. Examples from other species and fossil data are included in this review where necessary to provide a more accurate representation of evolutionary trajectories.

FIGURE 2 | Comparing shoot development across model land plants. Comparison of shoot characteristics found in the gametophyte (pink) and sporophyte (blue) generations of select developmental/genetic model plant species, including examples from most extant land plant (embryophyte) lineages. Broader clade denominations between these lineages are indicated by green bars above (see **Figure 1**). It should be noted that not all model species are entirely representative of development within each lineage, in particular the mosses and liverworts. Examples from other species and fossil data are included in this review where necessary to provide a more accurate representation of evolutionary trajectories.

distance between these two ends of the land plant phylogeny has made interpretation of such comparisons extremely challenging and, in some cases, of little use. It should be noted that model species are not always wholly representative of their extant relatives, nor necessarily the ancestral state of that particular lineage. This is particularly problematic in bryophyte lineages such as the mosses, where the fossil evidence and diversity in extant species highlight the potential for confusion over the ancestral state (see Shoot Branching section). Although in this review we necessarily focus on the combined genetic and developmental data available from model species, where required

we highlight fossil data or examples from non-model species to better represent evolutionary trajectories or the ambiguity currently surrounding them.

One group of plants that is notably absent in most comparative studies is the monilophytes (ferns and their relatives). Monilophytes are the most closely related extant land plant lineage to seed plants (**Figure 1**; Pryer et al., 2001). As such, monilophytes are a highly informative phylogenetic node, both as outgroup to the seed plants and as an intermediate lineage to provide resolution for functional comparisons between homologous genes in bryophytes and angiosperms. That said,

the monilophytes themselves represent an ancient and highly diverse lineage, diverging from the seed plants 400 million years ago (Pryer et al., 2001) and encompassing a wide variety of growth habits including tree forms, aquatics and epiphytes (reviewed in Schuettpelz and Pryer, 2008; Watkins and Cardelús, 2012). The largest clade of ferns, the leptosporangiate ferns, account for approximately 80% of non-flowering vascular plant species (Schuettpelz and Pryer, 2009). A number of developmental innovations have occurred independently within the monilophyte lineage, including the evolution of lateral organs (fronds) and heterospory (**Figure 1**). However, to date our understanding of fern developmental genetics has been impeded by serious technical barriers that are only now being overcome. These barriers include typically very large genomes (Barker and Wolf, 2010; Bainard et al., 2011), an obstacle further complicated by frequent polyploidy (Wood et al., 2009), and a lack of any genetic transformation system. Two fern species are now coming to prominence as research vehicles: *Ceratopteris richardii*, a homosporous fern long-established in laboratories for developmental studies and teaching (Hickok et al., 1995); and the heterosporous aquatic fern *Azolla filiculoides*, a species potentially well-suited for industrial biomass production (Brouwer et al., 2014). Efforts are currently underway to sequence the genomes of both species (Sessa et al., 2014), and a wealth of transcriptome data is being generated in diverse fern species via the 1 KP project (Wickett et al., 2014). In addition, a number of stable genetic transformation methods have recently been reported, including methods that are suitable for *C. richardii* (Muthukumar et al., 2013; Plackett et al., 2014; Bui et al., 2015). In light of these advances, the study of ferns to aid our understanding of shoot evolution is being viewed with increasing enthusiasm (Bennett, 2014; Banks, 2015; Harrison, 2015). A review of our current understanding of the genetic regulation of shoot development across the land plants, including what little is already known about monilophytes, is thus timely and presents an opportunity to outline the key developmental questions that need to be answered.

THE EVOLUTION OF LAND PLANT SHOOTS

The alternation of multicellular haploid gametophyte and diploid sporophyte generations is a shared feature of all land plant lifecycles. However, the relative dominance of each generation changed as new land plant lineages evolved. In bryophytes (liverworts, mosses, and hornworts) the dominant generation of the lifecycle is the gametophyte. For example, the haploid spores of *P. patens* germinate to form filamentous gametophytes that transition into shoot-like structures (gametophores; **Figure 3A**) that produce leaf-like organs (phyllidia) and ultimately male and female gametangia (gamete-producing structures; reviewed in Kofuji and Hasebe, 2014). Upon fertilization, the diploid zygote undergoes a strictly determinate developmental program to become an unbranched sporophyte axis terminating in a single sporangium. Within vascular plants (tracheophytes) the role of the sporophyte generation increased at the

expense of the gametophyte, which fossil evidence suggests occurred at the base of the clade (reviewed in Gerrienne and Genez, 2011). Indeterminate branched sporophytes are found in all tracheophyte lineages, for example the lycophyte *S. kraussiana* (**Figure 3B**), whilst the *S. kraussiana* female and male gametophytes, respectively, produce a thallus inside the megaspore or directly generate gametangia upon spore germination (Robert, 1971, 1973). Similar development has been recorded in the related species *S. apoda* (Schulz et al., 2010). Gametophyte development in monilophytes is also reduced compared to bryophytes. The *C. richardii* gametophytes develop as a single cell-layered thallus comprising a few specialized cell types (reviewed in Banks, 1999). The subsequent sporophyte develops as an indeterminate shooting structure, producing fronds sequentially from a persistent post-embryonic shoot apex (**Figures 3C–E**; Johnson and Renzaglia, 2008). In angiosperms the sporophyte develops a highly complex, indeterminate body-plan from multiple post-embryonic shoot apical and axillary meristems (**Figure 3F**; Gifford and Foster, 1989), whereas the male and female gametophytes comprise just a few cells each (reviewed in McCormick, 2004; Yadegari and Drews, 2004). In bryophytes and angiosperms the sporophyte and gametophyte, respectively, are fully dependent on the dominant stage of the lifecycle for nutrition (matrotrophic), whereas in both lycophytes and ferns the gametophyte develops independently of the sporophyte and, beyond a transient period where the sporophyte embryo develops upon the gametophyte, the sporophyte is not nutritionally dependent upon the gametophyte (reviewed in Qiu et al., 2012).

From a developmental perspective, canonical plant shoots (as generally recognized in vascular plants) can be defined as a process, i.e., they develop iteratively from an apex to produce lateral organs. Using this definition, the gametophores of extant mosses and 'leafy' liverworts can also be classified as shoots, possessing an axial body-plan. In contrast other liverwort species (including *M. polymorpha*) and all hornworts develop a thalloid body-plan (comprising multiple cell layers), which possesses apical growth in common with shoots but lacks defined lateral organs (**Figure 2**; reviewed in Renzaglia et al., 2009; Ligrone et al., 2012; Villarreal and Renzaglia, 2015). Although presumably arising from a common origin, the precise evolutionary relationship between the axial and thalloid body-plans in early diverging embryophytes is not yet fully resolved (reviewed in Qiu et al., 2012), and so at present it is not possible to assess character polarity. It is, however, quite probable that shared developmental characters are underpinned by conserved genetic mechanisms (see examples given in the review below).

A second important component to this definition of the shoot is the concept of indeterminate growth. The *P. patens* sporophyte demonstrates apical growth but only transiently, terminating after just a few cell divisions in a sporangium (reviewed in Kato and Akiyama, 2005). Recent transcriptome data from developing liverwort and moss sporophytes indicates expression of meiosis-associated genes even during embryonic stages (Frank and Scanlon, 2015), further suggesting that these sporophytes lack indeterminacy and thus true shoot function.

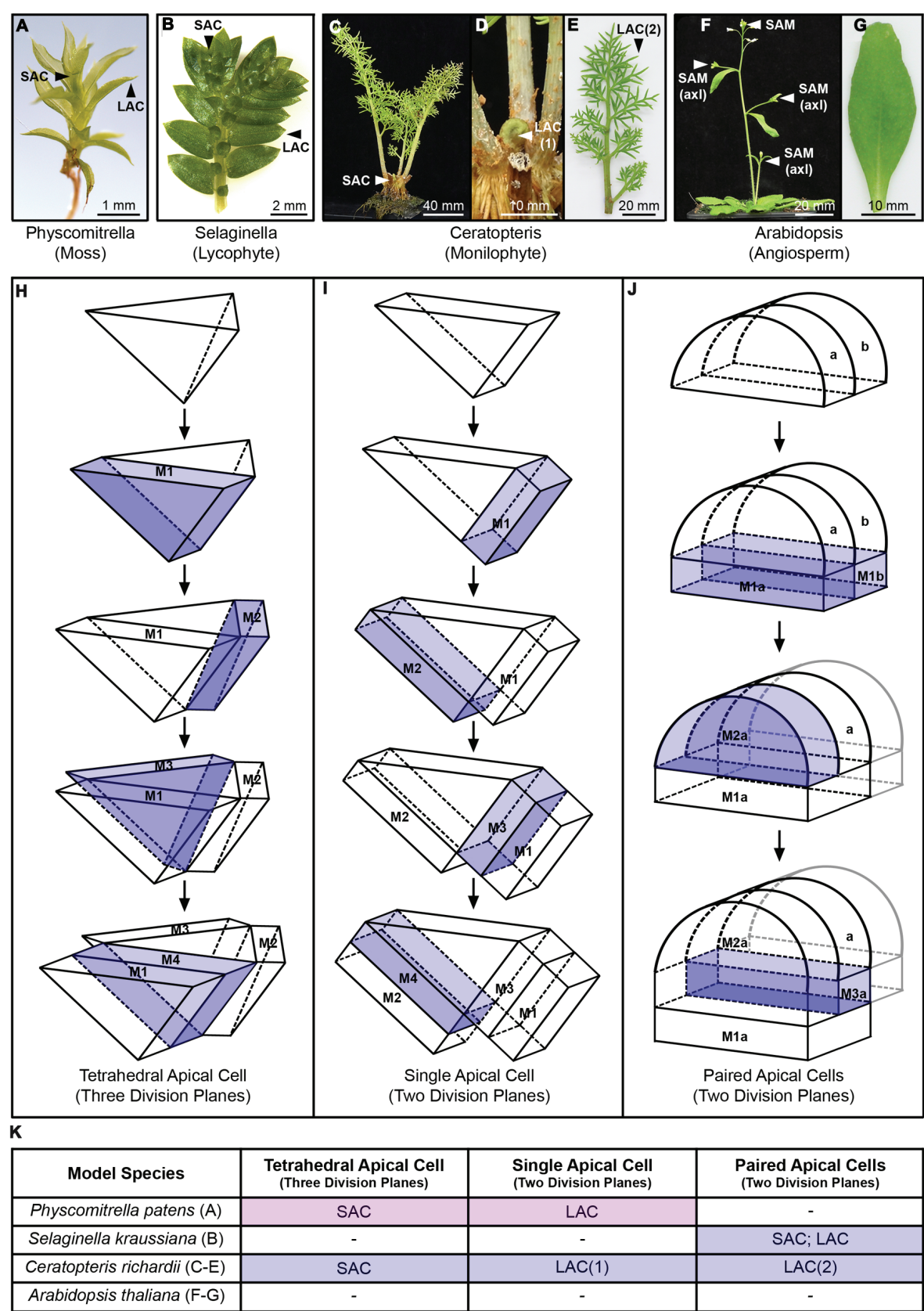


FIGURE 3 | Continued

Shoot apical activity across representative model land plants. (A–G) Shoots and lateral organs of representative model land plant species; **(A)** *Physcomitrella patens* gametophore; **(B)** *Selaginella kraussiana* shoot and microphylls, showing unequal apical branching; **(C–E)** *Ceratopteris richardii* sporophyte **(C)**, showing emergence of new frond from the shoot apex **(D)** and a fully developed reproductive frond with lateral pinnae **(E)**; **(F,G)** *Arabidopsis thaliana* sporophyte with axillary branches emerging **(F)**, showing a rosette leaf **(G)**. The position of shoot apical cells (SAC) and lateral apical cells (LAC) are marked. *A. thaliana* develops from a multicellular shoot apical meristem (SAM), and axillary branches develop from the activity of similar, axillary meristems [SAM(axl)]. *A. thaliana* leaves develop from a multicellular primordium and lack an apical cell (AC) or meristem. **(H,I)** Diagrammatic summary of different AC geometries and division patterns. **(H)** tetrahedral AC with three cutting faces; **(I)** single AC with two cutting faces; **(J)** adjacent paired ACS with two cutting faces each. Daughter cells (merophytes), generated through asymmetric divisions that reconstitute the AC, are marked M, and numbered in order of their production. For clarity, the most recently formed merophyte is highlighted in blue. In the case of paired ACs, these and their descendants are distinguished by 'a' and 'b' accordingly. In the interests of clarity, beyond the first division only the further divisions of AC 'a' are shown: these are mirrored by the activity of AC 'b'. The complex multicellular SAM of *A. thaliana* is not shown. **(K)** Table summarizing the developmental contexts in which the different shoot ACs in **(H–J)** are found, referring to the labels marked in **(A–G)** and distinguishing whether they occur in the gametophyte (pink) or sporophyte (blue) generation.

Similarly, the thalloid gametophytes of the ferns *C. richardii* and *Lygodium japonicum* initially grow from transient apical cells (ACs) that then terminate (Banks, 1999; Takahashi et al., 2015). Interestingly, in both species, growth of the chordate hermaphrodite thallus continues through proliferation of a second, distinct multicellular meristematic region (the 'notch meristem'), iteratively generating archegonia until successful fertilization has occurred (Banks, 1999; Takahashi et al., 2015). Development of the strap-like thalloid gametophyte of the epiphytic fern *Colysis decurrens* follows the same principles, with a single transient early AC followed by an indeterminate multicellular marginal meristem (Takahashi et al., 2009). The parallels with canonical shoot development are striking, but whether the notch/marginal meristem represents the reduction of an ancestral gametophytic shoot has yet to be determined.

The origins and early evolutionary trajectory of the vascular plant shoot are much debated, with several competing theories presented (reviewed in Tomescu et al., 2014), but there is general agreement on the key developmental innovations that occurred during shoot evolution: indeterminate apical activity, organogenesis, shoot branching, and developmental phase-change.

SHOOT INDETERMINACY- APICAL CELLS VERSUS MULTICELLULAR MERISTEMS

Cells with shoot apical function (i.e., having indeterminate cell fate) are present in all extant land plant lineages (Figure 2; Steeves and Sussex, 1989). In seed plants (gymnosperms and angiosperms) these are part of a highly organized, multicellular shoot apical meristem (SAM), whereas in non-seed plants (bryophytes, lycophytes, and monilophytes) they exist as a distinct single AC, or small cluster thereof (Figures 2 and 3). Although they vary in size, shape, and number of cutting planes, ACs can be defined as dividing asymmetrically to produce derivatives and replenish themselves.

The *P. patens* gametophore possesses a single persistent tetrahedral (pyramid-shaped) AC that cleaves sequentially in three planes to generate determinate leaf-like organs (Figure 3H; Harrison et al., 2009). Tetrahedral ACs are also found in the

gametophytic shoots produced by 'leafy' liverworts (Crandall-Stotler, 1980) but these undergo a different pattern of asymmetric cell division that could indicate convergent evolutionary origins, a suggestion supported by differing formative division planes during lateral organ formation (Crandall-Stotler, 1986). In contrast, the single ACs of thalloid bryophyte gametophytes display a different geometry, at maturity cleaving across four faces in both thalloid liverworts (Leitgeb, 1881; Kny, 1890) and hornworts (recently reviewed in Renzaglia, 1978). Extant sporophytes in all three bryophyte lineages exhibit entirely determinate development, developing from temporary ACs and/or intercalary basal meristems (reviewed in Bartlett, 1928; Crandall-Stotler, 1980; Kato and Akiyama, 2005). It is therefore possible that persistent ACs first arose in the gametophyte stage of the land plant lifecycle, and became incorporated into sporophyte development.

Whether a single AC truly represents the plesiomorphic state of the tracheophyte shoot apex is still debated (Banks, 2015; Harrison, 2015); a number of lycophyte and fern species possess multiple ACs at their apex (reviewed in White and Turner, 1995), whereas others such as the ferns *Nephrolepis exaltata* (Sanders et al., 2011) and *C. richardii* (Hou and Hill, 2002) develop from a single tetrahedral AC (Figure 3H). Evidence from histology and clonal analysis suggests that *S. kraussiana* shoots develop from two adjacent ACs (Figure 3J; Harrison et al., 2007; Harrison and Langdale, 2010), although single ACs can be observed in the early stages of minor branch formation (Harrison et al., 2007) and other authors have suggested that this condition persists (Jones and Drinnan, 2009). The multicellular SAM in seed plant shoots is at least superficially more complex in structure than these examples. The SAM comprises discrete functional zones, namely a central multicellular zone of pluripotent cells and a surrounding peripheral zone of cells from which lateral organ primordia are specified; both zones overlap distinct tissue layers derived from separate cell lineages (reviewed in Gaillochet and Lohmann, 2015). It has recently been proposed that lycophyte and monilophyte ACs, subtended by a transcriptionally distinct and rapidly proliferating 'core domain' of daughter cells, might be functionally equivalent to the central zone of the SAM (Frank et al., 2015). Laser capture microdissection (LCM)-RNAseq comparison between apices from *S. moellendorffii*, the monilophyte *Equisetum arvense* and the angiosperm *Zea*

mays (maize) found disparate expression profiles between the lycophyte and monilophyte ACs, but the core domain of both species expressed numerous genes associated with developmental regulation in the maize SAM (Frank et al., 2015). As such, envisaging the AC alone as functionally equivalent to a SAM may be too simplistic.

Sufficient data is now available to examine how homologs of genes with important functions in the *A. thaliana* SAM function in other land plant groups. A number of distinct modules are crucial to maintaining SAM identity and indeterminacy (summarized in Figure 4). The CLAVATA/WUSCHEL (CLV/WUS) pathway regulates the size of the apical initial domain within the multicellular SAM (Schoof et al., 2000; Bäurle and Laux, 2005), and the Class I KNOTTED1-like HOMEODOMAIN/ASYMMETRIC LEAVES, ROUGH SHEATH, PHANTASTICA (KNOX/ARP) pathway regulates indeterminate cell fate versus specification of the determinate leaf development program (Schneeberger et al., 1998; Timmermans et al., 1999; Tsiantis et al., 1999; Byrne et al., 2000; Guo et al., 2008; reviewed in Gaillochet and Lohmann, 2015). In some angiosperms Class 1 KNOX expression is later reactivated in established leaf primordia to generate compound leaves (see Phase Change section). Both the CLV/WUS and KNOX/ARP pathways require intercellular communication, which is mediated at least in part by movement of the component proteins between cells (Lucas et al., 1995; Lenhard and Laux, 2003; Yadav et al., 2011). A third family of transcription factors, Class III homeodomain-leucine zipper (HD-Zip), is also required for SAM formation and maintenance, the function of which is antagonized by the KANADI (KAN) genes (Emery et al., 2003; reviewed in Floyd and Bowman, 2007).

Although a WUSCHEL-related HOMEODOMAIN (WOX) gene is preferentially expressed in the *P. patens* gametophyte AC (Frank and Scanlon, 2015), the CLV1 and CLV2 gene families are absent from the *P. patens* genome (Banks et al., 2011), precluding the existence of the WUS-CLV regulatory module. In contrast, Class III HD-Zip and KAN homologs have been identified in *P. patens* (Sakakibara et al., 2001; Floyd and Bowman, 2007; Banks et al., 2011) but none are enriched in the AC (Frank and Scanlon, 2015). The expression of Class I KNOX genes has been reported in the *P. patens* gametophore AC (Frank and Scanlon, 2015) but no loss-of-function mutant phenotypes have been detected in the gametophyte (Singer and Ashton, 2007; Sakakibara et al., 2008), and ARP genes are absent from the *P. patens* genome (Banks et al., 2011). As such, there is currently no molecular evidence to support the suggestion that these components of the SAM regulatory network were established in gametophyte shoots.

The ancestral role for KNOX proteins is sporophytic, as inferred from studies in chlorophyte algae where heterodimerization of a KNOX and a BELLRINGER protein facilitates zygote formation (Lee et al., 2008). In angiosperms, Class 1 KNOX expression is essential for SAM maintenance (Long et al., 1996; Vollbrecht et al., 2000; Belles-Boix et al., 2006). In the *P. patens* sporophyte, Class 1 KNOX genes are expressed transiently in the AC during phases of cell proliferation. However, loss-of-function mutants demonstrated that although Class I KNOX genes promote sporophytic cell divisions and regulate

their orientation, they are not essential for apical activity *per se* (Sakakibara et al., 2008). Within the tracheophytes, Class 1 KNOX expression has been detected in the shoot apices of both *S. kraussiana* and *S. moellendorffii* with expression localized to either the AC (Frank et al., 2015) or cells immediately subtending it (Harrison et al., 2005). In both cases, transcripts were absent from newly developing organ primordia, a pattern similar to that seen in the angiosperm SAM (e.g., Jackson et al., 1994). Class 1 KNOX activity in angiosperm leaf primordia is repressed by ARP gene function, as a consequence of which the two components display mutually exclusive expression patterns (Timmermans et al., 1999; Tsiantis et al., 1999; Byrne et al., 2000; Guo et al., 2008). Class 1 KNOX expression is seen in both the shoot apex and young organ primordia of the ferns *Osmunda regalis* (Harrison et al., 2005), *Annogramma chaeophylla* (Bharathan et al., 2002) and *C. richardii* (Sano et al., 2005), although transcripts are absent from older primordia. ARP homologs are expressed in the organ primordia of both *S. kraussiana* and *O. regalis* (older primordia only), but they are also co-expressed with Class 1 KNOX at the shoot apex (Harrison et al., 2005). The ancestral function of Class 1 KNOX appears to relate to cell division in the land plant sporophyte, but in the absence of mutant phenotypes in lycophytes or monilophytes it is impossible to say at what stage it became essential for AC/meristem maintenance. Based on observed expression patterns, the evolution of the mutually exclusive KNOX/ARP expression pattern may have occurred coincident with the formation of the SAM in seed plants. Overexpression and complementation studies in *A. thaliana* suggest that Class 1 KNOX and ARP homologs from lycophytes and monilophytes, respectively, can provide some of the same functions as endogenous *A. thaliana* genes (Harrison et al., 2005; Sano et al., 2005), but these experiments are not informative about what these genes do in their native context. Currently there is no transgenic system available in a lycophyte species, but future functional studies in *C. richardii* should begin to resolve some of these functional questions.

At present, much less is known about the function and expression of other SAM gene homologs within non-seed vascular plants. Although one WOX gene is detected at the *S. moellendorffii* shoot apex, it is not expressed in the AC, instead transcripts accumulate within the core domain and in developing primordia (Frank et al., 2015). There may be greater conservation of HD-Zip function between lycophytes and angiosperms: a Class III HD-Zip homolog is strongly expressed in the *S. kraussiana* shoot AC (Floyd et al., 2006) and KAN expression is up-regulated in the core domain beneath them (Frank et al., 2015). In the *C. richardii* sporophyte, expression of Class III HD-Zip homologs has been detected (Aso et al., 1999; Floyd et al., 2006) but spatial expression data is not yet available. Thus, although known regulators may have a role in apical development within vascular plants, whether their specific functions are conserved or differ remains to be established.

Phytohormones are another important regulatory force within the angiosperm SAM, acting to integrate other developmental signals. Cytokinin (CK) maintains the indeterminate central

	Mosses		Lycophytes		Monilophytes		Angiosperms	
	<i>Physcomitrella patens</i>		<i>Selaginella</i> spp.		<i>Ceratopteris richardii</i>		<i>Arabidopsis thaliana</i>	
Gene Family	Expression Pattern	Gene Function	Expression Pattern	Gene Function	Expression Pattern	Gene Function	Expression Pattern	Gene Function
<i>WUSCHEL</i> -related <i>Homeobox</i> (<i>WOX</i>)	Gametophore (AC)	Unknown	Unknown	Unknown	Unknown	Unknown	SAM (Central zone)	SAM identity; indeterminate cell fate
<i>CLAVATA</i> (<i>CLV</i>)	Not present	-	Unknown	Unknown	Unknown	Unknown	SAM (Peripheral zone)	Antagonises <i>WUSCHEL</i> ; restricts central zone
Class I <i>KNOX</i>	Gametophore (AC); zygote; sporophyte (AC and intercalary meristem)	Sporophyte cell division; no gametophyte phenotype	Shoot apex (core domain); internodes (<i>S. kraussiana</i>)	Unknown	Shoot apex (AC); frond primordium; vasculature; gametophyte	Unknown	SAM	SAM identity
<i>ASYMMETRIC LEAVES1</i> , <i>ROUGH SHEATH</i> , <i>PHANTASTICA</i> (<i>ARP</i>)	Not present	-	Shoot apex; microphyll primordia (<i>S. kraussiana</i>)	Unknown	Unknown	Unknown	Leaf primordium	Antagonises <i>KNOX</i> ; leaf primordia identity
<i>PINFORMED</i> auxin efflux carrier (<i>PIN</i>)	Gametophore shoot apex; sporophyte	Organogenesis; sporophyte development and branching	Shoot apex (core domain) (<i>S. moellendorffii</i>)	Microphyll AC boundary formation; rhizophore root fate specification (<i>S. kraussiana</i>)	Unknown	Unknown	SAM (Peripheral zone)	Organogenesis; leaf primordia outgrowth
Class III Homeodomain-Leucine Zipper (HD-Zip)	Detected in gametophyte and sporophyte (RT-PCR)	Unknown	Shoot apex (AC); microphyll primordium (adaxial); vasculature (<i>S. kraussiana</i>)	Unknown	Unknown	Unknown	SAM; leaf primordium (adaxial); vasculature	SAM initiation and maintenance; leaf adaxial identity; vasculature specification
<i>KANADI</i> (<i>KAN</i>)	Unknown	Unknown	Shoot apex (core domain) (<i>S. moellendorffii</i>)	Unknown	Unknown	Unknown	SAM; leaf primordia (abaxial); vasculature	Antagonises HD-Zip; leaf abaxial identity; vasculature specification
<i>LEAFY</i> (<i>LFY</i>)	Gametophore (shoot apex); archegonium; zygote; sporophyte	Zygotic cell division; no gametophyte phenotype	Unknown	Unknown	Detected in gametophyte and sporophyte (shoot apex and reproductive frond primordia) (RT-PCR)	Unknown	SAM; IM; FM	Phase change (vegetative to flowering); FM identity and early development
MADS-box (<i>MIKC</i> ^c)	Gametophyte all tissues (protonema, gametophore, gametangia); sporophyte all tissues	Gametophore gametangia and phyllidia development; sporophyte development	Unknown	Unknown	Detected in gametophyte (RT-PCR); sporophyte shoot AC; frond primordia; frond apex and vasculature; developing sporangia	Unknown	FM; floral organs	Floral whorl identity and floral organ development

FIGURE 4 | Conservation of the genetic regulators of shoot apical meristem (SAM) function across model land plant species. Summary table comparing known data about expression patterns and gene function of homologs of important regulators of the *A. thaliana* SAM across different land plant model species. Higher phylogenetic relationships between the model species are indicated by color coding (see **Figures 1 and 2**). In the case of *Selaginella*, genetic and developmental data come from both *S. kraussiana* and *S. moellendorffii*, as specified. Gene families highlighted in gray are absent from the genome of that particular species.

zone by promoting *WUS* expression in a complex multiple feedback loop (reviewed in Gaillochet and Lohmann, 2015), and CK biosynthesis is up-regulated by the Class 1 *KNOX* gene *SHOOTMERISTEMLESS* (*STM*; Jasinski et al., 2005; Yanai et al., 2005), thus linking Class 1 *KNOX* and *WUS* activity. *STM* also represses biosynthesis of gibberellin (GA) within the SAM (Jasinski et al., 2005), which otherwise promotes tissue growth and differentiation. Interestingly, a homolog of the GA response-repressing *DELLA* transcription factor is up-regulated in the *S. moellendorffii* shoot apex (Frank et al., 2015), suggesting a conserved requirement for GA suppression, although the same was not found in the monilophyte *E. arvense*. In *P. patens*, CK signaling-related transcripts are up-regulated in the gametophore AC (Frank and Scanlon, 2015), and exogenous CK promotes AC identity, causing increased branching and the development of ectopic meristematic cells in callus-like tissue (Coudert et al., 2015). Application of CK is also sufficient to induce callus tissue at the shoot apex of *C. richardii* sporophytes (Plackett et al., 2014). Collectively these observations point to an ancestral and conserved function for CK in regulating AC function and shoot development. Importantly, loss of Class 1 *KNOX* function does not perturb the expression of CK biosynthesis gene homologs in the *P. patens* sporophyte (Sakakibara et al., 2008), indicating that functional links between Class 1 *KNOX* and CK emerged in the tracheophyte lineage.

A second hormone, auxin, functions to promote pluripotency in the central zone of the angiosperm SAM by enhancing CK signaling (reviewed in Gaillochet and Lohmann, 2015), and both CK and auxin signaling are present in all land plant lineages (Wang et al., 2015). Disruption to polar auxin transport (*PAT*) in *S. kraussiana* causes the shoot apex to terminate, supporting a conserved function for auxin in indeterminate cell fate in lycophyte and angiosperm shoot apices (Sanders and Langdale, 2013). Notably, *in situ* analysis of a *PIN* auxin transporter in *S. moellendorffii* detected *PIN* expression surrounding the shoot AC, with a concomitant increase in expression of an *AUXIN RESPONSE FACTOR* (*ARF*) in the AC (Frank et al., 2015). This suggests the presence of an auxin maximum (peak in concentration) at the lycophyte AC. Presumably perturbed *PAT* therefore leads to a decrease in auxin levels in the AC, and hence to the observed termination (Sanders and Langdale, 2013). Recent analysis in *M. polymorpha* also found the greatest concentration of auxin in apical/meristematic regions (Eklund et al., 2015). Conversely, phenotypic analysis of *pin* mutants in *P. patens* identified a role for *PAT* in maintaining gametophore AC function by preventing auxin accumulation at the AC (Bennett et al., 2014). This apparent contradiction may relate to the inhibitory role of auxin in suppressing axillary branching in the moss gametophore, which is not found in lycophytes or monilophytes (see Shoot Branching section). Despite this difference, interactions between the auxin and CK signaling pathways are thought to promote AC fate in the *P. patens* gametophore (reviewed in Kofuji and Hasebe, 2014), as they do in angiosperm SAMs. As such, the auxin-CK signaling module likely became associated with AC function and shoot indeterminacy in the earliest diverging land plants.

ORGANOGENESIS AND LATERAL ORGAN DEVELOPMENT

It is often assumed that the morphology of the majority of extant bryophyte sporophytes is representative of ancestral sporophytes, being single axes with no lateral outgrowths (reviewed in Kato and Akiyama, 2005). It has been proposed that the transition from an unbranched shoot axis to a complex, indeterminate shoot branching system occurred in a stepwise fashion (reviewed in Tomescu et al., 2014). One of the most significant steps in this trajectory was the evolution of lateral organs, i.e., 'leaves.' Leafless fossils have been assigned to each of the lycophyte, monilophyte, and seed plant clades (Kenrick and Crane, 1997), and subsequent analysis of fossil characters against extant species strongly suggest that lateral organs evolved independently within the three tracheophyte lineages (Boyce and Knoll, 2002; Sanders et al., 2009; Tomescu, 2009). To avoid confusion, the term 'megaphylls' (describing both fern fronds and seed plant leaves) is not used in this review because of their probable independent origins. The term 'frond' is instead used to distinguish monilophyte lateral organs from the seed plant 'leaf.' Importantly for developmental studies, it has been proposed that lycophyte lateral organs ('microphylls') have an independent evolutionary origin to both fronds and leaves, arising as tissue outgrowths which later became vascularized (the enation theory; Bower, 1935) as opposed to being modified lateral branches of vascularized shoots (the telome theory; Zimmermann, 1952). Notably, a comparison of genetic mechanisms operating in these different lateral organs provided evidence for *KNOX/ARP* function in the formation of microphylls, monilophyte fronds and seed plant leaves (Harrison et al., 2005). This suggests that the same pathway was recruited to distinguish determinate lateral organs from indeterminate shoots during the evolution of both microphylls and megaphylls.

Lateral organs arise sequentially from the shoot apex across all lineages, but through different generative mechanisms. Angiosperm lateral organ primordia develop from multicellular populations of founder cells specified at the periphery of the SAM whereas lateral organs in non-seed plant lineages arise from a single or a few initials derived from the shoot AC (reviewed in Steeves and Sussex, 1989; Gaillochet et al., 2015). Lateral organ ACs are for the most part morphologically distinct from their corresponding shoot AC (with the exception of *S. kraussiana*), comprising wedge or lenticular shapes with only two cutting faces (**Figure 3K**). Similar AC shapes are found in early stages of bryophyte thallus development (reviewed in Ligrone et al., 2012). The phyllidia of moss gametophores each arise from a single AC that is specified within two cell divisions of the shoot AC (Harrison et al., 2009). Sector analysis demonstrated that microphylls arising from *S. kraussiana* shoots initiate from a pair of adjacent ACs (Harrison et al., 2007), strikingly mirroring the two-celled nature of the shoot apex. Fern fronds typically (but not always) arise from a single AC (reviewed in Vasco et al., 2013). Non-seed plant lateral organ growth is therefore largely driven by ordered patterns of cell division at the tip of the structure (**Figure 3**), whereas in seed plant leaves cell divisions occur across the organ

and morphogenesis is co-ordinated by non-cell autonomous 'supracellular' mechanisms (reviewed in Dengler and Tsukaya, 2001; Fleming, 2002).

Whilst a great deal is now understood about the specification of angiosperm leaf primordia, in which positional signals such as transient auxin maxima are critical (see below), very little is known about the specification of lateral organ initials. In both *P. patens* gametophores (Harrison et al., 2009) and *S. kraussiana* shoots (Harrison et al., 2007), cells arising from shoot and leaf initials follow predictable fates. Although these patterns could indicate cell-autonomous mechanisms for specification, it has been demonstrated in similarly predictable systems that perturbations to division patterns do not change cell fate specification (e.g., van den Berg et al., 1995, 1997), and thus that non-cell autonomous signals can at least compensate for loss of any lineage-based mechanisms. Studies of fern development have found that new frond and pinna initials are specified within distinct merophytes, i.e., groups of related cells descended from a single daughter cell of the AC, in a manner similar to that seen in *P. patens* and *S. kraussiana* (Hou and Hill, 2002; Sanders et al., 2011). However, a role for non-cell autonomous signals is more evident in this case because newly arisen frond primordia develop as shoots if grown in isolation from the shoot apex and older fronds, demonstrating that frond identity is specified by the apex and/or other fronds (reviewed in Vasco et al., 2013). In *C. richardii*, specification of the frond initial is increasingly delayed after cleavage from the shoot AC as development progresses (Hou and Hill, 2002), and patterning of sporangia on reproductive pinnae is dependent on cell position (Hill, 2001). Thus, although fern frond development displays tip-based acropetal growth in common with bryophytes and lycophytes, there is also evidence for non cell-autonomous regulation of developmental patterning in common with angiosperms.

In the case of bryophytes, lycophytes, and seed plants, development of individual lateral organs is determinate. In contrast, fern frond development is iterative, with further subordinate ACs arising from the products of the frond AC, resulting in the outgrowth of pinnae (Figures 3C–E). Interestingly, pinna development on *C. richardii* reproductive fronds is driven by the activity of two adjacent ACs (Figure 3J; Hill, 2001), rather than the single AC seen at the apex of vegetative fronds (Figure 3I; Hou and Hill, 2002), suggesting a functional distinction between the two hierarchical levels (Figure 3K). Frond development is fully indeterminate in some fern species (Vasco et al., 2013), and fossil fronds of early monilophytes also contain indeterminate characters (Sanders et al., 2009). Together with fossil analysis that shows shoot branching in both fern and seed plant lineages prior to the emergence of lateral organs (Sanders et al., 2009), these observations suggest that fern fronds originated as modified shoots.

Polar auxin transport is an essential component of organogenesis at the angiosperm SAM (Reinhardt et al., 2003), where transient auxin maxima in the peripheral zone specify the site of each incipient lateral organ (Vernoux et al., 2011). Similarly, blocking PAT in the *P. patens* gametophore disrupts phyllidia outgrowth and development (Bennett et al., 2014;

Viaene et al., 2014), with extreme examples lacking lateral organs entirely. PAT is also necessary for correct boundary formation between the shoot ACs and microphyll initials in *S. kraussiana*, but microphyll initiation *per se* is unaffected by inhibition of PAT (Sanders and Langdale, 2013). The functions of PAT in the fern sporophyte remain to be investigated, but microsurgical experiments found that primordia do not arise independently, in that each primordium influences the positioning of subsequent primordia at the shoot apex (reviewed in Vasco et al., 2013). At least superficially, this reflects what happens in the angiosperm shoot apex. These observations suggest a conserved role for auxin and PAT in specifying which cells at the apex contribute to lateral organs, and a more divergent role in organ initiation and outgrowth. PAT also has a conserved role in specifying which cells within the lateral organ will form vascular tissue, influencing venation patterns in angiosperm leaves (Scarpella et al., 2006), fronds of the fern *Matteucia struthiopteris* (Ma and Steeves, 1992) and microphylls of *S. kraussiana* (Sanders and Langdale, 2013). More detailed analysis of auxin and PAT function in *C. richardii* shoot development would determine the extent to which these different auxin functions are each conserved within the vascular plants.

As in the case of auxin, two aspects of HD-Zip function appear to be differentially conserved in vascular plants. In addition to functions within the SAM, Class III HD-Zip transcription factors in *A. thaliana* specify adaxial fate and sites of vascular development in newly formed leaf primordia (Prigge et al., 2005; reviewed in Floyd and Bowman, 2007). Expression patterns in two gymnosperms indicate a conserved role for specifying adaxial leaf fate in seed plants (Floyd and Bowman, 2006), but no expression is found in newly formed microphyll primordia of *S. kraussiana* (Floyd and Bowman, 2006). In contrast, expression patterns support a conserved role in the developing vasculature of *S. kraussiana* microphylls (Floyd and Bowman, 2006; Floyd et al., 2006). Given that vasculature evolved in the tracheophytes prior to lateral organs, it is likely that HD-Zips and auxin were first recruited to specify veins, a role that is conserved in extant lycophytes, ferns, and seed plants. When lateral organs subsequently evolved in each of the three lineages, HD-Zip function was modified for specification of leaf polarity in seed plants but not in lycophytes. Closer analysis of HD-Zips in fern frond development would determine whether a role in leaf polarity was independently adopted in monilophytes, and functional analysis in *P. patens* should reveal the ancestral role in non-vascular plants.

SHOOT BRANCHING

The ability to branch is a key innovation in sporophyte shoot development. Two distinct branching systems are found in tracheophytes: apical branching, where the shoot apex bifurcates; and the outgrowth of lateral (axillary) meristems produced in association with lateral organs (reviewed in Sussex and Kerk, 2001). Apical branching is found across the tracheophytes, including lycophytes (such as *S. kraussiana*; Harrison et al., 2007), monilophytes (although not *C. richardii*; Bierhorst,

1977) and in some seed plants (reviewed in Gola, 2014). The existence of tracheophyte fossils such as *Cooksonia*, which have determinate, branched sporophytes (reviewed in Boyce, 2010), suggests that apical branching is the ancestral sporophytic branching mechanism. In addition, the existence of non-vascular polysporangiate fossils (recently reviewed in Edwards et al., 2014) indicates that sporophyte branching emerged prior to the divergence of the first tracheophytes. Most extant bryophyte sporophytes comprise a single axis, but examples of sporophyte apical branching have also been reported in extant mosses and liverworts (Leitgeb, 1876; Györfy, 1929; Bower, 1935). Apical branching is also seen during thallus development of liverwort and hornwort gametophytes (Schuster, 1984a,b). Thus the capacity for sporophyte branching presumably first originated prior to the divergence of bryophytes and vascular plants, but whether branched sporophytes represent an ancestral state in any of the bryophyte lineages is unknown.

Multiple different cellular mechanisms for apical branching have been described across land plants (reviewed in Gola, 2014), such as the proliferation of existing ACs in *S. kraussiana* to establish new axes without interruption (Harrison et al., 2007) or the loss of a single AC followed by initiation of multiple new branch initials, as seen in some leptosporangiate ferns (Héban-Mauri, 1993). The genetic mechanisms underlying apical branching are poorly understood. Experiments in *P. patens* demonstrated that disturbance of PAT or *LEAFY* (*LFY*) gene function can induce sporophyte branching and the production of two terminal sporangia (Tanahashi et al., 2005; Fujita et al., 2008; Bennett et al., 2014). In *S. kraussiana* and fern shoot apices, branching occurs in a regular pattern after a fixed number of lateral organs have been initiated (Bierhorst, 1977; Harrison et al., 2007), suggesting the involvement of a time or distance-dependent regulatory mechanism. Excising *S. kraussiana* shoot tips from a parent plant disrupts this mechanism, resulting in a far greater interval before branching re-initiates, and thus implying that branching is regulated by a mobile signal (Sanders and Langdale, 2013). Auxin is an important branching regulator in seed plants, imposing apical dominance by inhibiting outgrowth of axillary buds through basipetal PAT (reviewed in Müller and Leyser, 2011). *S. kraussiana* shoots exhibit basipetal PAT, but inhibiting auxin transport did not affect the branching interval (Sanders and Langdale, 2013). The different apical branching modes could reflect either convergent evolution of different mechanisms or subsequent diversification from an ancestral branching mechanism. Further study in all non-seed plant lineages is necessary to resolve this.

Early tracheophyte sporophyte fossils exhibit equal (dichotomous) branching (Boyce, 2010), but in subsequent lineages shoot architecture is more complex, with unequal branch growth and apical dominance. Apical branching in *S. kraussiana* is unequal (**Figure 3B**): one branch becomes the major growth axis because of an unequal partitioning of the AC population at the time of branching (Harrison et al., 2007). Unequal branch growth has been proposed as an important component in the origins of lateral organs as part of the telome theory (reviewed in Sussex and Kerk, 2001), with a progression from equal (dichotomous) branching to an asymmetric

branching structure with a dominant shoot apex. The regulatory mechanisms underpinning the evolution of unequal growth have so far not been investigated. Shoot growth in seed plants is regulated by the hormone GA (reviewed in Fleet and Sun, 2005), triggering degradation of the DELLA transcription factors that otherwise restrict growth (reviewed in Ueguchi-Tanaka and Matsuoka, 2010). Functional GA signaling evolved after the divergence of the bryophytes (Hirano et al., 2007; Yasamura et al., 2007), although evidence from *S. moellendorffii* suggests that GA originally regulated reproductive development and not vegetative shoot growth (Aya et al., 2011). The advent of unequal branch growth in the tracheophyte sporophyte might therefore be linked with the co-option of GA signaling as a regulator of vegetative growth.

In contrast to other tracheophytes, shoot architecture in seed plants is dominated by axillary branching (reviewed in Sussex and Kerk, 2001). *De novo* lateral meristems arise in the axils of leaves after lateral organ formation, a process that requires a local depletion of auxin followed by a 'pulse' of CK (Wang et al., 2014). Basipetal PAT from the SAM inhibits axillary bud outgrowth by maintaining high local auxin concentrations, whereas CK promotes their activation by antagonizing auxin function (reviewed in Müller and Leyser, 2011). Axillary branching is also found in moss gametophores where lateral branches arise from single initials re-specified from epidermal cells (Berthier, 1972 and references therein; reviewed in La Farge-England, 1996). Experiments in *P. patens* show that, in striking similarity to seed plant shoots, apically synthesized auxin creates a zone of branching inhibition equivalent to apical dominance (although basipetal PAT is not involved), whilst CK correspondingly promotes branching (Coudert et al., 2015). The degree to which this represents convergent evolution is unclear. Chemical and genetic manipulation of auxin levels in *M. polymorpha* indicate a role in apical dominance and branching in the liverwort thallus (Kaul et al., 1962; Binns and Maravolo, 1972; Davidonis and Munroe, 1972; Maravolo, 1976; Flores-Sandoval et al., 2015), indicating a potential ancestral role for auxin in apical dominance at the base of the land plants. A third hormone, strigolactone (SL), is an important repressor of branch outgrowth in angiosperms (reviewed in Janssen et al., 2014). SL biosynthesis and signaling are thought to have originated prior to the evolution of land plants (Delaux et al., 2012; Wang et al., 2015), and SL has been shown to similarly repress branching in the *P. patens* gametophore (Coudert et al., 2015). Thus, the hormonal regulation of axillary branching is strongly similar in bryophyte gametophytes and seed plant sporophytes.

The precise origins of axillary branching remain unknown. Interestingly, branch points in *Selaginella* species generate *de novo* structures termed 'angle meristems' (Cusick, 1954; Jernstedt et al., 1992). These can develop into aerial roots or shoots, with shoot fate promoted through increased CK or inhibition of PAT (Sanders and Langdale, 2013). The nature of fern fronds is still not fully resolved, but they bear a superficial resemblance to the axillary shooting structure in seed plants, in that ACs are initiated at the frond margin in a hierarchical manner to produce pinnae (Sanders et al., 2011). Many fern

species (including *C. richardii*) also develop *de novo* foliar buds on the adaxial surface of otherwise differentiated lateral organs, which are capable of becoming independent sporophytes (reviewed in Vasco et al., 2013). From these observations, it can be hypothesized that axillary branching in seed plants may have been derived from mechanisms of lateral apical development similar to that seen in fern fronds, potentially relating back to ancestral apical branching mechanisms. However, given that axillary buds are derived from the adaxial surface of the developing lateral organ primordia in angiosperms (McConnell and Barton, 1998), it is perhaps more likely that the axillary branching mechanisms were co-opted from those operating to form *de novo* shoots in the context of monilophyte foliar buds and/or lycophyte rhizophores. In either case, a greater understanding of fern shoot and frond development will be highly informative in addressing this question.

PHASE CHANGE- MODIFYING APICAL AND LATERAL ORGAN DEVELOPMENT

It can be inferred from extant bryophytes that the ancestral sporophyte was purely reproductive in nature, consisting entirely of a stalked sporangium. In contrast, all tracheophyte sporophytes precede reproduction with a vegetative phase that can be short or prolonged depending on the combined activity of endogenous developmental cues and external environmental signals. The developmental origins of this vegetative phase are unclear, and theories to explain its appearance include sterilization of sporangia or interpolation of a novel vegetative structure prior to development of the ancestral reproductive sporophyte (reviewed in Tomescu et al., 2014). Expression analysis of embryonic liverwort and moss sporophytes found evidence of meiosis-associated gene function, even prior to visible sporangium formation (Frank and Scanlon, 2015). It has been proposed that repression of these genetic programs in the early sporophyte led to indeterminate development and, ultimately, the emergence of a vegetative phase. This evidence is consistent with the hypothesis of Bower (1908) that proposed that the elaboration of the sporophyte was driven by selective pressure to delay meiosis (recently reviewed in Qiu et al., 2012).

The vegetative phase can be further sub-divided into juvenile and adult phases, the transitions distinguishable in angiosperms through changes in leaf shape and properties such as leaf hairs and cuticle composition (reviewed in Huijser and Schmid, 2011). Similarly, *S. kraussiana* exhibits a developmental phase change, distinguishable as a change from juvenile spiral phyllotaxy to adult dorsiventral asymmetry (Harrison et al., 2007). Consistent with this, *C. richardii* fronds also undergo a strong and gradual heteroblastic change in morphology during vegetative development, progressing from simple, spade-shaped lamina to highly dissected forms (Hou and Hill, 2002). In angiosperms these phase changes, including the transition to reproductive development (see below) are regulated by two microRNAs, miR156 and miR172 (reviewed in Huijser and Schmid, 2011). miR156 expression has been detected in non-seed plants including mosses and ferns, although corresponding

gene function is not known, whereas conservation of miR172 outside of the angiosperms is still subject to debate. Downstream of these regulators, changes in leaf shape between the juvenile and adult phases in angiosperms are caused by a number of diverse, independently originating mechanisms (reviewed in Bar and Ori, 2015), including reactivation of Class 1 *KNOX* gene expression in leaf primordia (e.g., in tomato; Hareven et al., 1996), ectopic expression of *LFY* (e.g., in pea; Hofer et al., 1997), or through the activity of the REDUCED COMPLEXITY (RCO) homeodomain protein (e.g., in brassicas; Vlad et al., 2014). Activity of these transcription factors, along with the organization of discrete auxin maxima along the leaf margin, promote localized cell divisions in the leaf that convert entire leaf blades into more complex structures with serrated, dissected or compound morphology. Whether the underlying mechanisms driving changes in *C. richardii* frond morphology are conserved with those regulating phase transitions in seed plants is currently unknown.

The most evident phase change during tracheophyte shoot development is the transition to reproductive growth (reviewed in Huijser and Schmid, 2011). In angiosperms, the SAM converts to an inflorescence meristem (IM) that produces floral meristems (FMs) subtended by bracts at its periphery. This transition is promoted by the *LFY* transcription factor, which is up-regulated in the SAM (Weigel et al., 1992). *LFY* also plays a role in the reproductive transition in gymnosperms (Mouradov et al., 1998). In *P. patens*, however, *LFY* instead regulates the first division of the zygote (Tanahashi et al., 2005), a function clearly distinct from its known role in seed plants. *P. patens LFY* (*PpLFY*) homologs are expressed in gametophore shoot apices throughout development, and also in the developing archegonium and the developing sporophyte (Tanahashi et al., 2005). However, no loss-of-function mutant phenotype is obvious in the gametophyte. *C. richardii LFY* (*CrLFY*) homologs are also expressed in both gametophytic and sporophytic tissues (Himi et al., 2001), but no functional data have yet been reported. These observations suggest ancestral functions for the *LFY* gene family in tracheophytes that cannot be predicted from our current knowledge of seed plants.

Whether *LFY* has a role in reproductive transitions in lycophytes and monilophytes is not known, but this question is particularly pertinent in ferns where sporangia develop on 'fertile' fronds rather than as distinct structures arising from the shoot apex. Transcriptional analysis of the fern *A. filiculoides* found that homologs of genes associated with the angiosperm floral transition were up-regulated in sporogenous tissues, including *FLOWERING TIME* (*FT*) and *LFY* (Brouwer et al., 2014). Consistent with a role in reproductive development, *CrLFY* expression has also been detected in the shoot apex and in developing reproductive fronds (Himi et al., 2001). The functional divergence of *LFY* between angiosperms and bryophytes is reflected in changes in protein structure that alter target specificity (Sayou et al., 2014). As a consequence, *PpLFY* homologs cannot complement *lfy* loss-of-function mutants in *A. thaliana* (Maizel et al., 2005). A *CrLFY* homolog (*CrLFY2*) can partially complement the *A. thaliana lfy* mutant (Maizel et al., 2005), indicating some functional conservation but equally that

CrLFY function is not identical to that in angiosperms. Notably, *LFY* expression in angiosperms is promoted by GA to induce flowering (Blázquez et al., 1998) and GA treatment of *Ceratopteris thalictroides* accelerates the production of fertile fronds (Stein, 1971). These results suggest that at least some reproductive functions of *LFY* might be conserved between monilophytes and seed plants, although in light of changing target specificity the downstream mechanisms could vary (see below).

Floral meristem development represents a modification of organogenesis and shoot development within the angiosperms, producing modified lateral (floral) organs in successive, concentric whorls and then terminating the meristem (reviewed in Irish, 2010). Most closely studied in *A. thaliana*, floral organ identity and shoot determinacy are governed by MADS box transcription factors. Two major types of MADS box genes are found across eukaryotes, with type I MADS box genes having received least attention. Type I genes are involved in female gametophyte development and post-zygotic lethality of interspecific hybrids, and mutants generally display phenotypes that are only subtly different from wild-type (Alvarez-Buylla et al., 2000). Type II MADS box genes, with a role in gamete formation in representatives of the sister lineage to land plants (Tanabe et al., 2005), underwent a duplication after the transition to land, diverging into an MIKC⁺ clade, implicated mainly in male gametophyte development (Zobell et al., 2010; Kwantes et al., 2012) and an MIKC^c clade, functioning mostly in the sporophyte. In seed plants, MIKC^c MADS box genes are expressed, with one exception, exclusively in the sporophyte generation (Zobell et al., 2010), whereas they are expressed in both generations in ferns and mosses (Münster et al., 1997; Hasebe et al., 1998; Quodt et al., 2007). It is therefore presumed that MIKC^c MADS box gene function became canalized from an ancestral role in gametophyte development to sporophyte reproduction in seed plants (Nishiyama et al., 2003).

In a revealing analogy to how HOX genes organize the animal body plan (Shubin et al., 1997), floral MIKC^c MADS box genes were first described in angiosperms for their patterning role during flower development, as part of the ABCE model (Bowman et al., 1989; Schwarz-Sommer et al., 1990; Coen and Meyerowitz, 1991; Pelaz et al., 2000). The eye-catching examples whereby homeotic mutations swapped organs such as legs and antenna in the fly *Drosophila melanogaster* or petals and stamens in *A. thaliana* flowers propelled and popularized the field of evo-devo. But beyond the in-depth studies conducted originally in the model flowers of *A. thaliana* and *Antirrhinum majus*, and the necessary modifications to the ABCE model when delving into other branches of the flowering plant phylogeny (Litt and Kramer, 2010), little is known today about the function of ancestral MADS box genes, prior to the evolution of seed plants.

Interrogation of the *S. moellendorffii* genome and subsequent analyses conclude that at least two Type II genes were present in the common ancestor of vascular plants (Banks et al., 2011; Gramzow et al., 2012) but the ABCE class genes are seed plant-specific. Gymnosperms have orthologs of B and C class MIKC^c genes; the expression of the former during male cone development and the later during female and male cone development, points to a conserved role in sporophyte

reproductive structures across seed plants (Tandre et al., 1995, 1998; Rutledge et al., 1998; Mouradov et al., 1999; Shindo et al., 1999; Sundström et al., 1999; Winter et al., 1999; Jager et al., 2003; Zhang et al., 2004). The origin of the different clades of angiosperm-specific MIKC^c genes, including the floral homeotic genes, is presumed to trace back to the seed plant ancestor after the evolution of ferns (Gramzow and Theissen, 2015). This hypothesis is consistent with the absence of floral homeotic gene orthologs from the genomes of *P. patens* and *S. moellendorffii* (Rensing et al., 2008; Banks et al., 2011; Gramzow et al., 2012). Similarly, homologs of MIKC^c genes identified from ferns cannot be assigned to particular subclades of ABCE class genes. In *C. richardii*, at least eight MIKC^c MADS box genes belonging to three main clades have been reported, representing an independent line of evolution from the seed plant MADS box genes and occupying an intermediate position between those of moss and of the major clades of seed plants (Münster et al., 1997; Hasebe et al., 1998; Gramzow et al., 2012; Kwantes et al., 2012). Expression of these genes has been detected in the shoot apex and in both developing vegetative and reproductive fronds (Figure 4; Hasebe et al., 1998). Within the bryophytes, *P. patens* has six MIKC^c-type MADS box genes, which are expressed in both the gametophyte and sporophyte generation, and development of both is impaired upon down-regulation (Quodt et al., 2007; Singer et al., 2007).

In light of the documented substantial bias toward seed plants (mostly angiosperms) in the study of MIKC^c MADS box genes, the prospect of investigating the role of these important regulators of plant development in an evolutionary intermediate plant lineage such as ferns is timely. Unfortunately, the current lack of a transgenic system in lycophytes hinders any immediate prospects of learning about their function in the earliest lineage of vascular plants. Yet the time is ripe for functional studies in a representative of the fern lineage. Such studies will offer a unique glimpse into the function of these genes before they evolved their important role in the development of the flower as a key innovation.

Within the angiosperms the MIKC^c MADS box genes are directly activated by LFY (reviewed in Irish, 2010). While the regulatory relationship between *LFY* and the floral homeotic genes is an established fact in angiosperms, and possibly conserved within seed plants, it is unknown how early in land plant evolution this module was established. Coincident patterns of gene expression between the gymnosperm *LFY* homolog *NEEDLY* (*NDLY*) and the MIKC^c MADS box genes, and the ability of *NDLY* to largely rescue the *A. thaliana* mutant *lfy-1* suggested that LFY-mediated regulation of floral MADS box orthologs was already present in the ancestor of seed plants (Mouradov et al., 1998). Whether *LFY* regulates MADS box genes outside of seed plants is unclear. Non-overlapping patterns of expression of *C. richardii* *LFY* and MADS box genes suggest that *LFY* homologs may not have functioned as regulators of MIKC^c MADS box genes prior to seed plants (Münster et al., 1997; Hasebe et al., 1998; Himi et al., 2001). Functional analysis of *CrLFY* in *C. richardii* will facilitate the reconstruction of the *LFY* gene response network in general, and of its relationship to the MADS box genes in particular.

CONCLUSION AND PERSPECTIVES

In seed plants, shoots and organs develop from the co-ordinated activity of multiple cells, requiring complex intercellular communication to co-ordinate development. In contrast, non-seed plant shoots typically develop from single or multiple distinct ACs. With little known about the genetic pathways underlying AC function, based on morphology they were considered to be functionally divergent from the SAM. However, recent cell-specific expression analysis has suggested that it is inappropriate to think of ACs as single-cell shoot apices, with a number of regulatory mechanisms associated with seed plant SAMs expressed in tissues immediately surrounding them. A comparison between ACs from different land plant lineages suggests a gradual accumulation of conserved shoot apical regulatory mechanisms, a number of which (e.g., Class 1 KNOX, CK and auxin) are associated with apical function in earliest diverging bryophyte lineages. Beyond the shoot apex, however, greater divergence is evident in the regulation of organogenesis and subsequent lateral organ development. This is perhaps to be expected, as it reflects the independent origins of lateral organs within each lineage of vascular plants: the lycophytes, monilophytes, and seed plants.

Close scrutiny of the monilophyte (fern) shoot system highlights differences compared to both the moss and flowering plant developmental models, *P. patens* and *A. thaliana*. Arising from single or paired ACs, *C. richardii* shoot and frond development nevertheless shows indications of complex supracellular regulation more similar to flowering plants than to moss. However, the evidence to date points to independent origins for fern fronds and seed plant leaves, with frond development more equivalent to flowering plant shoots than to leaves. In short, ferns are not more elaborate mosses or slightly simpler flowering plants, but possess a distinct and complex developmental identity of their own.

In most aspects of shoot development covered in this review, relatively clear trajectories can be inferred between the bryophytes and the tracheophytes, including the conservation or adaptation of several ancestral regulatory mechanisms. Although major evolutionary changes occurred during the bryophyte-tracheophyte transition, equally significant alterations to shoot development occurred within the vascular plant lineages, and the questions regarding these are often intractable based on current data. As sister group to the seed plants, exploring fern development has the potential to dramatically improve our understanding of seed plant evolution and to

fully resolve the broader evolutionary trajectories that have occurred in land plants as a whole. Throughout this review we have highlighted numerous specific examples where further information regarding gene function in a fern would be invaluable. Given the broad diversity of the monilophytes, it is probable that numerous model species will ultimately be required from within this clade to fully understand different adaptive aspects of their development. Despite its derived aquatic adaptations, *C. richardii* can be considered a good candidate as an initial model species because, as a leptosporangiate fern, it represents a major clade within the monilophytes. Of crucial advantage, however, are the facts that it has already been established for laboratory use and the tools for genetic analysis in this species have now been developed. Forward genetic analysis (i.e., identification of unknown genes involved in a particular developmental process through mutants) has been exploited successfully in *C. richardii* to elucidate the pathway regulating sex-determination during gametophyte development (Warne and Hickok, 1991; Banks, 1994, 1997; Strain et al., 2001), and a double-haploid mapping population between two *C. richardii* ecotypes was used to create a genetic linkage map (Nakazato et al., 2006). Until now, however, such approaches have been severely limited by a lack of resources, with large-scale mutant libraries such as those available for *A. thaliana* not yet established. The recent development of methods to genetically transform *C. richardii* is therefore an important milestone, as it will allow investigation of gene function in a monilophyte via a reverse genetics approach (manipulation of candidate gene expression or function), and hopefully also provide the impetus to improve the other genetic resources available for this and other fern species.

FUNDING

JL and AP were supported by grants from the ERC (AdG - EDIP) and the Gatsby Charitable Foundation to JL. VD was supported by the Royalty Research Fund, University of Washington and National Science Foundation grant IOS-1121669.

ACKNOWLEDGMENT

The authors are grateful to Laura Moody for supplying plant material for photography.

REFERENCES

- Alvarez-Buylla, E. R., Pelaz, S., Liljgren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., et al. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5328–5333. doi: 10.1073/pnas.97.10.5328
- Aso, K., Kato, M., Banks, J. A., and Hasebe, M. (1999). Characterization of homeodomain-leucine zipper genes in the fern *Ceratopteris richardii* and the evolution of the homeodomain-leucine zipper gene family in vascular plants. *Mol. Biol. Evol.* 16, 544–552. doi: 10.1093/oxfordjournals.molbev.a026135
- Aya, K., Hiwatashi, Y., Kojima, M., Sakakibara, H., Ueguchi-Tanaka, M., Hasebe, M., et al. (2011). The gibberellin perception system evolved to regulate a pre-existing GAMYB-mediated system during land plant evolution. *Nat. Commun.* 2, 544. doi: 10.1038/ncomms1552
- Bainard, J. D., Henry, T. A., Bainard, L. D., and Newmaster, S. G. (2011). DNA content variation in monilophytes and lycophytes: large genomes that are not endopolyploid. *Chromosome Res.* 19, 763–775. doi: 10.1007/s10577-011-9228-1
- Banks, J. A. (1994). Sex-determining genes in the homosporous fern *Ceratopteris. Development* 120, 1949–1958.

- Banks, J. A. (1997). The TRANSFORMER genes of the fern *Ceratopteris* simultaneously promote meristem and archegonia development and repress antheridia development in the developing gametophyte. *Genetics* 147, 1885–1897.
- Banks, J. A. (1999). Gametophyte development in ferns. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 163–186. doi: 10.1146/annurev.arplant.50.1.163
- Banks, J. A. (2009). *Selaginella* and 400 million years of separation. *Annu. Rev. Plant Biol.* 60, 223–238. doi: 10.1146/annurev.arplant.59.032607.092851
- Banks, J. A. (2015). The evolution of the shoot apical meristem from a gene expression perspective. *New Phytol.* 207, 486–487. doi: 10.1111/nph.13525
- Banks, J. A., Nishiyama, T., Hasebe, M., Bowman, J. L., Gribskov, M., dePamphilis, C., et al. (2011). The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332, 960–963. doi: 10.1126/science.1203810
- Bar, M. S., and Ori, N. (2015). Compound leaf development in model plant species. *Curr. Opin. Plant Biol.* 23, 61–69. doi: 10.1016/j.pbi.2014.10.007
- Barker, M. S., and Wolf, P. G. (2010). Unfurling fern biology in the genomics age. *Bioscience* 60, 177–185. doi: 10.1525/bio.2010.60.3.4
- Bartlett, E. M. (1928). The comparative study of the development of the sporophyte in the Anthocerotaceae, with special reference to the genus *Anthoceros*. *Ann. Bot.* 42, 409–430.
- Bäurle, I., and Laux, T. (2005). Regulation of WUSCHEL transcription in the stem cell niche of the *Arabidopsis* shoot meristem. *Plant Cell* 17, 2271–2280. doi: 10.1105/tpc.105.032623
- Belles-Boix, E., Hamant, O., Witiak, S. M., Morin, H., Traas, J., and Pautot, V. (2006). KNAT6: an *Arabidopsis* homeobox gene involved in meristem activity and organ separation. *Plant Cell* 18, 1900–1907. doi: 10.1105/tpc.106.041988
- Bennett, T. A. (2014). PIN proteins and the evolution of plant development. *Trends Plant Sci.* 20, 498–507. doi: 10.1016/j.tplants.2015.05.005
- Bennett, T. A., Liu, M. M., Aoyama, T., Bierfreund, N. M., Braun, M., Coudert, Y., et al. (2014). Plasma membrane-targeted PIN proteins drive shoot development in a moss. *Curr. Biol.* 24, 2776–2785. doi: 10.1016/j.cub.2014.09.054
- Berthier, J. (1972). Recherches sur la structure et le développement de l'apex du gamétophyte feuillé des mousses. *Rev. Bryol. Lichénol.* 38, 421–551.
- Bharathan, G., Goliber, T. E., Moore, C., Kessler, S., Pham, T., and Sinha, N. R. (2002). Homologies in leaf form inferred from KNOX1 gene expression during development. *Science* 296, 1858–1860. doi: 10.1126/science.1070343
- Bierhorst, D. W. (1977). On the stem apex, leaf initiation and early leaf ontogeny in filiclean ferns. *Am. J. Bot.* 64, 125–152. doi: 10.2307/2442101
- Binns, A. N., and Maravolo, N. C. (1972). Apical dominance, polarity and adventitious growth in *Marchantia polymorpha*. *Am. J. Bot.* 59, 691–696. doi: 10.2307/2441140
- Blázquez, M. A., Green, R., Nilsson, O., Sussman, M. R., and Weigel, D. (1998). Gibberellins promote flowering of *Arabidopsis* by activating the LEAFY promoter. *Plant Cell* 10, 791–800. doi: 10.1105/tpc.10.5.791
- Bower, F. O. (1908). *The Origin of a Land Flora: A Theory Based Upon the Facts of Alternation*. London: MacMillan Press.
- Bower, F. O. (1935). *Primitive Land Plants*. London: MacMillan Press.
- Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1989). Genes directing flower development in *Arabidopsis*. *Plant Cell* 1, 37–52. doi: 10.1105/tpc.1.1.37
- Boyce, C. K. (2010). The evolution of plant development in a paleontological context. *Curr. Opin. Plant Biol.* 13, 102–107. doi: 10.1016/j.pbi.2009.10.001
- Boyce, C. K., and Knoll, A. H. (2002). Evolution of developmental potential and the multiple independent origins of leaves in Paleozoic vascular plants. *Paleobiology* 28, 70–100. doi: 10.1666/0094-8373(2002)028<0070:EODPAT>2.0.CO;2
- Brouwer, P., Bräutigam, A., Külahoglu, C., Tazelaar, A. O. E., Kurz, S., Nierop, K. G. J., et al. (2014). Azolla domestication towards a biobased economy? *New Phytol.* 202, 1069–1082. doi: 10.1111/nph.12708
- Bui, L. T., Cordle, A. R., Irish, E. E., and Cheng, C. L. (2015). Transient and stable transformation of *Ceratopteris richardii* gametophytes. *BMC Res. Notes* 8:214. doi: 10.1186/s13104-015-1193-x
- Byrne, M. E., Barley, R., Curtis, M., Arroyo, J. M., Dunham, M., Hudson, A., et al. (2000). Asymmetric leaves1 mediates leaf patterning and stem cell function in *Arabidopsis*. *Nature* 408, 967–971. doi: 10.1038/35050091
- Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0
- Coudert, Y., Palubicki, W., Ljung, K., Novak, O., Leyser, O., and Harrison, C. J. (2015). Three ancient hormonal cues co-ordinate shoot branching in a moss. *Elife* 4, e06808. doi: 10.7554/eLife.06808
- Crandall-Stotler, B. (1980). Morphogenetic designs and a theory of bryophyte origins and divergence. *Bioscience* 30, 580–585. doi: 10.2307/1308107
- Crandall-Stotler, B. (1986). Morphogenesis, developmental anatomy and bryophyte phylogenetics: contraindications of monophyly. *J. Bryol.* 14, 1–23. doi: 10.1179/jbr.1986.14.1.1
- Cusick, F. (1954). Experimental and analytical studies of pteridophytes. XXV. Morphogenesis in *Selaginella willdenovii* Baker. II. Angle-meristems and angle-shoots. *Ann. Bot.* 18, 171–181.
- Davidonis, G. H., and Munroe, M. H. (1972). Apical dominance in *Marchantia*—correlative inhibition of neighbor lobe growth. *Bot. Gaz.* 133, 177–184. doi: 10.1086/336631
- Delaux, P. M., Xie, X., Timme, R. E., Puech-Pages, V., Dunand, C., Lecompte, E., et al. (2012). Origin of strigolactones in the green lineage. *New Phytol.* 195, 857–871. doi: 10.1111/j.1469-8137.2012.04209.x
- Dengler, N. G., and Tsukaya, H. (2001). Leaf morphogenesis in dicotyledons: current issues. *Int. J. Plant Sci.* 162, 459–464. doi: 10.1101/sqb.2013.77.015883
- Edwards, D., Morris, J. L., Richardson, J. B., and Kenrick, P. (2014). Cryptospores and cryptophytes reveal hidden diversity in early land floras. *New Phytol.* 202, 50–78. doi: 10.1111/nph.12645
- Eklund, D. M., Ishizaki, K., Flores-Sandoval, E., Kikuchi, S., Takebayashi, Y., Tsukamoto, S., et al. (2015). Auxin produced by the indole-3-pyruvic acid pathway regulates development and gemmae dormancy in the liverwort *Marchantia polymorpha*. *Plant Cell* 27, 1650–1669. doi: 10.1105/tpc.15.00065
- Emery, J. F., Floyd, S. K., Alvarez, J., Eshed, Y., Hawker, N. P., Izhaki, A., et al. (2003). Radial patterning of *Arabidopsis* shoots by Class III HD-ZIP and KANADI genes. *Curr. Biol.* 13, 1768–1774. doi: 10.1016/j.cub.2003.09.035
- Fleet, C. M., and Sun, T. P. (2005). A DELLAcate balance: the role of gibberellin in plant morphogenesis. *Curr. Opin. Plant Biol.* 8, 77–85. doi: 10.1016/j.pbi.2004.11.015
- Fleming, A. J. (2002). The mechanism of leaf morphogenesis. *Planta* 216, 17–22. doi: 10.1007/s00425-002-0864-8
- Flores-Sandoval, E., Eklund, D. M., and Bowman, J. L. (2015). A simple auxin transcriptional response system regulates multiple morphogenetic processes in the liverwort *Marchantia polymorpha*. *PLoS Genet.* 11:e1005207. doi: 10.1371/journal.pgen.1005207
- Floyd, S. K., and Bowman, J. L. (2006). Distinct developmental mechanisms reflect the independent origins of leaves in vascular plants. *Curr. Biol.* 16, 1911–1917. doi: 10.1016/j.cub.2006.07.067
- Floyd, S. K., and Bowman, J. L. (2007). The ancestral developmental tool kit of land plants. *Int. J. Plant Sci.* 168, 1–35. doi: 10.1086/509079
- Floyd, S. K., Zalewski, C. S., and Bowman, J. L. (2006). Evolution of class III homeodomain-leucine zipper genes in streptophytes. *Genetics* 173, 373–388. doi: 10.1534/genetics.105.054239
- Frank, M. H., Edwards, M. B., Schultz, E. R., McKain, M. R., Fei, Z., Sørensen, I., et al. (2015). Dissecting the molecular signatures of apical cell-type shoot meristems from two ancient land plant lineages. *New Phytol.* 207, 893–904. doi: 10.1111/nph.13407
- Frank, M. H., and Scanlon, M. J. (2015). Transcriptomic evidence for the evolution of shoot meristem function in sporophyte-dominant land plants through concerted selection of ancestral gametophytic and sporophytic genetic programs. *Mol. Biol. Evol.* 32, 355–367. doi: 10.1093/molbev/msu303
- Fujita, T., Sakaguchi, H., Hiwatashi, Y., Wagstaff, S. J., Ito, M., Deguchi, H., et al. (2008). Convergent evolution of shoots in lands plants: lack of auxin polar transport in moss shoots. *Evol. Dev.* 10, 176–186. doi: 10.1111/j.1525-142X.2008.00225.x
- Gaillochet, C., Daum, G., and Lohmann, J. U. (2015). O cell, where art thou? The mechanisms of shoot meristem patterning. *Curr. Opin. Plant Biol.* 23, 91–97. doi: 10.1016/j.pbi.2014.11.002
- Gaillochet, C., and Lohmann, J. U. (2015). The never-ending story: from pluripotency to plant developmental plasticity. *Development* 142, 2237–2249. doi: 10.1242/dev.117614
- Gerrienne, P., and Genez, P. (2011). Early evolution of life cycles in embryophytes: a focus on the fossil evidence of gametophyte/sporophyte size and morphological complexity. *J. Syst. Evol.* 49, 1–16. doi: 10.1111/j.1759-6831.2010.00096.x

- Gifford, E. M., and Foster, A. S. (1989). *Morphology and Evolution of Vascular Plants*. San Francisco, CA: W.H. Freeman and Company.
- Gola, E. M. (2014). Dichotomous branching: the plant form and integrity upon the apical meristem bifurcation. *Front. Plant Sci.* 5:263. doi: 10.3389/fpls.2014.00263
- Gramzow, L., Barker, E., Schulz, C., Ambrose, B., Ashton, N., Theißen, G., et al. (2012). *Selaginella* genome analysis—entering the “homoplasy heaven” of the MADS world. *Front. Plant Sci.* 3:214. doi: 10.3389/fpls.2012.00214
- Gramzow, L., and Theißen, G. (2015). Phylogenomics reveals surprising sets of essential and dispensable clades of MIKCC-group MADS-box genes in flowering plants. *J. Exp. Zool. B Mol. Dev. Evol.* 324, 353–362. doi: 10.1002/jez.b.22598
- Guo, M., Thomas, J., Collins, G., and Timmermans, M. C. P. (2008). Direct repression of KNOX loci by the ASYMMETRIC LEAVES1 complex of *Arabidopsis*. *Plant Cell* 20, 48–58. doi: 10.1105/tpc.107.056127
- Györfy, I. (1929). Monstruoses sporophyten von tetraplodon bryoides aus suomi. *Ann. Soc. Zool.* 9, 299–319.
- Hareven, D., Gutfinger, T., Parnis, A., Eshed, Y., and Lifschitz, E. (1996). The making of a compound leaf: genetic manipulation of leaf architecture in tomato. *Cell* 84, 735–744. doi: 10.1016/S0092-8674(00)81051-X
- Harrison, C. J. (2015). Shooting through time: new insights from transcriptomic data. *Trends Plant Sci.* 20, 468–470. doi: 10.1016/j.tplants.2015.06.003
- Harrison, C. J., Corlet, S. B., Moylan, E. C., Alexander, D. L., Scotland, R. W., and Langdale, J. L. (2005). Independent recruitment of a conserved developmental mechanism during leaf evolution. *Nature* 434, 509–514. doi: 10.1038/nature03410
- Harrison, C. J., and Langdale, J. A. (2010). The developmental pattern of shoot apices in *Selaginella kraussiana* (Kunze) A. Braun. *Int. J. Plant Sci.* 171, 690–692. doi: 10.1086/653134
- Harrison, C. J., Rezvani, M., and Langdale, J. A. (2007). Growth from two transient apical initials in the meristem of *Selaginella kraussiana*. *Development* 134, 881–889. doi: 10.1242/dev.001008
- Harrison, C. J., Roeder, A. H., Meyerowitz, E. M., and Langdale, J. A. (2009). Local cues and asymmetric cell divisions underpin body plan transitions in the moss *Physcomitrella patens*. *Curr. Biol.* 19, 461–471. doi: 10.1016/j.cub.2009.02.050
- Hasebe, M., Wen, C. K., Kato, M., and Banks, J. A. (1998). Characterization of MADS homeotic genes in the fern *Ceratopteris richardii*. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6222–6227. doi: 10.1073/pnas.95.11.6222
- Héban-Mauri, R. (1993). Caudine meristems in leptosporangiate ferns: structure, lateral appendages, and branching. *Can. J. Bot.* 71, 1612–1624. doi: 10.1139/b93-196
- Hickok, L. G., Warne, T. R., and Fribourg, R. S. (1995). Biology of the fern *Ceratopteris* and its use as a model system. *Int. J. Plant Sci.* 156, 332–345. doi: 10.1086/297255
- Hill, J. (2001). Meristem development at the sporophyll pinna apex in *Ceratopteris richardii*. *Int. J. Plant Sci.* 16, 235–247. doi: 10.1086/319576
- Himi, S., Sano, R., Nishiyama, T., Tanahashi, T., Kato, M., Ueda, K., et al. (2001). Evolution of MADS-box gene induction by FLO/LFY genes. *J. Mol. Evol.* 53, 387–393. doi: 10.1007/s002390010228
- Hirano, K., Nakajima, M., Asano, K., Nishiyama, T., Sakakibara, H., Kojima, M., et al. (2007). The GID1-mediated gibberellin perception mechanism is conserved in the lycophyte *Selaginella moellendorffii* but not in the bryophyte *Physcomitrella patens*. *Plant Cell* 19, 3058–3079. doi: 10.1105/tpc.107.051524
- Hofer, J., Turner, L., Hellens, R., Ambrose, M., Matthews, P., Michael, A., et al. (1997). UNIFOLIATA regulates leaf and flower morphogenesis in pea. *Curr. Biol.* 7, 581–587. doi: 10.1016/S0960-9822(06)00257-0
- Hou, G. C., and Hill, J. P. (2002). Heteroblastic root development in *Ceratopteris richardii* (Parkeriaceae). *Int. J. Plant Sci.* 163, 341–351. doi: 10.1086/339156
- Huijser, P., and Schmid, M. (2011). The control of developmental phase transitions in plants. *Development* 138, 4117–4129. doi: 10.1242/dev.063511
- Irish, V. F. (2010). The flowering of *Arabidopsis* flower development. *Plant J.* 61, 1014–1028. doi: 10.1111/j.1365-3113X.2009.04065.x
- Jackson, D., Veit, B., and Hake, S. (1994). Expression of maize KNOTTED1 related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development* 120, 405–413.
- Jager, M., Hassanin, A., Manuel, M., Le Guyader, H., and Deutsch, J. (2003). MADS-box genes in *Ginkgo biloba* and the evolution of the AGAMOUS family. *Mol. Biol. Evol.* 20, 842–854. doi: 10.1093/molbev/msg089
- Janssen, B. J., Drummond, R. S. M., and Snowden, K. C. (2014). Regulation of axillary shoot development. *Curr. Opin. Plant Biol.* 17, 28–35. doi: 10.1016/j.pbi.2013.11.004
- Jasinski, S., Piazza, P., Craft, J., Hay, A., Woolley, L., Rieu, I., et al. (2005). KNOX action in *Arabidopsis* is mediated by coordinate regulation of cytokinin and gibberellin activities. *Curr. Biol.* 15, 1560–1565. doi: 10.1016/j.cub.2005.07.023
- Jernstedt, J. A., Cutter, E. G., Gifford, E. M., and Lu, P. (1992). Angle meristem origin and development in *Selaginella martensii*. *Ann. Bot.* 69, 351–363.
- Johnson, G. P., and Renzaglia, K. S. (2008). Embryology of *Ceratopteris richardii* (Pteridaceae, tribe Ceratopterideae), with emphasis on placental development. *J. Plant Res.* 121, 581–592. doi: 10.1007/s10265-008-0187-3
- Jones, C. S., and Drinnan, A. N. (2009). The developmental pattern of shoot apices in *Selaginella kraussiana* (Kunze) A. Braun. *Int. J. Plant Sci.* 170, 1009–1018. doi: 10.1086/605118
- Karol, K. G., McCourt, R. M., Cimino, M. T., and Delwiche, C. F. (2001). The closest living relatives of land plants. *Science* 294, 2351–2353. doi: 10.1126/science.1065156
- Kato, M., and Akiyama, H. (2005). Interpolation hypothesis for origin of the vegetative sporophyte of land plants. *Taxon* 54, 443–450. doi: 10.2307/25065371
- Kaul, K. N., Mitra, G. C., and Tripathi, B. K. (1962). Responses of *Marchantia* in aseptic culture to well-known auxins and antiauxins. *Ann. Bot.* 26, 177–184.
- Kenrick, P., and Crane, P. R. (1997). *The Origin and Early Diversification of Land Plants*. Washington, DC: Smithsonian Institution Press.
- Kny, L. (1890). *Botanische Wandtafeln Mit Erläuterndem Text*. Berlin: Wiegandt, Hempel & Parey.
- Kofuji, R., and Hasebe, M. (2014). Eight types of stem cells in the life cycle of the moss *Physcomitrella patens*. *Curr. Opin. Plant Biol.* 17, 13–21. doi: 10.1016/j.pbi.2013.10.007
- Kwantes, M., Liebsch, D., and Verelst, W. (2012). How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. *Mol. Biol. Evol.* 29, 293–302. doi: 10.1093/molbev/msr200
- La Farge-England, C. (1996). Growth form, branching pattern, and perichaetial position in mosses: cladocarp and pleurocarpy redefined. *Bryologist* 99, 170–186. doi: 10.2307/3244546
- Lee, J. H., Lin, H., Joo, S., and Goodenough, U. (2008). Early sexual origins of homeoprotein heterodimerization and evolution of the plant KNOX/BELL family. *Cell* 133, 829–840. doi: 10.1016/j.cell.2008.04.028
- Leitgeb, H. (1876). Ueber verzweigte Moosporogonien. *Mitt. Naturwiss. Ver. Steiermark* 13, 3–20.
- Leitgeb, H. (1881). *Untersuchungen über die Lebermoose. Heft VI, Die Marchantieen*. Jene: Leuschner & Lubensky.
- Lenhard, M., and Laux, T. (2003). Stem cell homeostasis in the *Arabidopsis* shoot meristem is regulated by intercellular movement of CLAVATA3 and its sequestration by CLAVATA1. *Development* 130, 3163–3173. doi: 10.1242/dev.00525
- Lewis, L. A., and McCourt, R. M. (2004). Green algae and the origin of land plants. *Am. J. Bot.* 91, 1535–1556. doi: 10.3732/ajb.91.10.1535
- Ligrone, R., Duckett, J. G., and Renzaglia, K. S. (2012). Major transitions in the evolution of early land plants: a bryological perspective. *Ann. Bot.* 109, 851–871. doi: 10.1093/aob/mcs017
- Litt, A., and Kramer, E. M. (2010). The ABC model and the diversification of floral organ identity. *Semin. Cell Dev. Biol.* 21, 129–137. doi: 10.1016/j.semcdb.2009.11.019
- Long, J. A., Moan, E. I., Medford, J. I., and Barton, M. K. (1996). A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of *Arabidopsis*. *Nature* 379, 66–69. doi: 10.1038/379066a0
- Lucas, W. J., Bouché-Pillon, S., Jackson, D. P., Nguyen, L., Baker, L., Ding, B., et al. (1995). Selective trafficking of KNOTTED1 homeodomain protein and its mRNA through plasmodesmata. *Science* 270, 1980–1983. doi: 10.1126/science.270.5244.1980
- Ma, Y., and Steeves, T. A. (1992). Auxin effects on vascular differentiation in ostrich fern. *Ann. Bot.* 70, 277–282.

- Maizel, A., Busch, M. A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M., et al. (2005). The floral regulator *LEAFY* evolves by substitutions in the DNA binding domain. *Science* 308, 260–263. doi: 10.1126/science.1108229
- Maravolo, N. C. (1976). Polarity and localization of auxin movement in the hepatic, *Marchantia polymorpha*. *Am. J. Bot.* 63, 526–531. doi: 10.2307/2441816
- McConnell, J. R., and Barton, M. K. (1998). Leaf polarity and meristem formation in *Arabidopsis*. *Development* 125, 2935–2942.
- McCormick, S. (2004). Control of male gametophyte development. *Plant Cell* 16, S142–S153. doi: 10.1105/tpc.016659
- Mouradov, A., Glassick, T., Hamdorf, B., Murphy, L., Fowler, B., Marla, S., et al. (1998). NEEDLY, a *Pinus radiata* ortholog of *FLORICAULA/LEAFY* genes, expressed in both reproductive and vegetative meristems. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6537–6542. doi: 10.1073/pnas.95.11.6537
- Mouradov, A., Hamdorf, B., Teasdale, R. D., Kim, J. T., Winter, K. U., and Theissen, G. (1999). A DEF/GLO-like MADS-box gene from a gymnosperm: *Pinus radiata* contains an ortholog of angiosperm B class floral homeotic genes. *Dev. Genet.* 25, 245–252. doi: 10.1002/(SICI)1520-6408(1999)25:3<245::AID-DVG7>3.0.CO;2-N
- Müller, D., and Leyser, O. (2011). Auxin, cytokinin and the control of shoot branching. *Ann. Bot.* 107, 1203–1212. doi: 10.1093/aob/mcr069
- Münster, T., Pahnke, J., Di Rosa, A., Kim, J. T., Martin, W., Saedler, H., et al. (1997). Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proc. Natl. Acad. Sci. U.S.A.* 94, 2415–2420. doi: 10.1073/pnas.94.6.2415
- Muthukumar, B., Joyce, B. L., Elless, M. P., and Stewart, C. N. (2013). Stable transformation of ferns using spores as targets: *Pteris vittata* and *Ceratopteris thalictroides*. *Plant Physiol.* 163, 648–658. doi: 10.1104/pp.113.224675
- Nakazato, T., Jung, M. K., Houseworth, E. A., Rieseberg, L. H., and Gastony, G. J. (2006). Genetic map-based analysis of genome structure in the homosporous fern *Ceratopteris richardii*. *Genetics* 173, 1585–1597. doi: 10.1534/genetics.106.055624
- Niklas, K. J., and Kutschera, U. (2010). The evolution of the land plant life cycle. *New Phytol.* 185, 27–41. doi: 10.1111/j.1469-8137.2009.03054.x
- Nishiyama, T., Fujita, T., Shin-I, T., Seki, M., Nishide, H., Uchiyama, I., et al. (2003). Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8007–8012. doi: 10.1073/pnas.0932694100
- Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405, 200–203. doi: 10.1038/35012103
- Plackett, A. R. G., Huang, L., Sanders, H. L., and Langdale, J. A. (2014). High-efficiency stable transformation of the model fern species *Ceratopteris richardii* via microparticle bombardment. *Plant Physiol.* 165, 3–14. doi: 10.1104/pp.113.231357
- Prigge, M. J., Otsuga, D., Alonso, J. M., Ecker, J. R., Drews, G. N., and Clark, S. E. (2005). Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic and distinct roles in *Arabidopsis* development. *Plant Cell* 17, 61–76. doi: 10.1105/tpc.104.026161
- Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S., et al. (2001). Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* 409, 618–622. doi: 10.1038/35054555
- Qiu, Y. L., Li, L., Wang, B., Chen, Z., Knoop, V., Growth-Malonek, M., et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15511–15516. doi: 10.1073/pnas.0603335103
- Qiu, Y. L., Taylor, A. B., and McManus, H. A. (2012). Evolution of the life cycle in land plants. *J. Syst. Evol.* 50, 171–194. doi: 10.1111/j.1759-6831.2012.00188.x
- Quodt, V., Faigl, W., Saedler, H., and Münster, T. (2007). The MADS-domain protein PPM2 preferentially occurs in gametangia and sporophytes of the moss *Physcomitrella patens*. *Gene* 400, 25–34. doi: 10.1016/j.gene.2007.05.016
- Reinhardt, D., Pesce, E. R., Stieger, P., Mandel, T., Baltensperger, K., Bennett, M., et al. (2003). Regulation of phyllotaxis by polar auxin transport. *Nature* 426, 255–260. doi: 10.1038/nature02081
- Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319, 64–69. doi: 10.1126/science.1150646
- Renzaglia, K. S. (1978). A comparative morphology and developmental anatomy of the Anthocerotophyta. *J. Hattori Bot. Lab.* 44, 31–90.
- Renzaglia, K. S., Villarreal, J. C., and Duff, R. J. (2009). “New insights into morphology, anatomy and systematics of hornworts,” in *Bryophyte Biology*, eds B. Goffinet and J. Shaw (Cambridge: Cambridge University Press), 139–171.
- Robert, D. (1971). The female gametophyte of *Selaginella kraussiana*. Part 2: histological organization of the reproductive tissue and the main aspects of cellular de-differentiation preparatory to oogenesis. *Rev. Cytol. Biol. Vegetales* 34, 189–211.
- Robert, D. (1973). The male gametophyte of *Selaginella kraussiana*. Organization and development: an electron microscopic study. *Ann. Sci. Nat. Bot. Biol. Vegetale* 14, 465–504.
- Ruhfel, B. R., Gitzendanner, M. S., Soltis, P. S., Soltis, D. E., and Burleigh, J. G. (2014). From algae to angiosperms— inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* 14:23. doi: 10.1186/1471-2148-14-23
- Rutledge, R., Regan, S., Nicolas, O., Fobert, P., Côte, C., Bosnich, W., et al. (1998). Characterization of an AGAMOUS homologue from the conifer black spruce (*Picea mariana*) that produces floral homeotic conversions when expressed in *Arabidopsis*. *Plant J.* 15, 625–634. doi: 10.1046/j.1365-313x.1998.00250.x
- Sakakibara, K., Nishiyama, T., Deguchi, H., and Hasebe, M. (2008). Class I KNOX genes are not involved in shoot development in the moss *Physcomitrella patens* but do function in sporophyte development. *Evol. Dev.* 10, 555–566. doi: 10.1111/j.1525-142X.2008.00271.x
- Sakakibara, K., Nishiyama, T., Kato, M., and Hasebe, M. (2001). Isolation of homeodomain-leucine zipper genes from the moss *Physcomitrella patens* and the evolution of homeodomain-leucine zipper genes in land plants. *Mol. Biol. Evol.* 18, 491–502. doi: 10.1093/oxfordjournals.molbev.a003828
- Sanders, H. L., Darrah, P. R., and Langdale, J. A. (2011). Sector analysis and predictive modelling reveal iterative shoot-like development in fern fronds. *Development* 138, 2925–2934. doi: 10.1242/dev.065888
- Sanders, H. L., and Langdale, J. A. (2013). Conserved transport mechanisms but distinct auxin responses govern shoot patterning in *Selaginella kraussiana*. *New Phytol.* 198, 419–428. doi: 10.1111/nph.12183
- Sanders, H. L., Rothwell, G. W., and Wyatt, S. E. (2009). Key morphological alterations in the evolutions of leaves. *Int. J. Plant Sci.* 170, 860–868. doi: 10.1086/600135
- Sano, R., Juárez, C. M., Hass, B., Sakakibara, K., Ito, M., Banks, J. A., et al. (2005). KNOX homeobox genes potentially have similar function in both diploid unicellular and multicellular meristems, but not in haploid meristems. *Evol. Dev.* 7, 69–78. doi: 10.1111/j.1525-142X.2005.05008.x
- Sayou, C., Monniaux, M., Nanao, M. H., Moyroud, E., Brockington, S. F., Thévenon, E., et al. (2014). A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* 343, 645–648. doi: 10.1126/science.1248229
- Scarpella, E., Marcos, D., Friml, J., and Berleth, T. (2006). Control of leaf vascular patterning by polar auxin transport. *Genes Dev.* 20, 1015–1027. doi: 10.1101/gad.1402406
- Schneeberger, R., Tsiantis, M., Freeling, M., and Langdale, J. A. (1998). The rough sheath2 gene negatively regulates homeobox gene expression during maize leaf development. *Development* 125, 2857–2865.
- Schoof, H., Lenhard, M., Haecker, A., Mayer, K. F. X., Jürgens, G., and Laux, T. (2000). The stem cell population of *Arabidopsis* shoot meristems is maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. *Cell* 100, 635–644. doi: 10.1016/S0092-8674(00)80700-X
- Schuettpelz, E., and Pryer, K. M. (2008). “Fern phylogeny,” in *Biology and Evolution of Ferns and Lycophytes*, eds T. A. Ranker and C. H. Haufler (Cambridge: Cambridge University Press), 395–416.
- Schuettpelz, E., and Pryer, K. M. (2009). Evidence for a Cenozoic radiation of ferns in an angiosperm-dominated canopy. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11200–11205. doi: 10.1073/pnas.0811136106
- Schulz, C., Little, D. P., Stevenson, D. W., Bauer, D., Moloney, C., and Stützel, T. (2010). An overview of the morphology, anatomy, and life cycle of a new model species: the lycophyte *Selaginella apoda* (L.) Spring. *Int. J. Plant Sci.* 171, 693–712. doi: 10.1086/654902

- Schuster, R. M. (1984a). "Comparative anatomy and morphology of the Hepaticae," in *New Manual of Bryology*, Vol. 2, ed. R. M. Schuster (Nichinan: Hattori Botanical Laboratory), 760–891.
- Schuster, R. M. (1984b). "Morphology, phylogeny and classification of the Anthocerotae," in *New Manual of Bryology*, Vol. 2, ed. R. M. Schuster (Nichinan: Hattori Botanical Laboratory), 1071–1092.
- Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250, 931–936. doi: 10.1126/science.250.4983.931
- Sessa, E. B., Banks, J. A., Barker, M. S., Der, J. P., Duffy, A. M., Graham, S. W., et al. (2014). Between two fern genomes. *Gigascience* 3, 15. doi: 10.1186/2047-217X-3-15
- Shindo, S., Ito, M., Ueda, K., Kato, M., and Hasebe, M. (1999). Characterization of MADS genes in the gymnosperm *Gnetum parvifolium* and its implication on the evolution of reproductive organs in seed plants. *Evol. Dev.* 1, 180–190. doi: 10.1046/j.1525-142x.1999.99024.x
- Shubin, N., Tabin, C., and Carroll, S. (1997). Fossils, genes and the evolution of animal limbs. *Nature* 388, 639–648. doi: 10.1038/41710
- Singer, S. D., and Ashton, N. W. (2007). Revelation of ancestral roles of KNOX genes by a functional analysis of *Physcomitrella* homologues. *Plant Cell Rep.* 26, 2039–2054. doi: 10.1007/s00299-007-0312-0
- Singer, S. D., Krogan, N. T., and Ashton, N. W. (2007). Clues about the ancestral roles of plant MADS-box genes from a functional analysis of moss homologues. *Plant Cell Rep.* 26, 1155–1169. doi: 10.1007/s00299-007-0312-0
- Steeves, T. A., and Sussex, I. M. (1989). *Patterns in Plant Development*. New York, NY: Cambridge University Press.
- Stein, D. B. (1971). Gibberellin-induced fertility in the fern *Ceratopteris thalictroides* (L.) Brongn. *Plant Physiol.* 48, 416–418. doi: 10.1104/pp.48.4.416
- Strain, E., Hass, B., and Banks, J. A. (2001). Characterization of mutations that feminize gametophytes of the fern *Ceratopteris*. *Genetics* 159, 1271–1281.
- Sundström, J., Carlsbecker, A., Svensson, M. E., Svenson, M., Johanson, U., Theißen, G., et al. (1999). MADS-box genes active in developing pollen cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. *Dev. Genet.* 25, 253–266. doi: 10.1002/(SICI)1520-6408(1999)25:3<253::AID-DVG8>3.0.CO;2-P
- Sussex, I. M., and Kerk, N. M. (2001). The evolution of plant architecture. *Curr. Opin. Plant Biol.* 4, 33–37. doi: 10.1016/S1369-5266(00)00132-1
- Takahashi, N., Hashino, M., Kami, C., and Imaichi, R. (2009). Developmental morphology of strap-shaped gametophytes of *Colysis decurrens*: a new look at meristem development and function in fern gametophytes. *Ann. Bot.* 104, 1353–1361. doi: 10.1093/aob/mcp245
- Takahashi, N., Kami, C., Ota, I., Morita, N., and Imaichi, R. (2015). Developmental morphology of the typical cordate gametophyte of a homosporous fern, *Lygodium japonicum* (Lygodiaceae), focussing on the initial cell behaviour of two distinct meristems. *Am. J. Bot.* 102, 197–207. doi: 10.3732/ajb.1400333
- Tanabe, Y., Hasebe, M., Sekimoto, H., Nishiyama, T., Kitani, M., Henschel, K., et al. (2005). Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2436–2441. doi: 10.1073/pnas.0409860102
- Tanahashi, T., Sumikawa, N., Kato, M., and Hasebe, M. (2005). Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss *Physcomitrella patens*. *Development* 132, 1727–1736. doi: 10.1242/dev.01709
- Tandre, K., Albert, V. K., Sundås, A., and Engström, P. (1995). Conifer homologues to genes that control floral development in angiosperms. *Plant Mol. Biol.* 27, 69–78. doi: 10.1007/BF00019179
- Tandre, K., Svenson, M., Svensson, M. E., and Engström, P. (1998). Conservation of gene structure and activity in the regulation of reproductive organ development of conifers and angiosperms. *Plant J.* 15, 615–623. doi: 10.1046/j.1365-3113x.1998.00236.x
- Timmermans, M. C. P., Hudson, A., Becraft, P. W., and Nelson, T. (1999). ROUGH SHEATH2: a myb protein that represses knox homeobox gene in maize lateral organ primordia. *Science* 284, 151–153. doi: 10.1126/science.284.5411.151
- Tomescu, A. M. F. (2009). Megaphylls, microphylls and the evolution of leaf development. *Trends Plant Sci.* 14, 5–12. doi: 10.1016/j.tplants.2008.10.008
- Tomescu, A. M. F., Wyatt, S. E., Hasebe, M., and Rothwell, G. W. (2014). Early evolution of the vascular plant body plan—the missing mechanisms. *Curr. Opin. Plant Biol.* 17, 126–136. doi: 10.1016/j.pbi.2013.11.016
- Tsiantis, M., Schneeberger, R., Golz, J. F., Freeling, M., and Langdale, J. A. (1999). The maize rough sheath2 gene and leaf development programs in monocot and dicot plants. *Science* 284, 154–156. doi: 10.1126/science.284.5411.154
- Ueguchi-Tanaka, M., and Matsuoka, M. (2010). The perception of gibberellins: clues from receptor structure. *Curr. Opin. Plant Biol.* 13, 1–6. doi: 10.1016/j.pbi.2010.08.004
- van den Berg, C., Williamsen, V., Hage, W., Weisbeek, P., and Scheres, B. (1995). Cell fate in the *Arabidopsis* root meristem determined by directional signalling. *Nature* 378, 62–65. doi: 10.1038/378062a0
- van den Berg, C., Williamsen, V., Hendriks, G., Weisbeek, P., and Scheres, B. (1997). Short-range control of cell differentiation in the *Arabidopsis* root meristem. *Nature* 390, 287–289. doi: 10.1038/36856
- Vasco, A., Moran, R. C., and Ambrose, B. A. (2013). The evolution, morphology and development of fern leaves. *Front. Plant Sci.* 4:345. doi: 10.3389/fpls.2013.00345
- Vernoux, T., Brunoud, G., Farcot, E., Morin, V., Van den Daele, H., Legrand, J., et al. (2011). The auxin signalling network translates dynamic input into robust patterning at the shoot apex. *Mol. Syst. Biol.* 7, 508. doi: 10.1038/msb.2011.39
- Viaene, T., Landberg, K., Thelander, M., Medvecka, E., Pederson, E., Feraru, E., et al. (2014). Directional auxin transport mechanisms in early diverging land plants. *Curr. Biol.* 24, 2786–2791. doi: 10.1016/j.cub.2014.09.056
- Villarreal, J. C., and Renzaglia, K. S. (2015). The hornworts: important advancements in early plant evolution. *J. Bryol.* 37, 157–170. doi: 10.1179/1743282015Y.00000000016
- Vlad, D., Kierzkowski, D., Rast, M. I., Vuolo, F., Dello Ioio, R., Galinha, C., et al. (2014). Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. *Science* 343, 780–783. doi: 10.1126/science.1248384
- Vollbrecht, E., Reiser, L., and Hake, S. (2000). Shoot meristem size is dependent on inbred background and presence of the maize homeobox gene, knotted1. *Development* 127, 3161–3172.
- Wang, C., Liu, Y., Li, S. S., and Han, G. Z. (2015). Insights into the origin and evolution of the plant hormone signaling machinery. *Plant Physiol.* 167, 872–886. doi: 10.1104/pp.114.247403
- Wang, Y., Wang, C., Shi, B., Yu, T., Qi, J., Meyerowitz, E. M., et al. (2014). The stem cell niche in leaf axils is established by auxin and cytokinin in *Arabidopsis*. *Plant Cell* 26, 2055–2067. doi: 10.1105/tpc.114.123083
- Warne, T. R., and Hickok, L. G. (1991). Control of sexual development in gametophytes of *Ceratopteris richardii*: antheridiogen and abscisic acid. *Bot. Gaz.* 152, 148–153. doi: 10.1086/337874
- Watkins, J. E., and Cardelús, C. L. (2012). Ferns in an angiosperm world: cretaceous radiation into the epiphytic niche and diversification on the forest floor. *Int. J. Plant Sci.* 173, 695–710. doi: 10.1073/pnas.0811136106
- Weigel, D., Alvarez, J., Smyth, D. R., Yanofsky, M. F., and Meyerowitz, E. M. (1992). LEAFY controls floral meristem identity in *Arabidopsis*. *Cell* 69, 843–859. doi: 10.1016/0092-8674(92)90295-N
- White, R. A., and Turner, M. D. (1995). Anatomy and development of the fern sporophyte. *Bot. Rev.* 61, 281–305. doi: 10.1007/BF02912620
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4859–E4868. doi: 10.1073/pnas.1323926111
- Winter, K. U., Becker, A., Münster, T., Kim, J. T., Saedler, H., and Theissen, G. (1999). MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 96, 7342–7347. doi: 10.1073/pnas.96.13.7342
- Wodniok, S., Brinkmann, H., Glöckner, G., Heide, A. J., Philippe, H., Melkonian, M., et al. (2011). Origin of land plants: do conjugating green algae hold the key? *BMC Evol. Biol.* 11:104. doi: 10.1186/1471-2148-11-104

- Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., and Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13875–13879. doi: 10.1073/pnas.0811575106
- Yadav, R. K., Perales, M., Gruel, J., Girke, T., Jönsson, H., and Reddy, V. (2011). WUSCHEL protein movement mediates stem cell homeostasis in the *Arabidopsis* shoot apex. *Genes Dev.* 25, 2025–2030. doi: 10.1101/gad.17258511
- Yadegari, R., and Drews, G. N. (2004). Female gametophyte development. *Plant Cell* 16, S133–S141. doi: 10.1105/tpc.018192
- Yanai, O., Shani, E., Dolezal, K., Tarkowski, P., Sablowski, R., Sandberg, G., et al. (2005). *Arabidopsis* KNOX1 proteins activate cytokinin biosynthesis. *Curr. Biol.* 15, 1566–1571. doi: 10.1016/j.cub.2005.07.060
- Yasamura, Y., Crumpton-Taylor, M., Fuentes, S., and Harberd, N. P. (2007). Step-by-step acquisition of the gibberellin-DELLA growth-regulatory mechanism during land-plant evolution. *Curr. Biol.* 17, 1225–1230. doi: 10.1016/j.cub.2007.06.037
- Zhang, P., Tan, H. T. W., Pwee, K. H., and Kumar, P. P. (2004). Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. *Plant J.* 37, 566–577. doi: 10.1046/j.1365-3113X.2003.01983.x
- Zimmermann, W. (1952). Main results of the ‘Telome Theory’. *Palaeobotanist* 1, 456–470.
- Zobell, O., Faigl, W., Saedler, H., and Münster, T. (2010). MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. *Mol. Biol. Evol.* 27, 1201–1211. doi: 10.1093/molbev/msq005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Plackett, Di Stilio and Langdale. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Sequenced genomes and rapidly emerging technologies pave the way for conifer evolutionary developmental biology

Daniel Uddenberg¹, Shirin Akhter², Prashanth Ramachandran¹, Jens F. Sundström^{2*} and Annelie Carlsbecker^{1*}

¹ Physiological Botany, Department of Organismal Biology and Linnean Centre for Plant Biology, Uppsala BioCenter, Uppsala University, Uppsala, Sweden, ² Department of Plant Biology and Linnean Centre for Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences, Uppsala, Sweden

OPEN ACCESS

Edited by:

Rainer Melzer,
University College Dublin, Ireland

Reviewed by:

Lydia Gramzow,
Friedrich-Schiller-University Jena,
Germany

Francisco Vergara-Silva,
Universidad Nacional Autónoma de
México, Mexico

Giorgio Casadoro,
University of Padova, Italy

*Correspondence:

Annelie Carlsbecker
annelie.carlsbecker@ebc.uu.se;
Jens F. Sundström
jens.sundstrom@slu.se

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 31 July 2015

Accepted: 22 October 2015

Published: 03 November 2015

Citation:

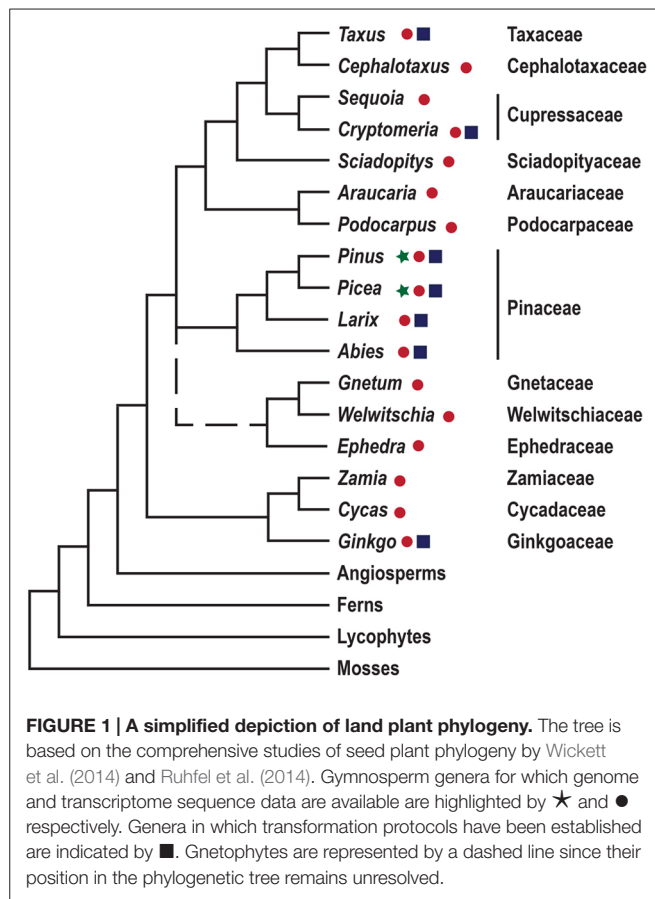
Uddenberg D, Akhter S,
Ramachandran P, Sundström JF
and Carlsbecker A (2015) Sequenced
genomes and rapidly emerging
technologies pave the way for conifer
evolutionary developmental biology.
Front. Plant Sci. 6:970.
doi: 10.3389/fpls.2015.00970

Conifers, Ginkgo, cycads and gnetophytes comprise the four groups of extant gymnosperms holding a unique position of sharing common ancestry with the angiosperms. Comparative studies of gymnosperms and angiosperms are the key to a better understanding of ancient seed plant morphologies, how they have shifted over evolution to shape modern day species, and how the genes governing these morphologies have evolved. However, conifers and other gymnosperms have been notoriously difficult to study due to their long generation times, inaccessibility to genetic experimentation and unavailable genome sequences. Now, with three draft genomes from spruces and pines, rapid advances in next generation sequencing methods for genome wide expression analyses, and enhanced methods for genetic transformation, we are much better equipped to address a number of key evolutionary questions relating to seed plant evolution. In this mini-review we highlight recent progress in conifer developmental biology relevant to evo-devo questions. We discuss how genome sequence data and novel techniques might allow us to explore genetic variation and naturally occurring conifer mutants, approaches to reduce long generation times to allow for genetic studies in conifers, and other potential upcoming research avenues utilizing current and emergent techniques. Results from developmental studies of conifers and other gymnosperms in comparison to those in angiosperms will provide information to trace core molecular developmental control tool kits of ancestral seed plants, but foremost they will greatly improve our understanding of the biology of conifers and other gymnosperms in their own right.

Keywords: gymnosperms, plant developmental biology, plant evo-devo, next-generation sequencing, plant transformation

CAN WE ESTABLISH A CONIFER MODEL SPECIES FOR DEVELOPMENTAL STUDIES?

Conifers are of great ecological and economic importance; they dominate the forests of the northern hemisphere, and comprise two thirds of extant gymnosperms (Wang and Ran, 2014). Seed plants, constituting gymnosperms and angiosperms, evolved 300–350 million years ago, and their appearance is defined by the evolution of the ovule. The subsequent evolution of seed plants



resulted in the elaboration of reproductive organ morphologies, including the innovation of the flower and carpel in the angiosperm lineage, but is also associated with, e.g., variations in embryo morphologies and water and assimilate conducting tissues (Taylor et al., 2009). To understand the evolution of novel morphologies we need to put these traits into a phylogenetic context. However, the deep branches of the seed plant phylogeny have been notoriously difficult to resolve, and the relative position of conifers, gnetophytes, cycads and *Ginkgo*, remains a focus for research (Figure 1; Wang and Ran, 2014). Over the last decades evolutionary developmental biology (evo-devo), has surfaced as an approach adding to traditional systematic efforts. Evo-devo studies rely on comparative analyses of the genetic mechanisms underlying the development of certain morphological traits, as exemplified by the evolution of reproductive structures in seed plants, see Mathews and Kramer (2012).

Currently, we have extensive knowledge on developmental genetic mechanisms mainly from a handful of angiosperm model species, primarily *Arabidopsis thaliana* (*Arabidopsis*). This is because ideal models typically are small, self-fertile, have short generation times, small genomes sizes and are amenable to genetic transformation (The *Arabidopsis* Genome Initiative, 2000). Gymnosperms, on the other hand, comprise long-lived perennial woody species with large population sizes, high degree of heterozygosity, and huge genomes sizes, and therefore lack model organism characteristics. The current

phylogenetically narrow focus on selected pines and spruces has instead largely been the result of geo-economical decisions. However, recent advances in molecular techniques have laid the foundation for a knowledge leap toward revealing the underlying genetic mechanisms controlling important traits also in species that lack the typical characteristics of model species.

NEXT GENERATION SEQUENCES AND GENETIC TRANSFORMATION—CONIFER DEVELOPMENTAL BIOLOGY STUDIES MADE POSSIBLE

The development of next-generation sequencing (NGS) techniques has surfaced as one of the most important technological breakthrough in current biology (Wang et al., 2009), making genomes and transcriptomes available from both model and non-model species. For non-model plants, such resources include draft sequences of the 20–30 Gigabase genomes from *Picea abies*, *P. glauca*, and *Pinus taeda* (Figure 1; Birol et al., 2013; Nystedt et al., 2013; Neale et al., 2014). These initiatives revealed that although the genomes are huge, largely owing to accumulation of long-terminal repeat transposable elements, the numbers of protein-coding sequences are similar to angiosperms. The draft conifer genome sequences and accompanying transcriptome data can be found in dedicated, constantly updated databases, aiming to help researchers navigate this vast amount of data^{1,2}. These data and corresponding databases will serve as an essential foundation for future studies.

The development and improvement of single-cell “omics” will probably drive the next advancement in genetic and transcriptomic research (Junker and van Oudenaarden, 2014), and, moreover, methods to retain positional information of the cell (*in situ* “omics”) promise to shed light also on the spatial regulation (Crosetto et al., 2015). Although technology development in this area still is in its early days and in large remains to be adapted to plants, this second avenue of NGS techniques will open up for more fine-tuned systems biology approaches, allowing computational and mathematical modeling of, e.g., transcription factor and signaling pathways.

Functional studies are crucial to test hypotheses of biochemical activity and forward genetic screens have therefore been imperative in identifying novel key developmental regulators in angiosperms. Previously, this relied on mapping using recombinant mapping populations, but NGS now allows sequencing of entire genomes, thus dramatically speeding up cloning of the causal mutation in model systems, and potentially making forward genetic screens possible in non-model systems (Schneeberger, 2014). Techniques that allow for NGS of particular genomic regions or transcribed loci, i.e., exome sequencing, may also help to overcome problems of genome complexity and SNP discovery in non-model systems (Neves et al., 2014).

Assessments of gene function require the generation of mutants or transgenes with altered gene activity, caused by knock-out,

¹<http://dendrome.ucdavis.edu/>

²<http://congenie.org/>

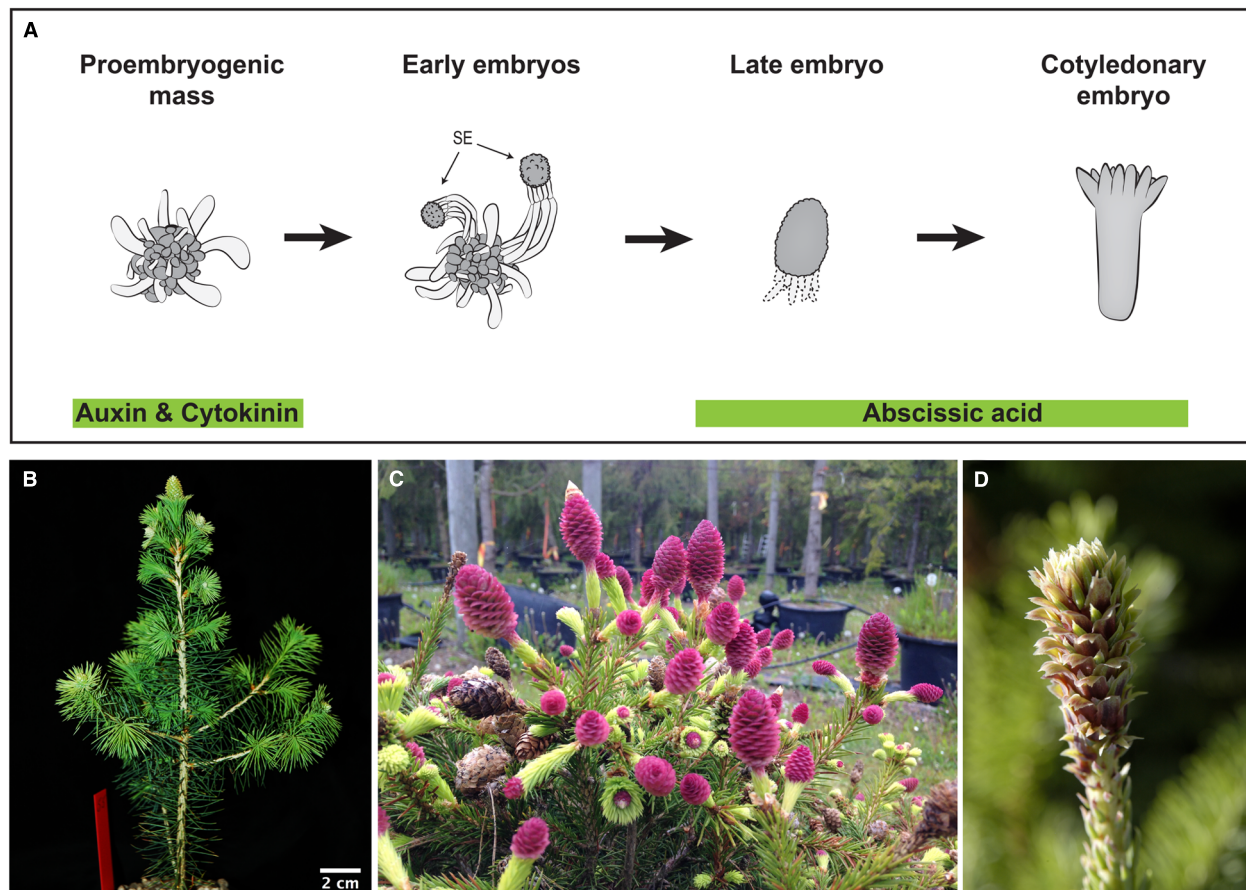


FIGURE 2 | Emerging tools for conifer functional studies: Embryogenic cultures and the rapid cycling *P. abies acrocona* mutant. (A) Schematic representation of somatic embryogenesis in *P. abies*: Embryogenic cultures are routinely established from zygotic embryo explants via the addition of auxin and cytokinin to culture media. Proliferating cultures consist of proembryogenic masses (PEMs), a mixture of densely cytoplasmic meristematic cells and large vacuolated cells. Withdrawal of auxin and cytokinin stimulates differentiation of early somatic embryos (SE) from PEMs. Early somatic embryos consist of apically located meristematic cells of the embryo proper, embryonal tube cells in the central region and a tier of terminally differentiated suspensor cells. Further embryo development and maturation requires the addition of abscissic acid. During late embryogeny apical meristems are formed and the suspensor cells undergo programmed cell death. Mature cotyledonary embryos are formed after 4–8 weeks on maturation medium. Image after Filonova et al. (2000). (B) Early cone setting in an inbred *acrocona* plant after three growth cycles (1 year) (Uddenberg et al., 2013). (C) Massive cone production in an older inbred *acrocona* plant. (D) *Acrocona* vegetative branch transitioning into female reproductive state. Needles gradually convert into bracts and ovuliferous scales appear in their axils. This transition is accompanied by the onset of a number of putative key reproductive developmental regulators (Carlsbecker et al., 2013).

knock-down, or over-expression of specific loci. In species with long life cycles such as gymnosperms genetic transformation over seed generations is not possible. However, this can be circumvented by utilizing somatic embryogenesis, in which proliferating embryogenic tissue is transformed by direct DNA delivery or via bacteria-mediated horizontal gene transfer. This method has been employed to generate transgenic conifer tissues for many years (Figure 2A; Tang and Newton, 2003). Some conifer and gymnosperm species are more recalcitrant to genetic transformation; however, the efficacy of transformation has greatly improved, mainly by using hypervirulent *Agrobacterium* strains and improved protocols, now facilitating the generation of stably transformed plants from many conifer species (Levee et al., 1997; Wenck et al., 1999; Klimaszewska et al., 2001; Le et al., 2001; Alvarez and Ordás, 2013). The use of embryo explants during transformation, followed by selective tissue culture and

plant regeneration, provide an alternative for recalcitrant species (Tang et al., 2014).

CONIFER SOMATIC EMBRYOS ENABLE FUNCTIONAL EVOLUTIONARY DEVELOPMENTAL BIOLOGY

Somatic embryogenesis is used in certain conifer species as a method for large-scale clonal propagation, facilitating long-term storage of germplasm, and as a tool in breeding programs. This technique also offers an efficient and versatile tool to study the morphology and underlying molecular regulation of conifer embryonal traits. Somatic embryo systems allow closer studies of the establishment of the plant basal body plan, including apical-basal specification, formation of the apical meristems and patterning of the dermal, ground and procambial tissues

(Smertenko and Bozhkov, 2014). Studies in *Arabidopsis* have shown that these processes depend on distinct spatio-temporal action of certain transcription factors and local biosynthesis and polar transport of the plant hormone auxin (Ten Hove et al., 2015). Studies on the effect of chemical inhibition of polar auxin transport during somatic embryo development of *P. abies* show increased levels of endogenous auxin, decreased programmed cell death (PCD) activity and abnormal suspensor differentiation during early stages of embryogenesis. Later stages treated with the chemical display both basal and apical aberrations, including fused cotyledons and unorganized meristems (Larsson et al., 2007; Hakman et al., 2009), suggesting a conserved role for auxin in basic embryo formation in seed plants. Comparative studies of homologs to angiosperm key factors for embryo patterning and polarity such as *WUSCHEL-RELATED HOMEODOMAIN* (*WOX*) genes and class I *KNOTTED1-like homeobox* (*KNOX1*) genes, using transgenic conifer somatic embryos, suggest considerable conservation but also functional divergence (Belmonte et al., 2007; Zhu et al., 2014; Alvarez et al., 2015). Potentially, Less biased methods such as global gene expression analyses during both somatic and zygotic embryogenesis indicate a significant overlap in transcript profiles of developmental regulators between conifers and angiosperms, but also reveal many genes of unknown function active during embryogenesis, emphasizing the need for future comparative functional studies (Vestman et al., 2011; de Vega-Bartol et al., 2013).

Interestingly, the first plant metacaspase involved in PCD was originally discovered in *P. abies*. Functional studies suppressing the type II metacaspase, *mCII-Pa*, in somatic embryos of *P. abies* showed that it is an essential component of vacuolar cell death, required for normal development and degradation of suspensors during early embryogenesis (Suárez et al., 2004) and that it acts via an autophagy-related pathway (Minina et al., 2013). Further studies, initiated in *P. abies* have also demonstrated that plant PCD share common genetic components with PCD in animals and humans (Sundström et al., 2009). Hence, conifer somatic embryogenesis provides an excellent system, not only for comparative studies, but also to identify novel regulators of general developmental processes.

GYMNOSPERM REPRODUCTIVE DEVELOPMENT THROUGH A GENOMIC LENS: ABC OR ONLY BC?

The evolution of the flower remains a major unresolved question in biology, since transition forms have not been reliably identified in the fossil record and extant gymnosperms are only distantly related to the angiosperms (Frohlich and Chase, 2007). While the angiosperms and gymnosperms are united by the feature of producing ovules, their reproductive organs are distinct: In contrast to the hermaphroditic angiosperm flower with the stamens and carpels surrounded by a sterile perianth of sepals and petals, the reproductive organs in gymnosperms are formed from separate meristems. Furthermore, the gymnosperm organs carrying the ovules have very distinct morphologies compared to the angiosperm carpel, preventing reliable inferences of organ homologies based on their morphology. However, despite

morphological diversity, evo-devo-studies show that molecular mechanisms controlling the development of the reproductive organs of angiosperms and gymnosperms are at least partially conserved (Melzer et al., 2010; Mathews and Kramer, 2012).

The identities of the flower organs are based on conserved key regulatory transcription factors, and were summarized in the ABC-model: A-function specify sepals, A together with B specifying petals, B and C specify stamens, whereas C alone specifies the carpel (Coen and Meyerowitz, 1991). The ABC-model was based on studies in *Arabidopsis* and has at least in part, been shown valid for most angiosperms, although the A-function have been assigned to floral meristem identity rather than to sepal identity in some angiosperms (Litt and Irish, 2003). Support for conserved molecular mechanisms for reproductive organ identity determinations among the seed plants came with the identification of putative orthologs to B- and C-genes in several gymnosperm species, along with the finding that both the B- and C-homologs are active specifically in developing male cones (Mouradov et al., 1999; Sundström et al., 1999; Winter et al., 1999) whereas C-function homologs also are active during the formation of the ovule bearing organs of the female cones (Tandre et al., 1995, 1998; Rutledge et al., 1998), leading to the hypothesis that B and C together specifies male reproductive identity, and C alone female reproductive identity in all seed plants.

Most gymnosperm female cones have a compound architecture, with ovule-bearing structures subtended by bracts, and neither female nor male cones have structures with apparent homology to the sterile perianth (sepals and petals). In line with this, PCR-based methods, aimed at identifying a broad range of MADS-box genes, failed to identify gymnosperm genes orthologous to the A-type MADS-box genes (Shindo et al., 1999; Winter et al., 1999; Carlsbecker et al., 2013). For a long time, it was considered an established fact that gymnosperms lacked both perianth-like organs and associated regulatory genes. In the first analysis of the *P. abies* genome, however, Nystedt et al. (2013) observed a remarkable expansion of the MADS-box gene family. Among the staggering 249 Type II MADS-box genes in the *P. abies* genome at least one gene group in a clade including both angiosperm A-function and FLOWERING LOCUS C-genes (Gramzow et al., 2014), calling for a reexamination of a potential A-function in conifers.

TEENS FOR DECADES—CAN WE OVERCOME THE LONG GENERATION TIME OF GYMNASPERMS TO FACILITATE DEVELOPMENTAL GENETIC STUDIES?

Gymnosperms are in general perennial trees, or shrubs, and most take decades until they enter the reproductive phase. Therefore, all functional evidence of any gene active in reproductive development in gymnosperms comes from testing their effect on the development of rapid cycling angiosperms (e.g., Tandre et al., 1998; Winter et al., 2002; Karlgren et al., 2011). Thus, there is a great need to better understand the molecular control of juvenile–adult and vegetative–reproductive transitions in gymnosperms, and if possible establish a more rapidly cycling

model (Uddenberg et al., 2013). Currently, most knowledge of developmental transitions comes from the annual plant *Arabidopsis*, although studies of perennial angiosperm trees, in particular poplar, promise to add important knowledge for comparative analyses with gymnosperms (Böhlenius et al., 2006; Wang et al., 2011). Not surprisingly, transitions may be controlled by distinct mechanisms in annuals and perennials, and in angiosperms and gymnosperms. In angiosperms, key regulators of the transition from vegetative to reproductive phase are orthologs of *FLOWERING LOCUS T (FT)* from *Arabidopsis* (Wigge et al., 2005). Although conifers possess *FT* homologs, studies in Norway spruce indicated that they lack *FT* orthologs (Karlgrén et al., 2011; Klintenäs et al., 2012), a notion confirmed with the sequencing of the spruce genome (Nystedt et al., 2013). Another conserved angiosperm key regulator acting upstream of the ABC-genes is *LEAFY (LFY)* (Moyroud et al., 2010). While gymnosperms do have an apparent *LFY* ortholog they also have a paralogous gene, called *NEEDLY* (Mellerowicz et al., 1998; Mouradov et al., 1998; Vazquez-Lobo et al., 2007). Currently available data is not informative to reveal if these genes may confer similar functions as their angiosperm counterpart.

Much of what we know about reproductive development in angiosperms is based on functional analysis of individual genes using mutants, either in forward or reverse genetic approaches. Interestingly, several varieties of conifers with peculiar reproductive structures or other phenotypes are available in arboreta (Rudall et al., 2011), and natural variants are alternatives to classic forward genetic screens (Dosmann and Groover, 2012). A naturally occurring mutant of *P. abies*, called *acrocona*, produces cones frequently, even in years when surrounding trees rarely set cones (**Figures 2B–D**). Inbred crosses show that a quarter of the segregating siblings initiate cones extremely early, already during their second growth season (**Figure 2B**; Uddenberg et al., 2013), and a single locus of importance for the cone setting phenotype has been mapped to a specific chromosome (Achery et al., 2004). Hence, the segregation pattern suggests that the early cone setting phenotype is caused by a monogenic loci and further analyses of its phenotype that it is likely semidominant (Uddenberg et al., 2013). NGS of *acrocona* transcriptomes (RNA-seq) identified a candidate gene related to the angiosperm floral integrator *SUPPRESSOR OF THE OVEREXPRESSION OF CONSTANS 1 (SOC 1)*; Lee and Lee, 2010) that may be involved in the early cone-setting phenotype (Uddenberg et al., 2013).

In addition to the early and frequent cone-setting, the *acrocona* mutant produces vegetative shoots transformed into reproductive cones, by initiation of ovuliferous scales in the axil of needles (**Figures 2C,D**). Detailed expression analysis using mRNA *in situ* hybridization have been used to study regulatory genes with putative functions in reproductive initiation, organ identity and pattern formation in wild type, male and female cones as well as in the *acrocona* transition shoots, as a means of testing of hypotheses of function by assessing gene expression correlation with the initiation and formation of ectopic female structures (Carlsbecker et al., 2013). Hence, already now, without knowing the nature of the causal mutation, *acrocona* allows further studies of putative reproductive development genes. These may include MADS-box

genes hypothesized to control phase transitions (Carlsbecker et al., 2003, 2004), or the newly identified putative A-class homolog (Gramzow et al., 2014). Like their angiosperm homologs (Litt and Irish, 2003), these genes may initiate reproduction in *P. abies*, and their activity can be analyzed in the mutant background. NGS of dissected tissues in various developmental phases in wild types and mutants will allow detailing such studies further.

FEEDING CONIFER DEVELOPMENTAL BIOLOGY INTO BREEDING PROGRAMS

In addition to the key evolutionary position occupied by the gymnosperms, a strong driving force to further our knowledge about their development, reflected in the two *Picea* and one *Pinus* species chosen for full genome sequences, is the economic importance of conifer wood. Wood is formed from the vascular meristem, the cambium. Although its activity is essential for all tree species, determining growth rate, wood formation and quality, it is among meristems the least understood. Most of our understanding of cambium activity and wood formation comes from studies of *Arabidopsis* and Poplar. These studies have revealed transcriptional and hormonal control mechanisms for cambium and wood formation as well as the biosynthesis pathways for cellulose, hemicellulose, and lignin (Lucas et al., 2013). Promisingly, comparative studies of genomes and transcriptomes have revealed a substantial conservation of regulatory mechanisms for cambium and wood formation between angiosperms and conifers (Li et al., 2010; Carvalho et al., 2013). Now, systems biology approaches will likely rapidly enhance our knowledge of conifer wood formation, beyond a mere comparison with more tractable angiosperm models. These approaches include co-expression analyses, transcription factor–promoter and protein–protein interaction analyses (Duval et al., 2014), in combination with assessments of transgenic seedlings with perturbed putative key regulators (Bomal et al., 2008). In addition, analyses of naturally occurring mutants, such as the cinnamyl alcohol dehydrogenase mutant defective in lignin formation (Ralph et al., 1997), will be important to connect wood properties and growth parameters. Knowledge gained could be used to generate computational models for vascular development, increase our understanding of specific features distinguishing angiosperm and gymnosperm secondary development and improve early stage identification of desirable traits important for breeding of economically important conifer species.

OUTLOOK

As costs for current sequencing methods decrease and third and fourth generation techniques such as nanopore sequencing are taken into general use (Feng et al., 2015), we can envision a more diversified sampling of sequenced organisms within the gymnosperm lineage, together with the assembly of high quality genomes. Better sequence information can be used to develop denser maps of short nucleotide polymorphisms (SNPs), enabling genome-wide marker-based selection and allowing more efficient breeding, as well as utilization of natural variation in studies

of developmental control mechanisms. Emerging quality updates on reference genomes will also most likely facilitate the establishment of methods such as CRISPR/Cas9 (Belhaj et al., 2015), greatly increasing the possibility to generate single and multiple mutants. A continued development of efficient techniques to generate inducible genes, the development of strong fluorescent reporters coupled with better detection techniques will likely revolutionize functional studies of at least early stages of conifer development.

The long generation time of most gymnosperms makes any attempt to perform functional studies of adult characters or even simple breeding efforts a time-consuming endeavor. Once the causing mutation of the early cone-setting phenotype of the *acrocona* mutant is known, it will be a potentially powerful tool to generate rapid cycling lines not only in *P. abies*, but perhaps also in other transformable conifers and gymnosperms. This

would enable functional studies of regulatory genes implicated in juvenile–adult transition, as well as reproductive initiations and reproduction organ specification. It may also allow the transfer of introduced traits from primary transformants to consecutive generations. Hence, new and emerging technologies promise a blooming future for conifer developmental biology, as well as for evo-devo studies in gymnosperms.

ACKNOWLEDGMENTS

Research on conifer developmental biology in the groups of Carlsbecker and Sundström is supported by grants from the Swedish Research Council FORMAS. We acknowledge that due to the condensed format of this mini-review not all original papers have been cited; when relevant we have referred to recent comprehensive reviews.

REFERENCES

- Achere, V., Faivre-Rampant, P., and Jeandroz, S. (2004). A full saturated linkage map of *Picea abies* including AFLP, SSR, ESTP, 5S rDNA and morphological markers. *Theor. Appl. Genet.* 108, 1602–1613. doi: 10.1007/s00122-004-1587-y
- Alvarez, J. M., and Ordás, R. J. (2013). Stable Agrobacterium-mediated transformation of maritime pine based on kanamycin selection. *ScientificWorldJournal* 2013, 681792. doi: 10.1155/2013/681792
- Alvarez, J. M., Sohlberg, J., Engström, P., Zhu, T., Englund, M., Moschou, P. N., et al. (2015). The *WUSCHEL-RELATED HOMEBOX 3* gene *PaWOX3* regulates lateral organ formation in Norway spruce. *New Phytol.* doi: 10.1111/nph.13536 [Epub ahead of print].
- Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N. J., and Nekrasov, V. (2015). Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* 32, 76–84. doi: 10.1016/j.copbio.2014.11.007
- Belmonte, M. F., Tahir, M., Schroeder, D., and Stasolla, C. (2007). Overexpression of HBK3, a class I KNOX homeobox gene, improves the development of Norway spruce (*Picea abies*) somatic embryos. *J. Exp. Bot.* 58, 2851–2861. doi: 10.1093/jxb/erm099
- Birol, I., Raymond, A., Jackman, S. D., Pleasance, S., Coope, R., Taylor, G. A., et al. (2013). Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics* 29, 1492–1497. doi: 10.1093/bioinformatics/btt178
- Bomal, C., Bedon, F., Caron, S., Mansfield, S. D., Levasseur, C., Cooke, J. E. K., et al. (2008). Involvement of *Pinus taeda* MYB1 and MYB8 in phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative in planta analysis. *J. Exp. Bot.* 59, 3925–3939. doi: 10.1093/jxb/ern234
- Böhlenius, H., Huang, T., Charbonnel-Campaa, L., Brunner, A. M., Jansson, S., Strauss, S. H., et al. (2006). CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* 312, 1040–1043. doi: 10.1126/science.1126038
- Carlsbecker, A., Sundström, J. F., Englund, M., Uddenberg, D., Izquierdo, L., Kvarnheden, A., et al. (2013). Molecular control of normal and *acrocona* mutant seed cone development in Norway spruce (*Picea abies*) and the evolution of conifer ovule-bearing organs. *New Phytol.* 200, 261–275. doi: 10.1111/nph.12360
- Carlsbecker, A., Sundström, J., Tandré, K., Englund, M., Kvarnheden, A., Johanson, U., et al. (2003). The DAL10 gene from Norway spruce (*Picea abies*) belongs to a potentially gymnosperm-specific subclass of MADS-box genes and is specifically active in seed cones and pollen cones. *Evol. Dev.* 5, 551–561. doi: 10.1046/j.1525-142X.2003.03060.x
- Carlsbecker, A., Tandré, K., Johanson, U., Englund, M., and Engström, P. (2004). The MADS-box gene DAL1 is a potential mediator of the juvenile-to-adult transition in Norway spruce (*Picea abies*). *Plant J.* 40, 546–557. doi: 10.1111/j.1365-3113X.2004.02226.x
- Carvalho, A., Paiva, J., Louzada, J., and Lima-Brito, J. (2013). The transcriptomics of secondary growth and wood formation in conifers. *Mol. Biol. Int.* 2013, 974324. doi: 10.1155/2013/974324
- Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls—genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0
- Crosetto, N., Bienko, M., and van Oudenaarden, A. (2015). Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* 16, 57–66. doi: 10.1038/nrg3832
- de Vega-Bartol, J. J., Simões, M., Lorenz, W. W., Rodrigues, A. S., Alba, R., Dean, J. F., et al. (2013). Transcriptomic analysis highlights epigenetic and transcriptional regulation during zygotic embryo development of *Pinus pinaster*. *BMC Plant Biol.* 13:123. doi: 10.1186/1471-2229-13-123
- Dosmann, M., and Groover, A. (2012). The importance of living botanical collections for plant biology and the “next generation” of evo-devo research. *Front. Plant Sci.* 3:137. doi: 10.3389/fpls.2012.00137
- Duval, L., Lachance, D., Giguère, I., Bomal, C., Morency, M.-J., Pelletier, G., et al. (2014). Large-scale screening of transcription factor–promoter interactions in spruce reveals a transcriptional network involved in vascular development. *J. Exp. Bot.* 65, 2319–2333. doi: 10.1093/jxb/eru116
- Feng, Y., Zhang, Y., Ying, C., Wang, D., and Du, C. (2015). Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics* 13, 4–16. doi: 10.1016/j.gpb.2015.01.009
- Filonova, L. H., Bozhkov, P. V., and Arnold, von, S. (2000). Developmental pathway of somatic embryogenesis in *Picea abies* as revealed by time-lapse tracking. *J. Exp. Bot.* 51, 249–264. doi: 10.1093/jexbot/51.343.249
- Frohlich, M. W., and Chase, M. W. (2007). After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450, 1184–1189. doi: 10.1038/nature06393
- Gramzow, L., Weilandt, L., and Theissen, G. (2014). MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Ann. Bot.* 114, 1407–1429. doi: 10.1093/aob/mcu066
- Hakman, I., Hallberg, H., and Palovaara, J. (2009). The polar auxin transport inhibitor NPA impairs embryo morphology and increases the expression of an auxin efflux facilitator protein PIN during *Picea abies* somatic embryo development. *Tree Physiol.* 29, 483–496. doi: 10.1093/treephys/tpn048
- Junker, J. P., and van Oudenaarden, A. (2014). Every cell is special: genome-wide studies add a new dimension to single-cell biology. *Cell* 157, 8–11. doi: 10.1016/j.cell.2014.02.010
- Karlgrén, A., Gyllenstrand, N., Kallman, T., Sundström, J. F., Moore, D., Lascoux, M., et al. (2011). Evolution of the PEBP gene family in plants: functional diversification in seed plant evolution. *Plant Physiol.* 156, 1967–1977. doi: 10.1104/pp.111.176206
- Klimaszewska, K., Lachance, D., Pelletier, G., Lelu, M. A., and Séguin, A. (2001). Regeneration of transgenic *Picea glauca*, *P. mariana*, and *P. abies* after cocultivation of embryogenic tissue with *Agrobacterium tumefaciens*. *In Vitro Cell. Dev. Biol. Plant* 37, 748–755. doi: 10.1007/s11627-001-0124-9
- Klintonäs, M., Pin, P. A., Benlloch, R., Ingvarsson, P. K., and Nilsson, O. (2012). Analysis of conifer FLOWERING LOCUS T/TERMINAL FLOWER1-like genes

- provides evidence for dramatic biochemical evolution in the angiosperm FT lineage. *New Phytol.* 196, 1260–1273. doi: 10.1111/j.1469-8137.2012.04332.x
- Larsson, E., Sitbon, F., Ljung, K., and Arnold, von, S. (2007). Inhibited polar auxin transport results in aberrant embryo development in Norway spruce. *New Phytol.* 177, 356–366. doi: 10.1111/j.1469-8137.2007.02289.x
- Lee, J., and Lee, I. (2010). Regulation and function of SOC1, a flowering pathway integrator. *J. Exp. Bot.* 61, 2247–2254. doi: 10.1093/jxb/erq098
- Levee, V., Lelu, M. A., Jouanin, L., Cornu, D., and Pilate, G. (1997). *Agrobacterium tumefaciens*-mediated transformation of hybrid larch (*Larix kaempferi* × *L. decidua*) and transgenic plant regeneration. *Plant Cell Rep.* 16, 680–685.
- Le, V. Q., Belles-Isles, J., Dusabenyagasani, M., and Tremblay, F. M. (2001). An improved procedure for production of white spruce (*Picea glauca*) transgenic plants using *Agrobacterium tumefaciens*. *J. Exp. Bot.* 52, 2089–2095. doi: 10.1093/jxb/52.364.2089
- Li, X., Wu, H. X., and Southerton, S. G. (2010). Comparative genomics reveals conservative evolution of the xylem transcriptome in vascular plants. *BMC Evol. Biol.* 10:190. doi: 10.1186/1471-2148-10-190
- Litt, A., and Irish, V. (2003). Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* 165, 821–833.
- Lucas, W. J., Groover, A., Lichtenberger, R., Furuta, K., Yadav, S. R., Helariutta, Y., et al. (2013). The plant vascular system: evolution, development and functions. *J. Int. Plant Biol.* 55, 294–388. doi: 10.1111/jipb.12041
- Mathews, S., and Kramer, E. M. (2012). The evolution of reproductive structures in seed plants: a re-examination based on insights from developmental genetics. *New Phytol.* 194, 910–923. doi: 10.1111/j.1469-8137.2012.04091.x
- Mellerowicz, E. J., Horgan, K., Walden, A., Coker, A., and Walter, C. (1998). PRFLL—a *Pinus radiata* homologue of FLORICAULA and LEAFY is expressed in buds containing vegetative shoot and undifferentiated male cone primordia. *Planta* 206, 619–629.
- Melzer, R., Wang, Y.-Q., and Theissen, G. (2010). The naked and the dead: the ABCs of gymnosperm reproduction and the origin of the angiosperm flower. *Semin. Cell Dev. Biol.* 21, 118–128. doi: 10.1016/j.semcdb.2009.11.015
- Minina, E. A., Filonova, L. H., Fukada, K., Savenkov, E. I., Gogvadze, V., Clapham, D., et al. (2013). Autophagy and metacaspase determine the mode of cell death in plants. *J. Cell Biol.* 203, 917–927. doi: 10.1083/jcb.201307082
- Mouradov, A., Glassick, T., Hamdorf, B., Murphy, L., Fowler, B., Marla, S., et al. (1998). NEEDLY, a *Pinus radiata* ortholog of FLORICAULA/LEAFY genes, expressed in both reproductive and vegetative meristems. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6537–6542.
- Mouradov, A., Hamdorf, B., Teasdale, R., Kim, J., Winter, K., and Theissen, G. (1999). A DEF/GLO-like MADS-Box gene from a gymnosperm: *pinus radiata* contains an ortholog angiosperm B class floral homeotic genes. *Dev. Genet.* 25, 245–252.
- Moyroud, E., Kusters, E., Monniaux, M., Koes, R., and Parcy, F. (2010). LEAFY blossoms. *Trends Plant Sci.* 15, 346–352. doi: 10.1016/j.tplants.2010.03.007
- Neale, D. B., Wegrzyn, J. L., Stevens, K. A., Zimin, A. V., Puiu, D., Crepeau, M. W., et al. (2014). Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol.* 15, R59. doi: 10.1186/gb-2014-15-3-r59
- Neves, L. G., Davis, J. M., Barbazuk, W. B., and Kirst, M. (2014). A high-density gene map of loblolly pine (*Pinus taeda* L.) based on exome sequence capture genotyping. *G3 (Bethesda)* 4, 29–37. doi: 10.1534/g3.113.008714
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolotto, A., Lin, Y.-C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584. doi: 10.1038/nature12211
- Ralph, J., MacKay, J. J., Hatfield, R. D., O'Malley, D. M., Whetten, R. W., and Sederoff, R. R. (1997). Abnormal lignin in a loblolly pine mutant. *Science* 277, 235–239.
- Rudall, P. J., Hilton, J., Vergara-Silva, F., and Bateman, R. M. (2011). Recurrent abnormalities in conifer cones and the evolutionary origins of flower-like structures. *Trends Plant Sci.* 16, 151–159. doi: 10.1016/j.tplants.2010.11.002
- Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E., and Burleigh, J. G. (2014). From algae to angiosperms-inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. *BMC Evol. Biol.* 14:23. doi: 10.1186/1471-2148-14-23
- Rutledge, R., Regan, S., Nicolas, O., Fobert, P., Côté, C., Bosnich, W., et al. (1998). Characterization of an AGAMOUS homologue from the conifer black spruce (*Picea mariana*) that produces floral homeotic conversions when expressed in *Arabidopsis*. *Plant J.* 15, 625–634.
- Schneeberger, K. (2014). Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat. Rev. Genet.* 15, 662–676. doi: 10.1038/nrg3745
- Shindo, S., Ito, M., Ueda, K., Kato, M., and Hasebe, M. (1999). Characterization of MADS genes in the gymnosperm *Gnetum parvifolium* and its implication on the evolution of reproductive organs in seed plants. *Evol. Dev.* 1, 180–190. doi: 10.1046/j.1525-142x.1999.99024.x
- Smertenko, A., and Bozhkov, P. V. (2014). Somatic embryogenesis: life and death processes during apical-basal patterning. *J. Exp. Bot.* 65, 1343–1360. doi: 10.1093/jxb/eru005
- Suárez, M. F., Filonova, L. H., Smertenko, A., Savenkov, E. I., Clapham, D. H., Arnold, von, S., et al. (2004). Metacaspase-dependent programmed cell death is essential for plant embryogenesis. *Curr. Biol.* 14, R339–R340. doi: 10.1016/j.cub.2004.04.019
- Sundström, J., Carlsbecker, A., Svensson, M., Svenson, M., Johanson, U., Theissen, G., et al. (1999). MADS-box genes active in developing pollen cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. *Dev. Genet.* 25, 253–266.
- Sundström, J. F., Vaculova, A., Smertenko, A. P., Savenkov, E. I., Golovko, A., Minina, E., et al. (2009). Tudor staphylococcal nuclease is an evolutionarily conserved component of the programmed cell death degradome. *Nat. Cell Biol.* 11, 1347–1354. doi: 10.1038/ncb1979
- Tandre, K., Albert, V., and Sundås, A. (1995). Conifer homologues to genes that control floral development in angiosperms. *Plant Mol. Biol.* 27, 69–78.
- Tandre, K., Svenson, M., Svensson, M., and Engström, P. (1998). Conservation of gene structure and activity in the regulation of reproductive organ development of conifers and angiosperms. *Plant J.* 15, 615–623.
- Tang, W., and Newton, R. J. (2003). Genetic transformation of conifers and its application in forest biotechnology. *Plant Cell Rep.* 22, 1–15. doi: 10.1007/s00299-003-0670-1
- Tang, W., Xiao, B., and Fei, Y. (2014). Slash pine genetic transformation through embryo cocultivation with *A. tumefaciens* and transgenic plant regeneration. *In Vitro Cell. Dev. Biol. Plant* 50, 199–209. doi: 10.1007/s11627-013-9551-7
- Taylor, T. N., Taylor, E. L., and Krings, M. (2009). Paleobotany: The Biology and Evolution of Fossil Plants, 2nd Edn. San Diego, CA: Academic Press.
- Ten Hove, C. A., Lu, K.-J., and Weijers, D. (2015). Building a plant: cell fate specification in the early *Arabidopsis* embryo. *Development* 142, 420–430. doi: 10.1242/dev.111500
- The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692
- Uddenberg, D., Reimegård, J., Clapham, D., Almqvist, C., Arnold, von, S., Emanuelsson, O., et al. (2013). Early cone setting in *Picea abies* acrocona is associated with increased transcriptional activity of a MADS box transcription factor. *Plant Physiol.* 161, 813–823. doi: 10.1104/pp.112.207746
- Vazquez-Lobo, A., Carlsbecker, A., Vergara-Silva, F., Alvarez-Buylla, E. R., Pinero, D., and Engström, P. (2007). Characterization of the expression patterns of LEAFY/FLORICAULA and NEEDLY orthologs in female and male cones of the conifer genera *Picea*, *Podocarpus*, and *Taxus*: implications for current evo-devo hypotheses for gymnosperms. *Evol. Dev.* 9, 446–459. doi: 10.1111/j.1525-142X.2007.00182.x
- Vestman, D., Larsson, E., Uddenberg, D., Cairney, J., Clapham, D., Sundberg, E., et al. (2011). Important processes during differentiation and early development of somatic embryos of Norway spruce as revealed by changes in global gene expression. *Tree Genet. Genomes* 7, 347–362. doi: 10.1007/s11295-010-0336-4
- Wang, J.-W., Park, M. Y., Wang, L.-J., Koo, Y., Chen, X.-Y., Weigel, D., et al. (2011). miRNA control of vegetative phase change in trees. *PLoS Genet.* 7:e1002012. doi: 10.1371/journal.pgen.1002012
- Wang, X.-Q., and Ran, J.-H. (2014). Evolution and biogeography of gymnosperms. *Mol. Phylogenet. Evol.* 75, 24–40. doi: 10.1016/j.ympev.2014.02.005
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484
- Wenck, A. R., Quinn, M., Whetten, R. W., Pullman, G., and Sederoff, R. (1999). High-efficiency *Agrobacterium*-mediated transformation of Norway spruce (*Picea abies*) and loblolly pine (*Pinus taeda*). *Plant Mol. Biol.* 39, 407–416. doi: 10.1023/A:1006126609534

- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4859–E4868. doi: 10.1073/pnas.1323926111
- Wigge, P. A., Kim, M. C., Jaeger, K. E., Busch, W., Schmid, M., Lohmann, J. U., et al. (2005). Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science* 309, 1056–1059. doi:10.1126/science.1114358
- Winter, K. U., Becker, A., Munster, T., Kim, J. T., Saedler, H., and Theissen, G. (1999). MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 96, 7342–7347.
- Winter, K.-U., Saedler, H., and 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*. *Plant J.* 31, 457–475. doi: 10.1046/j.1365-313X.2002.01375.x
- Zhu, T., Moschou, P. N., Alvarez, J. M., Sohlberg, J. J., and Arnold, von, S. (2014). Wuschel-related homeobox 8/9 is important for proper embryo patterning in the gymnosperm Norway spruce. *J. Exp. Bot.* 65, 6543–6552. doi: 10.1093/jxb/eru371

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Uddenberg, Akhter, Ramachandran, Sundström and Carlsbecker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Petunia, Your Next Supermodel?

Michiel Vandenbussche*, Pierre Chambrier, Suzanne Rodrigues Bento and Patrice Morel

Laboratoire de Reproduction et Développement des Plantes, UMR5667 CNRS, INRA, ENS Lyon, Université de Lyon, Lyon, France

OPEN ACCESS

Edited by:

Rainer Melzer,
University College Dublin, Ireland

Reviewed by:

Bart Jan Janssen,
New Zealand Institute for Plant &
Food Research, New Zealand
Marcelo Carnier Domelas,
Universidade Estadual de Campinas,
Brazil

Ronald Koes,
University of Amsterdam, Netherlands

*Correspondence:

Michiel Vandenbussche
michiel.vandenbussche@ens-lyon.fr

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 15 September 2015

Accepted: 15 January 2016

Published: 02 February 2016

Citation:

Vandenbussche M, Chambrier P,
Rodrigues Bento S and Morel P
(2016) Petunia, Your Next
Supermodel? *Front. Plant Sci.* 7:72.
doi: 10.3389/fpls.2016.00072

Plant biology in general, and plant evo–devo in particular would strongly benefit from a broader range of available model systems. In recent years, technological advances have facilitated the analysis and comparison of individual gene functions in multiple species, representing now a fairly wide taxonomic range of the plant kingdom. Because genes are embedded in gene networks, studying evolution of gene function ultimately should be put in the context of studying the evolution of entire gene networks, since changes in the function of a single gene will normally go together with further changes in its network environment. For this reason, plant comparative biology/evo–devo will require the availability of a defined set of ‘super’ models occupying key taxonomic positions, in which performing gene functional analysis and testing genetic interactions ideally is as straightforward as, e.g., in *Arabidopsis*. Here we review why petunia has the potential to become one of these future supermodels, as a representative of the Asterid clade. We will first detail its intrinsic qualities as a model system. Next, we highlight how the revolution in sequencing technologies will now finally allows exploitation of the petunia system to its full potential, despite that petunia has already a long history as a model in plant molecular biology and genetics. We conclude with a series of arguments in favor of a more diversified multi-model approach in plant biology, and we point out where the petunia model system may further play a role, based on its biological features and molecular toolkit.

Keywords: petunia, model system, genome sequence, transposon mutagenesis, functional genomics, evolution, plants, evo–devo

INTRODUCTION

Since the beginning of the 1990s, there has been an extreme focus in plant molecular biology on one particular model system, the small weed *Arabidopsis thaliana*. *Arabidopsis* offers a combination of characteristics that made it in many ways the perfect model to study plant biology. Besides its obvious advantages as a laboratory model, one of the others was certainly the fact that *Arabidopsis* has a very small genome compared to many other plant species. This was a very important issue in the pre-Next Generation Sequencing (NGS) era, since the sequencing of even a small genome like that of *Arabidopsis* represented at that time a multi-million dollar investment and several years of large scale collaborative efforts (Arabidopsis Genome Initiative, 2000). Moreover, its easy transformation method (Clough and Bent, 1998), and the generation of a large scale functional genomics platform (Alonso et al., 2003) further accelerated the steep rise of *Arabidopsis* to become the gold standard in plant biology.

For many reasons (Scutt and Vandenbussche, 2014) it is obvious that most if not all research subjects studied in plant molecular biology would benefit from the availability of a broader range

of experimental systems. With the advent of NGS technology, the sequencing of entire plant genomes and transcriptomes has become technically and financially feasible even for individual research teams, removing one of the major obstacles in the development of new model systems. Consequently, nowadays a wealth of genome sequence data is becoming available, sampled from species throughout plant phylogeny. While a lot can be learned based on the analysis of genomes and transcriptomes alone, the ultimate understanding of the molecular basis of a biological process and its evolutionary origin still relies on classic determination of gene function by loss- and/or gain-of-function approaches. Therefore, the development of gene functional analysis tools for an array of species will be crucial to fully exploit this novel goldmine of sequence information. In recent years, considerable progress has been made with the development of Virus Induced Gene Silencing (VIGS; Burch-Smith et al., 2004; Senthil-Kumar and Mysore, 2011), and TILLING (Wang et al., 2012), allowing comparison of individual gene functions across a broad species pallet of the plant kingdom. Moreover, the recently developed CRISPR/Cas9 technology is creating a revolution in gene functional analysis (Lozano-Juste and Cutler, 2014). The results of such comparative functional studies will form the basis for the understanding of the molecular origin of the diversity of life, one of the most fundamental questions in plant biology and life sciences in general. However, individual genes do not act in an isolated fashion, but are embedded in gene networks. Therefore, studying evolution of gene function ultimately should be put in the context of studying the evolution of the entire gene network, as changes in the function of one gene will in many cases go together with changes in its network environment. Inspired by the famous essay of T. Dobzhansky (Nothing in biology makes sense except in the light of evolution), along the same line one could state that “Comparative gene functional analysis does not make sense except in the light of gene network evolution”. For this reason, novel comparative bio-informatics and systems biology approaches should become progressively more embedded in resolving evolutionary questions. In addition, plant comparative biology/evo-devo will require the availability of a defined set of ‘super’ models occupying key taxonomic positions, in which performing gene functional analysis and testing genetic interactions is as straightforward as, e.g., in *Arabidopsis*, allowing ease of study and comparison of the function of all members of entire gene networks and their genetic interactions.

Here we review why petunia has the potential to become one of these future supermodels. We will first detail its intrinsic qualities as a laboratory model system. Next, we highlight why it is only now that the benefits of the petunia system will become fully exploitable, despite that petunia has already a long history as a model in plant molecular biology and genetics (Gerats and Vandenbussche, 2005). We conclude with a series of arguments in favor of a more diversified multi-model approach in plant biology, and we point out where the petunia model system may further play a role, based on its biological features and (future) molecular genetics toolkit.

PETUNIA: LAB-MODEL CHARACTERISTICS

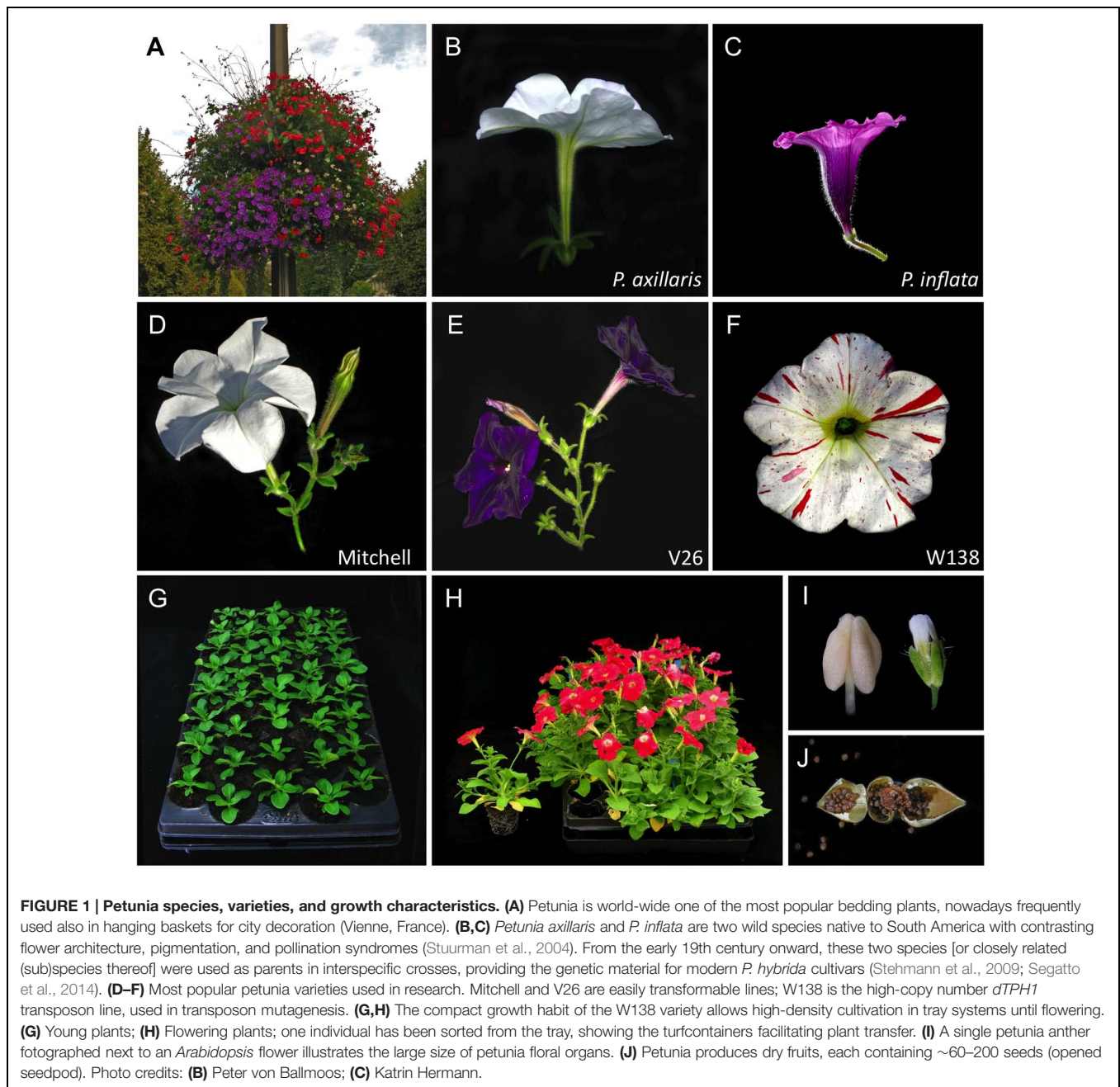
The cultivated garden petunia, with its big colorful flowers and diverse morphology is worldwide one of the most popular bedding flowers (**Figure 1A**). The genus *Petunia*, established as a genus by Jussieu in 1803, originates from South America, and belongs to the family of the Solanaceae (Stehmann et al., 2009). Commercial petunia cultivars as well as the standard laboratory lines have a hybrid origin (therefore called *Petunia hybrida*). Although there has been some discussion on the exact origin of *P. hybrida*, it is generally accepted that crosses in the early 19th century between the white hawkmoth-pollinated *P. axillaris* species (**Figure 1B**) and a member(s) of the purple bee-pollinated *P. integrifolia* group (containing a small number of closely related species, including *P. inflata*; **Figure 1C**) created the basis of the selection material from which all modern *P. hybrida* cultivars are derived (Stehmann et al., 2009; Segatto et al., 2014). Note that because species barriers in the *Petunia* genus are mainly prezygotic, *Petunia* species can be perfectly crossed with each other and yield normal diploid offspring. Thus, *P. hybrida* varieties have the same 2n chromosome numbers as the parental species, and therefore do not suffer from associated genetic complications found in hybrids that are (allo)tetraploid. The most popular petunia lines used in research are V26 and Mitchell, both renowned for their high transformation capacity, and W138, the high-copy number *dTPH1* transposon line used for transposon mutagenesis (**Figures 1D–F**). Petunia laboratory lines display a number of qualities that make petunia ideally suited as a plant model.

A Short Lifecycle and Easy Culture Conditions

In optimal circumstances, a lifecycle from seed to seed of only 3–3.5 months can easily be obtained, allowing the growth of up to four generations a year. It has been shown for petunia that the rate in progress to flowering increased linearly with increasing temperatures (Adams et al., 1998). In practice, we found that lower temperatures have a double impact on generation time, since they not only delay flowering, but also strongly delay ripening of the seedpods. To reach such a short generation time, we culture petunia at relatively high temperatures (e.g., 27–30°C daytime; 23–25°C night time; long day conditions: 15–16 h light/day). Note that the specific growth conditions mentioned above are usually not applied in petunia horticultural production, where energy cost considerations and the aim to obtain a specific plant architecture demand for different growth parameters.

Space usually is a limiting factor in laboratory growth chambers. In contrast to the Mitchell line, which grows very tall and has an inflorescence characterized by very long internodes, the W138 transposon line exhibits a more compact growth habit. It is therefore particularly well suited to cultivation in growth chambers, and can be easily grown in high-density trays until flowering (**Figures 1G,H**). We plant petunia seedlings in trays¹

¹<http://www.jiffygroup.com/>



(55 cm × 32 cm) containing 40 plants, using turf containers that allow easy repotting of selected individuals later on. Once flowering, we transfer plants of interest in individual containers (0,5 l) and regularly cut side-branches to favor vertical growth, again to save place. With a regular fertilization, plants can in this way easily be maintained for a long period (1–2 years), and may be cut back regularly, while they continue to flower. Petunia grows well in growth chambers, standard green houses, and during spring and summer also outside in simple plastic greenhouse tunnels without any artificial light or temperature control. The latter provides an extremely cheap and feasible solution in case large populations need to be grown, such as for

forward genetics screens, and for populations for reverse genetics purposes.

Easy Propagation, Both Sexual and Asexual

In the wild, petunia depends on animals (bees, hawkmoths, humming birds, depending on the species) for pollination (Stuurman et al., 2004). As a consequence, petunia plants normally do not set seed spontaneously in growth chambers or greenhouses devoid of insects. However, the large petunia flowers and floral organs (Figure 1I) make manual pollination (either selfing or crossing) extremely easy. Pollinating a flower

requires only a few seconds, each time resulting in a capsule, from which 3–4 weeks later ~60–200 seeds can be harvested (**Figure 1J**). Asexual propagation by cuttings (massively used in the petunia horticulture) is also very straightforward. In research, this is particularly useful since it allows the creation of stocks of identical genetic material that can be challenged simultaneously under different conditions. It also allows individual plants of interest to be maintained indefinitely without the need of resowing. Petunia can also be easily grafted (Napoli, 1996), creating a powerful tool to study long distance signaling. Furthermore, propagation by callus culture or plant regeneration starting from leaf explants or protoplasts is also possible.

Stable/Transient Transformation and Biochemical Analysis

Petunia was among the first plant species that were successfully used to create stable transgenic plants (Horsch et al., 1985). Petunia is classically transformed using a leaf-disk transformation protocol, and a defined set of varieties exists that are particularly easy to transform, such as Mitchell and V26, and the wild species *P. axillaris*. Moreover, any F1 hybrid derived from crossing different petunia varieties displays superior transformation capacity. The latter is routinely applied to transform mutants that arose in W138 (see further), since the pure W138 line is very recalcitrant to transformation.

Agrobacterium infiltration in tobacco leaves is widely used in transient assays (Yang et al., 2000). This technique works fine in petunia leaves as well. In addition, due to its large flowers, petals can also be used for infiltration assays (Verweij et al., 2008). Furthermore, an efficient protocol has been developed for the isolation and transformation of protoplasts derived from petals (Faraco et al., 2011). As in tobacco, VIGS works very efficiently in petunia (Chen et al., 2004; Broderick and Jones, 2014). In plant species that are not possible to transform, VIGS technology often offers the only possible way for gene functional analysis. While VIGS in petunia might offer a rapid way to identify interesting phenotypes, the existing alternatives for gene functional analysis (stable transformation; transposon insertion mutagenesis) might be preferred as a final proof of function. Finally, because of its large leaves and flowers, petunia is particularly suited for biochemical analysis, which often requires large quantities of plant material.

The Petunia *dTPH1* Transposable Element System in the W138 Line: A Powerful Tool for Forward and Reverse Genetics Approaches

Insertion mutagenesis remains one of the methods of choice to obtain mutants in genes of interest or in hitherto unknown genes. The cloning of the petunia *dTPH1* transposable element (Gerats et al., 1990, 2013) opened the way for insertion mutagenesis approaches in petunia. Interestingly, since this is a completely natural mutagenesis system, the obtained mutants are non-transgenic and therefore their use is not constrained by GMO rules. It turned out that the biology of the *dTPH1* system in the petunia W138 line is extremely well suited for forward and

reverse genetics approaches (Koes et al., 1995; Van den Broeck et al., 1998; Vandenbussche et al., 2003, 2008, 2013). The petunia *dTPH1* element is a non-autonomous *hAT*-like transposon that induces a target site duplication of 8 bp upon integration (**Figure 2A**). Thanks to the small size of the *dTPH1* element (284 bp), genotyping *dTPH1* insertions for segregation analyses is extremely straightforward, and is done in one single PCR reaction using a gene-specific primer pair flanking the insertion site, followed by agarose gel electrophoresis (**Figure 2B**). Note that in practice (partial) excision of the *dTPH1* transposon potentially may complicate the interpretation of PCR genotyping results. Especially homozygous mutants in which partial excision occurs may be wrongly considered as heterozygous plants, since the excision allele is close to WT size. However, this is less of a problem when using 4% agarose gels and choosing segregation primers that generate a WT fragment between ~90 and 120 bp, resulting in a resolution power that in the majority of the cases clearly distinguishes excision footprints from true WT fragments (**Figure 2B**). Excision of the *dTPH1* transposon from an insertion site and subsequent repair may have a variable outcome, resulting in different classes of excision alleles (van Houwelingen et al., 1999). Depending on the desired effect, this might be exploited in two different ways (**Figure 2A**). Firstly, excision alleles can be selected that cause an out-of-frame mutation in the reading frame, and thus result in a fully stabilized mutant allele. Secondly, although more rarely, excision alleles may be produced that result in the restoration of the reading frame, and possibly gene function. Note that forward phenotypic screening for revertants possibly may require the growing of a large number of plants. Such a revertant analysis associated with the characterization of footprint size can be used to further proof the causal relationship between a *dTPH1* insertion and a phenotype (**Figure 2A**). Note that for insertions in coding sequences in particular, in practice it may often not be needed to stabilize the insertion, since excision events that lead to restoration of gene function are relatively rare. Likewise, the risk of accidentally losing an identified mutation is low.

The *dTPH1* element was isolated from the inbred line W138 (Doodeman et al., 1984), which produces high numbers of new mutations each generation. Interestingly, it was found that the large majority of these new mutations in W138 were caused by *dTPH1* insertions (van Houwelingen et al., 1998; Spelt et al., 2000), despite the presence of other types of transposable elements in petunia. This enormously facilitates the cloning of mutated genes by transposon tagging, since the nature of the mutation (insertion of a *dTPH1* element) can be assumed a priori. For such forward genetics approaches, the *dTPH1* transposon display technique (**Figure 2C**; Van den Broeck et al., 1998; Vandenbussche et al., 2013), or related approaches have been developed, and have led to the identification of many interesting novel genes (e.g., Stuurman et al., 2002; Tobena-Santamaria et al., 2002; Quattrocchio et al., 2006; Cartolano et al., 2007; Rebocho et al., 2008; Verweij et al., 2008; Vandenbussche et al., 2009; Rich et al., 2015).

Furthermore, *dTPH1* inserts preferentially in genic regions and in the W138 line, up to 20–40 novel insertions may arise per individual plant and per generation (Koes et al.,

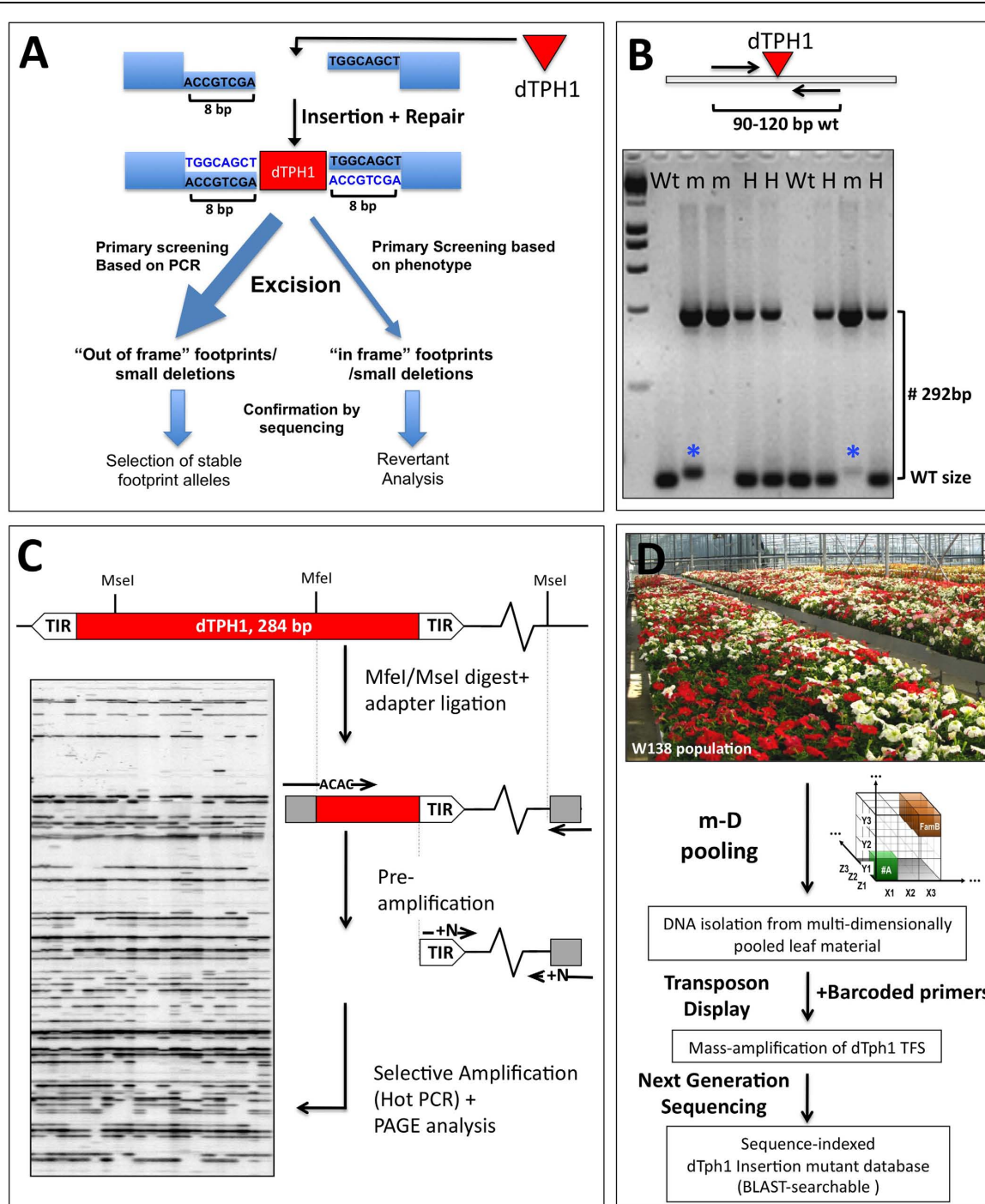


FIGURE 2 | Petunia *dTPH1* transposon biology and mutagenesis. (A) Petunia *dTPH1* creates a 8 bp target site duplication upon insertion. Excision of the *dTPH1* transposon from an insertion site and subsequent repair may have a variable outcome, resulting in different classes of excision alleles (van Houwelingen et al., 1999). For insertions in the coding sequence, excision will result in an out of frame mutation in most cases, while more rarely, this may result in restoration of gene function (revertant), in case the remaining footprint/deletion is a multiple of three. Note that revertant analysis can be used as a strategy to provide independent proof of gene function. **(B)** The small size of the *dTPH1* transposon allows easy genotyping in a single PCR reaction. A 4% agarose gel is shown loaded with PCR products resulting from amplification with a gene-specific primer pair flanking the insertion site. Fragments containing the *dTPH1* element are 292 bp larger than the WT fragment (284 bp of the transposon + 8 bp target site duplication). Genotypes are indicated Wt = homozygous wild-type; H = heterozygous; m = homozygous mutant. The blue asterisks indicate partial excision of the transposon, resulting in a fragment slightly larger than WT. **(C)** *dTPH1* Transposon Display procedure, used in the forward cloning of *dTPH1* tagged mutants in petunia forward genetics. See Van den Broeck et al. (1998) and Vandenbussche et al. (2013) for more details. **(D)** Procedure for the massive parallel amplification and sequencing of *dTPH1* flanking sequences, derived from populations of 1000–4000 individuals. The resulting sequence database can be BLAST-searched for reverse genetics purposes. For experimental details, see Vandenbussche et al. (2008).

1995; Vandenbussche et al., 2003). These two characteristics together make the W138 line, besides its application in forward genetics, also extremely well suited for reverse genetics mutagenesis: Despite the large (~1.3 GB) genome size of *Petunia* (Arumuganathan and Earle, 1991), relatively small mutant populations may be sufficient to saturate the genome with genic insertions.

Since in 1990s, W138 populations (varying from 1000 to 4000 plants) have been regularly grown by a handful of petunia groups, providing a source of *dTPH1* mutants for the community, both for forward and reverse genetics screens. For years, reverse genetics screenings of these populations were performed by PCR on a gene-per-gene basis, making the whole procedure slow and labor-intensive (Koes et al., 1995; Vandenbussche et al., 2003). Meanwhile in *Arabidopsis*, a completely different approach was being developed: Insertion sites of a large collection of T-DNA lines were systematically characterized at the sequence level. The resulting publicly available collections of insertion site-sequenced T-DNA lines (Alonso et al., 2003) revolutionized reverse-genetics approaches, since databases with insertion site flanking sequences can be *in silico* searched for mutants of interest based on gene ID or sequence homology, instead of having to perform laborious PCR-based assays. However, because the large scale Sanger sequencing of insertion flanking sequences is very costly, such an approach was the exclusive domain of model organisms financially supported by a large scientific community, such as *Arabidopsis*. From 2005 onward, the first generation of massive parallel sequencing methods started to become available (Margulies et al., 2005), creating a true paradigm shift in molecular biology by bringing large scale sequencing projects financially within reach of individual research teams.

Based on the early GS20 (454) sequencing technology (Margulies et al., 2005), we developed a concept (**Figure 2D**) that allows to mass amplification and sequencing of *dTPH1* transposon flanking sequences (TFS) simultaneously from an entire population, and that permits automatic assignment of TFS to individuals within the same population (Vandenbussche et al., 2008). With this approach, we were able to identify and sequence around 10000 different *dTPH1* insertion loci simultaneously amplified from a population of 1000 individuals. While these results were certainly encouraging and provided for the first time a small blast-searchable mutant collection for petunia, the high costs of GS20 sequencing and limited sequencing capacity were still constraining a large-scale application needed to saturate the genome with *dTPH1* insertions.

PETUNIOMICS: PETUNIA EMBRACES GENOMICS

Creation of a Large *dTPH1* Transposon Flanking Sequence Database for Reverse Genetics in Petunia

Since their conception, sequencing capacity of high-throughput sequencing methods has increased exponentially, combined with

a steep drop in costs. In particular the Illumina sequencing technology (Bentley et al., 2008) proved to be well adapted to further develop our mass *dTPH1* TFS sequencing principle, leading to a method with unprecedented efficiency, accuracy, and capacity. We are currently preparing a manuscript detailing the protocol and the resulting *dTPH1* TFS collection (Morel et al., unpublished). Analysis of the new *dTPH1* transposon flanking sequence database indicates a good coverage of the petunia genome with *dTPH1* insertions, since we are able to identify (usually multiple) candidate insertions for the large majority of the genes screened for. This collection will revolutionize functional genomics in petunia, in the same way as the SALK collection has revolutionized *Arabidopsis* research. Besides the obvious benefit for research teams using petunia as a model system, this mutant collection can also be of interest for the petunia horticultural industry. Many valuable traits (affecting growth habit, plant architecture, floral architecture) can be obtained by loss-of-function approaches. The non-transgenic mutants identified in our collection can be directly used for crossing with commercial petunia varieties.

The *Petunia* Genome Sequence(s)

For years, petunia molecular biology research has been slowed down by the unavailability of a sequenced genome. The small size of the petunia scientific community combined with the large size of the *Petunia* genome (~1.3 GB) rendered a genome sequencing project based on classical Sanger sequencing completely out of reach. As for our insertion mutagenesis program, the advent of NGS technologies suddenly made a petunia genome project feasible. A few years ago, members of the Petunia Platform (see further) joined forces to launch a petunia genome sequencing initiative, in collaboration with BGI (Beijing Genomics Institute, China). To cover the complete gene content of all petunia cultivars, the petunia genome sequencing initiative chose to sequence the genomes of the parental species *P. axillaris* and *P. inflata* (see Petunia: Lab-Model Characteristics), rather than sequencing a few of the many existing hybrids. Sequencing of both *Petunia* species is now finished, and a manuscript is currently being finalized (The Petunia Genome Consortium, in preparation). Consequently, public release of the *Petunia* genome sequences may be expected in the near future. Finally, a number of petunia teams are currently performing a detailed RNAseq-based characterization of the petunia transcriptome in a variety of tissues and processes, which will greatly enhance annotation quality of the genome. Some examples were recently published (Broderick et al., 2014; Villarino et al., 2014; Guo et al., 2015), but many more studies are to be expected.

IMPORTANCE OF DEVELOPING AND MAINTAINING A BROAD RANGE OF PLANT MODELS, AND THE POTENTIAL ROLE OF PETUNIA

While the impact of *Arabidopsis* research on plant biology cannot be overestimated, plant biology will strongly benefit from the

development and maintenance of a broader range of plant models. This applies even for research subjects that have been already heavily studied in *Arabidopsis*. Below, we mention five arguments in favor of a more diversified multi-model approach in plant biology. While the first two arguments are so obvious that they do not need further explanation, the last three arguments might be less trivial. Because of the focus of this paper, we provide further support for the last three arguments with specific examples coming from petunia research, but obviously many other examples may be found based on research in other models.

- (1) A major challenge in (plant) biology remains to understand the molecular basis of evolution and resulting diversity of lifeforms. It is more than obvious that this question only can be efficiently tackled by comparing gene functions in a diverse set of model species occupying key taxonomic positions and/or exhibiting key evolutionary novelties.
- (2) There is an extensive list of biological processes that are too divergent between *Arabidopsis* and other species or that simply do not occur in *Arabidopsis*, and thus require other plant models to analyze. Good examples are fleshy fruit development (tomato), flower color (petunia), mycorrhization, and nodulation (Medicago, Lotus, petunia), leaf and flower development in grasses (rice, maize, *Brachypodium*), floral asymmetry (Antirrhinum), and tuber development (potato) just to name a few.
- (3) Plant genomes have a complex origin, shaped by whole and partial genome duplication events that occurred during evolution of the plant lineage, and that were followed by further local rearrangements and duplications. This has been first demonstrated at the whole genome scale in *Arabidopsis* (Arabidopsis Genome Initiative, 2000), and has been confirmed in later sequenced plant genomes. These duplication events have often led to (partial) genetic redundancy. Consequently, many gene functions cannot simply be uncovered by classical forward genetics screens. This is probably the main reason why even now in *Arabidopsis*, despite having been intensively studied for more than two decades, the majority of the genes still remain to be functionally characterized. Note that genetic redundancy in many cases can be efficiently tackled through reverse genetics strategies, but it requires much more effort, and without the guarantee that the result will answer the specific biological question a researcher is interested in. Fortunately, gene duplications and all of the processes that have given a plant genome its current gene content are completely random in nature, meaning that distantly related species will most likely display very different sets of redundant and unique gene functions. Therefore, repeating identical forward genetics screens in distantly related species has the strong potential to uncover completely novel functions that may have remained hidden in redundancy in other species. Good illustrations for this from petunia research are the cases of the *HAM* (*HAIRY MERISTEM*) gene, a promotor of shoot indeterminacy (Stuurman et al., 2002), the auxin biosynthetic gene *FLOOZY* (Tobena-Santamaria et al., 2002), and *MAW* (*MAEWEST*), a *WOX* gene involved in blade development (Vandenbussche et al., 2009). These genes were all identified based on single mutant phenotypes in forward genetics screens, while similar phenotypes in *Arabidopsis* were only obtained after creating double or higher order mutants (Cheng et al., 2006; Vandenbussche et al., 2009; Engstrom et al., 2011).
- (4) Similar to duplications, gene losses may follow independent patterns in different plant lineages. Consequently, some model species may have lost a number of genes that play important roles in other species. An interesting case is at the heart of the ABC model of floral development, of which the corresponding B-class mutants *ap3* (*apetala3*) in *Arabidopsis* (Jack et al., 1992) and *def* (*deficiens*) in snapdragon (Sommer et al., 1990) display homeotic conversions of petals into sepals and stamens into carpels. Remarkably, null mutants for the petunia ortholog *PhDEF* (also known as *GREEN PETALS, GP*) display only a homeotic conversion of petals into sepals, while stamen development remains unaffected (van der Krol et al., 1993; Vandenbussche et al., 2004). The molecular basis of this one whorled phenotype turned out to be caused not by a difference in the function of *PhDEF* itself, but by the presence of the ancestral B-class gene *TM6* in petunia genome (Rijpkema et al., 2006). *TM6* genes are found in many plant species, but have been lost both in *Arabidopsis* and snapdragon.
- (5) Even though the end result of a biological process might be very similar in different species, the underlying regulatory gene networks may have diverged more than generally thought. Some very clear examples of network divergence between *Arabidopsis* and petunia can be found in the mechanisms that pattern homeotic gene expression in the flower (Cartolano et al., 2007), and in the regulation of floral meristem identity genes (Kusters et al., 2015). Thus, even if clear orthologs for all individual components of a regulatory gene network identified in one species can be found back in another species, it does not guarantee that a similar network architecture will control the same biological process in the two species. A better knowledge of the evolution (and divergence) of gene networks may strongly improve the success rate of translating basic knowledge from model organisms to applications in crop species. An improved understanding of gene network evolution could come from the functional comparison of all members of entire gene networks and their genetic interactions in a diverse set of species. Such an approach requires the availability of several 'super' models, in which performing gene functional analysis and testing genetic interactions ideally would be as straightforward as in *Arabidopsis*.

While petunia was already renowned as a very convenient plant model, its upcoming large sequence-indexed mutant collection and genome sequence will enormously facilitate gene function analysis at a large scale. Ongoing experiments in our lab involving the comparative functional analysis of ~30 key floral regulators selected from the *Arabidopsis* gene network indicate that comparative analyses of large regulatory gene networks are indeed feasible in petunia: We succeeded

to identify null mutations for >90% of the genes, usually obtaining multiple insertion alleles per gene. Moreover, the straightforward genotyping, crossing and short generation time allow to easily test genetic interactions between mutants: we now regularly obtain double, triple, quadruple and even quintuple mutants.

Arabidopsis and petunia belong to the Rosids and Asterids, respectively, which represent the two major groups within the eudicot species, and are thought to have diverged approximately 100 million years ago (Moore et al., 2010). Together with the input from other models, the characterization of gene regulatory networks in petunia and the comparison with *Arabidopsis* will therefore help to reveal the degree of gene network divergence within the higher eudicots.

Petunia belongs to the Solanaceae, which harbors several species that are major (food) crops (potato, tomato, pepper, eggplant, tobacco), while others are mainly cultivated as ornamentals (e.g., petunia, Calibrachoa, Datura, Schizanthus, and many others; Saerkinen et al., 2013). Moreover, some of these food crops have been developed into highly performing plant models, such as tomato (Tomato Genome Consortium, 2012) and potato (Xu et al., 2011). Petunia may be an excellent comparative genetic model to understand the molecular basis and origin of some aspects of the developmental diversity in this family of major agronomical importance. For example,

the advanced molecular genetics toolkits available both in petunia and tomato could help to elucidate the molecular mechanisms that determine the difference between dry (petunia) and fleshy (tomato) fruit development (Pabon-Mora and Litt, 2011).

A BROAD RANGE OF INTERESTING COMPARATIVE AND UNIQUE RESEARCH TOPICS IN PETUNIA

Nowadays, with its advanced molecular genetics toolkit, petunia is a very attractive model to study a number of subjects (reviewed in the book “Petunia: Evolutionary, Developmental and Physiological Genetics”; Gerats and Strommer, 2009), some of which are difficult to analyze in other species. Reasons for this may be either technical (e.g., other possible species not amenable to reverse and forward genetics) or biological (e.g., a developmental process not occurring in other species).

Petunia development differs in many interesting ways with *Arabidopsis*, which forms an excellent basis for comparative studies, and evo-devo oriented research. The most eye-catching differences are (1) its cymose inflorescence architecture (Kusters et al., 2015), compared to a raceme in *Arabidopsis*; (2) its large, fused and brightly colored petals ideally suited to study flower

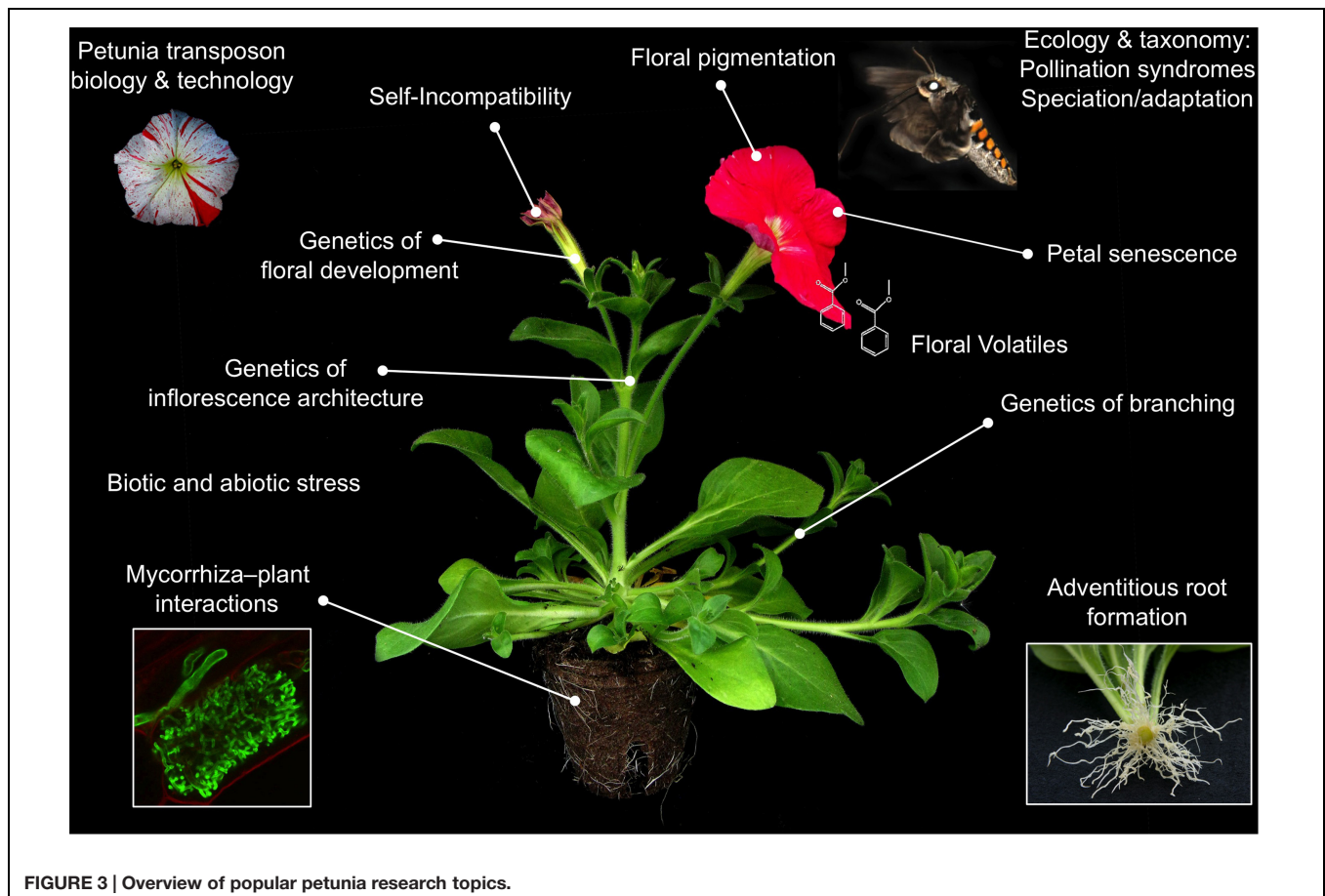


FIGURE 3 | Overview of popular petunia research topics.

color (Faraco et al., 2014) and sympetaly (Vandenbussche et al., 2009); and (3) its central placentation topology (Colombo et al., 2008) compared to parietal placentation in *Arabidopsis*. Other interesting differences include the presence of a gametophytic self-incompatibility system (Williams et al., 2015), the existence of different pollination syndromes (Hermann and Kuhlemeier, 2011), its abundant and clock regulated floral volatile production (Verdonk et al., 2003, 2005; Fenske et al., 2015), and being a suited host for mycorrhiza colonization (Rich et al., 2015). In addition, petunia exhibits a more diverse range in forms and ecological niches (Stehmann et al., 2009). Interestingly, since species barriers in petunia are mainly prezygotic, genetic analysis above the ecotype level is much more straightforward compared to *Arabidopsis* (Nasrallah et al., 2000), and thus allows the integration most of the available molecular tools into ecological studies. Today, research in petunia covers a broad range of topics, many of which have an impact beyond the model system. While a detailed review of the current petunia scientific literature goes beyond the scope of this manuscript, we provide a non-exhaustive list of the most popular subjects in **Figure 3**, which may form a basis for further literature searches. An important part of the research groups working with petunia worldwide are associated with the Petunia Platform², a community driven platform that aims to promote petunia research and to facilitate collaborations among its members. More information on the research performed in petunia can be found on the website, where keyword descriptions of the research of each group are presented, together with links to their respective websites. An ideal introduction into the petunia scientific community is participating in the “World Petunia days” (WPDs), a scientific congress organized every 18 months. Seminar topics traditionally cover molecular mechanisms controlling inflorescence architecture and flower development (floral architecture and floral organ identity), petal senescence, floral scent production, gametophytic self-incompatibility, flower pigmentation, evolution of pollination syndromes, root development (adventitious root formation, mycorrhiza interactions), and petunia genomics, but also new subjects are warmly welcomed. Traditionally, the WPDs have a friendly and informal character, stimulating scientists to exchange ideas, materials, techniques, and unpublished data without any inhibition. Organization of the WPDs is done by volunteering members of the Petunia Platform. Place, date, and organizer of the next edition are announced on the Petunia Platform website.

² <http://flower.ens-lyon.fr/PetuniaPlatform/PetuniaPlatform.html>

REFERENCES

- Adams, S. R., Hadley, P., and Pearson, S. (1998). The effects of temperature, photoperiod, and photosynthetic photon flux on the time to flowering of *Petunia* ‘Express Blush Pink’. *J. Am. Soc. Hortic. Sci.* 123, 577–580.
- Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657. doi: 10.1126/science.1086391

CONCLUSION

For several decades already, petunia has been successfully explored as a model system in plant molecular biology by a relatively small but productive scientific community. Petunia displays a number of characteristics that combined have contributed to its survival as a model system, during a period in which many experimental plant systems were abandoned mainly in favor of *Arabidopsis*. These characteristics include a short generation time, an easy growth habit, its endogenous highly active transposon system with a strong potential for forward and reverse genetics, an easy transformation protocol and an amenity for biochemical analysis because of its large leaves and flowers. Yet, despite all these advantages, the absence of a genome sequence and the lack of a functional genomics platform equivalent to the *Arabidopsis* SALK collection made working with petunia sometimes feel like driving a F1 racecar without engine. Thanks to NGS technology, both the petunia genome sequence and a large functional genomics platform will become available in the near future. This will finally allow exploitation of the petunia system to its full potential. As a representative of the Asterids, it may be a powerful model to compare the function of entire gene regulatory networks with those of, e.g., *Arabidopsis*, a representative of the Rosids. Belonging to the Solanaceae, petunia may serve as a comparative genetic model in the exploration of the molecular origin of some of the developmental diversity in this family of major agronomical importance. In addition, petunia is expected to further excel in specific research topics that are hard to address in other models.

AUTHOR CONTRIBUTIONS

MV wrote the manuscript; PM, PC, and SB contributed to experiments referred to in the manuscript, and commented on the manuscript.

ACKNOWLEDGMENTS

Our team was financially supported by a CNRS/ATIP AVENIR award and currently by the Agence Nationale de Recherche (ANR BLANC). Peter von Ballmoos, Katrin Hermann, Cris Kuhlemeier, Didier Reinhardt, and the Leibniz Institute of Vegetable and Ornamental Crops, are acknowledged for image contributions to **Figures 1 and 3**.

- Arabidopsis* Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692
- Arumuganathan, K., and Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9, 208–218. doi: 10.1007/BF02672069
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59. doi: 10.1038/nature07517

- Broderick, S. R., and Jones, M. L. (2014). An optimized protocol to increase virus-induced gene silencing efficiency and minimize viral symptoms in *Petunia*. *Plant Mol. Biol. Rep.* 32, 219–233. doi: 10.1007/s11105-013-0647-3
- Broderick, S. R., Wijeratne, S., Wijeratn, A. J., Chapin, L. J., Meulia, T., and Jones, M. L. (2014). RNA-sequencing reveals early, dynamic transcriptome changes in the corollas of pollinated *Petunias*. *BMC Plant Biol.* 14:307. doi: 10.1186/s12870-014-0307-2
- Burch-Smith, T. M., Anderson, J. C., Martin, G. B., and Dinesh-Kumar, S. P. (2004). Applications and advantages of virus-induced gene silencing for gene function studies in plants. *Plant J.* 39, 734–746. doi: 10.1111/j.1365-313X.2004.02158.x
- Cartolano, M., Castillo, R., Efremova, N., Kuckenberg, M., Zethof, J., Gerats, T., et al. (2007). A conserved microRNA module exerts homeotic control over *Petunia hybrida* and *Antirrhinum majus* floral organ identity. *Nat. Genet.* 39, 901–905. doi: 10.1038/ng2056
- Chen, J. C., Jiang, C. Z., Gookin, T. E., Hunter, D. A., Clark, D. G., and Reid, M. S. (2004). Chalcone synthase as a reporter in virus-induced gene silencing studies of flower senescence. *Plant Mol. Biol.* 55, 521–530. doi: 10.1007/s11103-004-0590-7
- Cheng, Y., Dai, X., and Zhao, Y. (2006). Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in *Arabidopsis*. *Genes Dev.* 20, 1790–1799. doi: 10.1101/gad.1415106
- Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x
- Colombo, L., Battaglia, R., and Kater, M. M. (2008). *Arabidopsis* ovule development and its evolutionary conservation. *Trends Plant Sci.* 13, 444–450. doi: 10.1016/j.tplants.2008.04.011
- Doodeman, M., Boersma, E. A., Koomen, W., and Bianchi, F. (1984). Genetic analysis of instability in *Petunia hybrida*: a highly unstable mutation induced by a transposable element in the An1 locus for flower colour. *Theor. Appl. Genet.* 67, 345–355. doi: 10.1007/BF00272873
- Engstrom, E. M., Andersen, C. M., Gumalak-Smith, J., Hu, J., Orlova, E., Sozzani, R., et al. (2011). *Arabidopsis* homologs of the *Petunia* hairy meristem gene are required for maintenance of shoot and root indeterminacy. *Plant Physiol.* 155, 735–750. doi: 10.1104/pp.110.168757
- Faraco, M., Di Sansebastiano, G. P., Spelt, K., Koes, R. E., and Quattrocchio, F. M. (2011). One protoplast is not the other! *Plant Physiol.* 156, 474–478. doi: 10.1104/pp.111.173708
- Faraco, M., Spelt, C., Blik, M., Verweij, W., Hoshino, A., Espen, L., et al. (2014). Hyperacidification of vacuoles by the combined action of two different P-ATPases in the tonoplast determines flower color. *Cell Rep.* 6, 32–43. doi: 10.1016/j.celrep.2013.12.009
- Fenske, M. P., Hewett-Hazelton, K. D., Hempton, A. K., Shim, J. S., Yamamoto, B. M., Riffell, J. A., et al. (2015). Circadian clock gene LATE ELONGATED HYPOCOTYL directly regulates the timing of floral scent emission in *Petunia*. *Proc. Natl. Acad. Sci. U.S.A.* 112, 9775–9780. doi: 10.1073/pnas.1422875112
- Gerats, A. G., Huits, H., Vrijlandt, E., Marana, C., Souer, E., and Beld, M. (1990). Molecular characterization of a nonautonomous transposable element (dTph1) of *Petunia*. *Plant Cell* 2, 1121–1128. doi: 10.2307/3869264
- Gerats, T., and Strommer, J. (2009). “*Petunia*: evolutionary, developmental and physiological genetics,” in *Petunia*, eds T. Gerats and J. Strommer (New York, NY: Springer), 1–433.
- Gerats, T., and Vandenbussche, M. (2005). A model system for comparative research: *Petunia*. *Trends Plant Sci.* 10, 251–256. doi: 10.1016/j.tplants.2005.03.005
- Gerats, T., Zethof, J., and Vandenbussche, M. (2013). Identification and applications of the *Petunia* class II Act1/dTph1 transposable element system. *Methods Mol. Biol.* 1057, 223–237. doi: 10.1007/978-1-62703-568-2_16
- Guo, Y., Wiegert-Rininger, K. E., Vallejo, V. A., Barry, C. S., and Warner, R. M. (2015). Transcriptome-enabled marker discovery and mapping of plastochron-related genes in *Petunia* spp. *BMC Genomics* 16:726. doi: 10.1186/s12864-015-1931-4
- Hermann, K., and Kuhlmeier, C. (2011). The genetic architecture of natural variation in flower morphology. *Curr. Opin. Plant Biol.* 14, 60–65. doi: 10.1016/j.pbi.2010.09.012
- Horsch, R. B., Fry, J. E., Hoffmann, N. L., Eichholtz, D., Rogers, S. G., and Fraley, R. T. (1985). A simple and general-method for transferring genes into plants. *Science* 227, 1229–1231. doi: 10.1126/science.227.4691.1229
- Jack, T., Brockman, L. L., and Meyerowitz, E. M. (1992). The homeotic gene APETALA3 of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* 68, 683–697. doi: 10.1016/0092-8674(92)90144-2
- Koes, R., Souer, E., van Houwelingen, A., Mur, L., Spelt, C., Quattrocchio, F., et al. (1995). Targeted gene inactivation in *Petunia* by PCR-based selection of transposon insertion mutants. *Proc. Natl. Acad. Sci. U.S.A.* 92, 8149–8153. doi: 10.1073/pnas.92.18.8149
- Kusters, E., Della Pina, S., Castel, R., Souer, E., and Koes, R. (2015). Changes in cis-regulatory elements of a key floral regulator are associated with divergence of inflorescence architectures. *Development* 142, 2822–2831. doi: 10.1242/dev.121905
- Lozano-Juste, J., and Cutler, S. R. (2014). Plant genome engineering in full bloom. *Trends Plant Sci.* 19, 284–287. doi: 10.1016/j.tplants.2014.02.014
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.
- Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4623–4628. doi: 10.1073/pnas.0907801107
- Napoli, C. (1996). Highly branched phenotype of the *Petunia* dad1-1 mutant is reversed by grafting. *Plant Physiol.* 111, 27–37.
- Nasrallah, M. E., Yogeewaran, K., Snyder, S., and Nasrallah, J. B. (2000). *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiol.* 124, 1605–1614. doi: 10.1104/pp.124.4.1605
- Pabon-Mora, N., and Litt, A. (2011). Comparative anatomical and developmental analysis of dry and fleshy fruits of Solanaceae. *Am. J. Bot.* 98, 1415–1436. doi: 10.3732/ajb.1100097
- Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J., and Koes, R. (2006). PH4 of *Petunia* is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell* 18, 1274–1291. doi: 10.1105/tpc.105.034041
- Rebocho, A. B., Blik, M., Kusters, E., Castel, R., Procissi, A., Roobeek, I., et al. (2008). Role of EVERGREEN in the development of the cymose *Petunia* inflorescence. *Dev. Cell* 15, 437–447. doi: 10.1016/j.devcel.2008.08.007
- Rich, M. K., Schorderet, M., Bapaume, L., Falquet, L., Morel, P., Vandenbussche, M., et al. (2015). The *Petunia* GRAS transcription factor ATA/RAM1 regulates symbiotic gene expression and fungal morphogenesis in arbuscular mycorrhiza. *Plant Physiol.* 168, 788–797. doi: 10.1104/pp.15.00310
- Rijkema, A. S., Royaert, S., Zethof, J., van der Weerden, G., Gerats, T., and Vandenbussche, M. (2006). Analysis of the *Petunia* TM6 MADS box gene reveals functional divergence within the DEF/AP3 lineage. *Plant Cell* 18, 1819–1832. doi: 10.1105/tpc.106.042937
- Saerkinen, T., Bohs, L., Olmstead, R. G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol. Biol.* 13:214. doi: 10.1186/1471-2148-13-214
- Scutt, C. P., and Vandenbussche, M. (2014). Current trends and future directions in flower development research. *Ann. Bot.* 114, 1399–1406. doi: 10.1093/aob/mcu224
- Segatto, A. L., Ramos-Fregonezi, A. M., Bonatto, S. L., and Freitas, L. B. (2014). Molecular insights into the purple-flowered ancestor of garden *Petunias*. *Am. J. Bot.* 101, 119–127. doi: 10.3732/ajb.1300223
- Senthil-Kumar, M., and Mysore, K. S. (2011). New dimensions for VIGS in plant functional genomics. *Trends Plant Sci.* 16, 656–665. doi: 10.1016/j.tplants.2011.08.006
- Sommer, H., Beltran, J. P., Huijser, P., Pape, H., Lonnig, W. E., Saedler, H., et al. (1990). Deficiens, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J.* 9, 605–613.
- Spelt, C., Quattrocchio, F., Mol, J. N., and Koes, R. (2000). anthocyanin1 of *Petunia* encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. *Plant Cell* 12, 1619–1632. doi: 10.1105/tpc.12.9.1619
- Stehmann, J. O., Lorenz-Lemke, A., Freitas, L., and Semir, J. O. (2009). “The genus *Petunia*: Evolutionary, Developmental and Physiological Genetics,” eds T. Gerats and J. Strommer (New York, NY: Springer), 1–28.

- Stuurman, J., Hoballah, M. E., Broger, L., Moore, J., Basten, C., and Kuhlemeier, C. (2004). Dissection of floral pollination syndromes in *Petunia*. *Genetics* 168, 1585–1599. doi: 10.1534/genetics.104.031138
- Stuurman, J., Jaggi, F., and Kuhlemeier, C. (2002). Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. *Genes Dev.* 16, 2213–2218. doi: 10.1101/gad.230702
- Tobena-Santamaria, R., Bliet, M., Ljung, K., Sandberg, G., Mol, J. N., Souer, E., et al. (2002). FLOOZY of *Petunia* is a flavin mono-oxygenase-like protein required for the specification of leaf and flower architecture. *Genes Dev.* 16, 753–763. doi: 10.1101/gad.219502
- Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641. doi: 10.1038/nature11119
- Van den Broeck, D., Maes, T., Sauer, M., Zethof, J., De Keuleleire, P., D'Hauw, M., et al. (1998). Transposon display identifies individual transposable elements in high copy number lines. *Plant J.* 13, 121–129.
- Vandenbussche, M., Horstman, A., Zethof, J., Koes, R., Rijpkema, A. S., and Gerats, T. (2009). Differential recruitment of WOX transcription factors for lateral development and organ fusion in *Petunia* and *Arabidopsis*. *Plant Cell* 21, 2269–2283. doi: 10.1105/tpc.109.065862
- Vandenbussche, M., Janssen, A., Zethof, J., van Orsouw, N., Peters, J., van Eijk, M. J., et al. (2008). Generation of a 3D indexed *Petunia* insertion database for reverse genetics. *Plant J.* 54, 1105–1114. doi: 10.1111/j.1365-313X.2008.03482.x
- Vandenbussche, M., Zethof, J., and Gerats, T. (2013). Transposon display: a versatile method for transposon tagging. *Methods Mol. Biol.* 1057, 239–250. doi: 10.1007/978-1-62703-568-2_17
- Vandenbussche, M., Zethof, J., Royaert, S., Weterings, K., and Gerats, T. (2004). The duplicated B-class heterodimer model: whorl-specific effects and complex genetic interactions in *Petunia hybrida* flower development. *Plant Cell* 16, 741–754. doi: 10.1105/tpc.019166
- Vandenbussche, M., Zethof, J., Souer, E., Koes, R., Tornielli, G. B., Pezzotti, M., et al. (2003). Toward the analysis of the *Petunia* MADS box gene family by reverse and forward transposon insertion mutagenesis approaches: B, C, and D floral organ identity functions require SEPALLATA-like MADS box genes in *Petunia*. *Plant Cell* 15, 2680–2693. doi: 10.1105/tpc.017376
- van der Krol, A. R., Brunelle, A., Tsuchimoto, S., and Chua, N. H. (1993). Functional analysis of *Petunia* floral homeotic MADS box gene pMADS1. *Genes Dev.* 7, 1214–1228. doi: 10.1101/gad.7.7a.1214
- van Houwelingen, A., Souer, E., Mol, J., and Koes, R. (1999). Epigenetic interactions among three dTph1 transposons in two homologous chromosomes activate a new excision-repair mechanism in *Petunia*. *Plant Cell* 11, 1319–1336. doi: 10.2307/3870752
- van Houwelingen, A., Souer, E., Spelt, K., Kloos, D., Mol, J., and Koes, R. (1998). Analysis of flower pigmentation mutants generated by random transposon mutagenesis in *Petunia hybrida*. *Plant J.* 13, 39–50.
- Verdonk, J. C., Haring, M. A., van Tunen, A. J., and Schuurink, R. C. (2005). ODORANT1 regulates fragrance biosynthesis in *Petunia* flowers. *Plant Cell* 17, 1612–1624. doi: 10.1105/tpc.104.028837
- Verdonk, J. C., Ric de Vos, C. H., Verhoeven, H. A., Haring, M. A., van Tunen, A. J., and Schuurink, R. C. (2003). Regulation of floral scent production in *Petunia* revealed by targeted metabolomics. *Phytochemistry* 62, 997–1008. doi: 10.1016/S0031-9422(02)00707-0
- Verweij, W., Spelt, C., Di Sansebastiano, G. P., Vermeer, J., Reale, L., Ferranti, F., et al. (2008). An H⁺-P-ATPase on the tonoplast determines vacuolar pH and flower colour. *Nat. Cell Biol.* 10, 1456–1462. doi: 10.1038/ncb1805
- Villarino, G. H., Bombarely, A., Giovannoni, J. J., Scanlon, M. J., and Mattson, N. S. (2014). Transcriptomic analysis of *Petunia hybrida* in response to salt stress using high throughput RNA sequencing. *PLoS ONE* 9:e94651. doi: 10.1371/journal.pone.0094651
- Wang, T. L., Uauy, C., Robson, F., and Till, B. (2012). TILLING in extremis. *Plant Biotechnol. J.* 10, 761–772. doi: 10.1111/j.1467-7652.2012.00708.x
- Williams, J. S., Wu, L., Li, S., Sun, P., and Kao, T. H. (2015). Insight into S-RNase-based self-incompatibility in *Petunia*: recent findings and future directions. *Front. Plant Sci.* 6:41. doi: 10.3389/fpls.2015.00041
- Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., et al. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475, 189–195. doi: 10.1038/nature10158
- Yang, Y., Li, R., and Qi, M. (2000). In vivo analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. *Plant J.* 22, 543–551. doi: 10.1046/j.1365-313x.2000.00760.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer Ronald Koes declared a collaboration with the authors Michiel Vandenbussche and Patrice Morel to the handling editor Rainer Melzer, who ensured that the process met the standards of a fair and objective review.

Copyright © 2016 Vandenbussche, Chambrier, Rodrigues Bento and Morel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evolution of Catkins: Inflorescence Morphology of Selected Salicaceae in an Evolutionary and Developmental Context

Quentin C. B. Cronk^{1*}, Isabelle Needham² and Paula J. Rudall²

¹ Department of Botany, University of British Columbia, Vancouver, BC, Canada, ² Royal Botanic Gardens, Kew, London, UK

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington, USA

Reviewed by:

Jill Christine Preston,
University of Vermont, USA
Madelaine Elisabeth Bartlett,
University of Massachusetts Amherst,
USA

*Correspondence:

Quentin C. B. Cronk
quentin.cronk@ubc.ca

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 27 August 2015

Accepted: 06 November 2015

Published: 07 December 2015

Citation:

Cronk QCB, Needham I
and Rudall PJ (2015) Evolution
of Catkins: Inflorescence Morphology
of Selected Salicaceae
in an Evolutionary and Developmental
Context. *Front. Plant Sci.* 6:1030.
doi: 10.3389/fpls.2015.01030

Poplars (*Populus* sp.) and willows (*Salix* sp.) are well known woody plants common throughout the northern hemisphere, both with fully sequenced genomes. They bear compact unisexual inflorescences known as “catkins.” Closely related genera of the “salicoid clade” within the family Salicaceae include the Asian genera *Bennettiodendron*, *Idesia*, *Itoa*, *Poliothyrsis*, and *Carrierea* and the Central American genera *Olmediella* and *Macrohasseltia*. Like willow and poplar, most of these genera are dioecious, but unlike willow and poplar they generally have loosely branched panicles rather than catkins, and less highly reduced flowers. However, the early developing inflorescences of *Carrierea* and *Idesia* show similarities to catkins which suggest possible pathways by which the salicoid catkin may have evolved.

Keywords: inflorescence evolution, heterochrony, synorganization, preformation, dioecy, floral reduction, inflorescence architecture, genome-enabled model system

INTRODUCTION

The Catkin and its Recurrent Evolution

The catkin is a type of compact or string-like inflorescence characterized by a single relatively stout axis on which unisexual sessile or subsessile apetalous flowers are clustered in a spiral or whorled arrangement. It is an extremely striking characteristic of many common trees, particularly of northern temperate regions. Notable among these are members of the order Fagales (oaks, walnuts, hazels, birches, and alders) and the relatively distantly related family Salicaceae s. str. (willows and poplars). The similarities between the catkins of these two groups led to them being classified together for a century (see below). It is now accepted that the presence of catkins in the two groups is the result of convergent evolution.

In this paper ‘catkin’ will be used in preference to the alternative term ‘ament.’ According to the Oxford English Dictionary, the word catkin came into English in 1578 when Henry Lyte (1529–1607) coined it in his translation of Dodoens’ New Herbal as a translation of the Dutch “katteken” (kitten) used for the downy inflorescences of willows and other plants (Dodoens, 1578). The botanical Latin equivalent, *amentum*, the Latin word for a thong or string, is less common. Its use in English dates from the late 18th century, sometimes anglicized as ament. The use of *amentum* in botanical Latin overlooks the Latin word for catkin, *iulus*, used as such by Pliny. However, apart from the occasional use of Juliflorae instead of Amentiflorae, this form never became established.

The Catkin and Taxonomy

The striking amentaceous inflorescences of many trees quickly attracted the attention of botanists, some of whom thought that the catkin-bearing trees formed a natural group (variously called Amentiflorae, Amentiferae, Amentales, or Amentaceae; Stern, 1973). Although the group name “Amentacea” was used by Gmelin, Linnaeus, and de Jussieu (Stern, 1973) and sporadically by later authors (Du Mortier, 1825), it was Eichler who was most influential in defining a ‘scientific’ Amentaceae. In the third edition of Eichler’s *Syllabus* (Eichler, 1883) the order Amentaceae comprised the Cupuliferae (i.e., Betulaceae and Fagaceae s.l.), Juglandaceae, Myricaceae, Salicaceae, and Casuarinaceae. This collection of families can be considered the canonical Amentiferae, although other groups have drifted in and out of the catkin-bearing alliance in various systems, e.g., Piperaceae, Urticales, *Leitneria*, Garryales (**Figure 1**). Remarkably, Eichler’s (1883) Amentaceae is a good natural group (providing of course that the unrelated Salicaceae is excised). In fact it corresponds almost exactly to the modern concept of the Fagales (Angiosperm Phylogeny Group, 2009), missing only *Rhoiptelea* and *Ticodendron*, both unknown to Eichler.

In the last century the realization grew that the Salicean branch of the Amentiferae was very different from the Fagalean

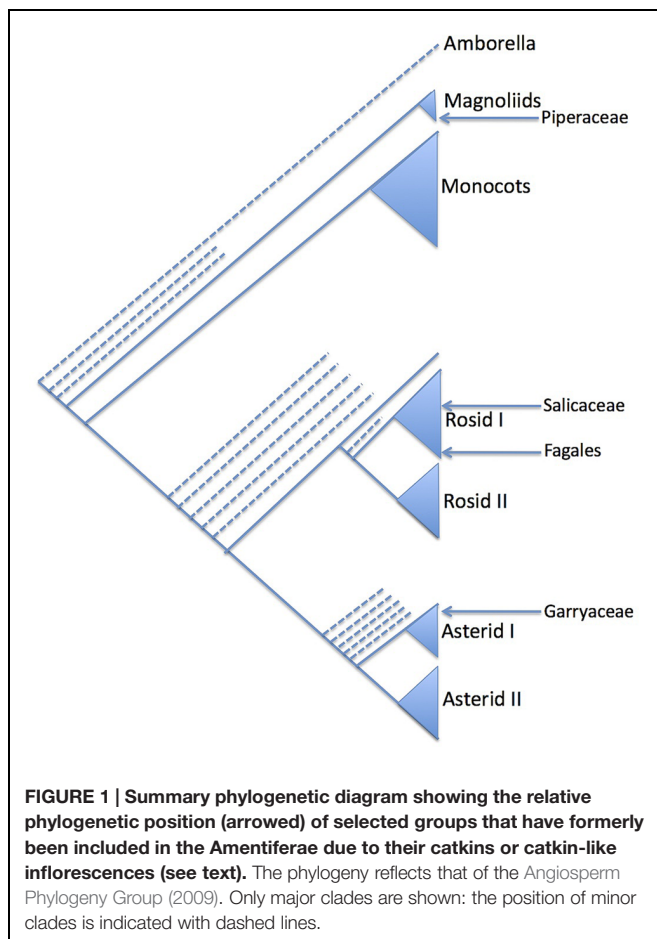
Amentiferae (**Figure 1**). This realization was confirmed by molecular phylogenetics and made clear in systems based on molecular phylogenetics such as the Angiosperm Phylogeny Group (APG) system (Angiosperm Phylogeny Group, 2009). An implication of the decomposition of the Amentiferae is that catkins, the most obvious unifying feature of the Amentiferae, have evolved in two very distinct lineages. This raises questions of convergent evolution: how the catkin evolved in each case and what the ancestral inflorescence form might be. Here we use comparative ontogenetic and anatomical observations as a basis to discuss these questions in one of the archetypal catkin-bearing groups, Salicaceae.

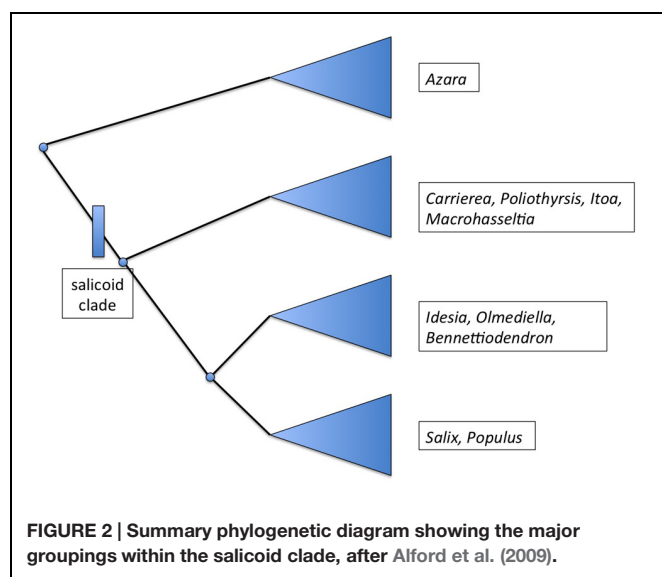
The Salicaceae, Classification and Morphology

When Eichler included Salicaceae within his order Amentaceae, the family was wholly amentiferous (i.e., catkin-bearing) comprising only the genera *Salix* and *Populus*. Molecular evidence, coupled with support from phytochemistry and morphology, has demonstrated a close relationship between *Salix* and *Populus* and many non-amentiferous genera that were formerly placed in the Flacourtiaceae (Leskinen and Alstrom-Rapaport, 1999; Chase et al., 2002; Alford, 2005). The heterogeneous family Flacourtiaceae is now dismembered, and its members are placed in other families, mainly the Salicaceae and Achariaceae. The family Salicaceae, as now circumscribed in the broad sense, is a more homogeneous group of about 1000 species in c. 55 genera. They are uniformly woody (trees or shrubs) with simple, usually alternate, leaves. The leaves are often dentate and the leaf teeth frequently glandular (characteristic ‘salicoid teeth’). The flowers are often inconspicuous and a perianth may be lacking in some genera. Inflorescence morphology in the family as a whole is highly variable. The sister family to the Salicaceae is probably the Lacistemataceae (Davis et al., 2005; Korotkova et al., 2009), and it is of interest that this family has also independently evolved catkins.

Salix and *Populus* are closely related sister genera which in turn are related to a group of seven other genera (Alford, 2005). Initial molecular phylogenetic evidence based on ITS and eight plastid regions suggests that “a clade consisting of *Bennettiodendron*, *Idesia*, and *Olmediella* are sister to *Salix* and *Populus* (**Figure 2**). Sister to that clade is a clade of the other four genera, *Carrierea*, *Itoa*, *Macrohasseltia*, and *Poliothyrsis*” (Alford et al., 2009).

These nine genera have been referred to as the “salicoid clade” of the family Salicaceae (Cronk, 2005). *Salix* and *Populus* are known to be palaeotetraploid (Sterck et al., 2005) with a primary chromosome number of $n = 19$ (Darlington and Wylie, 1955) whereas the haploid base number for the family is $n = 9$ or 11. For instance, *Azara serrata* Hook is $n = 9$ (Sanders et al., 1983). Very few chromosome counts exist for the genera of the salicoid clade but both *Olmediella* (Grill, 1990) and *Idesia* (Darlington and Wylie, 1955; Grill, 1990) appear to be tetraploid at $n = 22$. It is possible therefore that whole of the salicoid clade shares the same palaeotetraploidy event from $n = 11$ to $n = 22$, followed by (in *Populus* and *Salix*) reduction events to $n = 19$.





The Inflorescence of the Salicoid Clade of the Salicaceae

The genera of the salicoid clade have been described in various recent treatments (Sleumer, 1980; Fang et al., 1999; Yang and Zmartzty, 2007). They are all described as having paniculate inflorescences with the exception of *Populus* and *Salix* which have racemose inflorescences (catkins). A reduction of inflorescence branching (from panicle to raceme) does occur as a sexual dimorphism in *Itoa*, in which the male flowers are said to be in racemes and the female flowers in panicles (Table 1). Furthermore, in *Idesia* the paniculate inflorescences are long, pendulous and fairly narrow, superficially resembling racemes (Figure 3).

In the fossil record there is some blurring between raceme and panicle. *Pseudosalix*[†] (Boucher et al., 2003) is an Eocene fossil of Salicaceae that has leaves like willow (*Salix*) but somewhat paniculate inflorescences. Furthermore, the Eocene *Populus tidwellii*[†] Manchester, Judd & Handley (Boucher et al., 2003; Manchester et al., 2006) has catkins with some lateral branching near the base, placing it in an intermediate position in this character with the paniculate ancestors of *Populus*. Furthermore, Fisher (Fisher, 1928a,b) argued that the bract (with which the flower is associated), while appearing to be directly inserted on the main axis of the inflorescence, is in fact inserted on a minute lateral stump (which Fisher called the “internode”) at the top of which the flower is borne. She argued that this feature is an indication of the evolution of the Salicaceae catkin from a branched, paniculate antecedent.

In most of the salicoid clade, the inflorescence is terminal on the shoots, often (as in *Idesia*, *Carrierea*, and *Poliothyrsis*) terminating the shoots that appear after bud-break of the terminal bud. This condition contrasts with *Salix* and *Populus*, in which the inflorescences are produced in lateral buds (with very few exceptions: in *Salix* sect. *Chamaetia* they are often terminal). Again, *P. tidwellii*[†] (Manchester et al., 2006) is interesting in this regard as it has terminal inflorescences.

Floral Morphology of the Salicoid Clade

Apart from unisexuality, lack of petals and the presence of nectarial disk glands in some species, the flowers of the most genera in the salicoid clade are unexceptional.

The flowers all have a subtending bract. The highly reduced flowers in *Salix* and *Populus* prompted some early authors to suggest that the bracts in these taxa might be derived from the missing perianth. However, this interpretation was shown to be false by Fisher (1928a,b) who demonstrated that they have a foliar-type vascularization consistent with bract origin: these

TABLE 1 | Genera of the salicoid clade of Salicaceae.

Genus	No. of sp.	Sex of flowers	Distribution of sexes	Inflorescence	Perianth and disk
<i>Bennettio-dendron</i>	3 (Asia)	Unisexual	Dioecious	Terminal or axillary panicles	Sepals 3, petals 0, disk glands numerous, small
<i>Carrierea</i>	2 (Asia)	Unisexual	Dioecious	Short terminal or axillary panicles	Sepals 5, petals 0, disk glands 0
<i>Idesia</i>	1 (Asia)	Unisexual	Dioecious	Long raceme-like terminal or axillary panicles	Sepals c.5, petals 0, disk glands numerous among stamens or staminodes
<i>Itoa</i>	1 (Asia)	Unisexual	Dioecious (or partly monoecious?)	Terminal panicles (f) or terminal or axillary racemes (m)	Sepals 5, petals 0, disk glands 0
<i>Macrohasseltia</i>	1 (S. Am.)	Unisexual	Dioecious	Small panicles	Sepals 5 (reduced), petals 0, disk glands numerous (at base of each stamen)
<i>Olmediella</i>	1 (C. Am.)				
<i>Poliothyrsis</i>	1 (Asia)	Unisexual	Monoecious	Terminal panicle, upper flowers female	Sepals 5, petals 0, disk glands 0
<i>Populus</i>	c. 60 (wide)	Unisexual	Dioecious	Catkin	Sepals 0, petals 0, disk cupular
<i>Salix</i> (incl. <i>Chosenia</i> which has both disk glands absent)	c. 500 (wide)	Unisexual	Dioecious	Catkin	Sepals 0, petals 0, disk glands us. 2 (adaxial and abaxial; abaxial may be absent)

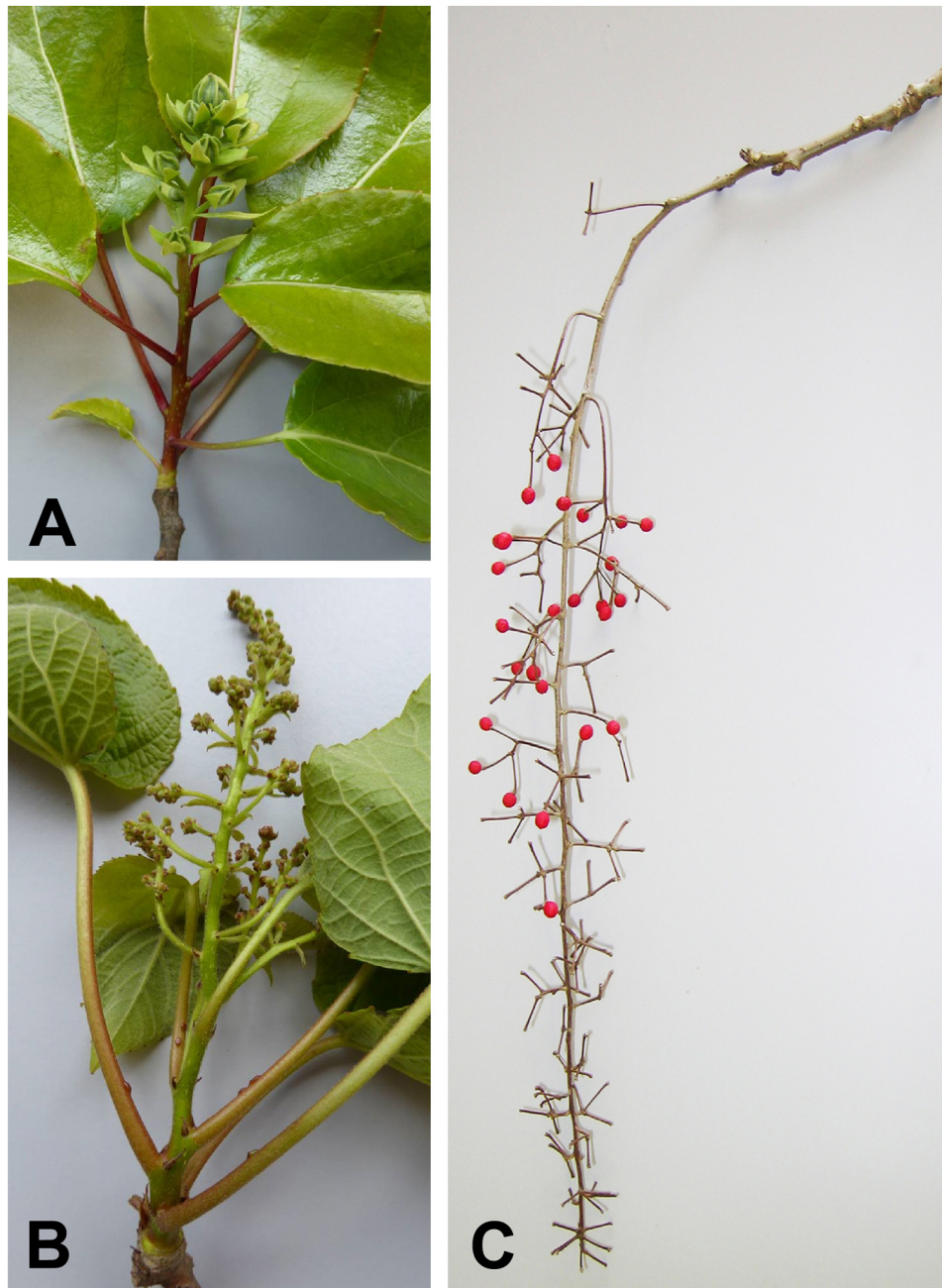


FIGURE 3 | Photographs of inflorescences. (A) Young inflorescence of *Carrierea calycina* Franch.) collected May 2015. The flowers are unopened. The floral bracts and calyces are visible. **(B)** Young male inflorescence of *Idesia polycarpa* Maxim., collected May 2015. The flower buds are still small but the branching of the paniculate inflorescence is evident. This inflorescence has developed in 3 months from an inflorescence meristem as shown in **Figure 5B**. **(C)** Previous year's female inflorescence (infructescence) of *I. polycarpa*, collected February 2015, showing the extensive branching of the inflorescence.

bracts are therefore simply homologous with the floral bracts of other members of the salicoid clade.

All genera have a calyx of (3-)5(-6) sepals, except *Salix* and *Populus* which lack an obvious perianth entirely. It should be noted that in *Olmediella* the calyx is reduced and quickly caducous (Alford, 2005). The staminate flowers have numerous stamens (in all except *Salix*, which generally has only 1–5

stamens) and a vestigial ovary (absent in *Salix* and *Populus*). The pistillate flowers have numerous small staminodes (except *Populus* and *Salix*). The presence of vestigial sexual organs in most species but their complete absence in *Salix* and *Populus* indicates the extent of the process of floral reduction in those genera. However, it is not known whether the developmental pathways for vestigial pistils in males (and staminodes in females)

has been completely removed as part of sex determination or merely reduced to the extent it no longer has anatomical consequence.

The nectarial disk glands are an important floral feature in many genera (Table 1). These glands are generally assumed to be outgrowths of the disk. However, they are consistently associated with the stamens and staminodes, appearing interspersed among the stamens (intrastaminal, as in *Idesia*), or at the bases of the stamens in *Olmediella* (Alford, 2005). This location raises the question of whether they might be staminodial in origin.

There is a further question of whether the disk glands (nectaries) of *Salix* are homologous with the disk glands of other genera. Fisher has argued convincingly (Fisher, 1928a,b) that they represent a modified perianth because there appears to be some vascularization. However, Fisher made this argument before the outgroups of *Salix* were known. Now that we know the close relationship between *Salix* and other genera with disk glands, it seems logical to assume their homology (Alford, 2005).

Another puzzle is the “cupular disk” of *Populus*. Fisher homologised this structure with perianth and with the disk glands of *Salix*: “The disk-shaped perianth of *Populus*, or its peripheral parts, is homologous with the nectary of *Salix*” (Fisher, 1928b). Skvortsov (1999) also had no difficulty homologizing the disk glands of *Salix* with the cupular disk of *Populus*, mainly because the disk glands in *Salix* are sometimes united and approach in morphology the cupular disk of *Populus*. He writes: [*Salix* has] “...one or two (or a few) nectariferous glands, which occasionally are connate into a lobed glandular disk. These glands are obviously homologous to the cup-shaped disk in the poplars (which is sometimes called perianth).” However, the cupular disk of *Populus* is vascularized, consistent with Fisher’s thesis that the disk glands of *Salix* have a perianth origin, assuming the two to have a common origin. However, it is also possible that as non-perianth disk glands evolved to increased complexity, vascularization was co-opted. Another possibility is that the cupular disk of *Populus* is indeed directly homologous with the calyx but not with the glands of *Salix* (which then have a non-perianth origin). Finally, the simplest explanation of all is that the cupular disk is merely an enlarged disk (i.e., receptacular in origin).

In this paper we seek to investigate whether the morphology of closely related non-catkin-bearing species can inform our understanding of the evolution of catkins in the Salicaceae. In particular we are interested in setting out the main ways in which *Salix* and *Populus* differ, in reproductive morphology and phenology, from their close relatives. Knowledge of the inflorescence morphology and flowering behavior of related plants allows the formulation of scenarios by which catkins evolved in this clade.

MATERIALS AND METHODS

Sample Collection

For *Idesia*, *Carrierea*, and *Poliothyrsis*, terminal resting buds were collected in spring before bud-break. In *Salix* and *Populus* lateral inflorescence buds were collected at the same time. A young inflorescence of *Olmediella* was collected at the same time from greenhouse-grown material. After budbreak young inflorescence shoots were also examined. A list of samples collected with accession numbers is given in Table 2.

Sample Preparation

Collections of inflorescence material were killed and fixed in formalin–acetic acid–alcohol (FAA) for approximately 1 week followed by storage in 70% ethanol. Some material was dehydrated through an ethanol series to 100% ethanol, transferred to HistoClear before embedding in Paraplast using standard protocols. The wax blocks were sectioned at 10 μ m thickness on a rotary microtome (Leica RM2155) and the resulting sections were stained in 0.5% (w/v) solution of toluidine blue before mounting on microscope slides in DPX mountant. Images were captured using a Zeiss AxioCam HRC camera attached to a Leica DMLB microscope.

Other material was dehydrated through an alcohol series into acetone and transferred to a critical-point drier (Tousimis Autosamdri 815B). Dried material was then sputter-coated with platinum in a sputter coater (Emitech K550). The material was examined on a Hitachi S-4700 II cold-field emission scanning electron microscope.

TABLE 2 | Details of material examined in this study.

Name	RBG Kew accession no.	Date collected
<i>Carrierea calycina</i> Franch. –F	2006-616	6.2.2015, 13.2.2015
<i>Idesia polycarpa</i> Maxim. – F	2006-332	6.2.2015, 13.2.2015
<i>Idesia polycarpa</i> – M	2008-416	6.2.2015, 13.2.2015
* <i>Olmediella betschleriana</i> (Göpp.) Loes. – F	1969-12335	16.2.2015
* <i>Poliothyrsis sinensis</i> Oliv. – Mon	1973-20904	7.2.2015, 13.2.2015
<i>Populus nigra</i> L. – M	1988-8331	2.2.2015
<i>Populus purdomii</i> Rehder – F	1973-6401	2.2.2015
<i>Populus wilsonii</i> C.K. Schneid. – F	1979-1110	18.2.2015
<i>Salix miyabeana</i> Seem. – M	1999 - 547	16.2.2015
<i>Salix</i> sp. – F	1999-548	16.2.2015

All material was collected from living material growing in the Royal Botanic Gardens Kew. Male and female plants are indicated as M and F, respectively (Mon, monoecious). An asterisk indicates those species not figured in the present study.

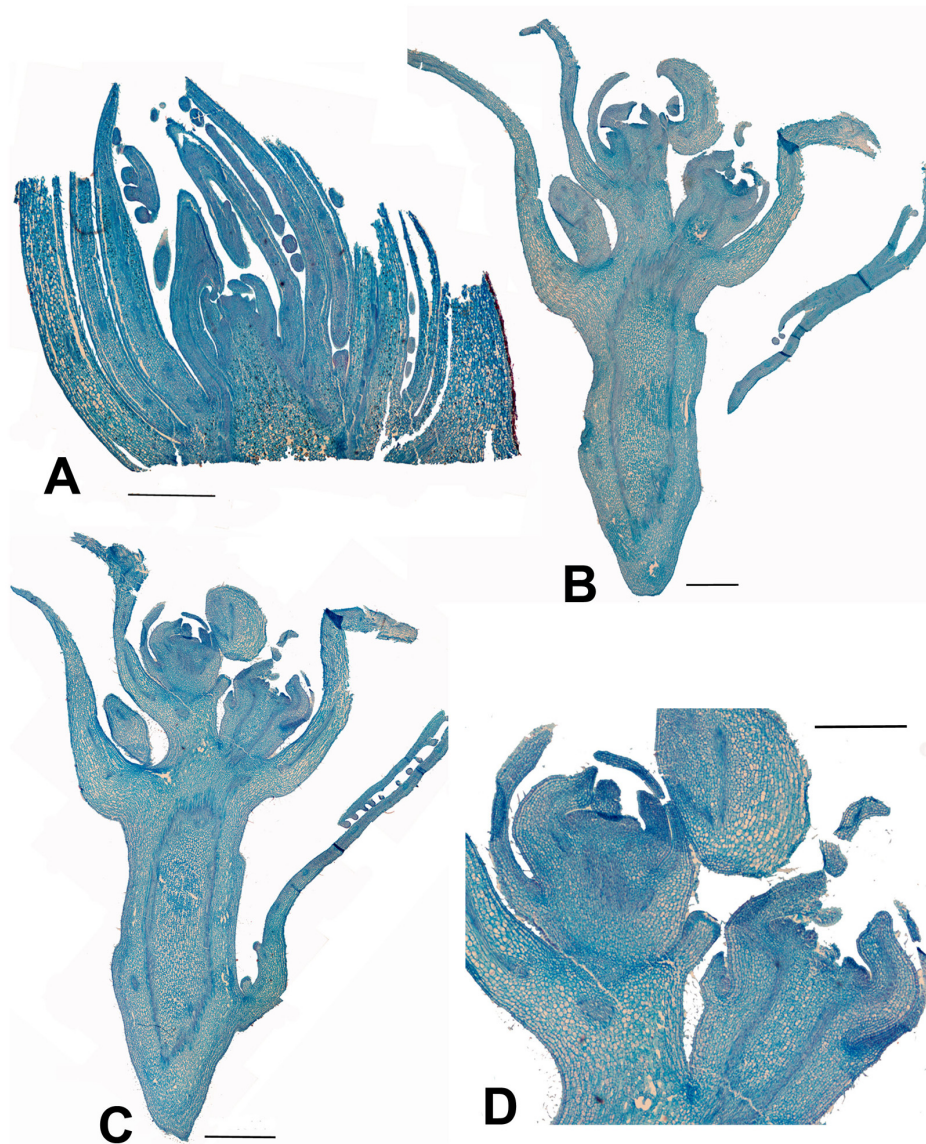


FIGURE 4 | *Carrierea calycina*. (A) Over-wintering bud with minute inflorescence meristem surrounded by protective bracts (bud scales) collected 13th February 2015. At this stage the inflorescence has not started developing. (B,C) Sequential longitudinal sections of dissected resting bud (pre-budbreak) collected 2nd April 2015, showing a fairly well-developed inflorescence with individual flowers differentiated. Floral bracts and calyx are visible but other organs have not formed and gender is not visible at this stage (bud scales removed). (D) Detail of developing flower from (C). Scale bars = 1 mm.

RESULTS

Phenology and Gross Morphology of Reproductive Shoots

In autumn *Poliothyrsis*, *Idesia*, and *Carrierea* set comparatively large terminal buds on all shoots of the previous year. The majority of these buds produce short shoots terminating in an inflorescence (Figure 3). Vegetative growth (and flowers of the following year) is therefore left to side shoots. This growth pattern corresponds with the “Modèle de Leeuwenberg” of Hallé and Oldeman (Hallé et al., 1978). *Populus* has the opposite tendency, with terminal buds tending to be vegetative and side shoots

(from axillary buds of the previous year) tending to contain catkins. In *Poliothyrsis*, nearly all growth is by terminal buds from side shoots of the previous year, these in turn terminate in inflorescences with between four and six leaves below them. In *Carrierea* there tend to be four to six leaves below each inflorescence and in *Idesia* four to five.

Developmental Anatomy

Carrierea (Figures 4A–D) and *Idesia* (Figures 5A,B) show very little development of the inflorescence when collected in February. In contrast, *Populus* and *Salix* (Figures 6 and 7) have fully formed flowers. The preformation and early development

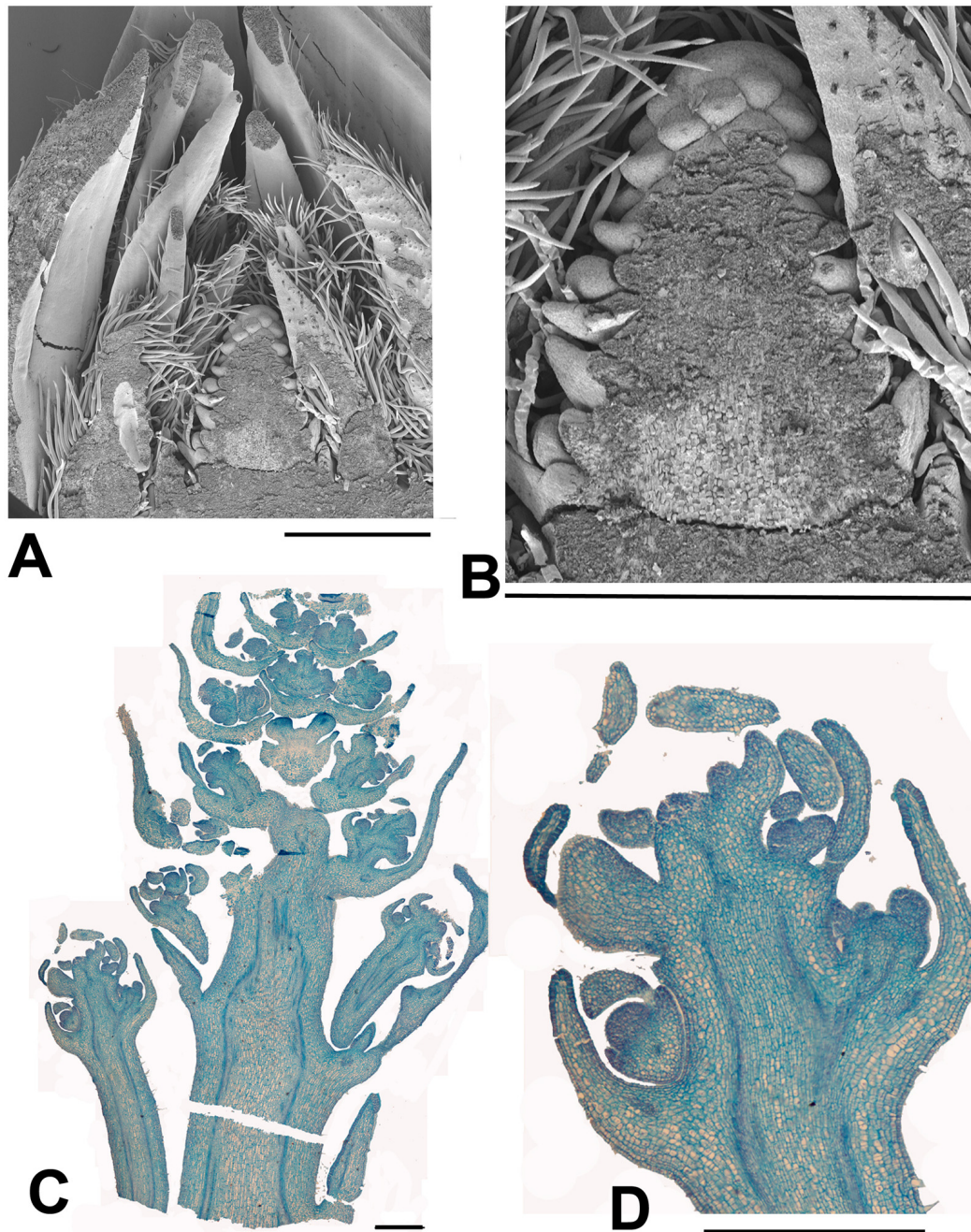


FIGURE 5 | *Idesia polycarpa*. (A) SEM of bisected over-wintering bud collected 13th February 2015 showing an inflorescence meristem surrounded by protective bracts (bud scales; (B) Detail of (A) showing just the inflorescence meristem. At this stage the inflorescence meristem shows inflorescence bract primordia but not floral primordia. Later branch primordia will form in the axils of the inflorescence bracts and develop into the panicle (i.e., branched) inflorescence. At the same stage *Salix* and *Populus* inflorescences are fully developed with completely formed flowers (Figures 6 and 7). (C) Longitudinal section of dissected over-wintering (pre-budbreak) bud collected 2nd April 2015 showing development of panicle (bud scales removed). Flowers can be seen in early development, with some development of the calyx but not other organs. At the same date *Salix* and *Populus* have finished flowering. (D) Enlarged portion of (C). Scale bars = 1 mm.

of inflorescences in *Populus* and *Salix* is well known, with inflorescences formed the previous year (Boes and Strauss, 1994; Kaul, 1995; Brunner et al., 2014). Figures 6 and 7 show the almost fully developed flowers inside the unopened buds enclosing catkins of *Populus* and *Salix* when sampled in early

February (well before bud opening and flowering in March). In contrast, developmental timing in *Idesia* was found to be very different. No identifiable inflorescence meristems were found in buds sampled in February, although well-developed leaf primordia were present (Figure 5). Resampling in April revealed

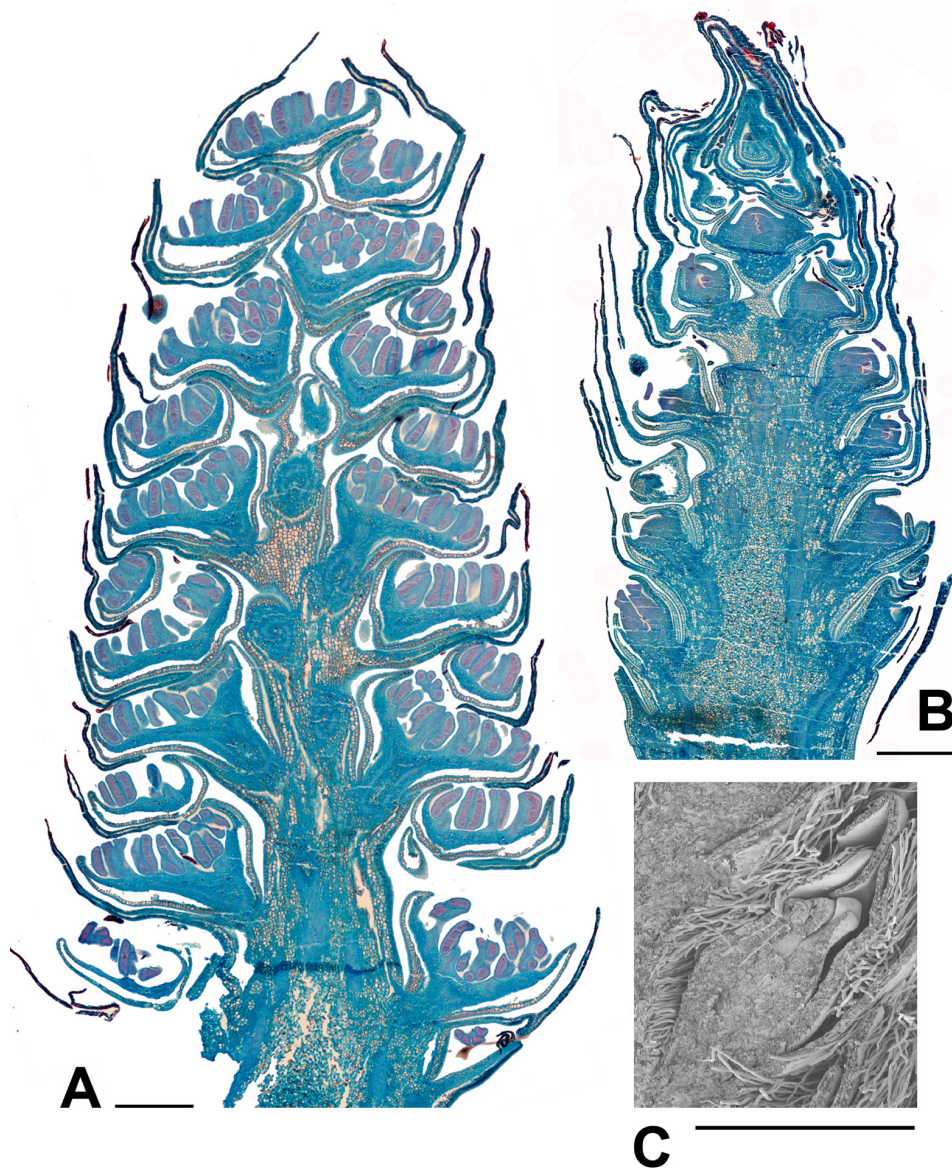


FIGURE 6 | *Populus*. (A) *Populus nigra* L., male inflorescence (catkin) with nearly fully developed flowers. Longitudinal section of catkin, collected 2nd February 2015, showing well-developed stamens with short filaments and longer anthers with red-staining developing pollen (bud scales removed). (B) *P. purdomii* Rehder, female inflorescence with flowers in a late stage of development. Longitudinal section of catkin collected 2nd February 2015, showing single ovary per flower. (C) *P. wilsonii* C.K.Schneid., SEM detail showing a well developed ovary from a female inflorescence, collected 2nd February 2015. Scale bars = 1 mm. Bud scales removed.

dramatic differences. A well-developed inflorescence meristem was found to be present, but no developed flowers (Figure 5). We conclude that inflorescence development in *Idesia* occurs in response to warming temperatures in the spring, although much of it is completed within the closed bud before bud break in May. At an early stage, the inflorescence meristem resembles a catkin in having numerous spirally arranged bracts and primordia on an axis. However, these primordia will develop into inflorescence branches and not individual flowers.

DISCUSSION

Evolutionary-developmental Mechanisms Implicated in Inflorescence Evolution in the Salicoid Clade of Salicaceae

When the highly reduced and specialized inflorescence of *Salix* (Figure 7) is compared with its outgroup genera, for instance *Idesia* (Table 3), there are several traits that are shared. Our

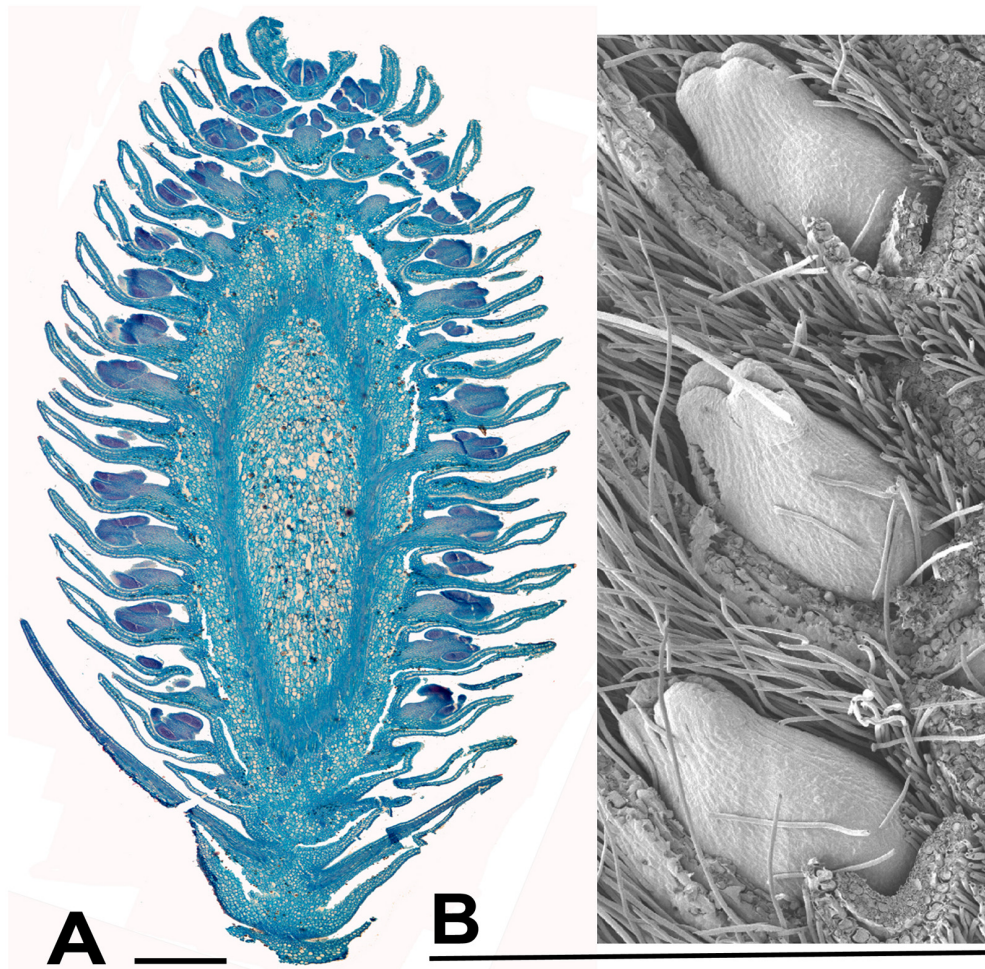


FIGURE 7 | *Salix*. (A) *Salix miyabeana* Seem., male inflorescence (catkin). Longitudinal section of over-wintering bud (bud scales removed), collected 16th February 2015, showing well-developed flowers, each with an apparent single stamen (actually two stamens completely united except at the anther). The anthers are well differentiated pre-anthesis. **(B) *Salix* sp., SEM showing three well developed female flowers in the axils of bracts, from an inflorescence in over-wintering bud collected 16th February 2015. Scale bars = 0.5 mm. Bud scales removed.**

comparative investigation indicates that these apparently pre-existing shared traits (unisexual flowers, dioecy and association of flowering with resting buds) likely pre-date the evolution of catkins rather than being a consequence of the evolution of catkins. In contrast to the monoecious Fagales, where typically male and female catkins occur on the same tree, dioecy is universal in the catkin-bearing Salicaceae and their immediate relatives. Although bisexual teratomorphs are sometimes found, bisexual species in the catkin-bearing Salicaceae are exceedingly rare, although they do occur as a derived condition (Rohwer and Kubitzki, 1984). Dioecy in the group is ancient and stable and under genetic control (Geraldes et al., 2015), even though the genetic mechanism is labile and has apparently undergone numerous shifts within the genome (Filatov, 2015; Geraldes et al., 2015).

A number of other traits, however, are specific to the catkin-bearing habit: preformation, precocity, bud dimorphism, inflorescence contraction, floral reduction, lateralization. These

traits represent necessary steps to the evolution of catkins in the Salicaceae. They will be discussed in turn.

Preformation

This is the formation of structures a long time before they become visible or functional. In *Salix* and *Populus* the inflorescence is initiated as soon as the resting buds form, which may be as early as May in the year preceding flowering. The early initiation allows time for the catkin to be fully formed by the time buds break in the spring. By contrast, other members of the salicoid clade complete inflorescence maturation only after bud break in the spring (contrast **Figures 5A,B** with **Figures 6** and **7**). Inflorescence development may start in the bud (partial preformation) but completes on the growing shoot (see results and **Figure 5**). Preformation is obviously a necessary precondition for precocity (below), and precocity may be what is driving full preformation. Preformation in resting buds involves a paradox. In a normal bud there is no leaf production and shoot growth has ceased. The bud

is in developmental stasis, at least vegetatively, in preparation for full dormancy. On the other hand with floral preformation there is much reproductive development in the bud with the formation of an inflorescence meristem and floral primordia and their development into flowers. The evolution of preformation implies increased developmental control in coupling two phase changes of the meristem: change from active growth to dormancy and change from vegetative to reproductive. In *Idesia* these phase changes seem to occur sequentially, first dormancy then inflorescence formation. In poplar and willow they appear to be coupled.

Precocity

The term precocity is here applied to the breaking of reproductive buds before the vegetative buds to allow flowering before the development of a canopy. It should be noted that precocity as discussed here is seasonal precocity, not flowering as a juvenile (also sometimes refers to as precocious flowering). Seasonal precocity implies bud dimorphism (below). Although most willows are strictly precocious, the genus *Salix* as a whole shows a great deal of variation in this trait. Some are what salicologists call “coetaneous” (an archaic word for contemporaneous), meaning that the catkins are produced at the same time (contemporaneously) with the leaves. More precisely this means that the catkins are not sessile on the previous years growth but on, bud-break, a leafy shoot is produced which the catkin terminates. In this respect it is directly analogous with *Idesia*, in which the inflorescence terminates a short leafy shoot. A third type of reproductive behavior, “serotiny” (from the latin *serotinus* = coming late), also occurs in willows. This is an extreme form of the coetaneous habit. In botany serotiny is more commonly applied to delayed seed dispersal, but in willows it refers to delayed flowering. This occurs when catkins are poorly developed in bud and complete their development post-budbreak, thus apparently flowering with the current season’s growth. Although formal analyses are lacking (partly due to continuing uncertainty over the phylogeny of *Salix*), it is likely

that coetaneous and serotinous willows are derived and may represent reversals. However, if the coetaneous habit is found to be primitive in *Salix* it provides a link to other genera of the salicoid clade.

Non-terminal Deletion

Closely associated with precocity, non-terminal deletion is the evolutionary loss of parts of an organ from the base rather than the tip. In this case, it refers to the loss of leaves below the terminal inflorescence, as in precocious *Salix* and *Populus*. For instance *Idesia* bears its inflorescences on leafy shoots whereas the catkins of *Populus* are not associated with vegetative leaves. Evolutionary loss at the end of a shoot may simply be the consequence of growth ceasing early, while evolutionary gain at the tip may be the result of growth continuing for a longer period. Loss (in this case of leaves) at the base of a shoot is more problematic. It requires that a late developmental program (in this case inflorescence production) is brought forward to replace early developmental programs (in this case leaf primordia production which would normally take place as the resting buds form). The concepts of terminal and non-terminal deletion have been used in evolutionary analyses of other botanical systems, including of fossils (Bateman, 1994). In our present system we can see that extreme inflorescence preformation, characteristic of *Populus* and *Salix*, brings forward inflorescence production to precisely the time when leaves would be forming during the development of the resting bud. Therefore precocity, preformation and non-terminal deletion, although separate concepts, may in fact be interlinked parts of a single evolutionary scenario.

Bud Dimorphism

In *Salix* and *Populus* there is a functional dimorphism between floral and vegetative buds. Precocity implies that floral and vegetative buds may have different temperature sensitivity, with inflorescence buds having a lower cumulative temperature requirement (heat sum, for instance in degree days) required for development. Bud dimorphism allows a marked “division of labor” between reproductive and vegetative meristems. In poplar and willow the catkin usually has no vegetative function whatsoever, and correspondingly the vegetative shoot has no reproductive function. In *Idesia* the distinction is blurred. Almost all shoots produce not only a terminal inflorescence but also numerous leaves below the inflorescence. Thus reproductive and vegetative functions are carried out by the same buds (reproductive vegetative buds).

Contraction

In most genera of the salicoid clade the inflorescence is a lax branched panicle with elongated rachises. The evolution of the catkin therefore requires evolutionary and developmental contraction of the inflorescence. The inflorescence meristem of *Idesia*, with primordia and associated bract primordia (Figure 5) gives a possible scenario of how the catkin could have evolved. The primordia would normally develop into panicle branches and then into flowers. If the floral developmental pathway were to be brought forward in developmental time then it is possible to see how the result would be a series

TABLE 3 | Comparison of *Idesia* and *Salix* in terms of putative processes and characteristics of inflorescence evolution and development.

Process/characteristic	<i>Idesia</i>	<i>Salix</i>
Unisexual flowers	Yes	Yes
Dioecy	Yes	Yes
Flowering linked to resting buds	Yes	Yes
Preformation	No (or short)	Yes (long)
Contraction	No (elongated branched rachis)	Yes (complete)
Non-terminal deletion	No	Yes (in most species)
Precocity	No	Yes
Bud dimorphism	No (buds producing leaves and inflorescence)	Yes (buds either vegetative or flowering)
Floral reduction	No	Yes
Lateralization	No (inflorescence terminal)	Yes (inflorescence lateral)

of bract associated flowers. This would be an example of heterochrony: a change in developmental timing (Bateman, 1994; Rudall and Bateman, 2004). In Solanaceae, it is suggested that minor changes in the maturation process of apical meristems can give rise to dramatic changes in reproductive shoot organization (Park et al., 2012, 2014). In grasses, more complex panicles can be formed by delaying the phase change from the indeterminate shoot meristem (SM) inflorescence building program to a determinate spikelet and floral meristem (FM) program (Kyojuka et al., 2014). In the evolution of catkins we propose the reverse: a simplification of the panicle by early phase change from an inflorescence building (SM) program to a determinate FM developmental program. It is of interest that Fisher (Fisher, 1928a,b) found, in the catkins of some species of *Salix*, microstructures that she interpreted as the vestigial branches of an ancestral branched inflorescence. This implies that catkin evolution proceeded via a progressive shortening of axes rather than a complete deletion of the branching pathway for inflorescence development. Furthermore, the floral developmental pathway has not been brought forward so far as to eliminate all trace of branch structure. Fisher's finding was all the more remarkable as it came long before the appropriate outgroups were known and it was easy to assume that the ancestral form was a simple raceme rather than a branched panicle.

Floral Reduction

Small flowers are an obvious consequence of the evolution of the catkin as there is no space for elaborate flowers in a highly condensed inflorescence. Additionally some of the functions of individual flowers are, in the catkin, taken over by the inflorescence as a whole. An example is floral protection in bud which is done by the calyx in *Idesia* but by the tight packing of the flowers and investing bracts in poplar and willow. A remaining question is whether the calyx has been lost completely or converted into disk glands (nectaries) in willow. For nearly a century this question has been considered closed with the consensus that the disk glands of willow and the cupular disk of poplar represent the lost perianth. However, the recent identification of the relatives of *Salix* and *Populus* followed by the realization that they have both disk glands and a calyx have cast some doubt on this consensus (Alford, 2005). In addition to the loss of calyx there has also been a reduction in stamens in the insect-pollinated *Salix* (down to one in some species). In the wind-pollinated *Populus*, large numbers of stamens have been retained, packed very tightly into the flowers in bud. This illustrates the constraint that the more pollen-wasteful process of wind pollination places on floral reduction in poplar.

Lateralization

In *Idesia*, the terminal buds on shoots produce inflorescences and the inflorescence terminates shoot growth. In poplar, terminal buds are never (or at least very rarely) reproductive. The catkin buds are all lateral (axillary) buds. The determinate growth of all terminal buds puts a constraint on the rate of height growth that can be attained by *Idesia* and its relatives which

tend to be relatively small trees. Poplar, however, because its inflorescences are lateral, can maintain indeterminate growth, resulting in poplars being generally the fastest growing and tallest dicotyledonous trees in the northern hemisphere. Catkins are also lateral in willow, but in willow the terminal shoot tends to abort rather than form a resting bud for continued growth the following year, hence willows also tend to be smaller in stature. In the Juglandaceae, lateralization of the catkins is only partial, as while staminate catkins are general lateral, pistillate catkins are usually terminal (Manning, 1938). This could suggest a physiological constraint between investment in large fruit (as in Juglandaceae) and inflorescence position.

Adaptive Significance of Inflorescence Evolution in Salicaceae

The compactness of the catkin allows inflorescence development to be completed within the inflorescence bud. This in turn allows for precocious flowering. Precocity has an obvious consequence for wind-pollinated plants (such as poplars, *Populus*) as it allows flowers to be pollinated before the emergence of the leafy canopy which may attenuate air movement among the branches. For insect-pollinated plants, precocity effectively removes competition for bees from other flowers. Willows (*Salix*) are generally insect pollinated (Karrenberg et al., 2002), particularly by bees of the genus *Andrena* (Knuth, 1909; Ostaff et al., 2015). They are also well known to be an important source of pollen and nectar for honey bees (*Apis mellifera* L.) early in the year when bees have few other food sources. However, a trade-off against the absence of competition for pollinators, is the fact that there may be fewer bees flying in early months of the year. This mechanism, of course, applies only to temperate regions with a pronounced cold season, and is not applicable to catkin-like inflorescences of tropical origin such as the related Lacistemataceae.

Another mechanism, related to precocity, is thermal protection (Tsukaya and Tsuge, 2001). The contraction of the flowers into a compact inflorescence allows the flowers to be protected by hairs on the margins of the bracts. These form, in some instances, a striking woolly insulating layer around the catkin. Indeed, the name 'catkin' alludes to this flocculence. The woolliness is equivalent to the woolly hairs of many alpine plants and by trapping air may allow flowers to survive the severe night frosts encountered as a consequence of precocity. A lax panicle, on the other hand, cannot be protected by hairs on its bracts.

A third mechanism that should be considered is reproductive efficiency. Poplars and willows produce large amounts of seed with little investment in inflorescence structures. Compare this with *Idesia*, which produces a modest amount of seed with a heavy investment in inflorescence rachis and flower stalks. Furthermore, individual flowers of *Idesia* are large. They have to be, as each one has to provide a sufficient landing surface for pollinating insects. By aggregating minute flowers together, *Salix* provides a landing platform for bees while minimizing investment in individual flowers. This is an example of synorganization, i.e., the provision of a novel or more efficient function by different plant organs working in concert, in this

case numerous small flowers organized into a larger unit that can function as a landing surface.

The Genome-enabled Family Salicaceae as a Promising System for Evolutionary Developmental Biology

The salicoid clade of the Salicaceae exhibits a promising range of ecologically important morphological traits (Cronk, 2005). It is also one of the best characterized clades of dicotyledons at the genome level. The poplar (*P. trichocarpa* Torr. & A.Gray) genome was the third plant genome to be released (Tuskan et al., 2006) and it has now been joined on the comparative genomics site Phytozome (Goodstein et al., 2012) by the genome of *Salix purpurea* L., an economically important basket and biofuel willow extensively used in breeding programs for crossing with other species. Complete genome sequencing projects are well advanced for other species of *Salix* and *Populus* and a plethora of genomic information will soon be available. This raises the possibility of a molecular approach to the evolution of many key traits in the salicoid clade, including inflorescence architecture. Importantly for reproductive traits, the genomic architecture of the sex locus in *P. trichocarpa* has recently been elucidated (Geraldes et al., 2015).

Inflorescence architecture is a economically important trait in many crop species. The grapevine (*Vitis*) is a good example, in which a compact or lax infructescence (caused by variation of rachis length) is a characteristic of commercial importance (Correa et al., 2014). For obvious reasons the genes underlying this trait are now attracting increased attention. A number of mutants are known in *Arabidopsis* that affect inflorescence traits. An example of genes of potential relevance to catkins are the compact inflorescence (CFL) genes (Goosey and Sharrock, 2001).

The rich genomic resources developing for *Salix* and *Populus* will greatly facilitate the development of genomic resources for other genera in the salicoid clade. A complete genome of *Idesia* would be particularly valuable as an outgroup for *Salix* and *Populus*. Similarly, a member of the Salicaceae that is more distant (such as *Azara*) would be useful as an outgroup for the salicoid clade as a whole. *Azara* is outside the palaeopolyploidy event

that has occurred in the salicoid clade, it would therefore be particularly useful to assess evolution of gene paralogues in *Salix* and *Populus* resulting from the whole genome duplication event.

CONCLUSION

The morphological richness of the Salicaceae coupled with the rapidly expanding genomic resources make this family, of all woody plant families, particularly promising for genome-enabled evolutionary developmental biology.

AUTHOR CONTRIBUTIONS

QC and PR planned the study, collected material, supervised the anatomical work and wrote the paper. IN carried out the anatomical work and photomicrography and contributed to writing the paper.

FUNDING

Funding for this study came from a Royal Society International Exchanges Scheme grant to PR and QC and from the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants Program (grant no. RGPIN-2014-05820) grant to QC.

ACKNOWLEDGMENTS

We thank the staff of the Horticulture Directorate of the Royal Botanic Gardens, Kew for maintaining living collections of the plants studied here. We also thank Irina Belyaeva (RBG, Kew) and the journal reviewers for helpful comments on the manuscript. QC gratefully acknowledges appointments as Visiting Professor at Queen Mary University of London, and as Honorary Research Associate at the Royal Botanic Gardens, Kew, which greatly facilitated this work.

REFERENCES

- Alford, M. (2005). *Systematic Studies in Flacourtiaceae*. Ph.D. thesis, Cornell University, Ithaca, NY.
- Alford, M. H., Brantley, R. J., Hernández, C. L., and Samarakoon, T. (2009). "What are the closest relatives of *Salix* and *Populus*?" in *Abstracts - Botany and Mycology 2009* (Snowbird, UT: Botanical Society of America).
- Angiosperm Phylogeny Group (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi: 10.1016/j.jep.2015.05.035
- Bateman, R. M. (1994). Evolutionary-developmental change in the growth architecture of fossil rhizomorphic lycopsids – scenarios constructed on cladistic foundations. *Biol. Rev.* 69, 527–597. doi: 10.1111/j.1469-185X.1994.tb01249.x
- Boes, T. K., and Strauss, S. H. (1994). Floral phenology and morphology of black cottonwood, *Populus trichocarpa* (Salicaceae). *Am. J. Bot.* 81, 562–567. doi: 10.2307/2445730
- Boucher, L. D., Manchester, S. R., and Judd, W. S. (2003). An extinct genus of Salicaceae based on twigs with attached flowers fruits, and foliage from the Eocene Green River Formation of Utah and Colorado, USA. *Am. J. Bot.* 90, 1389–1399. doi: 10.3732/ajb.90.9.1389
- Brunner, A. M., Evans, L. M., Hsu, C.-Y., and Sheng, X. (2014). Vernalization and the chilling requirement to exit bud dormancy: shared or separate regulation? *Front. Plant Sci.* 5:732. doi: 10.3389/fpls.2014.00732
- Chase, M. W., Zmarzty, S., Lledó, D., Wurdack, K., Swensen, S. M., and Fay, M. F. (2002). When in doubt, put it in Flacourtiaceae: a molecular phylogenetic analysis based on plastid rbcL DNA sequences. *Kew Bull.* 57, 141–181. doi: 10.2307/4110825
- Correa, J., Mamani, M., Munoz-Espinoza, C., Laborie, D., Munoz, C., Pinto, M., et al. (2014). Heritability and identification of QTLs and underlying candidate genes associated with the architecture of the grapevine cluster (*Vitis vinifera* L.). *Theor. Appl. Genet.* 127, 1143–1162. doi: 10.1007/s00122-014-2286-y

- Cronk, Q. C. B. (2005). Plant eco-devo: the potential of poplar as a model organism. *New Phytol.* 166, 39–48. doi: 10.1111/j.1469-8137.2005.01369.x
- Darlington, C. D., and Wylie, A. P. (1955). *Chromosome Atlas of Flowering Plants*. London: George Allen and Unwin Ltd.
- Davis, C. C., Webb, C. O., Wurdack, K. J., Jaramillo, C. A., and Donoghue, M. J. (2005). Explosive radiation of malpighiales supports a mid-Cretaceous origin of modern tropical rain forests. *Am. Nat.* 165, E36–E65. doi: 10.1086/428296
- Dodoens, R. (1578). *A Nieuwe Herball, or Historie of Plantes*. London: G. Dewes.
- Du Mortier, B.-C. (1825). *Verhandeling Over het Geslacht der Wilgen (Salix) en de Natuurlijke Familie der Amentaceae [Treatise on the Willow Genus, Salix and the Natural Family Amentaceae]*. Amsterdam: J. van der Hey.
- Eichler, A. W. (1883). *Syllabus der Vorlesungen über Spezielle und Medicinisch-Pharmaceutische Botanik*. Berlin: Gebrüder Borntraeger.
- Fang, Z., Fang, C., and Skvortsov, A. K. (1999). “Salicaceae,” in *Flora of China*, eds Z. Y. Wu and P. H. Raven (Beijing: Science Press), 162–274.
- Filatov, D. A. (2015). Homomorphic plant sex chromosomes are coming of age. *Mol. Ecol.* 24, 3217–3219. doi: 10.1111/mec.13268
- Fisher, M. J. (1928a). The morphology and anatomy of the flowers of the Salicaceae I. *Am. J. Bot.* 15, 307–326. doi: 10.2307/2435831
- Fisher, M. J. (1928b). The morphology and anatomy of the flowers of the Salicaceae II. *Am. J. Bot.* 15, 372–394. doi: 10.2307/2435831
- Geraldes, A., Hefer, C. A., Capron, A., Kolosova, N., Martinez-Núñez, F., Soolanayakanahally, R. Y., et al. (2015). Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). *Mol. Ecol.* 24, 3243–3256. doi: 10.1111/mec.13126
- Goodstein, D. M., Shu, S. Q., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Goosey, L., and Sharrock, R. (2001). The *Arabidopsis* compact inflorescence genes: phase-specific growth regulation and the determination of inflorescence architecture. *Plant J.* 26, 549–559. doi: 10.1046/j.1365-313x.2001.01053.x
- Grill, I. (1990). *Beiträge zur Karyosystematik der Flacourtiaceae*. Ph.D. dissertation, Universität Wien, Vienna.
- Hallé, F., Oldeman, R. A. A., and Tomlinson, P. B. (1978). *Tropical Trees and Forests — An Architectural Analysis*. Heidelberg: Springer-Verlag.
- Karrenberg, S., Kollmann, J., and Edwards, P. J. (2002). Pollen vectors and inflorescence morphology in four species of *Salix*. *Plant Syst. Evol.* 235, 181–188. doi: 10.1007/s00606-002-0231-z
- Kaul, R. B. (1995). Reproductive structure and organogenesis in a cottonwood, *Populus deltoides* (Salicaceae). *Int. J. Plant Sci.* 156, 172–180. doi: 10.1086/297238
- Knuth, P. (1909). *Handbook of Flower Pollination: Based upon Hermann Müller's Work 'The Fertilisation of Flowers by Insects'*, Vol. 3. Oxford: Clarendon Press.
- Korotkova, N., Schneider, J. V., Quandt, D., Worberg, A., Zizka, G., and Borsch, T. (2009). Phylogeny of the eudicot order Malpighiales: analysis of a recalcitrant clade with sequences of the petD group II intron. *Plant Syst. Evol.* 282, 201–228. doi: 10.1007/s00606-008-0099-7
- Kyozuka, J., Tokunaga, H., and Yoshida, A. (2014). Control of grass inflorescence form by the fine-tuning of meristem phase change. *Curr. Opin. Plant Biol.* 17, 110–115. doi: 10.1016/j.pbi.2013.11.010
- Leskinen, E., and Alstrom-Rapaport, C. (1999). Molecular phylogeny of Salicaceae and closely related Flacourtiaceae: evidence from 5.8 S, ITS 1 and ITS 2 of the rDNA. *Plant Syst. Evol.* 215, 209–227. doi: 10.1007/BF00984656
- Manchester, S. R., Judd, W. S., and Handley, B. (2006). Foliage and fruits of early poplars (Salicaceae: *Populus*) from the eocene of Utah, Colorado, and Wyoming. *Int. J. Plant Sci.* 167, 897–908. doi: 10.1086/503918
- Manning, W. E. (1938). The morphology of the flowers of the Juglandaceae. 1. The inflorescence. *Am. J. Bot.* 25, 407–419. doi: 10.2307/2436415
- Ostaf, D. P., Mosseler, A., Johns, R. C., Javorek, S., Klymko, J., and Ascher, J. S. (2015). Willows (*Salix* spp.) as pollen and nectar sources for sustaining fruit and berry pollinating insects. *Can. J. Plant Sci.* 95, 505–516. doi: 10.4141/cjps-2014-339
- Park, S. J., Eshed, Y., and Lippman, Z. B. (2014). Meristem maturation and inflorescence architecture—lessons from the Solanaceae. *Curr. Opin. Plant Biol.* 17, 70–77. doi: 10.1016/j.pbi.2013.11.006
- Park, S. J., Jiang, K., Schatz, M. C., and Lippman, Z. B. (2012). Rate of meristem maturation determines inflorescence architecture in tomato. *Proc. Natl. Acad. Sci. U.S.A.* 109, 639–644. doi: 10.1073/pnas.1114963109
- Rohwer, J., and Kubitzki, K. (1984). *Salix martiana*, a regularly hermaphrodite willow. *Plant Syst. Evol.* 144, 99–101. doi: 10.1007/BF00986668
- Rudall, P. J., and Bateman, R. M. (2004). Evolution of zygomorphy in monocot flowers: iterative patterns and developmental constraints. *New Phytol.* 162, 25–44. doi: 10.1111/j.1469-8137.2004.01032.x
- Sanders, R. W., Stuessy, T. W., and Rodriguez, R. (1983). Chromosome numbers from the flora of the Juan Fernandez Islands. *Am. J. Bot.* 70, 799–810. doi: 10.2307/2442929
- Skvortsov, A. K. (1999). *Willows of Russia and Adjacent Countries. Taxonomical and Geographical Revision*. Joensuu: Joensuu University.
- Sleumer, H. O. (1980). *Flacourtiaceae. Flora Neotropica Monograph 22*. New York, NY: New York Botanical Garden.
- Sterck, L., Rombauts, S., Jansson, S., Sterky, F., Rouze, P., and Van De Peer, Y. (2005). EST data suggest that poplar is an ancient polyploid. *New Phytol.* 167, 165–170. doi: 10.1111/j.1469-8137.2005.01378.x
- Stern, W. L. (1973). Development of the amentiferous concept. *Brittonia* 25, 316–333. doi: 10.2307/2805638
- Tsukaya, H., and Tsuge, T. (2001). Morphological adaptation of inflorescences in plants that develop at low temperatures in early spring: the convergent evolution of “downy plants.” *Plant Biol.* 3, 536–543. doi: 10.1055/s-2001-17727
- Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604.
- Yang, Q., and Zmartzy, S. (2007). “Flacourtiaceae,” *Flora of China: Clusiaceae through Araliaceae*, Vol. 13, eds Z. Wu and P. H. Raven (Beijing: Science Press), 112–138.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Cronk, Needham and Rudall. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Phenotypic and Genetic Underpinnings of Flower Size in Polemoniaceae

Jacob B. Landis^{1,2*}, Rebecca D. O'Toole², Kayla L. Ventura², Matthew A. Gitzendanner¹, David G. Oppenheimer^{1,3,4}, Douglas E. Soltis^{1,2,3,4} and Pamela S. Soltis^{2,3,4}

¹ Department of Biology, University of Florida, Gainesville, FL, USA, ² Florida Museum of Natural History, University of Florida, Gainesville, FL, USA, ³ Genetics Institute, University of Florida, Gainesville, FL, USA, ⁴ Plant Molecular and Cellular Biology Graduate Program, University of Florida, Gainesville, FL, USA

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington, USA

Reviewed by:

Sinead Drea,
University of Leicester, UK
Quentin Cronk,
University of British Columbia, Canada

*Correspondence:

Jacob B. Landis
jblandis@ufl.edu

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 05 October 2015

Accepted: 02 December 2015

Published: 05 January 2016

Citation:

Landis JB, O'Toole RD, Ventura KL,
Gitzendanner MA, Oppenheimer DG,
Soltis DE and Soltis PS (2016) The
Phenotypic and Genetic
Underpinnings of Flower Size in
Polemoniaceae.
Front. Plant Sci. 6:1144.
doi: 10.3389/fpls.2015.01144

Corolla length is a labile flower feature and has strong implications for pollinator success. However, the phenotypic and genetic bases of corolla elongation are not well known, largely due to a lack of good candidate genes for potential genetic exploration and functional work. We investigate both the cellular phenotypic differences in corolla length, as well as the genetic control of this trait, in *Saltugilia* (Polemoniaceae). Taxa in this clade exhibit a large range of flower sizes and differ dramatically in pollinator guilds. Flowers of each species were collected from multiple individuals during four stages of flower development to ascertain if cell number or cell size is more important in determining flower size. In *Saltugilia*, increased flower size during development appears to be driven more by cell size than cell number. Differences in flower size between species are governed by both cell size and cell number, with the large-flowered *S. splendens* subsp. *grantii* having nearly twice as many cells as the small-flowered species. Fully mature flowers of all taxa contain jigsaw cells similar to cells seen in sepals and leaves; however, these cells are not typically found in the developing flowers of most species. The proportion of this cell type in mature flowers appears to have substantial implications, comprising 17–68% of the overall flower size. To identify candidate genes responsible for differences in cell area and cell type, transcriptomes were generated for two individuals of the species with the smallest (*S. australis*) and largest (*S. splendens* subsp. *grantii*) flowers across the same four developmental stages visualized with confocal microscopy. Analyses identified genes associated with cell wall formation that are up-regulated in the mature flower stage compared to mid-stage flowers (75% of mature size). This developmental change is associated with the origin of jigsaw cells in the corolla tube of mature flowers. Further comparisons between mature flowers in the two species revealed 354 transcripts that are up-regulated in the large-flowered *S. splendens* subsp. *grantii* compared to the small-flowered *S. australis*. These results are likely broadly applicable to Polemoniaceae, a clade of nearly 400 species, with extensive variation in floral form and shape.

Keywords: flower epidermal cells, comparative transcriptomics, pollinator-mediated selection, Polemoniaceae, floral evo-devo, phylogeny, *Saltugilia*

INTRODUCTION

The vast morphological diversity observed in angiosperms is often considered to result in part from the interaction between flowers and their pollinators (e.g., Crepet, 1995; Waser et al., 1996; Crepet and Niklas, 2009; Van der Niet and Johnson, 2012; Kearns et al., 2015). Differences in morphological phenotypes are often associated with different pollinator classes to form pollination syndromes (Fenster et al., 2004) that may comprise many floral traits (Waser, 2006), with many studies showing the selection pressures certain pollinators place on these floral traits (Bruneau, 1997; Johnson et al., 1998; Fulton and Hodges, 1999; Schemske and Bradshaw, 1999; Whittall and Hodges, 2007; Cronk and Ojeda, 2008; Brunet, 2009).

Changes in floral traits associated with different pollination syndromes, especially traits involving size and number of organs, have been observed to be highly labile (Stebbins, 1974; Givnish, 2002; Lock et al., 2011; Alapetite et al., 2014) and are mostly attributed to shifts in timing, rates and/or patterns of gene expression (Pina et al., 2014). One important and labile floral trait associated with pollinator selection, corolla length, has been shown experimentally to be heritable and selectable (Mitchell and Shaw, 1993; Kaczorowski et al., 2008; Gómez et al., 2009). The length of the corolla is often more stable than the spread (width) of the corolla, with the spread being more highly modified under both internal and external stressors (Goodspeed and Clausen, 1915).

Corolla development is typically marked by two phases of growth: early growth involving cell division, followed by cell expansion in later growth (Irish, 2008; Hepworth and Lenhard, 2014). Where organ growth (i.e., corolla size) due to cell expansion is involved, enlargement of the cell wall must occur with either expansion of vacuoles or endoreduplication (Weiss et al., 2005; Anastasiou and Lenhard, 2007). However, the relative contribution of cell division and cell elongation to corolla growth is debated and likely varies among species (Martin and Gerats, 1993; Stuurman et al., 2004), with the shape of petal epidermal cells having been investigated in many angiosperm families (Kay et al., 1981; Ojeda et al., 2009, 2012). Recent experiments show that formation and extension of petal nectar spurs in both *Aquilegia* (Ranunculaceae) and *Centranthus* (Caprifoliaceae) are controlled primarily by cell elongation (Puzey et al., 2012; Mack and Davis, 2015), raising the hypothesis that corolla length may operate under the same genetic control mechanisms. However, the genetic components of flower size, specifically the control of cell number, cell size and cell shape, are not well understood (Glover, 2007; Ojeda et al., 2012).

Indirect evidence primarily from *Arabidopsis thaliana* and *Antirrhinum majus*, representing the evolutionarily divergent lineages of rosids and asterids, respectively, indicates several potential candidate genes that may be involved in corolla elongation, such as *JAGGED*, *AINTEGUMENTA*, *ARGOS*, *BFP*, *OPR3*, *ARF8*, *BIG BROTHER*, *KLH*, *DAI* and *GAST1* (Herzog et al., 1995; Elliott et al., 1996; Kotilainen et al., 1999; Mizukami and Fischer, 2000; Mizukami, 2001; Kim et al., 2002; Ben-Nissan et al., 2004; Disch et al., 2006; Krizek, 2009; Xu and Li, 2011; Chang et al., 2014). Additional studies in *Petunia*

have identified at least five QTL involved with corolla tube morphology and elongation (Stuurman et al., 2004; Galliot et al., 2006). Despite the identification of these candidate genes, further studies of expression and function involving any of these genes are lacking in non-model species. To elucidate the phenotypic, developmental and genetic components of flower size in the context of pollination biology, we are focusing on the phlox family, Polemoniaceae.

Polemoniaceae (26 genera, 387 species; Johnson et al., 1996, 2008; Prather et al., 2000; Ferguson and Jansen, 2002; Landis et al., in review) comprise annual and perennial plants native to North and South America, with the center of diversity in western North America (Grant, 1959). This family has a wide diversity of corolla length, varying from 3 mm in the mid-elevation selfing species *Lathrocasis tenerrima* to over 70 mm in species of *Cantua* and *Cobaea* (Schönenberger, 2009; Landis et al., in review). Polemoniaceae have long been a model for studies of pollination biology (e.g., Grant and Grant, 1965). More recent studies have addressed variation in corolla length, with bees selecting for longer tubes and funnel-shaped flowers in *Polemonium* (Galen and Cuba, 2001), increased moisture availability resulting in increased flower size in *Leptosiphon* (Lambrecht, 2013) and longer-flowered individuals excluding bees in four species of *Phlox* (Strakosh and Ferguson, 2005).

The focus of the current study is *Saltugilia*, which comprises four species—*S. australis*, *S. caruifolia*, *S. latimeri*, and *S. splendens* (Porter and Johnson, 2000; Johnson, 2007)—that differ primarily in corolla morphology (Weese and Johnson, 2005) and exhibit a 2.5-fold range in corolla size. These taxa also vary in observed pollinators, with *S. australis* and *S. latimeri* being autogamous, *S. caruifolia* pollinated by bees and the different subspecies of *S. splendens* pollinated by hummingbirds (*S. splendens* subsp. *grantii*) and bee flies (*S. splendens* subsp. *splendens*) (Grant and Grant, 1965; Weese and Johnson, 2005).

Our overall goal is to integrate phylogenetic, phenotypic and developmental analysis of floral traits associated with pollination syndromes and observed differences in corolla length, while linking the underlying genetic components of these differences to elucidate the mechanisms of pollinator-mediated selection in representatives of Polemoniaceae. Specifically, we investigated (1) the relationships of all currently recognized taxa in *Saltugilia* by reconstructing the phylogeny with both nuclear and plastid markers to provide a framework for the remaining goals; (2) the cellular component of differences in flower size to address the question of whether differences in flower size are predominantly controlled by cell size or cell number, or a combination of the two; and (3) the genetic underpinnings associated with differences in flower size in *Saltugilia*.

MATERIALS AND METHODS

Plant Material

Four taxa of *Saltugilia* and one of *Gilia* were grown in the greenhouses at the University of Florida from seeds obtained from Rancho Santa Ana Botanic Garden (Claremont, CA,

USA), the Ornamental Germplasm Center at The Ohio State University (Columbus, OH, USA) and Leigh Johnson (Brigham Young University; Provo, UT, USA): *G. brecciarum* subsp. *brecciarum* (W6 30785), *S. australis* (Johnson BYU), *S. caruifolia* (RSABG 19148), *S. splendens* subsp. *grantii* (RSABG 21757) and *S. splendens* subsp. *splendens* (RSABG 22676) (Table 1). Two additional samples of *Saltugilia* and one additional species of *Gilia* were collected in California in the spring of 2015: *G. stellata* (JOTR34200), *S. latimeri* (UCR261592) and a field accession of *S. splendens* subsp. *splendens* (JOTR32513), which differed morphologically from the accession obtained from Rancho Santa Ana Botanic Garden. Collections of *S. splendens* subsp. *splendens* and *G. stellata* were obtained with the help of Tasha LaDoux (University of California Riverside, Riverside, CA, USA), with the above accession numbers for those species representing previous collections from the same localities sampled here.

Phylogenetic Analyses

DNA was extracted from samples of all eight taxa following a modified CTAB extraction protocol (Doyle and Doyle, 1987). Rehydrated DNA was sonicated to a targeted length of 300-bp fragments using a Covaris S220 sonicator (Covaris, Inc., Woburn, MA, USA) following the manufacturer's suggested protocol. A total of 3–5 µg of sheared DNA from each sample was sent to RapidGenomics (Gainesville, FL, USA) for Illumina library preparation with dual indexed barcodes and targeted exon capture using MYbaits probes (MYcroarray; Ann Arbor, MI, USA). This approach was used to generate 100 single-copy nuclear genes for phylogenetic analysis. Probes were designed from transcriptomes of four species [*Fouquieria macdouglaii* (Fouquieriaceae), *Phlox drummondii* (Polemoniaceae), *Phlox* sp.,

and *Ternstroemia gymnanthera* (Pentaphylacaceae)] available in the IKP data set (www.onekp.com) and the *Arabidopsis* genome using MarkerMiner (Chamala et al., 2015) to find putative single-copy genes. One hundred nuclear genes were selected with exons of at least 300 bp and introns smaller than 600 bp. Probes were designed for 120 kmers and tiled 3x for additional capture efficiency. Capture products were pooled with additional samples and distributed across two runs: Illumina NextSeq 500 (2 × 150 bp) mid-throughput and an Illumina HiSeq 2000 (2 × 100 bp).

Raw reads were processed using two custom scripts (available on Github; https://github.com/soltislab/get_BLAT_reads). With this pipeline, reads were trimmed and filtered using cutadapt (Martin, 2011) and Sickle (Joshi and Fass, 2011), followed by a BLAT (Kent, 2002) analysis to isolate (1) plastid reads using the complete plastome reference of *Phlox amabilis* (provided by J. Mark Porter, unpublished) and (2) nuclear reads using a representative of each of the 100 nuclear genes. On-target reads were imported to Geneious (version 8.0.5; Biomatters Limited, Auckland, New Zealand) and mapped to the reference sequences (*P. amabilis* plastome for plastid coding genes and the complete plastome; one representative for each of the 100 loci for nuclear genes) (Supplemental Data 1) using the Geneious mapper with the following settings: medium sensitivity with 10 iterations, majority threshold to reduce ambiguities, with no coverage and coverage of less than 2 sequences coded as missing data.

Concatenated consensus reads for each taxon, consisting of 80 coding regions of the plastome, the complete plastome including coding genes and spacer regions and 90 nuclear loci for which there was sufficient coverage and overlap among taxa, were aligned using MAFFT (version 7.245; Katoh and Standley,

TABLE 1 | Plant material information, including herbarium accessions, locality of collection for field samples and source of seed material as well as sequencing coverage for phylogenetic analyses.

Species	Growth facility/Locality collected	Accession/Voucher	Plastid coding genes/80 total (bp)	Plastome length (bp)/percentage of plastome covered (coverage) and coverage	Nuclear genes/90 captured (bp)
<i>Gilia brecciarum</i> subsp. <i>brecciarum</i>	UF Greenhouse	OGPC W6 30785	80/80 (64,439)	152,899/99.4% (403x)	90/90 (143,769)
<i>Gilia stellata</i>	UF Greenhouse	JOTR34200	80/80 (64,495)	153,396/99.7% (926x)	90/90 (141,693)
<i>Saltugilia australis</i>	UF Greenhouse	L. Johnson Brigham Young University	73/80 (63,380)	124,673/81.0% (11.4x)	82/90 (90,130)
<i>Saltugilia caruifolia</i>	UF Greenhouse	RSABG 19148	62/80 (58,418)	96,715/62.9% (1.7x)	46/90 (34,942)
<i>Saltugilia latimeri</i>	Joshua Tree National Park/Big Horn Mountains	UCR261592	80/80 (64,491)	153,396/99.7% (825x)	90/90 (142,909)
<i>Saltugilia splendens</i> subsp. <i>grantii</i>	UF Greenhouse	RSABG 21757	74/80 (64,178)	132,303/86.0% (5.3x)	60/90 (64,429)
<i>Saltugilia splendens</i> subsp. <i>splendens</i>	UF Greenhouse	RSABG 22676	79/80 (64,178)	135,926/88.3% (11.2x)	87/90 (106,862)
<i>Saltugilia splendens</i> subsp. <i>splendens</i>	Joshua Tree National Park	JOTR32513	80/80 (64,497)	153,504/99.8% (1,379x)	90/90 (144,653)

2013) installed on the University of Florida Research Computing cluster. Individual alignments of plastid coding genes and nuclear genes were analyzed in PartitionFinder (version 1.1.1; Lanfear et al., 2012) to find the best partitions for maximum likelihood (ML) inference. Separate phylogenetic analyses of the two plastid data sets and nuclear genes were conducted using RAXML (version 8.2.2; Stamatakis, 2014) installed on the University of Florida Research Computing cluster. The two species of *Gilia* were used as outgroups.

Plastid coding regions and nuclear loci were analyzed with 1000 bootstrap replicates with the preferred partitions identified by PartitionFinder for all gene partitions, while the complete plastome was analyzed with 1000 bootstrap replicates using the GTR+G substitution model chosen by jModeltest (version 2.1.1; Darriba et al., 2012). The two separate analyses of the plastome data were conducted to assess the potential impact that missing data had on the phylogenetic reconstructions. Raw reads for each accession were deposited in GenBank's Short Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) and assembled plastomes and nuclear genes were submitted to GenBank (all accession numbers are given in Supplemental Table 1). Concatenated alignment files for each dataset are deposited in Figshare (DOI: 10.6084/m9.figshare.2007546, 10.6084/m9.figshare.2007549, and 10.6084/m9.figshare.2007555).

Following the completion of phylogenetic analyses, ancestral-state reconstructions of pollinators and flower size were conducted on the resulting topologies. Both a maximum parsimony (MP) and ML framework were implemented using Mesquite (version 3.03; Maddison and Maddison, 2015) for reconstructing pollinators. Pollinator states included autogamous (*G. brecciarum* subsp. *brecciarum*, *G. stellata*, *S. australis* and *S. latimeri*), bee pollination (*S. caruifolia*), hummingbird pollination (*S. splendens* subsp. *grantii*) and bee fly pollination (*S. splendens* subsp. *splendens*). For flower size, a continuous reconstruction was conducted in Mesquite using the average flower size of measured flowers, with flower length ranging from 0.8 to 2.5 cm.

Cell Morphology

Flower material of four developmental stages was collected from four individual plants per species when available (all stages and individuals could not be collected from all field samples due to insufficient flowering material on the plant): mature (flowers at anthesis), mid (75% of total flower length at anthesis), half (50% of total flower length at anthesis) and small (25% of total flower length at anthesis). Sample preparation and confocal microscopy were carried out following Landis et al. (2015). Imaging of cells was conducted using the methods of Landis et al. (2015), using a Zeiss Axiocam HRm camera mounted on a Zeiss Axioplan 2 Imaging microscope (Jena, Germany) and Axiovision software. Green fluorescence was obtained using Zeiss filter set 10 (excitation wavelengths, 45–490 nm; dichroic, 510 nm LP; emission wavelengths, 515–565 nm), with a 40x magnification lens and Apotome with optical sectioning with optical slice distance of 0.675 μm . Confocal images were imported into Fiji (<http://fiji.sc/Fiji>; Schindelin et al., 2012) and up to 50 cells per image were outlined and measured for area, perimeter, shape

descriptor (circularity value; Glasbey and Horgan, 1995) and bounding box length and width (the smallest box possible to encompass each cell).

Circularity values were calculated with the following formula: $\text{circularity} = 4\pi(\text{area}/\text{perimeter}^2)$. A value of 1.0 indicates a perfect circle, with values approaching 0 indicating increasingly elongated cells. One-way ANOVA and Tukey's honest significance tests were performed in R (version 3.2.1) with generation of box plots for the four developmental stages across all taxa, with data from all individuals per taxon pooled. Pooled data were also used to calculate mean and standard deviation of cell size for each cell type in each stage of development. Estimates of cell number were conducted for each cell type by taking the mean values of the width of the bounding box. The total length of each cell type was then divided by these values to obtain estimates of cell number.

Transcriptome Analysis

Total RNA was extracted from flowers at each of the four developmental stages of two individuals each of the small-flowered *S. australis* and the large-flowered *S. splendens* subsp. *grantii* using Tri-Reagent following the manufacturer's directions (Ambion, Austin, TX, USA). Library preparation of these 16 samples was conducted using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, Ipswich, MA, USA) following the manufacturer's directions and NEBNext Multiplex Oligos for Illumina Index Primers Set 1 barcodes (NEB, Ipswich, MA, USA). Transcriptome sequencing was performed at the University of Florida Interdisciplinary Center for Biotechnology Research using two runs on the Illumina NextSeq 500 (2 \times 150 bp) mid-throughput with 8 samples pooled per run. Raw reads were trimmed and filtered using the same pipeline described above minus the BLAT analysis. Following cleaning, *de novo* assemblies using the published Trinity pipeline were conducted (Grabherr et al., 2013; Haas et al., 2013).

A reference assembly was generated for each species with all reads from all stages and both individuals pooled. The *S. australis* reference was constructed with 66,850,418 paired-end reads and an additional 3,610,028 singleton reads, while the *S. splendens* subsp. *grantii* reference was constructed with 90,875,540 paired-end reads and 3,922,978 singleton reads. A final reference was created with all raw reads from both species for downstream analysis of gene expression between species. A BLAST analysis was conducted on each reference, *S. australis*, *S. splendens* subsp. *grantii* and the combined reference, to determine the number of known proteins that match a transcript by at least 80% coverage, with only the top transcript hit for each protein returned (UniProt Consortium, 2015). All raw reads for each species were then mapped back to the reference assemblies using RNA-Seq by Expectation-Maximization as implemented in Trinity, and filtered by 1 fpkm to remove artifacts that were formed in the *de novo* assemblies and to reduce the false discovery rate. Each filtered reference assembly was then used to map raw reads of each developmental stage for each of the two individuals for the respective species. These newly created fasta files were translated into open reading frames and protein sequence using the Transdecoder plugin in Trinity. Reference assemblies were

annotated with the Trinotate pipeline to identify GO categories in the Swiss-Prot database (UniProt Consortium, 2015).

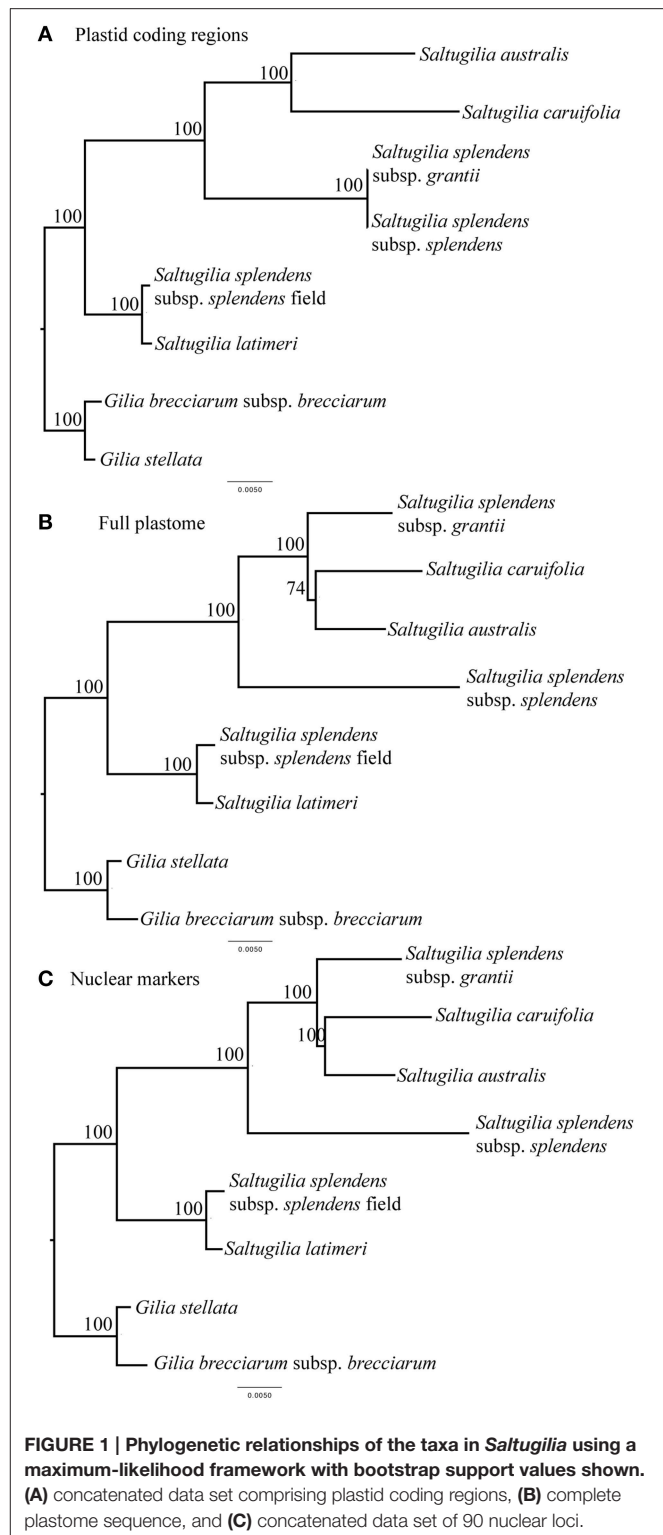
Translated files were then used in OrthoVenn (Wang et al., 2015) using default parameters to compare presence/absence for each stage and species, as well as to identify GO categories and any GO enrichments for different stages of development. Differential gene expression analyses were conducted using the EdgeR Bioconductor package (Robinson et al., 2010) for comparison within species at different developmental stages. For *S. australis*, two biological replicates for half (50%), mid (75%), and mature (100%) stages were used with a p-value cut-off of 0.05 and a 4-fold expression change. The small stage (25%) of *S. australis* was not included due to low confidence in one of the assemblies, in which only 13 genes met the 1 fpkm threshold used for filtering. Comparisons of developmental stages in *S. splendens* subsp. *grantii* were conducted in a different fashion due to a high false discovery rate when the preceding method was used. Similar problems in examining fold-changes in genes with low expression have been observed before (Butler et al., 2014). For *S. splendens* subsp. *grantii*, reads for the four stages were pooled prior to the differential expression analysis, resulting in only one input for each stage. This method leads to the inability to perform statistical analyses on the levels of expression; however, it provides evidence of candidate genes to compare to the results found in the other species (Butler et al., 2014). One final comparison was performed using the mature (100%) stages of both individuals of *S. australis* and *S. splendens* subsp. *grantii* to compare expression levels in the final stages of development between species using the reference assembly created using all raw reads. The goal of this last comparison was to investigate any expression differences between previously reported candidate genes.

RESULTS

Phylogenetic Analysis

Three separate analyses were performed: plastid coding regions only, complete or nearly complete plastome and nuclear markers. For the data set of the complete plastid genome, coverage of the reference plastome ranged from 62.9% (96,715 of 153,853 bp) for *S. caruifolia* to 99.8% for the field accession of *S. splendens* subsp. *splendens* (153,504 of 153,853 bp). Average coverage depth of the plastome ranged from 1.7x to 1379x among samples. For the nuclear genes, the number of loci for each individual ranged from 52 to 90 (Table 1). The plastid coding alignment consisted of 64,500 bp, with 11.5% missing data (cells were coded as either missing if data were lacking or Ns due to low coverage), while the complete plastome alignment had 157,884 bp with 20.2% missing data. For the 90 nuclear genes that had sufficient coverage and overlap between accessions, the concatenated alignment was 147,028 bp with 37.2% missing data.

The phylogenetic trees for all three analyses are compared in Figure 1. In the tree based on plastid coding regions, *S. australis* is sister to *S. caruifolia*, with that clade sister to *S. splendens* subsp. *grantii* and *S. splendens* subsp. *splendens*. This four-taxon clade is sister to *S. latimeri* and the field accession of *S. splendens*



subsp. *splendens*. Bootstrap support at all nodes is 100%. The trees derived from the complete plastome and the nuclear genes data sets are identical and give a slightly different topology than that from the plastid coding regions, with *S. splendens* subsp. *grantii* sister to *S. australis* and *S. caruifolia*, with *S. splendens*

subsp. *splendens* sister to these three taxa. Bootstrap support for all of these nodes is also 100%, except for the node leading to *S. australis* and *S. caruifolia*, which has a bootstrap value of 74% in the complete plastome tree. The relationships of *S. latimeri* and the field accession of *S. splendens* subsp. *splendens* are the same in the trees resulting from all three data sets. The incongruence between the tree derived from the plastid coding genes vs. those of the other two data sets may be the result of a lack of phylogenetic signal in the plastid coding genes. The branch lengths leading to *S. splendens* subsp. *grantii* and *S. splendens* subsp. *splendens* are very short. With the additional data obtained from the complete plastome and nuclear genes, these taxa do not form a clade.

Morphology

The flowers of *Saltugilia* range in size from 0.8 to 2.5 cm. The smallest flowers were observed in *S. australis* (0.8–1.1 cm) and *S. latimeri* (0.8–1.0 cm), followed by *S. caruifolia* (1.0–1.2 cm), the greenhouse accession of *S. splendens* subsp. *splendens* (0.9–1.3 cm) and the field accession of *S. splendens* subsp. *splendens* at (1.1–1.5 cm), with the largest flowers belonging to *S. splendens* subsp. *grantii* (2.3–2.5 cm). The two species of *Gilia* both had flowers with corollas of 0.8–0.9 cm (**Figures 2A–E**).

Four cell types with different shapes were characterized in petal material: conical, transition, elongated and jigsaw cells (**Figures 2E–I**). The first three cell types were found in all

developmental stages (25, 50, 75, and 100%) with jigsaw cells appearing predominantly in mature flowers. Exceptions to this pattern were observed in the field accessions of *S. splendens* subsp. *splendens* and *S. latimeri*, which have jigsaw cells in half, mid and mature flowers, and in the greenhouse accession of *S. splendens* subsp. *splendens*, which lacks elongated cells at the base of the petal tube.

As the flower develops, conical cells maintain their circular shape with a circularity value ranging from 0.85 to 0.95; however, these cells become larger with a 2.4- to 5.5-fold increase in size. The median observed increase in size is a 3.5-fold increase in cell size between small (25%) and mature flowers of *S. caruifolia*, *S. splendens* subsp. *grantii*, *S. splendens* subsp. *splendens* and the field accession of *S. splendens* subsp. *splendens*. The outlier is *S. australis* with a fold change of 5.5 between small (25%) flowers and mature flowers, with both *Gilia* accessions showing the smallest fold change of 2.4 (*G. brecciarum* subsp. *brecciarum*) and 2.5 (*G. stellata*). Along the lobe/tube margin is a distinct cell type defined here as transition cells. These cells are more elongated than conical cells with a circularity value of approximately 0.7 in mature stages, although in earlier stages the circularity values are sometimes very close to those of conical cells (**Figure 3**). In developing flowers, the corolla tube is composed of elongated cells with a circularity value of 0.7 in small (25%) flowers, then becoming more elongated toward the base of the corolla tube, with circularity values declining in the half stage (0.6)

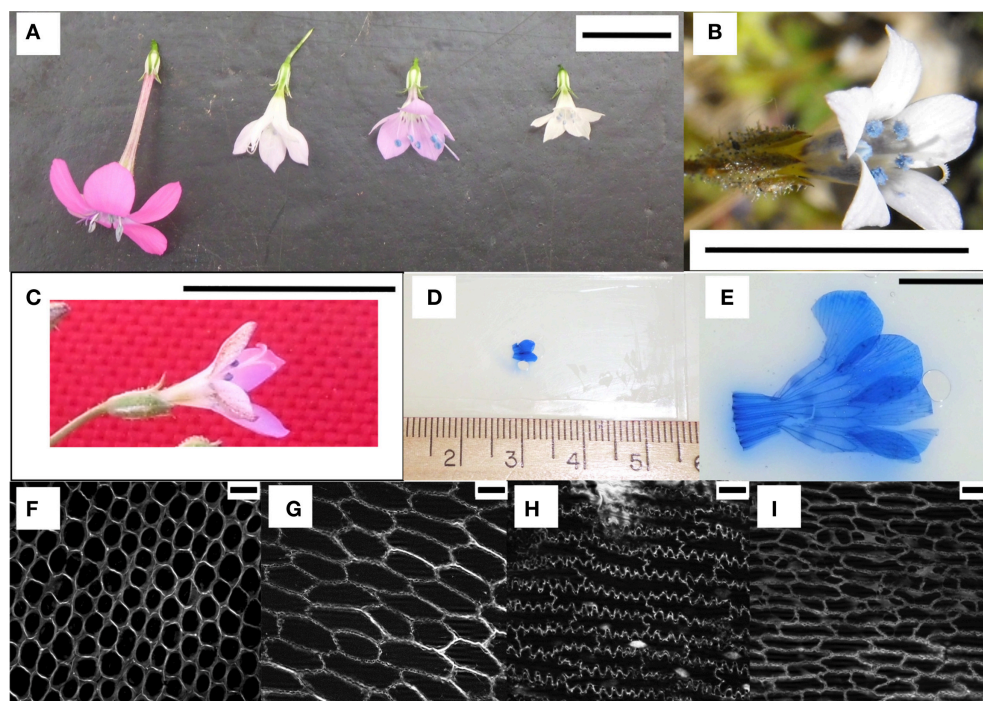
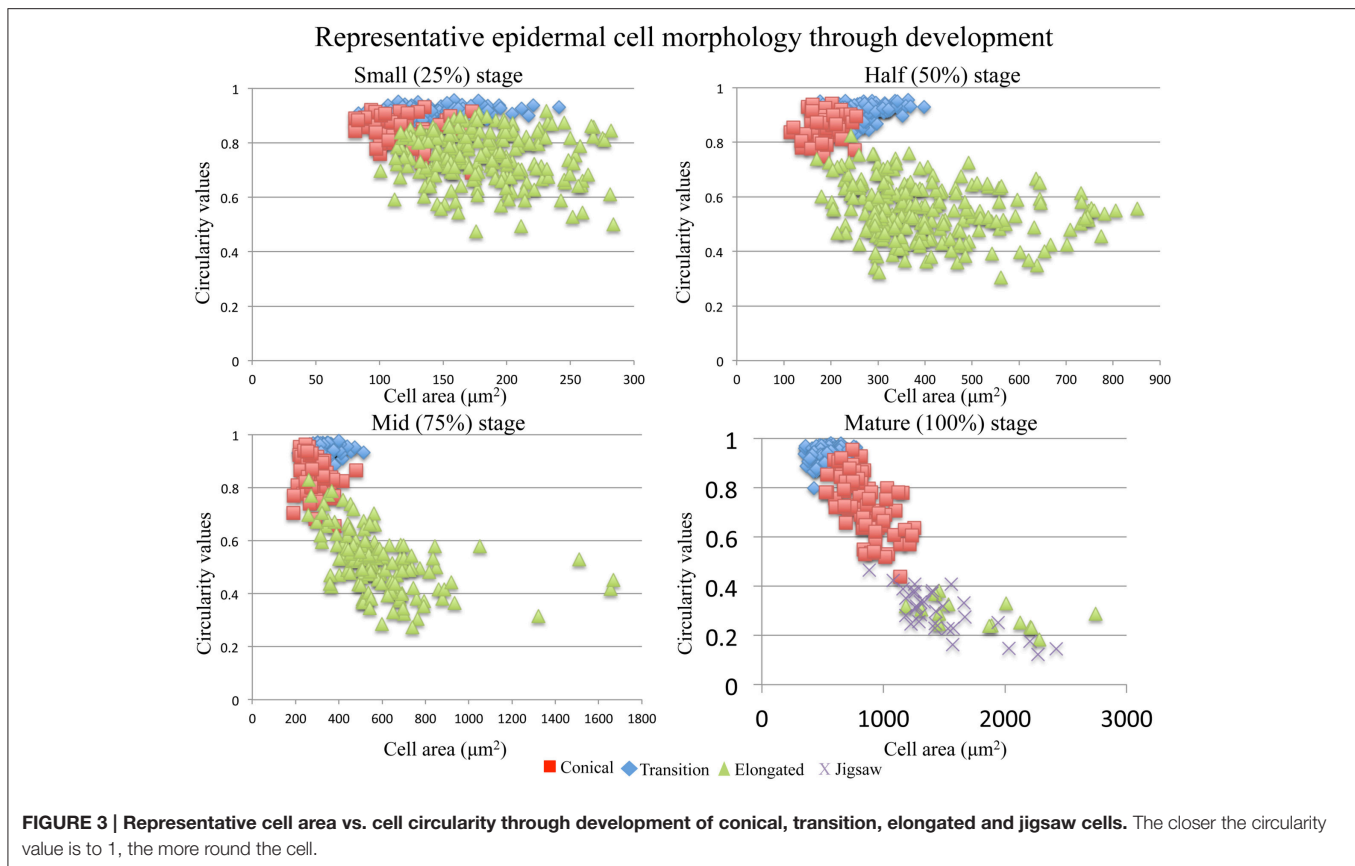


FIGURE 2 | Examples of flowers and cells visualized. (A) The four species of *Saltugilia* grown in the greenhouses at the University of Florida. From left to right: *S. splendens* subsp. *grantii*, *S. splendens* subsp. *splendens*, *S. caruifolia* and *S. australis*. **(B)** Flower of *Gilia stellata*, one of the species used as an outgroup. **(C)** *S. latimeri* collected in Joshua Tree National Park, CA, USA. **(D)** Prepped microscope slide of the small stage of *S. australis*. **(E)** Prepped microscope slide of the mature stage of *S. splendens* subsp. *grantii*. **(F–I)** Representatives of the four types of cells observed in flower material: conical, transition, jigsaw and elongated cells respectively. Scale bars in **(A–C,E)** are 1 cm, and **(F–I)** are 20 μ m.



and mid stage (0.5) and a circularity value of 0.2–0.4 in mature flowers. The average size of these cells increased by 5.2- to 9.6-fold between small and mature flowers, with the largest observed change in *S. splendens* subsp. *grantii* (Supplemental Table 2). In mature flowers, the corolla tube is composed of a second type of cells called jigsaw cells. These cells have a circularity of 0.2–0.3 and often have a slightly smaller area than elongated cells at the same developmental stage (Supplemental Figure 1).

As the flower develops, later stages of development have statistically larger cells than earlier developmental stages of the same cell type in all cases except in transition cells (ANOVA; $F = 542.6$, $p < 0.001$; Supplemental Figure 1). In *S. latimeri* and the field accession of *S. splendens* subsp. *splendens*, the transition cells between the half stage (50%) and the mid stage (75%) are not significantly different in size. In all taxa, the conical and transition cells do not differ significantly in size during the early developmental stages. Comparison of cell size between species was conducted for all four stages, with the mature comparisons visualized in **Figure 4**.

Conical cells in mature flowers overlap in size across the six accessions of *Saltugilia*, but are significantly different in size ($F = 128.8$, $p < 0.001$) except for cells of *S. splendens* subsp. *grantii*, which are not significantly different from *S. australis* ($p = 0.128$), *S. caruifolia* ($p = 0.999$) and the field accession of *S. splendens* subsp. *splendens* ($p = 0.807$). Jigsaw cells show two distinct

categories: the larger cells of *S. splendens* subsp. *grantii* and both accessions of *S. splendens* subsp. *splendens*, and the smaller cells found in *S. australis*, *S. caruifolia* and *S. latimeri* ($F = 38.79$, $p < 0.001$). The elongated cells of *S. caruifolia*, *S. splendens* subsp. *grantii* and the field accession *S. splendens* subsp. *splendens* are all similar in size, while the cells of *S. australis* are smaller ($F = 18.52$, $p = 0.001$). The elongated cells of *S. latimeri* are intermediate in size, being significantly larger than *S. australis* ($p < 0.001$) and significantly smaller than the field accession of *S. splendens* subsp. *splendens* ($p = 0.005$). However, no significant differences between the other comparisons of elongated cells were observed.

The relative abundance of each cell type changes across developmental stages, as well as among species (**Figure 5**). In early stages of development (small, 25%), conical cells make up roughly half of the total flower length, between 41.6 and 62.5%. As the flowers develop, the proportion of the flower composed of conical cells decreases to 25–42.8%, while the length of the corolla tube composed of conical cells increases. In *S. splendens* subsp. *splendens*, the proportion of the corolla composed of conical cells decreases from 57.1 to 31.3% through development, while the length of the corolla composed of conical cells increases from 0.23 to 0.38 cm. In general, the proportion of the corolla composed of elongated cells increases from the small to mid developmental stages, but then decreases substantially with the formation of jigsaw cells predominantly in mature flowers (Supplemental

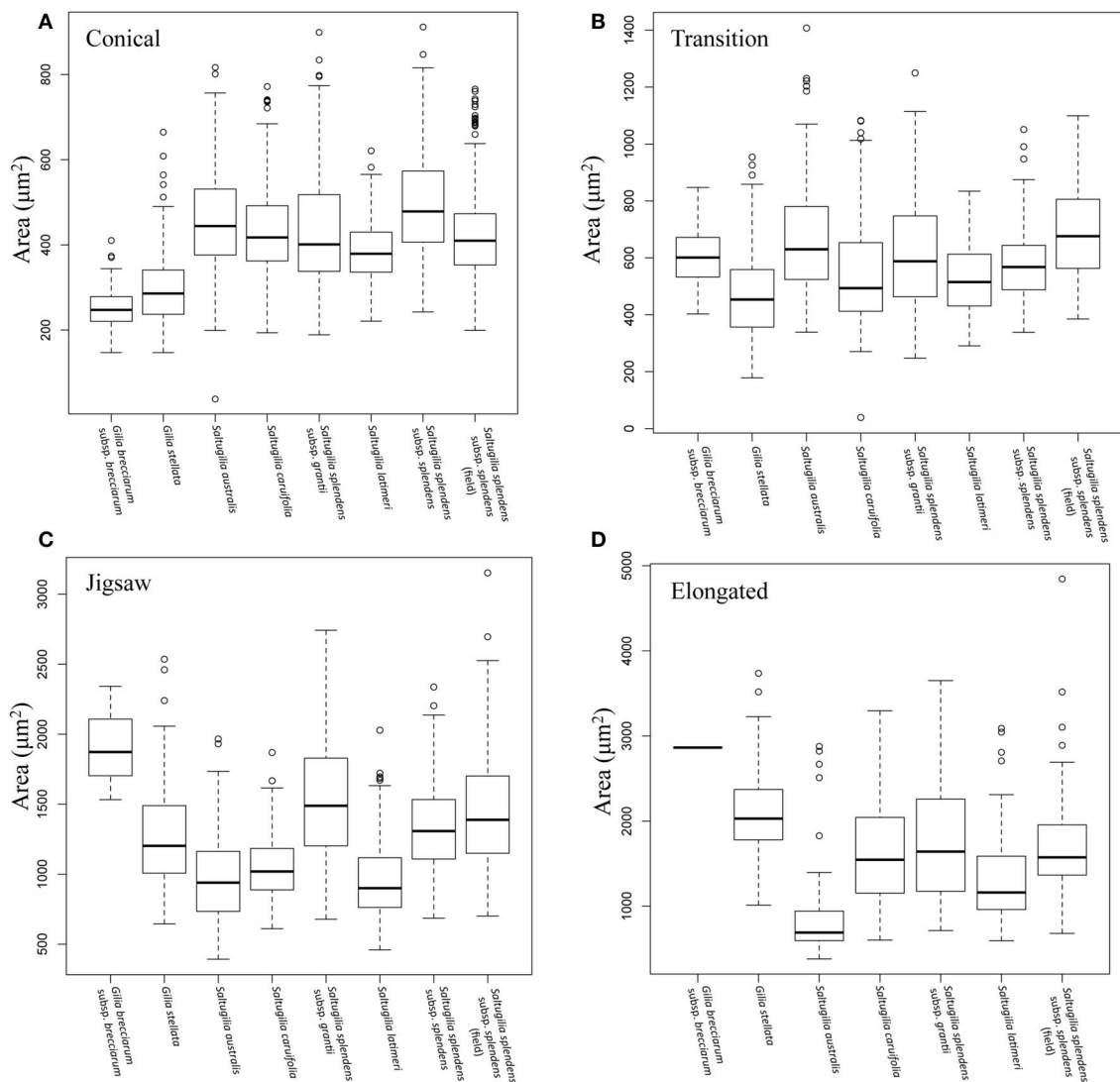


FIGURE 4 | Box plots comparing the area of each cell type in mature flowers of all species. (A) Mature conical cells, (B) mature transition cells, (C) mature jigsaw cells, (D) mature elongated cells.

Table 3). The two field-collected samples, *S. latimeri* and *S. splendens* subsp. *splendens*, both exhibit the formation of jigsaw cells starting at the half stage of development. *Gilia stellata* was also collected in the field, but does not exhibit jigsaw cells in any stage except the mature stage.

Estimates of total cell numbers, as well as the number of each cell type, were computed using the mean width of the bounding box around each measured cell (Supplemental Table 4). Only *S. splendens* subsp. *grantii* and the field accession of *S. splendens* subsp. *splendens* show increased numbers of cells between stages of development (Figure 6). These taxa have roughly 100 more cells in mature stages than at the small developmental stage. Estimates of the number of cells in mature flowers of the two *Gilia* species were similar, with a maximum of 219 cells in *G. brecciarum* subsp. *brecciarum* and 213 in *G. stellata*. The small-flowered species of *Saltugilia* have similar

estimated maximum cell numbers, 245 cells in *S. australis*, 236 in *S. latimeri* and 252 in *S. splendens* subsp. *splendens*. *Saltugilia caruifolia* has a maximum estimate of 344 cells, and the two taxa with the largest flowers, the field accession of *S. splendens* subsp. *splendens* and *S. splendens* subsp. *grantii*, have 350 and 567 cells, respectively.

Transcriptome Analysis

The filtered reference transcriptome of *S. australis* consists of 126,343 transcripts composed of 77,498,390 bases with a GC content of 44.7% and an n50 of 802 bp. The filtered reference for *S. splendens* subsp. *grantii* consists of 160,429 transcripts composed of 95,491,059 bases with a GC content of 44.5% and an n50 of 794 bp. The BLAST analysis of the *S. australis* transcripts against the SwissProt protein database yielded hits for 6128 transcripts covering at least 80% of the protein lengths of

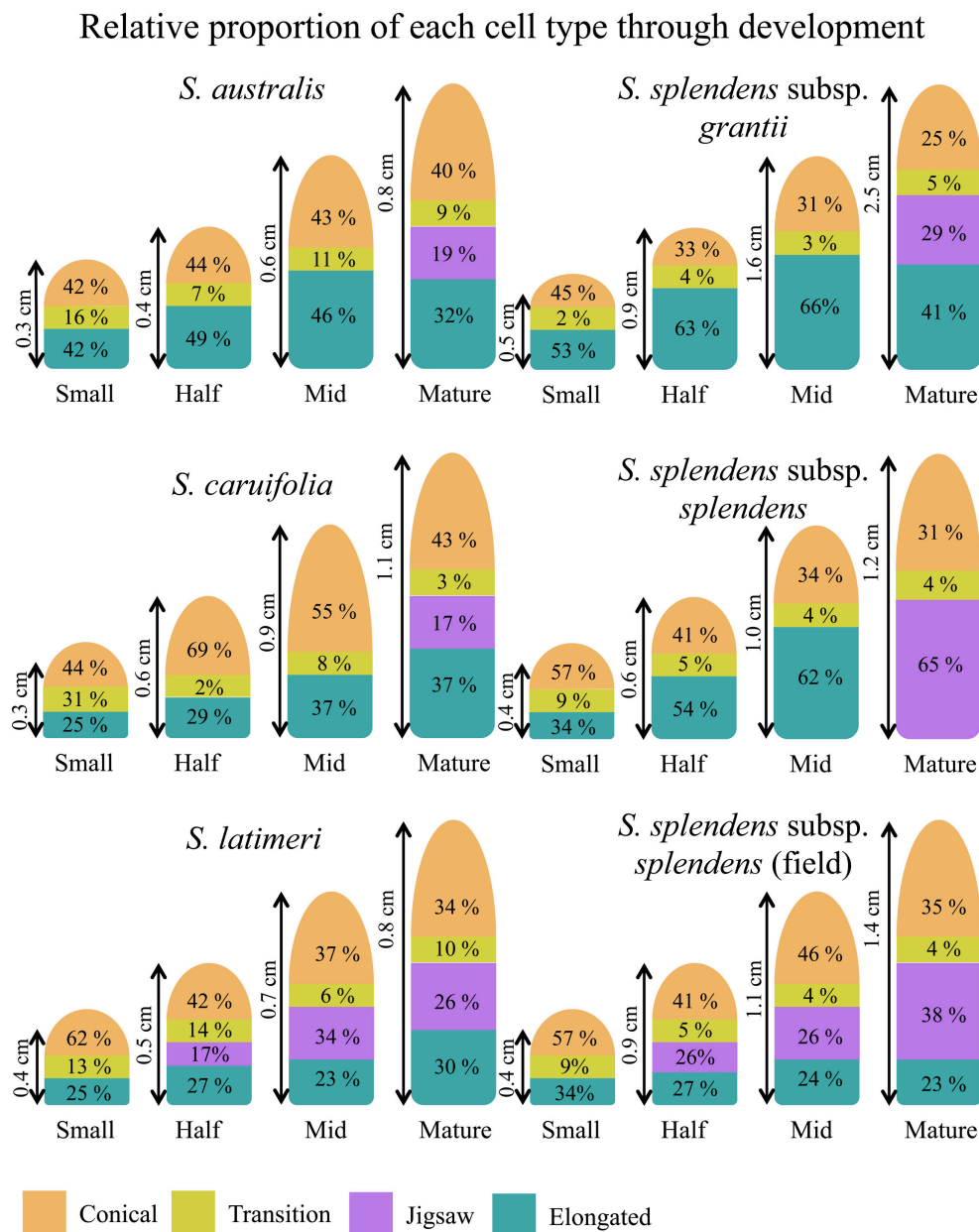
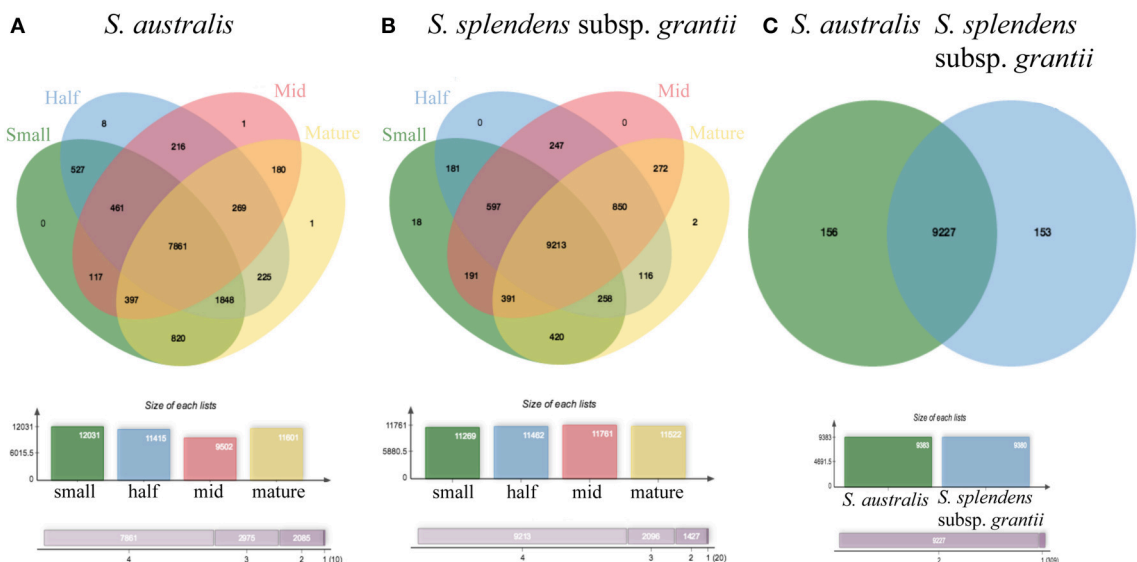
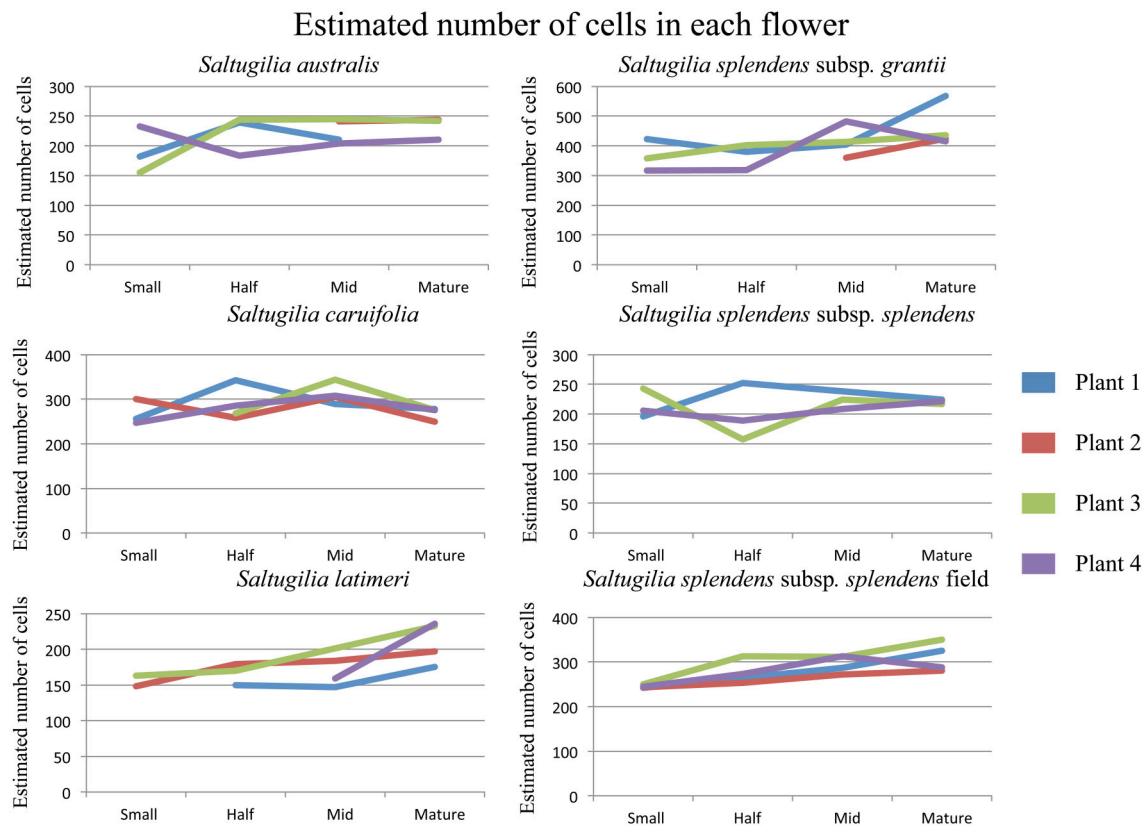


FIGURE 5 | Proportions of cell types through the four stages of development for all taxa investigated in one plant of each taxon.

15,136 unique known proteins. A BLAST analysis using the *S. splendens* subsp. *grantii* transcripts against SwissProt resulted in 5988 hits covering at least 80% of the protein lengths of 15,426 unique proteins. Functional annotation of gene ontology (GO) categories was also investigated for transcripts of all three reference transcriptomes. For *S. australis*, 41,874 transcripts were annotated with GO categories, whereas the *S. splendens* subsp. *grantii* transcriptome has 50,174 transcripts with GO category annotations. The combined reference, with all reads, has 62,253 transcripts annotated with GO categories (Supplemental Table 5).

Comparing the two reference transcriptomes using OrthoVenn, *S. australis* has 14,682 proteins in 9375 clusters, and *S. splendens* subsp. *grantii* has 14,545 proteins in 9362 clusters (Figure 7C). In all, 14,305 proteins in 9229 clusters are shared between the two taxa, with an additional 146 clusters unique to *S. australis* and 133 clusters unique to *S. splendens* subsp. *grantii*. The five GO categories with the largest numbers of proteins shared between the two taxa are biological process (2501 proteins), metabolic process (2073 proteins), cellular process (1814 proteins), cellular metabolic process (1743 proteins) and response to stimulus (1228 proteins). Other clusters of proteins



that might be relevant for investigating differences in flower size are developmental process (708 proteins), growth (149 proteins), cell wall organization (141 proteins), cell growth (112 proteins), cell division (80 proteins) and cell proliferation (28 proteins).

Of the protein clusters unique to *S. australis*, the top five GO categories are biological process (50 proteins), metabolic process (42 proteins), cellular process (39 proteins), cellular metabolic process (38 proteins) and nitrogen compound metabolic process (30 proteins). Additionally, 11 proteins are associated with developmental process, six with cell cycle, three with cell division, three with growth, two with cell proliferation, two with cell wall organization and two with cell growth. Categories that appear to be enriched only in *S. australis* are ammonium transmembrane transporter activity and DNA topoisomerase activity, both associated with the molecular function category. The top five protein clusters unique to *S. splendens* subsp. *grantii* are the same top five clusters shared between the two species. The only GO category enriched only in *S. splendens* subsp. *grantii* is maintenance of seed dormancy, a biological process.

Within each species, the number of protein clusters varies across developmental stages. In *S. australis*, the four stages exhibit 12,031 (small stage), 11,415 (half stage), 9502 (mid stage), and 11,601 (mature stage) protein clusters (**Figure 7A**). Overall, 7861 protein clusters are shared across the four developmental stages. The small stage has no unique clusters compared to the other stages, whereas the half stage has eight, and the mid and mature stages each have one unique protein cluster. The unique clusters in the half stage are the only clusters that have significantly enriched GO terms, and those are: polysaccharide binding, megasporogenesis and radial pattern formation. Most clusters shared between at least two stages of development in *S. australis* do not have any enriched GO terms except oligopeptide transporter activity shared between the half and mid stage, phosphatidylinositol-4-phosphate binding shared between the small and mid stages and plasma membrane ATP synthesis coupled proton transport shared between the small and half stages.

In *S. splendens* subsp. *grantii*, the number of protein clusters is very similar across stages, with 11,377 (small stage), 11,491 (half stage), 11,781 (mid stage), and 11,505 (mature stage) clusters. Most of these protein clusters (9198) are shared across all stages (**Figure 7B**). The small stage has the most unique clusters with eight, while the mid stage has one and the mature stage has two, with the half stage having no unique clusters. The unique clusters of the small-stage flowers have GO enrichment categories of regulation of growth rate and positive regulator of organ growth, while the unique cluster in the mid stage is enriched with the GO cellular component term nuclear speck. The unique clusters found in the mature stage are enriched for four terms: ATP synthesis coupled electron transport, phosphorelay signal transduction system, NADH dehydrogenase activity and respiratory chain. As seen in *S. australis*, most shared clusters do not have any GO-enriched terms except in a few specific cases. The clusters shared between small and half stages are enriched for condensing complex, while the clusters shared between the small, half and mid stages are enriched for cell wall

modification involved in multidimensional cell growth. There are four terms enriched in the clusters shared between the small and mature stages: sulfate transmembrane-transporting ATPase activity, valine-tRNA ligase activity, aminoacyl-tRNA editing activity and valyl-tRNA aminoacylation. Four additional terms are enriched in the clusters shared between the mid and mature stages: regulation of development (heterochronic), gene silencing by miRNA, histone deacetylation and histone deacetylase activity.

In *S. australis*, the greatest shift in differential expression occurs between the half and mature stages, in which 300 transcripts are differentially expressed, with 208 transcripts up-regulated in mature flowers (**Figure 8**; **Table 2**). Of these up-regulated transcripts, 133 are cellular component genes and seven molecular function genes. Of the cellular component transcripts, four are associated with the cell wall. The comparison between mature and mid stages resulted in 95 differentially expressed transcripts, with 67 of those up-regulated in the mid stage and 28 up-regulated in the mature stage. In comparison, only 43 transcripts are differentially expressed between the half and mid stages. (Comparisons involving flowers in small stages were not conducted due to lack of replication, with one of the transcriptomes removed because only 13 transcripts passed the 1 fpkm filtering used.)

In the large-flowered *S. splendens* subsp. *grantii*, in contrast, the greatest shift in expression occurs between the small to mid stages of flower development, with 1836 transcripts exhibiting potential differential expression, with 819 being up-regulated in mid stage flowers and 1017 up-regulated in small stage

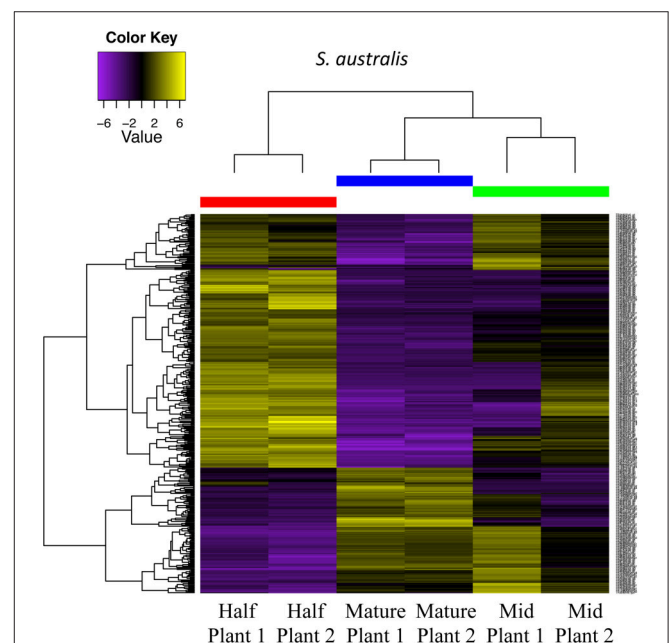


FIGURE 8 | Differential gene expression analysis among half, mid and mature stages of development for two plants of *S. australis*. Cutoff for differentially expressed genes was a 4-fold change in expression levels with a *p*-value of 0.05. Genes colored in yellow are up-regulated and genes colored purple are down-regulated compared to the other stages.

TABLE 2 | Results from differential gene expression analysis between stages of *Saltugilia australis* (Sa) and *S. splendens* subsp. *grantii* (Sg).

Comparison	Up-regulated transcriptome	Biological process	Cellular component	Molecular function	Total
Sa Mature vs. Sg Mature	Sa Mature	4	130	41	382
	Sg Mature	1	133	32	354
Sa Mature vs. Sa Half	Sa Mature	0	133	7	208
	Sa Half	1	46	15	92
Sa Half vs. Sa Mid	Sa Half	0	1	1	4
	Sa Mid	0	22	3	39
Sa Mature vs. Sa Mid	Sa Mature	1	13	7	28
	Sa Mid	1	31	4	67
Sg Mature vs. Sg Half	Sg Mature	2	176	77	427
	Sg Half	7	263	40	452
Sg Mature vs. Sg Mid	Sg Mature	2	184	98	506
	Sg Mid	4	166	40	280
Sg Mature vs. Sg Small	Sg Mature	3	220	44	474
	Sg Small	7	286	38	526
Sg Half vs. Sg Mid	Sg Half	0	8	1	19
	Sg Mid	0	12	5	27
Sg Half vs. Sg Small	Sg Half	6	375	65	743
	Sg Small	4	306	123	713
Sg Mid vs. Sg Small	Sg Mid	6	411	91	819
	Sg Small	5	457	160	1017

GO categories for the up-regulated transcripts from comparison are listed, as well as the total number of up-regulated transcripts in each transcriptome.

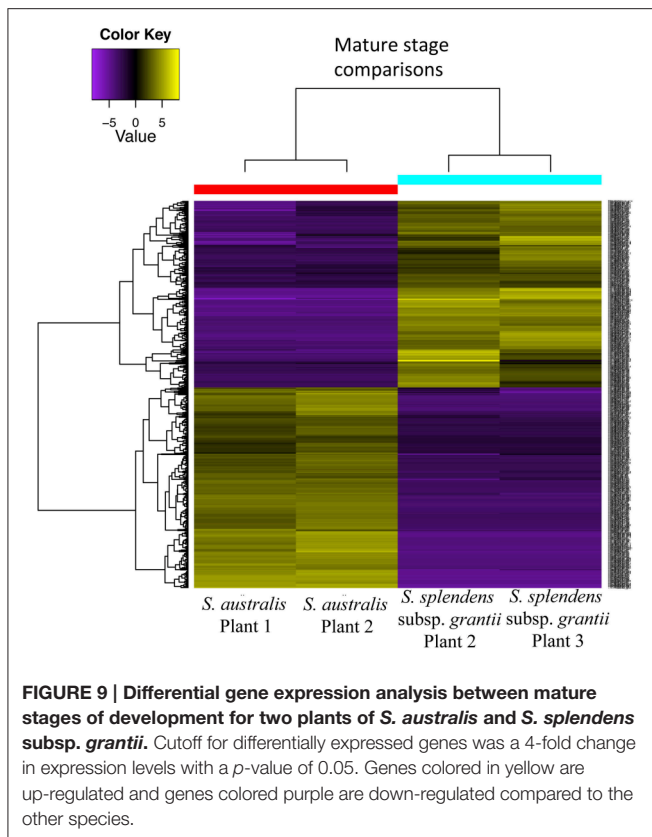
flowers. The second greatest shift in expression occurs between small to half stages of flower development with 1456 transcripts showing differential expression. In the transition between mid and mature flowers, the number of transcripts with differential expression in *S. splendens* subsp. *grantii* is nearly 8-fold greater than in the small-flowered *S. australis*, with 786 transcripts showing differential expression, 506 of which are up-regulated in mature flowers (compared with only 28 in *S. australis*). Of these, 184 transcripts represent cellular component genes, 98 molecular function genes and two biological function genes. Twelve of the cellular component genes are associated with the cell wall, including cell wall organization, growth polysaccharide catabolic process and trehalose metabolic process. One of the molecular function genes is associated with the cell wall: polysaccharide biosynthetic process. The two biological process genes that are up-regulated in the mature flower both involve cell division. When comparing mature flowers to half-stage flowers, *S. splendens* subsp. *grantii* (879 transcripts) showed a 4.2-fold increase in the number of differentially expressed transcripts compared to *S. australis* (300 transcripts). Similar to *S. australis*, the comparison of mid and half stages resulted in the fewest number of differentially expressed transcripts with 46.

Comparison between mature stages of *S. australis* and *S. splendens* subsp. *grantii* yields a total of 736 differentially expressed transcripts (Figure 9). Of these, 382 are up-regulated in *S. australis*, with 354 transcripts up-regulated in *S. splendens* subsp. *grantii* (Supplemental Table 6). Of the up-regulated transcripts in *S. australis*, 175 produce BLAST annotations to GO categories, with 130 of these representing cellular component genes, 41 molecular function genes and four

biological processes. Of the cellular component GO category transcripts, 15 represent transcripts associated with the cell wall, including specific associations with the vacuole and the regulation of cell proliferation. Three of the molecular function transcripts have GO categories associated with organization of the cell wall, with an additional four transcripts involved in pectin catabolic process. The four biological process category transcripts involve the auxin activated signaling pathway and responses to biotic stimulus and stress. Of the up-regulated transcripts in *S. splendens* subsp. *grantii*, 166 BLAST to known GO categories, with 133 of those cellular component genes, 32 molecular function genes and one biological process. Twenty-one of the cellular component genes are involved with cell wall formation, including carbohydrate metabolic process, pectin catabolic process and cell wall organization, with one of the molecular function genes associated with the cell wall and cell wall organization.

Phylogenetic Integration

Integrating the phylogeny of *Saltugilia* with the flower size data shows two independent shifts to larger corollas using either topology recovered (Figure 10). Using the topology based on the coding regions of the plastome, one case of elongation involves the clade of *S. splendens* subsp. *grantii* and *S. splendens* subsp. *splendens*, with respective sizes of 2.3–2.5 cm and 0.9–1.3 cm. The second transition is the field accession of *S. splendens* subsp. *splendens*, which has corollas that range in size from 1.1 to 1.5 cm. When using the complete plastome topology and nuclear genes, the same transitions are evident for *S. splendens* subsp. *grantii* and the field accession of *S. splendens* subsp. *splendens*. The difference



between the reconstructions is the corolla size inferred for some of the internal nodes.

Multiple shifts in pollinators are evident. When using MP, regardless of topology, the putative ancestor of *Saltugilia* is reconstructed as autogamous. Subsequent transitions to bee, hummingbird and bee fly (two transitions) are evident. Under an ML framework, the evolutionary trajectory is not as clear. In both topologies, the ancestral state for *Saltugilia* is equivocal, with equal likelihoods for each of the pollinator types, with subsequent transitions therefore difficult to determine.

There is also phylogenetic signal in cell type. Both *S. latimeri* and the field sample of *S. splendens* subsp. *splendens* possess jigsaw cells in the corolla during early stages of development, whereas all other taxa only possess these cells in mature flowers. These taxa are sisters in both topologies, and there appears to be a single evolutionary transition yielding jigsaw cells early in development.

DISCUSSION

Phylogenetic Relationships in *Saltugilia*

Previous phylogenetic analyses have included multiple members of *Saltugilia*, but those either lacked some taxa or were based on only a small number of markers (Weese and Johnson, 2005; Johnson, 2007; Landis et al., in review), and not all relationships were highly supported. In this study, analyses of three data

sets—the coding regions of the plastome, the complete plastome and 90 nuclear loci—yield high bootstrap values for all nodes. This concatenation approach for the coding regions of the plastome and the complete plastome has been utilized before in previous studies given that the plastome is inherited as a single locus (Small et al., 1998; Parks et al., 2009; Moore et al., 2010; Ruhfel et al., 2014). However, even though unlinked nuclear loci may have different coalescence histories, several recent studies have found that concatenation methods yield results similar to species tree reconciliation methods (e.g., Thompson et al., 2014; Tonini et al., 2015), supporting our use of a concatenated data set for the nuclear markers as well.

All three datasets suggest the same phylogenetic relatedness of species, except for the placement of the *S. splendens* taxa. If the topology based on both the complete plastome and nuclear genes is correct, taxonomic revisions are warranted, given the lack of monophyly of the accessions of *S. splendens*. Recognition of each accession as representing a distinct species would be consistent with both the phylogeny and the morphology of the three accessions, but further work is needed to investigate these taxa. In earlier phylogenetic analyses, *S. latimeri* was found to be sister to *S. australis* (Weese and Johnson, 2005; Johnson, 2007). However, these analyses were based on only one to five genes, and usually a combination of nuclear and plastid genes, which may show topological incongruence due to hybridization (Soltis and Kuzoff, 1995; Sang et al., 1997; Okuyama et al., 2005). Our analysis does not show this relationship between *S. latimeri* and *S. australis*, however, as we recovered *S. latimeri* to be sister to the field accession of *S. splendens* subsp. *splendens*. This result suggests that further phylogenetic and taxonomic work is needed, particularly because of high levels of morphological variability in *S. splendens*, making identification of subspecies difficult (Mark Porter, personal communication). *Gilia stellata* was deemed an appropriate outgroup because it has long been associated with *Saltugilia* taxonomically, with the proposal that it be transferred to *Saltugilia* (Grant, 1959; Grant and Grant, 1965), but sufficient evidence for this move has never been provided. Our analysis shows that it is sister to *G. brecciarum* based on both plastome and nuclear data, supporting its placement in *Gilia*.

Transitions from outcrossing to selfing are prevalent in many angiosperm groups (e.g., Stebbins, 1974; Bull and Charnov, 1985; Schoen et al., 1996; Barrett, 2003, 2008, 2013). In transitions involving changes in outcrossing vector, the most numerous transitions are bee to hummingbird pollination, with the reverse being observed less frequently (Barrett, 2013). Previous analysis of pollinators in Polemoniaceae, with sampling of nearly all species in the family, shows that the ancestor of *Saltugilia* was likely bee-pollinated (Landis et al., in review). If this is indeed the case, then with the increased sampling of *Saltugilia* in the present study, we observe two independent transitions to autogamy (*S. australis* and *S. latimeri*), one transition to hummingbird pollination (*S. splendens* subsp. *grantii*) and at least one (possibly two) transition to bee fly pollination (*S. splendens* subsp. *splendens*). However, with only *Saltugilia* and two species of *Gilia* represented in the current study, we reconstruct the ancestor of *Saltugilia* to be autogamous, with gains of bee,

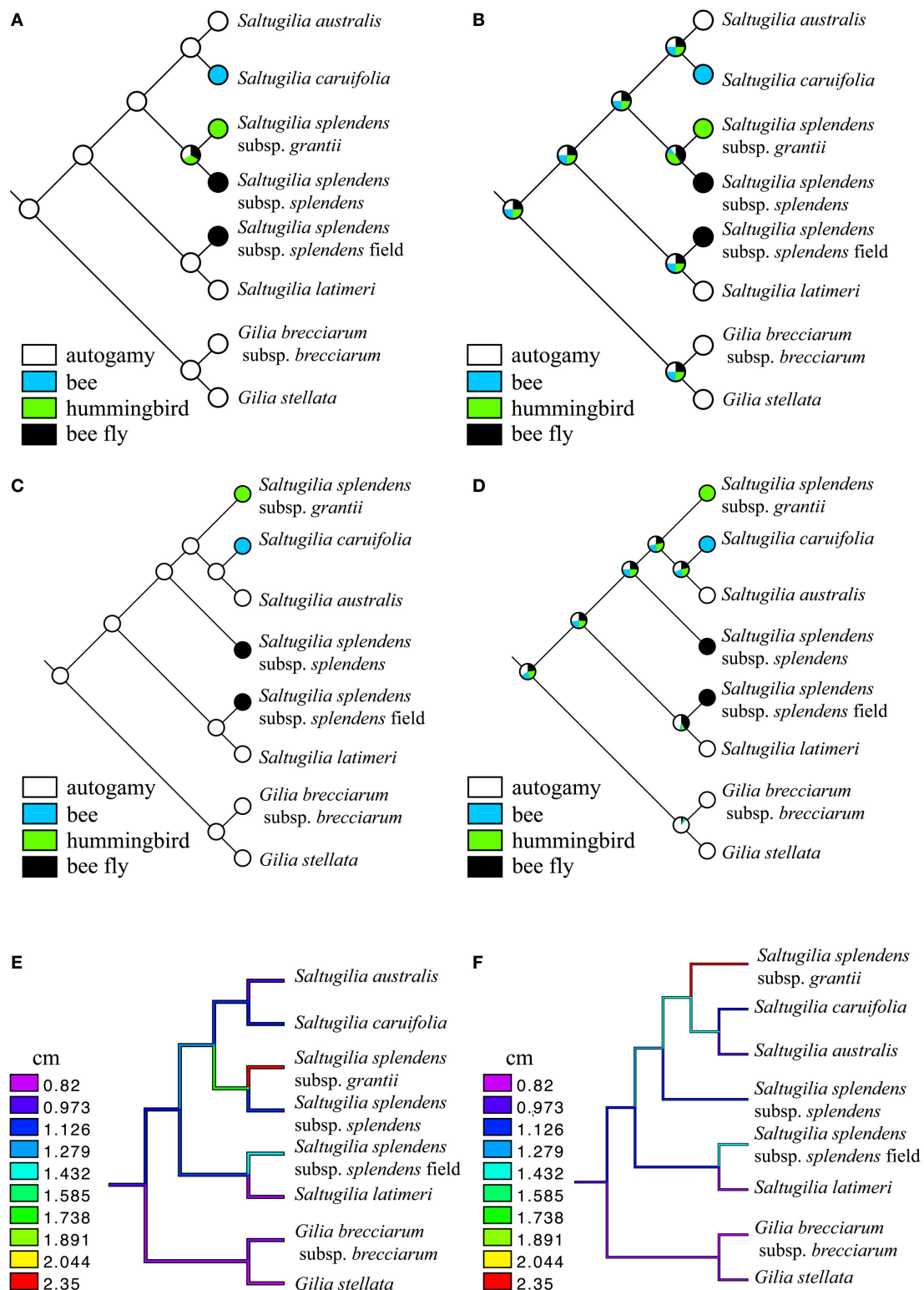


FIGURE 10 | Ancestral-state reconstructions of pollinators (A–D) and size (E,F) under a maximum parsimony (MP) and maximum-likelihood (ML) framework using the two competing topologies. (A) MP of pollinators using the topology from plastid coding genes, (B) ML of pollinators using the topology from plastid coding genes, (C) MP of pollinators using the topology from the complete plastome and nuclear genes, (D) ML of pollinators using the topology from the complete plastome and nuclear genes, (E) continuous reconstruction of flower size using the topology from plastid coding genes, and (F) continuous reconstruction of flower size using the topology from the complete plastome and nuclear genes.

hummingbird and bee fly pollination. This discrepancy may be attributed to taxon sampling, not in *Saltugilia* itself, but in its close relatives. Based on the larger analysis in Landis et al. (in review), the sister group to *Saltugilia* is a clade consisting of *Gilia*, *Collomia*, and *Navarretia*. In both analyses, the ancestor of *Gilia* is reconstructed as autogamous, but in the larger study, the combination of states results in the common ancestor of *Saltugilia* being bee-pollinated and here, autogamous. This highlights the importance of sufficient taxon sampling for character state reconstructions; otherwise, the reconstructions may be skewed to favor only those taxa sampled and not the best overall evolutionary hypothesis.

The importance of taxon sampling does not seem to be as significant when reconstructing flower size within *Saltugilia*. In the large analysis of Polemoniaceae (Landis et al., in review), the common ancestors of both *Saltugilia* and of *Gilia* were reconstructed as having flowers 0.8–1.2 cm in length. In both reconstructions of flower size in the current study, the common ancestor of *Saltugilia* had flowers of 0.97–1.13 cm in length. This concordance between the large and focused analyses may be due to the fact that the entire clade of *Saltugilia*, *Gilia*, *Collomia* and *Navarretia* is composed of relatively small flowers (Landis et al., in review), so there are no large outliers that affect the reconstruction of ancestral states. In investigations of flower size, especially in relation to shifts in pollinators, transitions from outcrossing to selfing are often accompanied by transitions to smaller flowers, with examples observed in *Aquilegia pawi* (Ranunculaceae; Martinell et al., 2011), *Arabis alpha* (Brassicaceae; Tedder et al., 2015) and *Camissoniopsis cheiranthifolia* (Onagraceae; Button et al., 2012). Our analyses of *Saltugilia* show that selfing species have smaller flowers than their outcrossing congeners, generally due to larger jigsaw and elongated cells comprising the corolla tube in many of the outcrossing species and twice as many cells in the large flowers of the outcrossers as in the small flowers of the selfers.

Cell Morphology and Cell Size

Four types of cells have been identified in developing flowers of species of *Saltugilia*: conical, transition, elongated and jigsaw cells. Jigsaw cells are evident mostly in mature flowers, but they appear at earlier stages in *S. latimeri* and *S. splendens* subsp. *splendens*. These jigsaw cells are similar in shape to the jigsaw pavement cells observed in leaves (Fu et al., 2009). Previous analyses of epidermal cell shape suggest that these cells are similar to multiple papillate cells (Kay et al., 1981) and tabular rugose cells (Ojeda et al., 2009). The development of leaf jigsaw pavement cells is fairly well characterized (Fu et al., 2005) and thought to be regulated by cell-cell signaling. The jigsaw pavement cells in leaves begin as circular cells and then become elongated before becoming jigsaw shaped (Fu et al., 2005). The cells in the corolla tube of flowers in Polemoniaceae exhibit similar developmental characteristics to these jigsaw pavement cells by starting out more circular in earlier stages, with later elongation and finally formation of jigsaw cells (Figure 3). However, jigsaw cells are not ubiquitous in corolla tubes across angiosperms, as they have not been identified in multiple species of *Petunia* (*P. axillaris*, *P. exserta*, and *P. integrifolia*;

Landis, unpublished data) but have been reported in the flowers of 13 families (Kay et al., 1981), including Caryophyllaceae, Onagraceae, Polygonaceae and Primulaceae, and are thought to be important for rapid petal expansion.

The shape of petal epidermal cells represents an evolutionarily labile trait, hypothesized to be regulated by a small subfamily of duplicate transcription factors (Glover et al., 2015). The petal epidermis, especially the epidermis located in the petal lobes, is composed of conical cells that enhance the attractiveness of the flower to potential pollinators by modifying flower color by focusing light on pigmented cells, affecting surface texture for pollinator grip and affecting floral surface temperature (Kevan and Lane, 1985; Noda et al., 1994; Comba et al., 2000; Whitney and Glover, 2007). In *Antirrhinum*, formation of these conical cells is regulated by a MYB-related transcription factor, *MIXTA*, and related genes *MIXTA-LIKE 1*, *MIXTA-LIKE 2*, and *MIXTA-LIKE 3* (Noda et al., 1994; Glover et al., 1998; Martin et al., 2002; Perez-Rodriguez, 2005; Baumann et al., 2007). Phylogenetic analyses by Brockington et al. (2013) of the closely related *MIXTA* MYB transcription factors showed multiple duplication events, including in the common ancestor of the eudicots and multiple lineage-specific duplications. This pattern of gene duplication and functional diversification suggests that changes in petal epidermal cells involving cell types other than conical cells may likewise be affected by similar patterns of divergence in transcription factors associated with gene duplication events.

In *Saltugilia*, the difference in flower size appears to be determined more by cell size than cell number, although cell numbers also increase in the large flowers in the genus. Given that organ sculpting via localized cell division is the key factor determining the initiation of nectar spurs in *Aquilegia* (Yant et al., 2015), with cell elongation responsible for variation in length of the spurs (Puzey et al., 2012), corolla development, in both shape and size, may involve multiple processes. More specifically, even the size of different parts of the flower (and/or leaves) may be controlled by different mechanisms. For example, in *Petunia*, the area of single cells begins to increase at the base of the petal tube and then gradually progresses toward the tip of the petal lobes, with distinct differences in cell growth and expansion in the bottom third of the corolla (Reale et al., 2002; Stuurman et al., 2004). Our results for *Saltugilia* follow this general pattern. In the earliest developmental stage investigated (25% of mature flower length), the sizes of conical, transition and elongated cells are similar. By the half and mid stages, the elongated cells have had a 2–5-fold change in size, while the conical cells have undergone a 1.5–2-fold increase in size.

In addition to these changes in cell size during development of the corolla in *Saltugilia*, the most obvious change in cell number is the appearance of jigsaw cells in most taxa between the mid and mature stages of flower development. With the formation of jigsaw cells, the number of elongated cells diminishes, in some cases between 20 and 140 cells (Supplemental Table 4). This is in stark contrast to development in *Petunia*, which exhibits no appreciable changes in cell number in any domain of the flower throughout development (Stuurman et al., 2004). In *Saltugilia*, changes in the number of conical and transition cells are much

less pronounced. The general trend appears to be a marginal increase in the number of conical cells and a decrease in the number of transition cells. When comparing estimates of total cell numbers across all flowers, the majority of flowers appear to be composed of 200–250 cells. In contrast, three of the four plants of the large-flowered *S. splendens* subsp. *grantii* have estimates ranging between 400 and 450 cells, with the fourth plant estimated to have 567 cells in the mature flower. The field-collected accession of *S. splendens* subsp. *splendens* has flowers with more than 300 cells. These two taxa bear the largest flowers in *Saltugilia* and are clearly non-monophyletic, indicating two independent evolutionary origins of large flowers, in both cases accomplished through increased cell number in conjunction with increased cell size.

Genetics of Flower Size

The most abundant GO categories detected in analyses of transcripts from both *S. australis* and *S. splendens* subsp. *grantii* are biological process, metabolic process, cellular process, cellular metabolic process and response to stimuli. However, those transcripts that were mostly up-regulated in *S. splendens* subsp. *grantii* relative to the smaller-flowered *S. australis* were cellular component genes, which were ranked 21st in the total number of genes found in the presence/absence comparison between species. Additionally, some molecular function genes and a small subset of biological process genes were also up-regulated in *S. splendens* subsp. *grantii*. With the formation of jigsaw cells, and their apparent importance in contributing to overall corolla length (Figure 5), genes associated with cell wall formation and organization may provide the genetic framework for larger flowers in *Saltugilia*.

Additional evidence supporting this hypothesis for the importance of cellular component genes comes from comparison of GO categories for differentially expressed transcripts in different developmental stages within species. With a single exception (the comparison of half and mid stages in *S. australis*), all comparisons of gene expression show that the largest category of up-regulated transcripts is for the cellular component genes. This trend is most apparent in the size comparisons of *S. splendens* subsp. *grantii*. Comparisons of the small and mid stage transcriptomes resulted in 411 cell component genes up-regulated in the mid stage and 457 up-regulated in the small stage. The large amount at the small stage may be attributed to the formation of new cells and cell walls, while the up-regulation in the mid stage may be the result of elongated cells starting to undergo the necessary changes to become jigsaw cells. Even though the small stage of *S. australis* could not be included in the developmental comparisons, when comparing the mature and mid stages of flower development to the half stage, transcripts attributed to cell component functions are up-regulated in the later stages, with 133 genes up-regulated in the mature stage compared to the half stage and 22 transcripts up-regulated in the mid stage compared to the half stage. The observation that such a large subset of cellular component transcripts, specifically cell wall modification genes, is up-regulated may not be surprising given that approximately 15% of the *Arabidopsis* genome is dedicated to cell wall formation and modification (Carpita et al.,

2001). Weiss et al. (2005) reviewed the importance of the process of cell wall deposition and how the cell wall is formed, producing important changes in floral organ size. Thus, the many genes involved in cell wall formation in general provide a large pool of genes for which expression may be modified to yield variation in cell size and cell number, effecting changes in flower size as well as other traits.

Many published studies have shown possible effects of candidate genes for floral size identified in the model systems *Arabidopsis thaliana* and *Antirrhinum majus* (Herzog et al., 1995; Elliott et al., 1996; Mizukami, 2001; Kim et al., 2002; Disch et al., 2006; Krizek, 2009; Xu and Li, 2011). However, none of the genes hypothesized to play a major role in overall size appears to be differentially expressed between the larger flowers of *S. splendens* subsp. *grantii* and the smaller flowers of *S. australis*. Orthologs of *JAGGED*, *OPR3*, *BIG BROTHER*, *KLH*, *GAST1* and *GASA* were identified in the reference assemblies for both species and were also detected in each of the developmental stages for both species. These candidate genes are expressed, but not differentially so, between the small-flowered *S. australis* and the large-flowered *S. splendens* subsp. *grantii*, or differentially expressed among stages within species. Because of this, their function and roles in *Saltugilia* are currently uncertain. Additional genes compiled by Hepworth and Lenhard (2014) have been identified to be associated with growth and expansion of leaves. Many of these additional genes, such as *AUXIN RESPONSE FACTOR2*, *TARGET-OF-RAPAMAYCIN* and *SPINDLY*, were found in the reference assemblies, but were not identified as genes differentially expressed between the two species or between stages within species. The lack of differential expression of these additional genes may not be surprising, however, given that Anastasiou and Lenhard (2007) found that QTL that affect leaf size and shape are largely distinct from those influencing floral organs, at least in *Arabidopsis* and tomato. Therefore, candidate genes from leaf growth and leaf size may not be good candidates for controlling differences in corolla size. Additional avenues for flower size may be the role of hormones, given that in the non-model species *Gaillardia grandiflora*, corolla growth was accomplished by an increase in gibberellin activity (Koning, 1984).

Transcriptome analysis of additional taxa of *Saltugilia* may contribute to the discovery of the genetic component(s) of flower size differences in *Saltugilia*. Comparing the transcriptomes of *S. latimeri* and the field accession of *S. splendens* subsp. *splendens*, especially at earlier stages of development, will be informative, given that jigsaw cells appear earlier in development in both taxa than in the other taxa investigated. Furthermore, functional analyses (using Virus Induced Gene Silencing) and additional expression studies using qPCR in *Saltugilia*, as well as in model species, are needed to evaluate the possible roles of the candidate genes identified here in determining flower size.

CONCLUSIONS

In *Saltugilia*, there are two independent evolutionary transitions from smaller to longer flowers, possibly driven by selective pressures of hummingbird and bee fly pollination. The longer

corollas of *S. splendens* subsp. *grantii* and the field accession of *S. splendens* subsp. *splendens* show significant increases in the cell area of jigsaw cells, as well as slightly longer elongated cells. In addition to increased cell size in *S. splendens* subsp. *grantii*, there is also a large increase in the estimated number of cells making up the corolla tube, especially of jigsaw cells. Comparisons of the transcriptome profiles across developmental stages of the taxa having the smallest and largest flowers, including within- and between-taxon comparisons, resulted in the identification of many genes associated with cell wall formation and organization that are up-regulated in the mature stages compared to earlier stages of development and between the small-flowered *S. australis* and the large-flowered *S. splendens* subsp. *grantii*. This shift in gene expression profiles coincides with the presence of jigsaw cells, which form a considerable proportion of the mature flowers in *Saltugilia*. None of the candidate genes known to affect cell size show differential expression between the small- and large-flowered species, but the transcriptome profiles suggest many possible candidates for controlling differences in corolla size in *Saltugilia*.

AUTHOR CONTRIBUTIONS

Primary project design was accomplished by JL, DS and PS, while data collection was done by JL, DO, KV and RO. Data analysis

was performed by JL, MG, DS and PS. All authors have approved the final version of the submitted manuscript and all agree to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

The authors thank Tasha LaDoux, Duncan Bell and Evan Meyers for help in obtaining flower material, Leigh Johnson for help in obtaining flower material as well as his many insights based on years of working with Polemoniaceae in the field, Milda Stanislaukas and Margarita Hernandez for assistance with lab work, Nicolas Garcia for bioinformatics help, Marissa Gredler for contributions to the analyses of cell shape and two reviewers for their comments and suggestions. This work was funded by the following grants: NSF DDIG DEB1406650 (PS and JL), BSA Graduate Student Research Grant (JL), Sigma Xi Grant in Aid of Research (JL) and a microMorph training grant (JL).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.01144>

REFERENCES

- Alapetite, E., Baker, W. J., and Nadot, S. (2014). Evolution of stamen number in Ptychospermatinae (Arecaceae): insights from a new molecular phylogeny of the subtribe. *Mol. Phylogent. Evol.* 76, 227–240. doi: 10.1016/j.ympev.2014.02.026
- Anastasiou, E., and Lenhard, M. (2007). Growing up to one's standard. *Curr. Opin. Plant Biol.* 10, 63–69. doi: 10.1016/j.pbi.2006.11.002
- Barrett, S. C. H. (2003). Mating strategies in flowering plants: the outcrossing-selfing paradigm and beyond. *Philos. Trans. R. Soc. Lond. B* 358, 991–1004. doi: 10.1098/rstb.2003.1301
- Barrett, S. C. H. (2008). Major evolutionary transitions in flowering plant reproduction: an overview. *Intl. J. Plant Sci.* 169, 1–5. doi: 10.1086/522511
- Barrett, S. C. H. (2013). The evolution of plant reproductive systems: how often are transitions irreversible? *Proc. Biol. Sci.* 280:20130913. doi: 10.1098/rspb.2013.0913
- Baumann, K., Perez-Rodriguez, M., Bradley, D., Venail, J., Bailey, P., Jin, H., et al. (2007). Control of cell and petal morphogenesis by R2R3 MYB transcription factors. *Development* 134, 1691–1701. doi: 10.1242/dev.02836
- Ben-Nissan, G., Lee, J.-Y., Borohov, A., and Weiss, D. (2004). GIP, a Petunia hybrida GA-induced cysteine-rich protein: a possible role in shoot elongation and transition to flowering. *Plant J.* 37, 229–238. doi: 10.1046/j.1365-3113X.2003.01950.x
- Brockington, S. F., Alvarez-Fernandez, R., Landis, J. B., Alcorn, K., Walker, R. H., Thomas, M. M., et al. (2013). Evolutionary analysis of the MIXTA gene family highlights potential targets for the study of cellular differentiation. *Mol. Biol. Evol.* 30, 526–540. doi: 10.1093/molbev/mss260
- Bruneau, A. (1997). Evolution and homology of bird pollination syndromes in Erythrina. *Am. J. Bot.* 84, 54–71. doi: 10.2307/2445883
- Brunet, J. (2009). Pollinators of the Rocky Mountain columbine: temporal variation, functional groups and associations with floral traits. *Ann. Bot.* 103, 1567–1578. doi: 10.1093/aob/mcp096
- Bull, J. J., and Charnov, E. L. (1985). On irreversible evolution. *Evolution* 39, 1149–1155. doi: 10.2307/2408742
- Butler, T., Dick, C., Carlson, M. L., and Whittall, J. B. (2014). Transcriptome analysis of a petal anthocyanin polymorphism in the arctic mustard, *Parrya nudicaulis*. *PLoS ONE* 9:e101338. doi: 10.1371/journal.pone.0101338
- Button, L., Villalobos, A. L., Dart, S. R., and Eckert, C. G. (2012). Reduced petal size and color associated with transitions from outcrossing to selfing in *Camissoniopsis cheiranthifolia* (Onagraceae). *Int. J. Plant Sci.* 173, 251–260. doi: 10.1086/663972
- Carpita, N., Tierney, M., and Campbell, M. (2001). Molecular biology of the plant cell wall: searching for the genes that define structure, architecture and dynamics. *Plant Mol. Biol.* 47, 1–5. doi: 10.1023/A:1010603527077
- Chamala, S., García, N., Godden, G. T., Krishnakumar, V., Jordon-Thaden, I. E., De Smet, R., et al. (2015). MarkerMiner 1.0: a new application for phylogenetic marker development using angiosperm transcriptomes. *Appl. Plant Sci.* 3:1400115. doi: 10.3732/apps.1400115
- Chang, L., Zhiyong, L., Xueling, Y., Yiheng, W., Qian, M., and Hui, F. (2014). Transcriptome analysis of the chinese cabbage (*Brassica rapa* ssp. *pekinensis*) petal using RNA-Seq. *J. Hortic.* 1:114. doi: 10.4172/2376-0354.1000114
- Comba, L., Corbet, S. A., Hunt, H., Outram, S., Parker, J. S., and Glover, B. J. (2000). The role of genes influencing the corolla in pollination of *Antirrhinum majus*. *Plant Cell Environ.* 23, 639–647. doi: 10.1046/j.1365-3040.2000.00580.x
- Crepet, W. L. (1995). Advanced (constant) insect pollination mechanisms: pattern of evolution and implications vis-à-vis angiosperm diversity. *Ann. Missouri Bot. Gard.* 6, 607–630.
- Crepet, W. L., and Niklas, K. J. (2009). Darwin's second "Abominable mystery": why are there so many angiosperm species? *Am. J. Bot.* 96, 366–381. doi: 10.3732/ajb.0800126
- Cronk, Q., and Ojeda, I. (2008). Bird-pollinated flowers in an evolutionary and molecular context. *J. Exp. Bot.* 59, 715–727. doi: 10.1093/jxb/ern009
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772. doi: 10.1038/nmeth.2109

- Disch, S., Anastasiou, E., Sharma, V. K., Laux, T., Fletcher, J. C., and Lenhard, M. (2006). The E3 ubiquitin ligase BIG BROTHER controls arabidopsis organ size in a dosage-dependent manner. *Curr. Biol.* 16, 272–279. doi: 10.1016/j.cub.2005.12.026
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Elliott, R. C., Betzner, A. S., Huttner, E., Oakes, M. P., Tucker, W. Q. J., Gerentes, D., et al. (1996). AINTEGUMENTA, an APETALA2-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. *Plant Cell* 8, 155–168. doi: 10.1105/tpc.8.2.155
- Fenster, C. B., Armbruster, W. S., Wilson, P., Dudash, M. R., and Thomson, J. D. (2004). Pollination syndromes and floral specialization. *Annu. Rev. Ecol. Evol. Syst.* 35, 375–403. doi: 10.1146/annurev.ecolsys.34.011802.132347
- Ferguson, C. J., and Jansen, R. K. (2002). A chloroplast DNA phylogeny of Eastern Phlox (Polemoniaceae): implications of congruence and incongruence with the ITS phylogeny. *Am. J. Bot.* 89, 1324–1335. doi: 10.3732/ajb.89.8.1324
- Fu, Y., Gu, Y., Zheng, Z., Wasteneys, G., and Yang, Z. (2005). *Arabidopsis* interdigitating cell growth requires two antagonistic pathways with opposing action on cell morphogenesis. *Cell* 120, 687–700. doi: 10.1016/j.cell.2004.12.026
- Fu, Y., Xu, T., Zhu, L., Wen, M., and Yang, Z. (2009). A ROP GTPase signaling pathway controls cortical microtubule ordering and cell expansion in *Arabidopsis*. *Curr. Biol.* 19, 1827–1832. doi: 10.1016/j.cub.2009.08.052
- Fulton, M., and Hodges, S. A. (1999). Floral isolation between *Aquilegia formosa* and *Aquilegia pubescens*. *Proc. Natl. Acad. Sci. U.S.A.* 96, 2493–2496. doi: 10.1098/rspb.1999.0915
- Galen, C., and Cuba, J. (2001). Down the tube: pollinators, predators, and the evolution of flower shape in the alpine skypilot, *Polemonium viscosum*. *Evolution* 55, 1963–1971. doi: 10.1111/j.0014-3820.2001.tb01313.x
- Galliot, C., Stuurman, J., and Kuhlmeier, C. (2006). The genetic dissection of floral pollination syndromes. *Curr. Opin. Plant Biol.* 9, 78–82. doi: 10.1016/j.pbi.2005.11.003
- Givnish, T. J. (2002). Ecological constraints on the evolution of plasticity in plants. *Evol. Ecol.* 16, 213–242. doi: 10.1023/A:1019676410041
- Glasbey, C., and Horgan, G. (1995). *Image analysis for the Biological Sciences*. Hoboken, NJ: Wiley Publishing.
- Glover, B. J. (2007). *Understanding Flowers and Flowering: An Integrated Approach*. Oxford, UK: Oxford University Press.
- Glover, B. J., Airoidi, C. A., Brockington, S. F., Fernández-Mazuecos, M., Martínez-Pérez, C., Mellers, G., et al. (2015). How have advances in comparative floral development influenced our understanding of floral evolution? *Int. J. Plant Sci.* 176, 307–323. doi: 10.1086/681562
- Glover, B. J., Perez-Rodriguez, M., and Martin, C. (1998). Development of several epidermal cell types can be specified by the same MYB-related plant transcription factor. *Development* 125, 3497–3508.
- Gómez, J. M., Abdelaziz, M., Muñoz-Pajares, J., and Perfectti, F. (2009). Heritability and genetic correlation of corolla shape and size in *Erysimum mediohispanicum*. *Evolution* 63, 1820–1831. doi: 10.1111/j.1558-5646.2009.00667.x
- Goodspeed, T. H., and Clausen, R. E. (1915). Factors influencing flower size in *Nicotiana* with special reference to questions of inheritance. *Am. J. Bot.* 2, 332–374. doi: 10.2307/2435145
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Grant, V. (1959). *Natural History of the Phlox Family*. Hague: Martinus Nijhoff.
- Grant, V., and Grant, K. A. (1965). *Flower Pollination in the Phlox Family*. New York, NY: Columbia University Press.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hepworth, J., and Lenhard, M. (2014). Regulation of plant lateral-organ growth by modulating cell number and size. *Curr. Opin. Plant Biol.* 17, 36–42. doi: 10.1016/j.pbi.2013.11.005
- Herzog, M., Dorne, A., and Grellet, F. (1995). GASA, a gibberellin-regulated gene family known from *Arabidopsis thaliana* related to the tomato GAST1 gene. *Plant Mol. Biol.* 1766, 743–752. doi: 10.1007/BF00020227
- Irish, V. F. (2008). The Arabidopsis petal: a model for plant organogenesis. *Trends Plant Sci.* 13, 430–436. doi: 10.1016/j.tplants.2008.05.006
- Johnson, L. A. (2007). Transfer of the Western North American species *Gilia splendens* to *Saltugilia* (Polemoniaceae), and the taxonomic affinities of *Gilia scopulorum*, *Gilia stellata*, and *Gilia yorkii*. *Novon* 503, 193–197. doi: 10.3417/1055-3177(2007)17[193:TOTWNA]2.0.CO;2
- Johnson, L. A., Chan, L. M., Weese, T. L., Busby, L. D., and McMurtry, S. (2008). Nuclear and cpDNA sequences combined provide strong inference of higher phylogenetic relationships in the phlox family (Polemoniaceae). *Mol. Phylogenet. Evol.* 48, 997–1012. doi: 10.1016/j.ympev.2008.05.036
- Johnson, L. A., Schultz, J. L., Soltis, D. E., and Soltis, P. S. (1996). Monophyly and Generic Relationships of Polemoniaceae Based on matK Sequences. *Am. J. Bot.* 83, 1207–1224. doi: 10.2307/2446205
- Johnson, S. D., Linder, H. P., and Steiner, K. E. (1998). Phylogeny and radiation of pollination systems in Disa (Orchidaceae). *Am. J. Bot.* 85, 402–411. doi: 10.2307/2446333
- Joshi, N. A., and Fass, J. N. (2011). *Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files*. version 1.33. Available online at <https://github.com/najoshi/sickle>
- Kaczorowski, R. L., Juenger, T. E., and Holtsford, T. P. (2008). Heritability and correlation structure of nectar and floral morphology traits in *Nicotiana glauca*. *Evolution* 62, 1738–1750. doi: 10.1111/j.1558-5646.2008.00400.x
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kay, Q. O. N., Daoud, H. S., and Stirton, C. H. (1981). Pigment distribution, light reflection and cell structure in petals. *Bot. J. Linn. Soc.* 83, 57–84. doi: 10.1111/j.1095-8339.1981.tb00129.x
- Kearns, C. A., Inouye, D. W., Waser, N. M., Kearns, C. A., Inouye, D. W., and Waser, N. M. (2015). Endangered mutualisms: the conservation of plant-pollinator interactions. *Annu. Rev. Ecol. Syst.* 29, 83–112. doi: 10.1146/annurev.ecolsys.29.1.83
- Kent, W. J. (2002). BLAT – the BLAST-like alignment tool. *Genome Res.* 12, 656–664. doi: 10.1101/gr.229202. Article published online before March 2002
- Kevan, P. G., and Lane, M. A. (1985). Flower petal microtexture is a tactile cue for bees. *Proc. Natl. Acad. Sci. U.S.A.* 82, 4750–4752. doi: 10.1073/pnas.82.14.4750
- Kim, G. T., Shoda, K., Tsuge, T., Cho, K. H., Uchimiya, H., Yokoyama, R., et al. (2002). The ANGUSTIFOLIA gene of Arabidopsis, a plant CtBP gene, regulates leaf-cell expansion, the arrangement of cortical microtubules in leaf cells and expression of a gene involved in cell-wall formation. *EMBO J.* 21, 1267–1279. doi: 10.1093/emboj/21.6.1267
- Koning, E. (1984). The roles of plant hormones in the growth of the corolla of *Gaillardia grandiflora* (Asteraceae) ray flowers. *Am. J. Bot.* 71, 1–8. doi: 10.2307/2443617
- Kotilainen, M., Helariutta, Y., Mehto, M., Pollanen, E., Albert, V., Elomaa, P., et al. (1999). GEG participates in the regulation of cell and organ shape during corolla and carpel development in gerbera hybrida. *Plant Cell* 11, 1093–1104. doi: 10.1105/tpc.11.6.1093
- Krizek, B. (2009). AINTEGUMENTA and AINTEGUMENTA-LIKE6 act redundantly to regulate Arabidopsis floral growth and patterning. *Plant Physiol.* 150, 1916–1929. doi: 10.1104/pp.109.141119
- Lambrecht, S. C. (2013). Floral water costs and size variation in the highly selfing leptosiphon bicolor (Polemoniaceae). *Int. J. Plant Sci.* 174, 74–84. doi: 10.1086/668230
- Landis, J. B., Ventura, K. L., Soltis, D. E., Soltis, P. S., and Oppenheimer, D. G. (2015). Optical Sectioning and 3D reconstructions as an alternative to scanning electron microscopy for analysis of cell shape. *Appl. Plant Sci.* 3:1400112. doi: 10.3732/apps.1400112
- Lanfear, R., Calcott, B., Ho, S. Y. W., and Guindon, S. (2012). Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29, 1695–1701. doi: 10.1093/molbev/mss020
- Lock, I. E., Sokoloff, D. D., and Remizowa, M. V. (2011). Morphogenetic liability of the *Ruppia maritima* (Ruppiaceae, Alismatales) reproduction organs: from two lateral flowers to a terminal flower. *Russ. J. Dev. Biol.* 42, 247–260. doi: 10.1134/S1062360411040084

- Mack, J. L. K., and Davis, A. R. (2015). The relationship between cell division and elongation during development of the nectar-yielding petal spur in *Centranthus ruber* (Valerianaceae). *Ann. Bot.* 115, 641–649. doi: 10.1093/aob/mcu261
- Maddison, W. P., and Maddison, D. R. (2015). *Mesquite: A Modular System for Evolutionary Analysis*. version 3.03. Available online at: <http://mesquiteproject.org>
- Martin, C., Bhatt, K., Baumann, K., Jin, H., Zachgo, S., Roberts, K., et al. (2002). The mechanics of cell fate determination in petals. *Philos. Trans. R. Soc. B* 357, 809–813. doi: 10.1098/rstb.2002.1089
- Martin, C., and Gerats, T. (1993). Control of pigment biosynthesis genes during petal development. *Plant Cell* 5, 1253–1264. doi: 10.1105/tpc.5.10.1253
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- Martinell, M. C., Rovira, A., Blanché, C., and Bosch, M. (2011). Shift towards autogamy in the extremely narrow endemic *Aquilegia pawi* and comparison with its widespread close relative *A. vulgaris* (Ranunculaceae). *Plant Syst. Evol.* 295, 73–82. doi: 10.1007/s00606-011-0463-x
- Mitchell, R. J., and Shaw, R. G. (1993). Heritability of floral traits for the perennial wild flower *Penstemon centranthifolius* (Scrophulariaceae): clones and crosses. *Heredity* 71, 185–192. doi: 10.1038/hdy.1993.123
- Mizukami, Y. (2001). A matter of size: developmental control of organ size in plants. *Curr. Opin. Cell Biol.* 4, 533–539. doi: 10.1016/S1369-5266(00)00212-0
- Mizukami, Y., and Fischer, R. L. (2000). Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 97, 942–947. doi: 10.1073/pnas.97.2.942
- Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4623–4628. doi: 10.1073/pnas.0907801107
- Noda, K., Glover, B. J., Linstead, P., and Martin, C. (1994). Flower color intensity depends on specialized cell-shape controlled MYB-related transcription factor. *Nature* 369, 661–664. doi: 10.1038/369661a0
- Ojeda, I., Francisco-Ortega, J., and Cronk, Q. C. B. (2009). Evolution of petal epidermal micromorphology in Leguminosae and its use as a marker of petal identity. *Ann. Bot.* 104, 1099–1110. doi: 10.1093/aob/mcp211
- Ojeda, I., Santos-Guerra, A., Caujape-Castells, J., Jaen-Molina, R., Marrero, A., and Cronk, Q. C. B. (2012). Comparative micromorphology of petals in Macaronesian *Lotus* (Leguminosae) reveals a loss of papillose conical cells during the evolution of bird pollination. *Int. J. Plant Sci.* 173, 365–374. doi: 10.1086/664713
- Okuyama, Y., Fujii, N., Wakabayashi, M., Kawakita, A., Ito, M., Watanabe, M., et al. (2005). Nonuniform concerted evolution and chloroplast capture: heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). *Mol. Biol. Evol.* 22, 285–296. doi: 10.1093/molbev/msi016
- Parks, M., Cronn, R., and Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 7:84. doi: 10.1186/1741-7007-7-84
- Perez-Rodriguez, M. (2005). Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of *Antirrhinum majus* flowers. *Development* 132, 359–370. doi: 10.1242/dev.01584
- Pina, S. D., Souer, E., and Koes, R. (2014). Arguments in the evo-devo debate: say it with flowers! *J. Exp. Bot.* 65, 2231–2242. doi: 10.1093/jxb/eru111
- Porter, J. M., and Johnson, L. A. (2000). A phylogenetic classification of Polemoniaceae. *Aliso* 19, 55–91.
- Prather, L. A., Ferguson, C. J., and Jansen, R. K. (2000). Polemoniaceae phylogeny and classification: implications of sequence data from the chloroplast gene *ndhF*. *Am. J. Bot.* 87, 1300–1308. doi: 10.2307/2656723
- Puzey, J. R., Gerbode, S. J., Hodges, S. A., Kramer, E. M., and Mahadevan, L. (2012). Evolution of spur-length diversity in *Aquilegia* petals is achieved solely through cell-shape anisotropy. *Proc. Biol. Sci.* 279, 1640–1645. doi: 10.1098/rspb.2011.1873
- Reale, L., Porceddu, A., Lanfalone, L., Moretti, C., Zenoni, S., Pezzotti, M., et al. (2002). Patterns of cell division and expansion in developing petals of *Petunia hybrida*. *Sex. Plant Reprod.* 15, 123–132. doi: 10.1007/s00497-002-0150-8
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E., and Burleigh, J. G. (2014). From algae to angiosperms—inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. *BMC Evol. Biol.* 14:23. doi: 10.1186/1471-2148-14-23
- Sang, T., Crawford, D. J., and Stuessy, T. F. (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Am. J. Bot.* 84, 1120–1136. doi: 10.2307/2446155
- Schemske, D. W., and Bradshaw, H. D. (1999). Pollinator preference and the in monkeyflowers (*Mimulus*) evolution of floral traits. *Proc. Natl. Acad. Sci. U.S.A.* 96, 11910–11915. doi: 10.1073/pnas.96.21.11910
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019
- Schoen, D. J., Morgan, M. T., and Bataillon, T. (1996). How does self-pollination evolve? Inferences from floral ecology and molecular genetic variation. *Phil. Trans. R. Soc. Lond. B.* 351, 1281–1290. doi: 10.1098/rstb.1996.0111
- Schönenberger, J. (2009). Comparative Floral Structure and Systematics of Fouquieriaceae and Polemoniaceae (Ericales). *Int. J. Plant Sci.* 170, 1132–1167. doi: 10.1086/605875
- Small, R. L., Ryburn, J. A., Cronn, R. C., Seelanan, T., and Wendel, J. F. (1998). The tortoise and the hare: choosing between noncoding plastome and nuclear *ADH* sequences for phylogeny reconstruction in a recently diverged plant group. *Am. J. Bot.* 85, 1301–1315. doi: 10.2307/2446640
- Soltis, D. E., and Kuzoff, R. K. (1995). Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49, 727–742. doi: 10.2307/2410326
- Stamatakis, A. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stebbins, G. L. (1974). *Flowering Plants: Evolution above the Species Level*. Cambridge, MA: Belknap Press.
- Strakosh, S. C., and Ferguson, C. J. (2005). Pollination biology of four Southwestern species of *Phlox* (Polmeoniaceae): insect visitation in relation to corolla tube length. *Southwest. Nat.* 50, 291–301. doi: 10.1894/0038-4909(2005)050[0291:PBOFSS]2.0.CO;2
- Stuurman, J., Hoballah, M. E., Broger, L., Moore, J., Basten, C., and Kuhlemeier, C. (2004). Dissection of floral pollination syndromes in *Petunia*. *Genetics* 168, 1585–1599. doi: 10.1534/genetics.104.031138
- Tedder, A., Carlial, S., Golebiewska, M., Kappel, C., Shimizu, K. K., and Stift, M. (2015). Evolution of the selfing syndrome in *Arabidopsis alpinus* (Brassicaceae). *PLoS ONE* 10:e012618. doi: 10.1371/journal.pone.0126618
- Thompson, A. W., Betancur-R, R., López-Fernández, H., and Ortí, G. (2014). A time-calibrated, multi-locus phylogeny of piranhas and pacus (Characiformes: Serrasalminidae) and a comparison of species tree methods. *Mol. Phylogenet. Evol.* 81, 242–257. doi: 10.1016/j.ympev.2014.06.018
- Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., and Ortí, G. (2015). Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. *PLoS Curr.* 7:ecurrents.tol.34260cc27551a527b124ec5f6334b6be. doi: 10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be
- UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acid Res.* 43, D204–D212. doi: 10.1093/nar/gku989
- Van der Niet, T., and Johnson, S. D. (2012). Phylogenetic evidence for pollinator-driven diversification of angiosperms. *Trends Ecol. Evol.* 27, 353–361. doi: 10.1016/j.tree.2012.02.002
- Wang, Y., Coleman-Derr, D., Chen, G., and Gu, Y. Q. (2015). OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 43, W78–W84. doi: 10.1093/nar/gkv487
- Waser, N. M. (2006). “Specialization and generalization in plant-pollinator interactions: a historical perspective,” in *Plant-Pollinator Interactions from*

- Specialization to Generalization*, eds N. Waser and J. Ollerton (Chicago, IL: University of Chicago Press), 3–17.
- Waser, N. M., Chittka, L., Price, M. V., Williams, N. M., and Ollerton, J. (1996). Generalization in pollination systems, and why it matters. *Ecology* 77, 1043–1060. doi: 10.2307/2265575
- Weese, T. L., and Johnson, L. A. (2005). Utility of NADP-dependent isocitrate dehydrogenase for species-level evolutionary inference in angiosperm phylogeny: a case study in *Saltugilia*. *Mol. Phylogenet. Evol.* 36, 24–41. doi: 10.1016/j.ympev.2005.03.004
- Weiss, J., Delgado-Benarroch, L., and Egea-Cortines, M. (2005). Genetic control of floral size and proportions. *Int. J. Dev. Biol.* 49, 513–525. doi: 10.1387/ijdb.051998jw
- Whitney, H. M., and Glover, B. J. (2007). Morphology and development of floral features recognised by pollinators. *Arthropod Plant. Interact.* 1, 147–158. doi: 10.1007/s11829-007-9014-3
- Whittall, J. B., and Hodges, S. A. (2007). Pollinator shifts drive increasingly long nectar spurs in columbine flowers. *Nature* 447, 706–709. doi: 10.1038/nature05857
- Xu, R., and Li, Y. (2011). Control of final organ size by mediator complex subunit 25 in *Arabidopsis thaliana*. *Development* 138, 4545–4554. doi: 10.1242/dev.071423
- Yant, L., Collani, S., Puzey, J., Levy, C., and Kramer, E. M. (2015). Molecular basis for three-dimensional elaboration of the *Aquilegia* petal spur. *Proc. R. Soc. B Biol. Sci.* 282:20142778. doi: 10.1098/rspb.2014.2778
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2016 Landis, O'Toole, Ventura, Gitzendanner, Oppenheimer, Soltis and Soltis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Loss of *YABBY2*-Like Gene Expression May Underlie the Evolution of the Laminar Style in *Canna* and Contribute to Floral Morphological Diversity in the Zingiberales

Kelsie Morioka¹, Roxana Yockteng^{1,2,3}, Ana M. R. Almeida^{1,4} and Chelsea D. Specht^{1*}

¹ Department of Plant and Microbial Biology, Department of Integrative Biology and the University and Jepson Herbaria, University of California at Berkeley, Berkeley, CA, USA, ² Corporación Colombiana de Investigación Agropecuaria (CORPOICA), Centro de Investigaciones Tibaitatá, Tibaitatá, Colombia, ³ Institut de Systématique, Évolution, Biodiversité, UMR 7205 Centre National de la Recherche Scientifique, Muséum National d'Histoire Naturelle, Paris, France, ⁴ Programa de Pós-graduação em Genética e Biodiversidade, Universidade Federal da Bahia, Salvador, Brazil

OPEN ACCESS

Edited by:

Jocelyn Hall,
University of Alberta, Canada

Reviewed by:

Bruce Veit,
AgResearch, New Zealand
Marcelo Camier Dornelas,
Universidade Estadual de Campinas,
Brazil

*Correspondence:

Chelsea D. Specht
cdspecht@berkeley.edu

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 15 August 2015

Accepted: 22 November 2015

Published: 16 December 2015

Citation:

Morioka K, Yockteng R, Almeida AMR
and Specht CD (2015) Loss of
YABBY2-Like Gene Expression May
Underlie the Evolution of the Laminar
Style in *Canna* and Contribute to
Floral Morphological Diversity in the
Zingiberales. *Front. Plant Sci.* 6:1106.
doi: 10.3389/fpls.2015.01106

The Zingiberales is an order of tropical monocots that exhibits diverse floral morphologies. The evolution of petaloid, laminar stamens, staminodes, and styles contributes to this diversity. The laminar style is a derived trait in the family Cannaceae and plays an important role in pollination as its surface is used for secondary pollen presentation. Previous work in the Zingiberales has implicated *YABBY2*-like genes, which function in promoting laminar outgrowth, in the evolution of stamen morphology. Here, we investigate the evolution and expression of Zingiberales *YABBY2*-like genes in order to understand the evolution of the laminar style in *Canna*. Phylogenetic analyses show that multiple duplication events have occurred in this gene lineage prior to the diversification of the Zingiberales. Reverse transcription-PCR in *Canna*, *Costus*, and *Musa* reveals differential expression across floral organs, taxa, and gene copies, and a role for *YABBY2*-like genes in the evolution of the laminar style is proposed. Selection tests indicate that almost all sites in conserved domains are under purifying selection, consistent with their functional relevance, and a motif unique to monocot *YABBY2*-like genes is identified. These results contribute to our understanding of the molecular mechanisms underlying the evolution of floral morphologies.

Keywords: *YABBY*, *YABBY2*, gene evolution, gene expression, floral development, plant evolution, Zingiberales, *Canna*

INTRODUCTION

Laminar outgrowth is a key process in the development of lateral organs, facilitating light capture and gas exchange (leaves), pollinator attraction (petals and sometimes other floral organs), and protection of the floral bud (sepals and bracts). It has been hypothesized that, at the molecular level, laminar outgrowth is promoted by the juxtaposition of abaxial and adaxial cell fates

(Waites and Hudson, 1995). Numerous studies have identified and characterized the genes involved in establishing abaxial-adaxial polarity and the promotion of laminar expansion of lateral organs. In *Arabidopsis*, studies of loss-of-function and gain-of-function mutants indicate that *KANADI*, *ARF3/ETT*, and *ARF4* genes play roles in specifying abaxial cell fate, while *HD-ZIP III* (*PHABULOSA*, *PHAVOLUTA*, and *REVOLUTA*) and *AS1/AS2* genes play roles in specifying adaxial cell fate (for more detailed review of these studies, see Bowman et al., 2002; Yamaguchi et al., 2012). *YABBY* genes are likely to act downstream in the adaxial-abaxial polarity regulatory network, promoting blade outgrowth at the abaxial-adaxial boundary (Husbands et al., 2009).

The *YABBY* gene family was believed to be specific to seed plants, but *YABBY* genes have been found in the green alga *Micromonas pusilla* (Worden et al., 2009). Since these genes have not been found in any other non-seed plants, it is unclear when this gene family evolved in plants. Members of this family encode transcription factors characterized by two domains, a zinc finger domain at the N terminus and a *YABBY* domain at the C terminus (Bowman and Smyth, 1999; Sawa et al., 1999a). The *YABBY* domain is similar in structure to the high mobility group (HMG) domain (Sawa et al., 1999a) and is necessary for DNA binding to occur (Kanaya et al., 2002). Four gene duplication events in this gene family have occurred prior to the diversification of the angiosperms (Bartholmes et al., 2012), leading to genes with both novel and redundant functions.

In angiosperms, *YABBY* genes have important roles in laminar expansion of lateral organs as well as in reproductive organ development and other processes. The *YABBY* gene family has six members in *Arabidopsis thaliana*. *FILAMENTOUS FLOWER* (*FIL*), *YABBY2*, and *YABBY3* are expressed in the abaxial domain of lateral organs and act redundantly to specify abaxial cell fate and ultimately to promote laminar outgrowth (Siegfried et al., 1999); *FIL* is also required for normal inflorescence and flower development (Chen et al., 1999; Sawa et al., 1999b). *CRABS CLAW* (*CRC*) and *INNER NO OUTER* (*INO*) have specialized functions in reproductive organ development: *CRC* is expressed in carpels and nectaries and is necessary for gynoecium elongation and nectary development (Alvarez and Smyth, 1999; Bowman and Smyth, 1999), and *INNER NO OUTER* (*INO*) is expressed in the abaxial domain of the outer integument and is required for normal outer integument development (Baker et al., 1997; Villanueva et al., 1999).

Studies in other eudicot species suggest that the abaxial expression pattern and specific function in promoting laminar outgrowth may be conserved across the eudicot lineage. In tomato, *LeYAB B* is expressed abaxially and may also specify abaxial identity (Kim et al., 2003). In *Antirrhinum majus*, *GRAMINIFOLIA* (*GRAM*), a *FIL*-like gene, and *PROLONGATA* (*PROL*), a *YABBY5*-like gene, are expressed abaxially and promote laminar outgrowth (Golz et al., 2004). *GRAM* has also been shown to be involved in the control of floral organ initiation and identity (Navarro et al., 2004). *TmFIL* in *Tropaeolum majus* and *SrGRAM* in *Streptocarpus rexii*, both *FIL*-like genes, are also expressed abaxially (Gleissberg et al., 2005; Tononi et al., 2010).

Studies in monocots, which have focused mainly on grass species (Poales), have demonstrated varied expression patterns leading to a diversity of proposed functions. In *Zea mays* (maize), the *FIL/YAB3*-like genes *ZYB9* and *ZYB14* are expressed adaxially and may play a role in lateral outgrowth (Juarez et al., 2004). In contrast, *OsYABBY1*, a *YABBY2*-like gene from *Oryza sativa* (rice), is expressed in precursor cells that give rise to abaxial sclerenchyma in the leaves, the mestome sheath in the large vascular bundle, and sclerenchymatous cells in the palea and lemma of the flower, and is thus proposed to specify differentiation of certain cell types in rice (Toriba et al., 2007). The *FIL*-like gene *OsYABBY4* is expressed in meristems and in developing phloem and thus may be involved in vasculature development in rice (Liu et al., 2007). The gene *AaCRC* from *Asparagus asparagoides* is expressed in the abaxial region of the ovary wall and leaf phloem (Nakayama et al., 2010), while *CRC/DL* orthologs in Poales species are expressed throughout the carpel and in the central region of the leaf and may specify carpel identity and midrib formation (Yamaguchi et al., 2004; Ishikawa et al., 2009). It seems that divergence in expression (and possibly function) of *YABBY* genes has occurred more so in the monocot lineage than in the eudicot lineage.

Studies from the early diverging angiosperms show variation in expression as well. *AmbF1*, a *YABBY2* homolog from *Amborella trichopoda*, is expressed adaxially in all floral organs, the shoot apex, and leaves (Yamada et al., 2004). *YABBY* genes from *Cabomba caroliniana* (*CcFIL*, *CcYAB5*, *CcINO*, and *CcCRC*), however, are expressed abaxially, as in eudicots (Yamada et al., 2011). In two *Nymphaea* species, *CcINO* is expressed in the outer epidermis of the outer integument, as is *INO* in *Arabidopsis*, but is also expressed in the inner integument and the tip of the nucellus (Yamada et al., 2003). *AmbCRC* from *A. trichopoda* is expressed in the abaxial carpel, maintaining a similar expression pattern to that observed in *Arabidopsis* (Fourquin et al., 2005).

Overall, abaxial expression seems to be conserved across eudicots, and *YABBY* function in controlling laminar outgrowth seems to be generally conserved across angiosperms, but shifts in expression have been observed, particularly in the monocots. Based on the expression patterns of *YABBY* genes across angiosperms, the strong expression of these genes in ectopic outgrowths on both abaxial and adaxial surfaces in polarity mutants, and the later timing of *YABBY* gene expression relative to that of other genes in the polarity gene network, Husbands et al. (2009) have proposed that the ancestral function of *YABBY* genes in angiosperms may have been to promote blade outgrowth at abaxial-adaxial boundaries.

The evolution of laminarity in the androecium and gynoecium contributes to diversity in floral morphology and the evolution of plant-pollinator interactions in the tropical monocot order Zingiberales (Specht et al., 2012). This group includes the four paraphyletic banana families (Musaceae, Strelitziaceae, Lowiaceae, and Heliconiaceae) and a monophyletic group of four ginger families (Cannaceae, Marantaceae, Zingiberaceae, and Costaceae). The banana lineages are characterized by five or six fertile stamens with radial filaments, and members of Heliconiaceae have one laminar staminode, while members of

the ginger clade exhibit a reduction in fertile stamen number—to one in Zingiberaceae and Costaceae and to one-half (i.e., fertile stamen with a single theca) in Cannaceae and Marantaceae—and are characterized by laminar, petaloid staminodes and fertile stamens. Members of Zingiberaceae and Costaceae possess a novel petaloid structure, the labellum, formed from two or four (Zingiberaceae) or five (Costaceae) fused laminar staminodes. The staminodes and the labellum are likely to function in pollinator attraction, making up most of the floral display in terms of showiness, coloration, and symmetry (Specht et al., 2012). In Cannaceae, the gynoeceum style is also laminar and is used for secondary pollen presentation: during flower development, pollen adheres to the lateral surface of the style and is then transferred to the bill of a hummingbird pollinator once the flower opens (Glinos and Cocucci, 2011).

The evolution of laminarity in different floral organs in the Zingiberales makes this group a useful system in which to investigate the evolution of this trait and the evolution of genes involved in the abaxial-adaxial polarity gene network. Results from a recent study on the evolution of stamen morphology in the Zingiberales (Almeida et al., 2014) implicated balanced expression of abaxial-adaxial polarity genes in the formation of laminar filaments in the ginger clade, while overexpression of a *YABBY2/5* gene was implicated in the formation of radial filaments in *Musa*. Interestingly, the authors also demonstrated a similar gene expression pattern in *Brassica rapa* radial filaments, suggesting that *YABBY2/5* genes are involved in the evolution of filament morphology in angiosperms (Almeida et al., 2014).

In order to better understand the evolution of the *YABBY2* gene subfamily in the Zingiberales and its role in the evolution of laminarity in the style of Cannaceae, we isolated homologs of *Arabidopsis YABBY2* from taxa across the order and performed phylogenetic and expression analyses to investigate the evolutionary history of the gene subfamily. We identified duplication events occurring prior to the diversification of the Zingiberales and more recent, lineage-specific duplication events within the order. These data, in combination with expression data from semi-quantitative RT-PCR, were used to describe a hypothesized role of *YABBY2*-like genes in the evolution of the laminar style. We also tested for selection along branches in the *YABBY2* gene subfamily and looked for motifs characteristic of *YABBY2*-like genes.

MATERIALS AND METHODS

Plant Material and cDNA Synthesis

Floral buds of Zingiberales species (Table 1) were collected and immediately frozen in liquid nitrogen. Floral tissue was stored at -80°C until RNA extraction. Floral organs of young flowers from *Costus spicatus*, *Canna indica*, and *Musa basjoo* were also dissected in order to extract organ-specific RNA. Organ-specific material was immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction. RNA extractions were performed with Plant RNA Reagent (Life Technologies) according to Yockteng et al. (2013). RNA was treated with Turbo DNase (Ambion) and cDNA was synthesized from

1.0 μg of DNase-treated RNA using iScript reverse transcriptase and oligo(dT) primers following the manufacturer's protocol (Bio-Rad Laboratories). As a control, a cDNA synthesis reaction without reverse transcriptase was set up for each sample.

Isolation of *YABBY2* in the Zingiberales Sequence Retrieval

Previously published sequences from across the *YABBY* gene family were retrieved from NCBI, with representatives from early diverging angiosperms, monocots, early diverging eudicots, core eudicots, and gymnosperms. *YABBY* sequences from NCBI were blasted against the genome of *Musa acuminata* (D'Hont et al., 2012) and against the whole flower transcriptomes of *Costus spicatus*, *Canna indica*, and *Musa basjoo* (unpublished; Roxana Yockteng, Ana M. R. Almeida, and Chelsea D. Specht)

TABLE 1 | Zingiberales taxa used in this study and obtained *YABBY2*-like sequences.

Species	Location	NCBI accession numbers
<i>Costus spicatus</i> (Jacq.) Sw.	2002-127 (NMNH)	KT795168-KT795183
<i>Canna indica</i> L.	2011-777 (UC Specht Lab)	KT795161
<i>Musa acuminata</i> Colla	2002-075 (NMNH)	KT795221-KT795230
<i>Musa basjoo</i> Siebold	89.0873 (UCBG)	KT795231-KT795238
<i>Zingiber officinale</i> Roscoe	MB0876 (UC Specht Lab)	KT795282-KT795284
<i>Strelitzia</i> sp. Banks	UC Specht Lab	KT795268-KT795281
<i>Marantochloa leucantha</i> (K. Schum.) Milne-Redh.	L-80.0376 (HLA)	KT795214-KT795220
<i>Orchidantha siamensis</i> K. Larsen	L-91.0308 (HLA)	KT795240-KT795247
<i>Phrynium oliganthum</i> Merr.	L-96.0226 (HLA)	KT795248-KT795251
<i>Schumannianthus virgatus</i> (Roxb.) Rolfe	L-83.0899 (HLA)	KT795258-KT795267
<i>Heliconia pendula</i> Wawra	71003-003 (McBryde)	KT795200-KT795207
<i>Kaempferia rubromarginata</i> (S. Q. Tong) R. J. Searle	L-2003.0153 (HLA)	KT795211-KT795213
<i>Etlingera corneri</i> Mood and Ibrahim	L-91.0443 (HLA)	KT795184-KT795187
<i>Pleuranthodium helwigii</i> (K. Schum.) R. M. Sm.	L-99.0492 (HLA)	KT795252-KT795257
<i>Globba laeta</i> K. Larsen	L-92.0182 (HLA)	KT795188-KT795195
<i>Heliconia</i> sp. L.	UC Specht Lab	KT795208-KT795210
<i>Canna jaegeriana</i> Urb.	MB0854 (UC)	KT795162-KT795167
<i>Heliconia caribaea</i> x <i>bihai</i> Lam. and L.	MB0862 (UC Specht Lab)	KT795196-KT795199
<i>Musa textilis</i> Née	1682/77 (NYBG)	KT795239

HLA, Lyon Arboretum, Oahu, Hawaii; McBryde, McBryde Botanical Garden, Kauai, Hawaii; UCBG, University of California Botanical Garden; UC, University of California; NMNH, Smithsonian Greenhouses; NYBG, New York Botanical Garden.

to retrieve YABBY genes from these taxa. We also used RNA-seq reads from the Monocot AToL (Angiosperm Tree of Life)¹ project to assemble transcriptomes from *Costus pulverulentus*, *Canna indica*, *Curcuma roscoeana*, *Orchidantha fimbriata*, *Heliconia collinsiana*, *Zingiber spectabile*, *Strelitzia reginae*, and *Maranta leuconeura* using Trinity r2013_08_14 (Grabherr et al., 2011). YABBY sequences were again blasted against these transcriptomes to retrieve YABBYs from these taxa. Source and accession numbers for all of the sequences used in this study are shown in Table S1.

Primer Design

The YABBY2 sequences from NCBI, the *Musa acuminata* genome, and the floral transcriptomes were aligned using the Geneious v5.6 algorithm (Drummond et al., 2012) and manually edited to further refine the alignment. This multiple sequence alignment was used to design forward and reverse primers at conserved sites of the gene—at the N-terminal end of the zinc finger domain and the C-terminal end of the YABBY domain, respectively—and these primers were used to amplify YABBY2 from taxa across the Zingiberales. Zingiberales sequences then obtained through cloning were edited and added to the alignment, along with the transcriptome contigs from MonAToL, and a preliminary tree was used to design clade-specific primers of Zingiberales sequences falling in the YABBY2 clade in order to amplify more sequences from Zingiberales taxa in each clade. All primers used in this study are found in Table S2.

Cloning of YABBY2 in the Zingiberales

YABBY2 orthologs from taxa across the eight families in the Zingiberales (Table 1) were amplified using 1.0 μ l of cDNA diluted 1:50 in water, 0.3 μ mol of each primer, and Phire II Hot Start DNA Polymerase (Thermo Scientific) as follows: 5 min at 98°C for initial denaturing; 40 cycles of 5 s at 98°C for denaturing, 5 s at a primer-specific annealing temperature for annealing, and 20 s at 72°C for extension; and 1 min at 72°C for final extension. PCR products were cloned into pJET1.2 blunt cloning vector (Thermo Scientific). Sequencing was performed using BigDye v3.1 on a 3730 Applied Biosystems DNA analyzer at the Museum of Vertebrate Zoology Evolutionary Genetics Laboratory at UC Berkeley.

Phylogenetic Analyses

All of the Zingiberales sequences obtained in this study and YABBY sequences retrieved from NCBI, the *Musa acuminata* genome, the floral transcriptomes listed above, and the MonAToL transcriptomes listed above were aligned using the Geneious v5.6 algorithm and manually edited using a codon-preserving approach to further refine the alignment. Unalignable regions outside of the zinc finger and YABBY domains were removed, resulting in a truncated alignment used for further analyses. The final truncated alignment was composed of a 132 nucleotide-long conserved region including the zinc finger domain and a 168 nucleotide-long conserved region including

the YABBY domain (Figure S1), with a total alignment length of 300 nucleotides.

jModeltest 2.1.1 (Darriba et al., 2012) was used for selection of the best fit model of nucleotide evolution, and indicated the TVMef+I+G model as most appropriate for the given nucleotide alignment according to the Bayesian information criterion (BIC).

Bayesian inference was used to infer a phylogeny using MrBayes 3.2.2 (Ronquist and Huelsenbeck, 2003) in the Cipres Science Gateway² using the model specified above. The MCMC was run for 10,000,000 generations and the output files were analyzed in Tracer (Rambaut et al., 2014) to check for convergence of the two chains.

A maximum likelihood phylogeny with 100 bootstrap replicates was reconstructed using PhyML 3.0 (Guindon et al., 2010) using the model specified in jModeltest.

Selection Tests

SLAC, FEL, and branch-site REL selection tests were implemented in HYPHY (Kosakovsky Pond et al., 2005; Kosakovsky Pond and Frost, 2005a) on the Datamonkey webserver (Kosakovsky Pond and Frost, 2005b) to identify sites under positive or negative (purifying) selection across the truncated multiple sequence alignment. The SLAC analysis was run using the same 300 nucleotide-long truncated multiple sequence alignment used for phylogenetic analyses, containing 133 sequences (Figure S1), and the maximum likelihood tree reconstructed using PhyML (Figure 2). The nucleotide model used was chosen by a Datamonkey model selection analysis. The global dN/dS value was estimated and ambiguities were averaged, and the significance level was set to 0.05. The FEL analysis was run using the same alignment, nucleotide model, and ML tree, and the significance level was set to 0.05. A branch-site REL analysis was also implemented in HYPHY using a reduced multiple sequence alignment containing 72 sequences representing all major clades in the YABBY gene tree, and a ML tree generated in PhyML.

We also tested for positive selection on codons along 9 branches of major clades in the YABBY2 subfamily, in separate analyses for each branch, using branch-site model A in codeml (model = 2 and NS sites = 2) implemented in PAML (Yang, 2007). For each analysis, a likelihood ratio test (LRT) was used to determine if the difference between the likelihood scores of the alternative model (model A, in which ω for the branch of interest is estimated and the background ω is set to 0) and the null model (in which the distribution of ω is set to 1) was significantly different (degrees of freedom = 1). For each LRT, a *p*-value of 0.05 or less was required for results to be considered significant.

Motif Identification Using MEME

MEME (Multiple Em for Motif Elicitation) version 4.10.0 (Bailey and Elkan, 1994) was used to identify ungapped motifs in translated YABBY2 protein sequences. Since this analysis does not require an alignment of sequences to identify conserved motifs, it could be used to search for motifs outside of the alignable regions. The “Normal” mode of motif discovery was

¹<http://www.botany.wisc.edu/givnish/Givnish/MonAToL.html>.

²<http://www.phylo.org/portal2/>.

used to search for motifs within the default width range (6–50 amino acids), and the site distribution for motifs was set to zero or one occurrence per sequence.

Semi-Quantitative RT-PCR

Reverse transcription PCR was used to determine presence or absence of expression of copies of *YABBY2* in the total flower, floral organs, and young leaves of *Costus spicatus*, *Musa basjoo*, and *Canna indica*. RNA extraction and cDNA synthesis were as described above. cDNA was synthesized from *C. spicatus* total flower, sepals, petals, labellum, theca, filament, gynoecium, style, and young leaves; *M. basjoo* total flower, free petal, floral tube (fused sepals and petals), theca, filament, gynoecium, style, and young leaves; and *C. indica* total flower, sepals, petals, petaloid fertile stamen, theca, staminodes, gynoecium, style, and young leaves. Primers for each gene copy were designed across intron-exon boundaries to avoid amplification of trace contaminating gDNA. All RT-PCRs were performed using 1.0 μ l of cDNA diluted 1:20 in water, 0.3 μ mol of copy-specific forward and reverse primers (Table S2), and Phire Hot-Start II DNA Polymerase (Thermo Scientific) as follows: 5 min at 98°C for initial denaturing followed by 35–37 cycles of 5 s at 98°C for denaturing, 5 s at a primer-specific annealing temperature for annealing, and 25 s at 72°C for elongation. The PCR products were run on 1% agarose gels.

Additionally, semi-quantitative RT-PCR was used to gauge relative levels of expression of *YABBY2* copies in the styles of *Costus spicatus*, *Musa basjoo*, and *Canna indica*. For each gene copy, five different reactions were run with different numbers of cycles (25, 27, 30, 32, and 35 cycles), while all other conditions were as described above. The PCR products were run on 1% agarose gels.

Products from the RT-PCRs were sequenced to confirm that the products being amplified were the targeted gene copy. At least three technical replicates were run for each reaction. *ACTIN* was used as an endogenous control; *ACTIN* primers used for each species are included in Table S2.

RESULTS

We obtained sequences for *YABBY2*-like genes from representative taxa across the Zingiberales in order to investigate the evolution of this gene subfamily and its potential role in the evolution of floral morphology in this order, particularly the evolution of laminarity in the gynoecium. To obtain sampling across the Zingiberales phylogeny, at least one taxon was sampled from each family. A total of 124 sequences were obtained from 19 taxa through cloning (Table 1). The sequences obtained in this study were blasted against the NCBI database to confirm that they do indeed blast to previously published *YABBY2* genes, and consensus sequences were made for sequences from the same taxa that shared at least 98% identity. Six contigs from the whole flower transcriptomes of *Costus spicatus* and *Musa basjoo* (unpublished; Roxana Yockteng, Ana M.R. Almeida, and Chelsea D. Specht) and four contigs from transcriptomes of *Canna indica* and *Costus pulverulentus* from MonAToL were also used in analyses. All

sequences obtained in this study have been deposited in GenBank (KT795161–KT795284).

Almost all sequences used in phylogenetic analyses contain the zinc finger and *YABBY* domains characteristic of genes of the *YABBY* gene family; some sequences are missing some of the zinc finger domain and/or the *YABBY* domain due to primer design limitations. Sequences from at least one taxon from each family were included in the final alignment used for phylogenetic analyses. This alignment also included sequences from other monocots, core eudicots, early diverging eudicots, early diverging angiosperms, and gymnosperms.

Evolution of *YABBY2*-Like Genes in the Zingiberales

The maximum likelihood reconstruction of the *YABBY* gene family is shown (Figure 2). Bootstrap values are shown at tree nodes, with bootstrap values greater than 50 in bold. When rooted with the gymnosperm group, the topology recovered is consistent with one of the previously proposed topologies for this gene family (Bartholmes et al., 2012), with the *YABBY2* and *YABBY5* subfamilies sister to the *INO*, *FIL*, and *CRC* subfamilies. The *FIL* and *CRC* subfamilies are sister to each other, and the *INO* subfamily is sister to both. The *YABBY2* and *YABBY5* subfamilies form sister clades.

Within the *YABBY2* subfamily, the *YABBY2*-like sequence from the early diverging angiosperm *Amborella trichopoda* is sister to all other *YABBY2*-like sequences from eudicots and monocots. The eudicot *YABBY2*-like sequences form a clade sister to a clade of monocot *YABBY2*-like sequences. The monocot *YABBY2*-like clade is further divided into two sister clades: one of these clades includes a clade of Poales sequences sister to three separate clades of Zingiberales sequences (*ZinYAB2-1*, *ZinYAB2-2*, and *ZinYAB2-3*), and the other monocot clade is composed of a clade of non-Zingiberales monocot sequences (including Poales sequences) sister to a clade of Zingiberales sequences (*ZinYAB2-4*) and one *Elaeis guineensis* sequence. Each of the Zingiberales clades includes at least one sequence from each family in the Zingiberales, with the exception of *ZinYAB2-1*, which does not have any sequences from Cannaceae. The Poales clade, the *ZinYAB2-1* clade, and the (*Elaeis guineensis* + *ZinYAB2-4*) clade are moderately well supported by the bootstrap analysis (with bootstrap values of 93, 85, and 68, respectively); bootstrap values for other clades are low, but have high or moderately high posterior probabilities in a Bayesian analysis (Figure S2).

Expression of *ZinYAB2* Genes

Reverse transcription PCR was used to evaluate the presence and absence of expression of *YABBY2*-like genes in the floral organs and young leaves of *Canna indica* (Cannaceae), *Costus spicatus* (Costaceae), and *Musa basjoo* (Musaceae). Semi-quantitative RT-PCR was also used to gauge relative expression of *YABBY2*-like genes in the styles of these species. These taxa span the Zingiberales phylogeny and were chosen to represent the diverse floral morphology found in the order (Figure 1). Differential expression between species could indicate changes in function of *YABBY2*-like genes in the development of floral organs.

Expression profiles are represented (**Figure 3**) and gel images for the RT-PCRs can be found in the Supplementary data (Figure S3).

In the Zingiberales, YABBY2-like genes show differential expression between gene copies, between species, and across floral organs and leaves (**Figure 3**). In *Musa basjoo*, *ZinYAB2-1* is expressed in all floral organs and in young leaves. In *Costus spicatus*, *ZinYAB2-1a* and *ZinYAB2-1b* are expressed in all floral organs except for stamen theca and style; *ZinYAB2-1b* is present in *C. spicatus* young leaves while *ZinYAB2-1a* is absent (**Figure 3**). In *C. indica*, *ZinYAB2-1* has been lost or has not been recovered in this study. *ZinYAB2-2* is expressed in all floral organs in both *M. basjoo* and *C. spicatus*, while in *C. indica* this copy is expressed in petal, petaloid filament, staminode, gynoecium, and young leaves and is absent from sepal, theca, and style (**Figure 3**). In *M. basjoo*, there are three copies of *ZinYAB2-3*, resulting from duplication events in the Musaceae lineage (**Figure 2**): *ZinYAB2-3a* is expressed in all floral organs except free petal and filament, and *ZinYAB2-3c* is expressed in all floral organs (**Figure 3**). *ZinYAB2-3c* is the only *ZinYAB2-3* copy in *M. basjoo* that is expressed in young leaves. *ZinYAB2-3* in both *C. spicatus* and *C. indica* is expressed in young leaves and in all floral organs except theca. In *M. basjoo*, *ZinYAB2-4* is expressed in free petal and floral tube and is absent from filament, theca, gynoecium, style, and young leaves. In *C. spicatus*, *ZinYAB2-4* is present in total flower (as confirmed by three technical replicates) but was not amplified by RT-PCR for any floral organs; it is possible that this copy is so lowly expressed that it could not be amplified from floral organ tissue. In *C. indica*, *ZinYAB2-4* is expressed in sepal, petal, and theca, and is absent from filament, staminode, gynoecium, style, and young leaves. *ZinYAB2-4* seems to be expressed only in the flower, since there is no expression in leaves (**Figure 3**).

Interestingly, *ZinYAB2-4* is absent from filament and gynoecium in all species considered; expression of this copy has diverged the most from that of the other YABBY2-like gene copies in the Zingiberales.

Across the Zingiberales, there is a pattern of reduction in the number of YABBY2-like gene copies expressed in the style: five gene copies are expressed in *M. basjoo* (*ZinYAB2-1*, *ZinYAB2-2*, *ZinYAB2-3a*, *ZinYAB2-3b*, and *ZinYAB2-3c*), two in *C. spicatus* (*ZinYAB2-2* and *ZinYAB2-3*), and one in *C. indica* (*ZinYAB2-3*). *Musa basjoo* and *C. spicatus* have the same style morphology—radial—but YABBY2-like expression between styles in these two species differs not only in the number of gene copies expressed, but also in the overall expression level of YABBY2-like genes (**Figure 3**). One copy, *ZinYAB2-2*, is expressed in *M. basjoo* and *C. spicatus* styles but is absent from the style of *C. indica*.

Selection in *ZinYAB2* Genes

Of the 100 sites tested, 87 sites were found to be under negative selection using the SLAC method and 90 sites were found to be under negative selection using the FEL test ($p = 0.05$). Neither test identified sites under positive selection. The branch-site REL analysis indicated that no branches are under episodic diversifying selection ($p \leq 0.05$).

Of the nine branches of major clades in the YABBY2 subfamily that were tested in codeml, none had codons that were found to be under positive selection once likelihood ratio tests were used to evaluate significance.

A Novel Motif Found in Monocot YABBY2-Like Genes

MEME identified one motif that is conserved across most monocot YABBY2-like sequences. It is 15 residues in width and occurs before the YABBY domain, in-between the zinc finger and

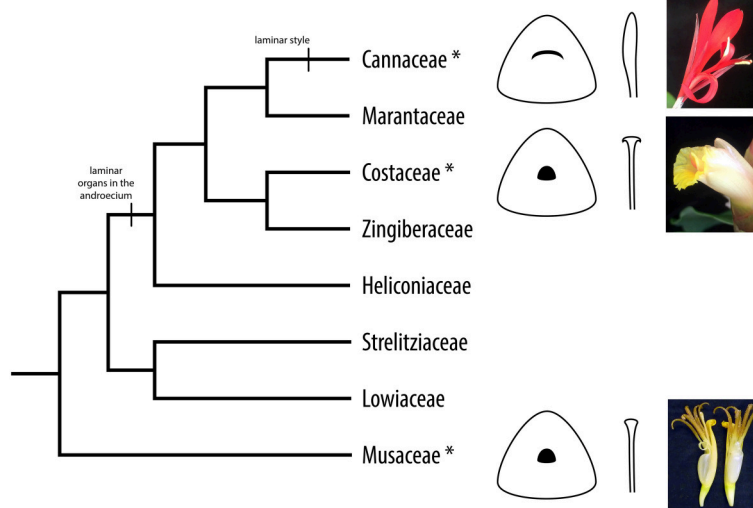


FIGURE 1 | Cladogram of the Zingiberales. Asterisks denote taxa used for expression analyses in this study. Evolution of laminar organs is indicated: flowers in Heliconiaceae have radial filaments and one laminar staminode; flowers in the ginger clade (Cannaceae, Marantaceae, Zingiberaceae, and Marantaceae) have laminar staminodes and fertile stamens—with a staminodial labellum in Zingiberaceae and Costaceae; and flowers in Cannaceae have a laminar style. Style morphology for Cannaceae, Costaceae, and Musaceae is illustrated and flowers from representative species are shown.

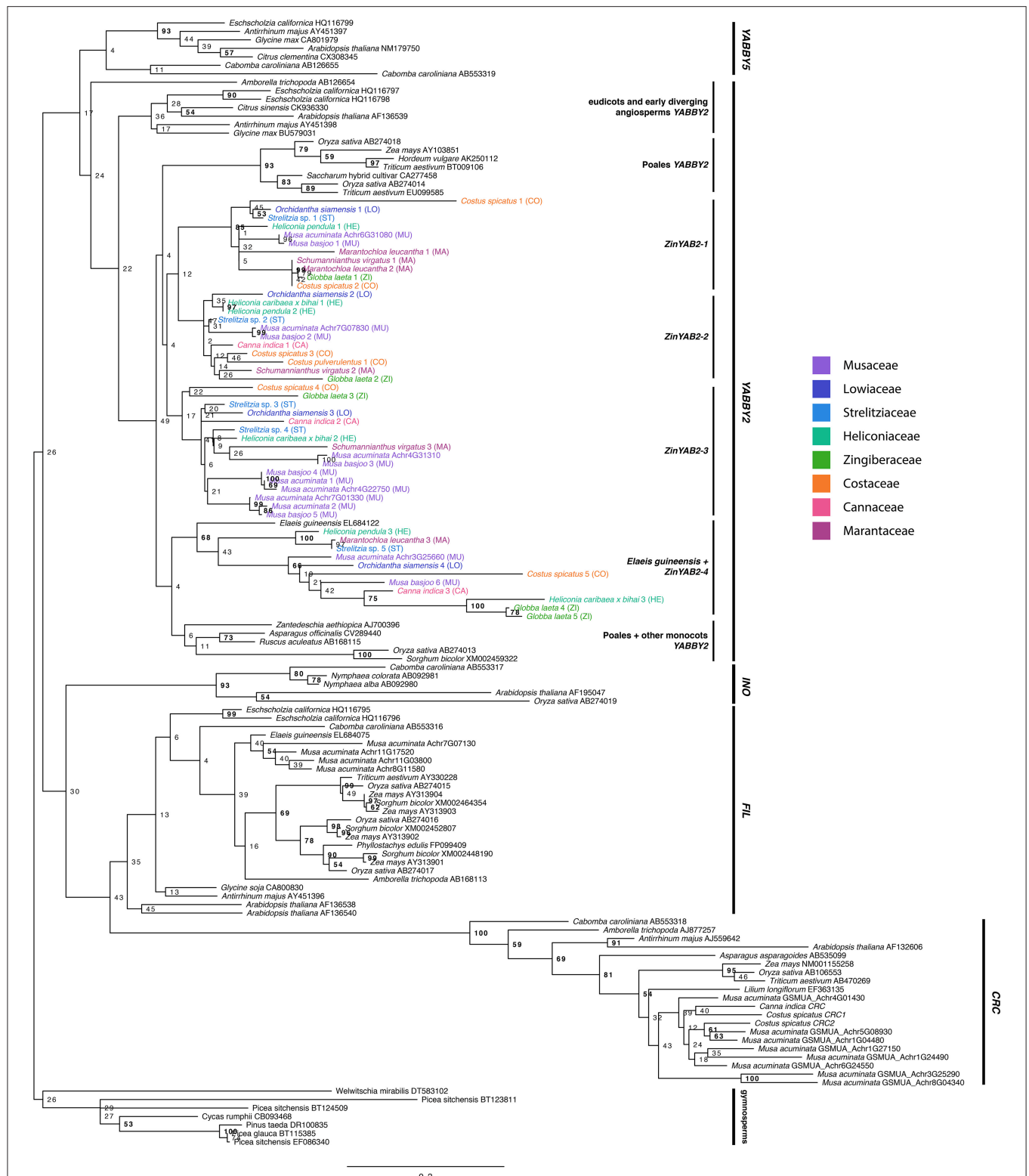


FIGURE 2 | Maximum likelihood tree of the YABBY gene family rooted with gymnosperm sequences. Bootstrap values are shown at nodes. Zingiberales YABBY2-like sequences are colored according to family.

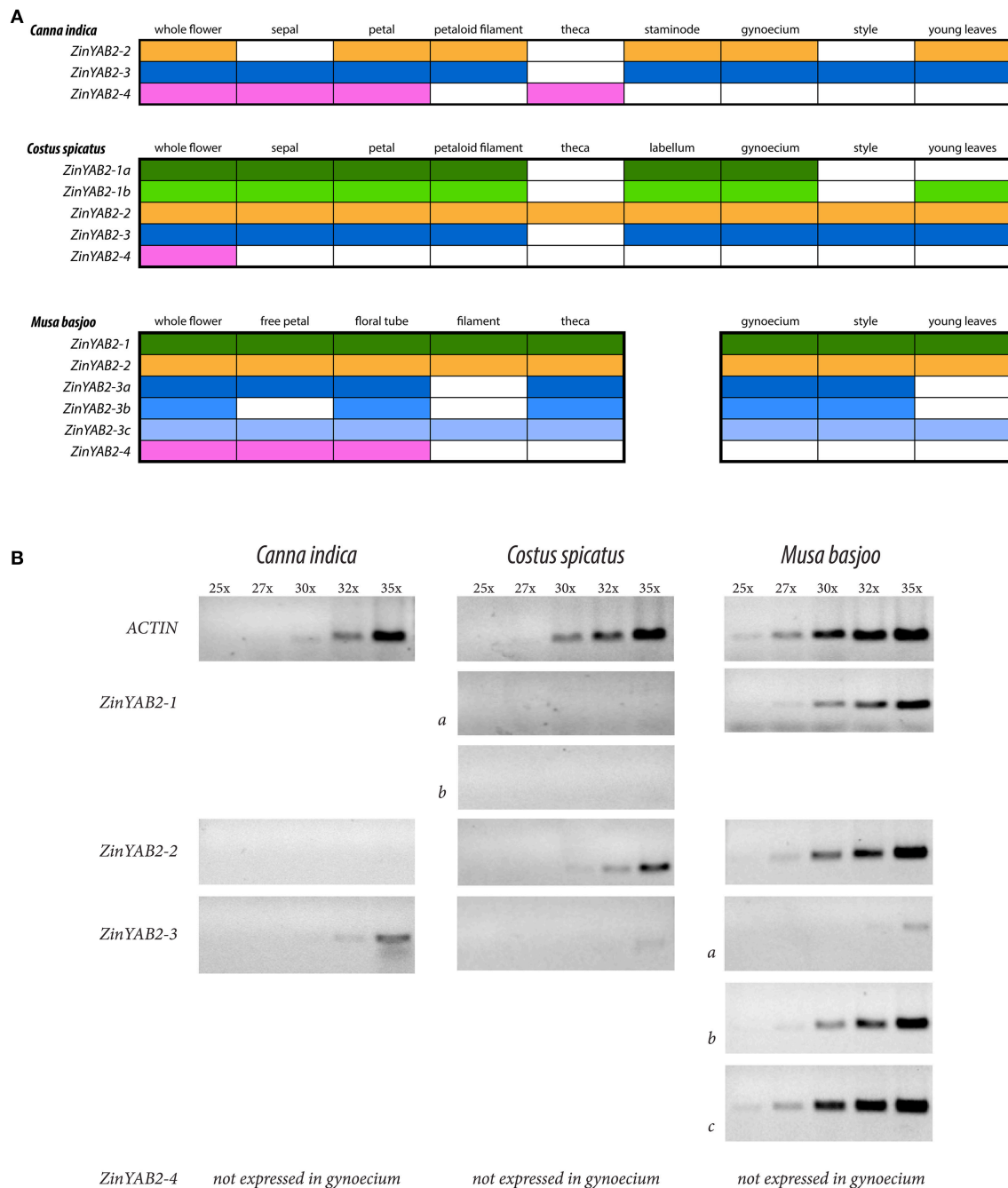
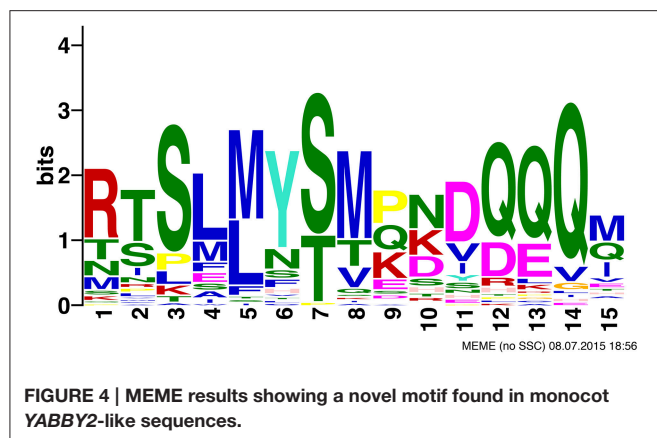


FIGURE 3 | Summary of RT-PCR results. (A) Summary of RT-PCR results from *Canna indica*, *Costus spicatus*, and *Musa basjoo*. Each color corresponds to a different copy of *ZinYAB2*; color indicates presence and lack of color indicates absence of the gene copy. **(B)** Summary of semi-quantitative RT-PCR results from flower styles of these taxa, showing results after 25, 27, 30, 32, and 35 cycles.

YABBY domains (Figure 4). Monocot YABBY2-like sequences that were identified as lacking the motif, when examined by eye, have some sequence similarity in this region, but have indels that make the conservation unrecognizable in the MEME analysis (which only recognizes ungapped motifs) or have amino acid changes that reduce similarity. This motif is not found in non-monocot YABBY2-like genes or in other YABBY subfamilies.

DISCUSSION

The Zingiberales is an order of tropical monocots that possesses a wide diversity of floral forms. This diversity can in part be attributed to the evolution of laminarity in the androecium and gynoecium. Members of the banana lineages (Musaceae, Heliconiaceae, Strelitziaceae, and Lowiaceae)



possess radial filaments, but taxa in Heliconiaceae have single laminar staminodes. The ginger clade (Cannaceae, Marantaceae, Costaceae, and Zingiberaceae) is characterized by laminar, petaloid staminodes and stamens. Staminodes in species of Costaceae (5) and Zingiberaceae (2 or 4) are fused to form a novel laminar structure called the labellum. In Cannaceae, the gynoeceum style has evolved to be laminar as well; all other families in the Zingiberales possess radial styles. Style laminarity in Cannaceae has functional importance because it facilitates secondary pollen presentation in this group: before a flower opens, pollen is transferred from the anthers to the laminar style. From there, the pollen is transferred to the bill of a hummingbird pollinator after the flower opens (Glinos and Cocucci, 2011).

In order to better understand the evolution of the laminar style in Cannaceae, we are interested in elucidating the molecular mechanisms involved in its development. In this study we have chosen to focus on one gene subfamily in particular, the YABBY2 gene subfamily. The YABBY gene family encodes transcription factors involved in the abaxial-adaxial polarity molecular network responsible for laminar expansion of lateral organs (and, conversely, radialization without laminar expansion). The YABBY2 gene subfamily was chosen because previous work has implicated the over-expression of a YABBY2-like gene as a mechanism for radialization in the filament of *Musa acuminata* and *Brassica rapa* (Almeida et al., 2014). To see if a similar mechanism may be involved in the evolution of the laminar style in Cannaceae, we have investigated the evolution of YABBY2-like genes in the Zingiberales and considered gene duplications and subsequent shifts in gene expression as a mechanism for morphological evolution.

Evolution and Diversification of YABBY2 in the Zingiberales

The YABBY2-like genes from monocots form two clades, each with a non-Zingiberales monocot clade (including Poales sequences) sister to a Zingiberales clade. This suggests that there may have been a gene duplication event prior to the divergence of the Poales and Zingiberales; however, more in-depth efforts

to isolate YABBY2-like genes from Poales species must be made to test this hypothesis. In one of these two monocot clades, the Zingiberales sequences are further divided into three separate clades, each with sequences from taxa from each the eight families in the Zingiberales—with the exception of *ZinYAB2-1*, which lacks representation from Cannaceae—suggesting that two additional gene duplication events occurred in this gene lineage before the diversification of the Zingiberales. Due to our extensive efforts to design *ZinYAB2-1*-specific primers and to clone *ZinYAB2-1* from *Canna*, we believe that *ZinYAB2-1* has likely been lost in Cannaceae, or it could not be isolated due to rapid sequence divergence. A copy of *ZinYAB2* was not found in either of the analyzed transcriptomes from *Canna*.

More recent gene duplication events have occurred within the Zingiberales. Two duplication events have occurred in Musaceae in the *ZinYAB2-3* lineage, as is evidenced by *Musa acuminata* and *Musa basjoo* sequences falling sister to each other in family-specific clades. Analysis of the *Musa acuminata* genome using the Genome Evolution analysis tool (GeVo) from the CoGe platform (Lyons and Freeling, 2008) shows that two of these gene copies, *ZinYAB2-3a* and *ZinYAB2-3c*, are located in syntenic regions on chromosomes 4 and 7 and are likely alpha paralogs from a whole genome duplication event (Figure 5). The third *Musa* paralog, *ZinYAB2-3b*, may be from an older or local duplication event.

Duplication events in other Zingiberales families may have also occurred. The two *Globba laeta* (Zingiberaceae) *ZinYAB2-4* sequences are recovered as being most closely related to each other and share 87.9% identity, while the two *Costus spicatus* (Costaceae) *ZinYAB2-1* sequences are recovered as being more closely related to sequences from other species than to each other and share only 66.9% identity. This latter case is also true for the two *Marantochloa leucantha* (Marantaceae) *ZinYAB2-1* sequences and the two *Strelitzia* sp. (Strelitziaceae) *ZinYAB2-3* sequences, which share 73.5% identity and 86.9% identity, respectively. Based on sequence identity, these pairs of sequences could be allelic variants, or they could be the result of recent duplication events in these lineages.

A Novel Motif Found in Monocot YABBY2-Like Sequences

The monocot YABBY2-like motif identified in the MEME analysis could be an artifact of the shared evolutionary history of these sequences, or it could be biologically relevant. Functional tests need to be performed to test whether this motif is biologically relevant and whether it results in new or altered molecular interactions for monocot YABBY2-like sequences. Regardless, this motif could be helpful to identify YABBY2-like genes from monocots in future analyses, when it is difficult to obtain support for placement of genes in the YABBY gene family due to the conservation in the alignable regions (the zinc finger and YABBY domains).

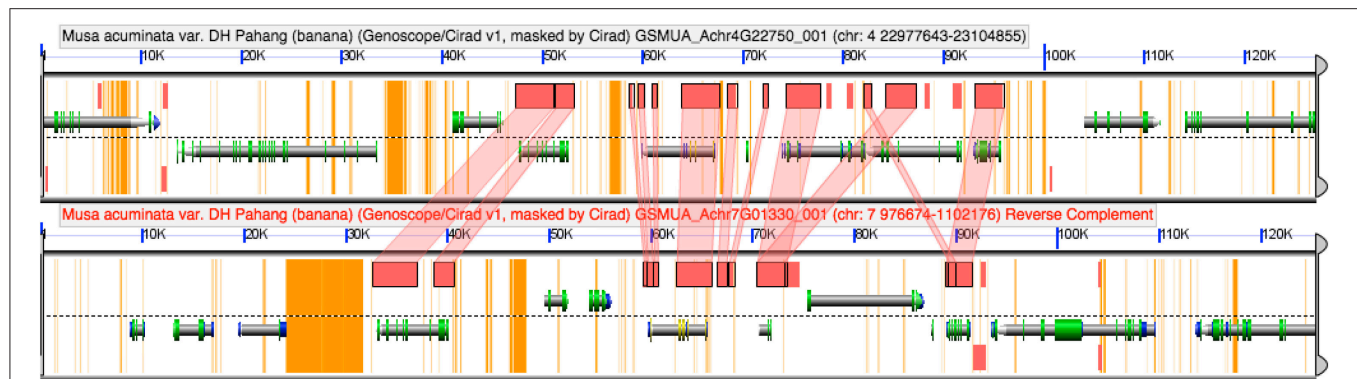


FIGURE 5 | COGE browser view of a portion of the *Musa acuminata* genome. *ZinYAB2-3a* and *ZinYAB2-3c* from *Musa* are found in syntenic regions in the *Musa acuminata* genome, suggesting that they are alpha paralogs from a recent whole genome duplication.

Expression Profiles Support Proposed Evolutionary Relationships for *ZinYAB2* Gene Copies

The absence of *ZinYAB2-4* from filament and gynoecium in *C. indica*, *C. spicatus*, and *M. basjoo* suggests that this copy is not involved in the gene network underlying laminar expansion (or radialization) in reproductive organs. The overall floral expression profiles of *ZinYAB2-1*, *ZinYAB2-2*, and *ZinYAB2-3* are more similar to each other than to that of *ZinYAB2-4*, which has a more restricted expression profile. These data are consistent with the hypothesis that *ZinYAB2-1*, *2-2*, and *2-3* arose from two subsequent Zingiberales-specific duplications, and that the ancestor of these three copies is sister to *ZinYAB2-4* (and thus *ZinYAB2-4* is more distantly related to the other three *ZinYAB2* copies than they are to one another).

A Potential Role of YABBY2-Like Genes in the Evolution of the Laminar Style in Cannaceae

It has previously been proposed that balanced expression of genes involved in the abaxial-adaxial polarity network facilitates laminar expansion, with evidence for gene expression imbalance, through high expression of a YABBY2-like gene, as a mechanism for radialization in *Musa acuminata* and *Brassica rapa* filaments (Almeida et al., 2014). *Musa basjoo* and *C. spicatus* share a radial style morphology, but have different levels of expression and copy number of YABBY2-like genes. It is possible that there is a threshold for total YABBY2 expression in the style at which the abaxial-adaxial polarity gene network is in balance and promotes laminar expansion, and above which the gene network is imbalanced, leading to radialization. Loss of expression of *ZinYAB2-2* (and, possibly, the loss of *ZinYAB2-1*) in the style of *C. indica* may have reduced total YABBY2-like gene expression to be in balance with other genes in the gene regulatory network, and thus facilitated the shift to laminar morphology in the style of *C. indica*.

Laminar expansion of the style is a derived trait in Cannaceae, and evolved separately from laminar expansion in stamens

and staminodes (in Heliconiaceae and the ginger clade) and laminar expansion in sepals and petals. Perhaps the mechanism of balanced expression for laminar expansion is shared across different floral organs, but the absolute levels of expression required for laminar expansion differ in different organs; this would explain why expression differs between the laminar style and other laminar floral organs (sepal, petal, petaloid filament, and staminode) in *C. indica*.

Future Directions

To better understand the molecular mechanisms underlying the evolution of the laminar style in the Zingiberales, *in situ* hybridization experiments are needed to fully characterize the exact locations of YABBY2-like gene expression during development, and more gene evolution studies and expression analyses for other genes involved in the abaxial-adaxial gene regulatory network should be done. Gene knockdown (virus-induced gene silencing) or gene knockout (CRISPR) experiments are also needed to test the hypotheses proposed here about the functions of YABBY2-like genes in Zingiberales floral morphology. By using the evolution of style laminarity in Cannaceae as a case study, we can elucidate the mechanisms underlying the evolution of novel laminar (or radial) morphology in lateral organs. In addition, YABBY genes in gymnosperms have yet to be studied, and our understanding of the evolution of this gene family and thus the evolution of the abaxial-adaxial gene regulatory network would benefit greatly from gene characterization and expression studies in gymnosperms and pteridophytic vascular plants.

AUTHOR CONTRIBUTIONS

KM, RY, AA, and CS contributed with conceptual and experimental design. KM contributed to data collection and analysis and drafted the manuscript. RY and AA contributed to data analysis and manuscript editing. CS edited the manuscript and provided financial support. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This work was supported by an NSF CAREER award (IOS 0845641) to CS, an NSF Doctoral Dissertation Improvement Grant (DEB 1110461) to CS and AA, and a UC Berkeley College of Natural Resources student-initiated Sponsored Projects for Undergraduate Research (SPUR) award to KM. AA was supported by a Jovem Talento fellowship (CAPES/CNPq). The authors thank the members of the Specht Lab for discussion and

technical advice and Jennifer Bates for help with creating the figures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.01106>

REFERENCES

- Alvarez, J., and Smyth, D. R. (1999). *CRABS CLAW* and *SPATULA*, two *Arabidopsis* genes that control carpel development in parallel with *AGAMOUS*. *Development* 126, 2377–2386.
- Bailey, T. L., and Elkan, C. (1994). “Fitting a mixture model by expectation maximization to discover motifs in biopolymers,” in *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* (Menlo Park, CA), 28–36.
- Baker, S. C., Robinson-Beers, K., Villanueva, J. M., Gaiser, J. C., and Gasser, C. S. (1997). Interaction among genes regulating ovule development in *Arabidopsis thaliana*. *Genetics* 145, 1109–1124.
- Bartholmes, C., Hidalgo, O., and Gleissberg, S. (2012). Evolution of the YABBY gene family with emphasis on the basal eudicot *Eschscholzia californica* (Papaveraceae). *Plant Biol.* 14, 11–23. doi: 10.1111/j.1438-8677.2011.00486.x
- Bowman, J. L., Eshed, Y., and Baum, S. F. (2002). Establishment of polarity in angiosperm lateral organs. *Trends Genet.* 18, 134–141. doi: 10.1016/S0168-9525(01)02601-4
- Bowman, J. L., and Smyth, D. R. (1999). *CRABS CLAW*, a gene that regulates carpel and nectary development in *Arabidopsis*, encodes a novel protein with zinc finger and helix-loop-helix domains. *Development* 126, 2387–2396.
- Chen, Q., Atkinson, A., Otsuga, D., Christensen, T., Reynolds, L., and Drews, G. N. (1999). The *Arabidopsis* *FILAMENTOUS FLOWER* gene is required for flower formation. *Development* 126, 2715–2726.
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772. doi: 10.1038/nmeth.2109
- de Almeida, A. M., Yockteng, R., Schnable, J., Alvarez-Buylla, E. R., Freeling, M., and Specht, C. D. (2014). Co-option of the polarity gene network shapes filament morphology in the angiosperms. *Nat. Sci. Rep.* 4:6194. doi: 10.1038/srep06194
- D’Hont, A., Denoeud, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488, 213–217. doi: 10.1038/nature11241
- Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., et al. (2012). *Geneious v5.6*. Available online at: <http://www.geneious.com>
- Fourquin, C., Vinauger-Douard, M., Fogliani, B., Dumas, C., and Scutt, C. P. (2005). Evidence that *CRABS CLAW* and *TOUSLED* have conserved their roles in carpel development since the ancestor of the extant angiosperms. *Proc. Natl. Acad. Sci.* 102, 4649–4654. doi: 10.1073/pnas.0409577102
- Gleissberg, S., Groot, E. P., Schmalz, M., Eichert, M., Kölsch, A., and Hutter, S. (2005). Developmental events leading to peltate leaf structure in *Tropaeolum majus* (Tropaeolaceae) are associated with expression domain changes of a YABBY gene. *Dev. Genes Evol.* 215, 313–319. doi: 10.1007/s00427-005-0479-8
- Glinos, E., and Cocucci, A. A. (2011). Pollination biology of *Canna indica* (Cannaceae) with particular reference to the functional morphology of the style. *Plant Syst. Evol.* 291, 49–58. doi: 10.1007/s00606-010-0379-x
- Golz, J. F., Roccaro, M., Kuzoff, R., and Hudson, A. (2004). *GRAMINIFOLIA* promotes growth and polarity of *Antirrhinum* leaves. *Development* 131, 3661–3670. doi: 10.1242/dev.01221
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Husbands, A. Y., Chitwood, D. H., Plavskin, Y., and Timmermans, M. C. (2009). Signals and prepatterns: new insights into organ polarity in plants. *Genes Dev.* 23, 1986–1997. doi: 10.1101/gad.1819909
- Ishikawa, M., Ohmori, Y., Tanaka, W., Hirabayashi, C., Murai, K., Ogihara, Y., et al. (2009). The spatial expression patterns of *DROOPING LEAF* orthologs suggest a conserved function in grasses. *Genes Genet. Syst.* 84, 137–146. doi: 10.1266/ggs.84.137
- Juarez, M. T., Twigg, R. W., and Timmermans, M. C. P. (2004). Specification of adaxial cell fate during maize leaf development. *Development* 131, 4533–4544. doi: 10.1242/dev.01328
- Kanaya, E., Nakajima, N., and Okada, K. (2002). Non-sequence-specific DNA binding by the *FILAMENTOUS FLOWER* protein from *Arabidopsis thaliana* is reduced by EDTA. *J. Biol. Chem.* 277, 11957–11964. doi: 10.1074/jbc.M108889200
- Kim, M., Pham, T., Hamidi, A., McCormick, S., Kuzoff, R. K., and Sinha, N. (2003). Reduced leaf complexity in tomato wiry mutants suggests a role for *PHAN* and *KNOX* genes in generating compound leaves. *Development* 130, 4405–4415. doi: 10.1242/dev.00655
- Kosakovsky Pond, S. L., and Frost, S. D. W. (2005a). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222. doi: 10.1093/molbev/msi105
- Kosakovsky Pond, S. L., and Frost, S. D. W. (2005b). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531–2533. doi: 10.1093/bioinformatics/bti320
- Kosakovsky Pond, S. L., Frost, S. D. W., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. doi: 10.1093/bioinformatics/bti079
- Liu, H., Xu, Y., Xu, Z., and Chong, K. (2007). A rice YABBY gene, *OsYABBY4*, preferentially expresses in developing vascular tissue. *Dev. Genes Evol.* 217, 629–637. doi: 10.1007/s00427-007-0173-0
- Lyons, E., and Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* 53, 661–673. doi: 10.1111/j.1365-313X.2007.03326.x
- Nakayama, H., Yamaguchi, T., and Tsukaya, H. (2010). Expression patterns of *AaDL*, a *CRABS CLAW* ortholog in *Asparagus asparagoides* (Asparagaceae), demonstrate a stepwise evolution of *CRC/DL* subfamily of YABBY genes. *Am. J. Bot.* 97, 591–600. doi: 10.3732/ajb.0900378
- Navarro, C., Efremova, N., Golz, J. F., Rubiera, R., Kuckenberger, M., Castillo, R., et al. (2004). Molecular and genetic interactions between *STYLOSA* and *GRAMINIFOLIA* in the control of *Antirrhinum* vegetative and reproductive development. *Development* 131, 3649–3659. doi: 10.1242/dev.01205
- Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). *Tracer v1.6*. Available online at: <http://beast.bio.ed.ac.uk/Tracer>
- Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180
- Sawa, S., Ito, T., Shimura, Y., and Okada, K. (1999b). *FILAMENTOUS FLOWER* controls the formation and development of *Arabidopsis* inflorescences and floral meristems. *Plant Cell* 11, 69–86. doi: 10.2307/3870839
- Sawa, S., Watanabe, K., Goto, K., Liu, Y. G., Shibata, D., Kanaya, E., et al. (1999a). *FILAMENTOUS FLOWER*, a meristem and organ identity gene of *Arabidopsis*, encodes a protein with a zinc finger and HMG-related domains. *Genes Dev.* 13, 1079–1088. doi: 10.1101/gad.13.9.1079

- Siegfried, K. R., Eshed, Y., Baum, S. F., Otsuga, D., Drews, G. N., and Bowman, J. L. (1999). Members of the YABBY gene family specify abaxial cell fate in *Arabidopsis*. *Development* 126, 4117–4128.
- Specht, C. D., Yockteng, R., Almeida, A. M., Kirchoff, B. K., and Kress, J. W. (2012). Homoplasmy, pollination, and emerging complexity during the evolution of floral development in the tropical ginger (Zingiberales). *Bot. Rev.* 78, 440–462. doi: 10.1007/s12229-012-9111-6
- Tononi, P., Möller, M., Bencivenga, S., and Spada, A. (2010). GRAMINIFOLIA homolog expression in *Streptocarpus rexii* is associated with the basal meristems in phyllocladus, a morphological novelty in Gesneriaceae. *Evol. Dev.* 12, 61–73. doi: 10.1111/j.1525-142X.2009.00391.x
- Toriba, T., Harada, K., Takamura, A., Nakamura, H., Ichikawa, H., Suzaki, T. et al. (2007). Molecular characterization of the YABBY gene family in *Oryza sativa* and expression analysis of OsYABBY1. *Mol. Genet. Genomics* 277, 457–468. doi: 10.1007/s00438-006-0202-0
- Villanueva, J. C., Broadhvest, J., Hauser, B. A., Meister, R. J., Schneitz, K., and Gasser, C. S. (1999). INNER NO OUTER regulates abaxial-adaxial patterning in *Arabidopsis* ovules. *Genes Dev.* 13, 3160–3169. doi: 10.1101/gad.13.23.3160
- Waites, R., and Hudson, A. (1995). Phantastica: a gene required for dorsoventrality of leaves in *Antirrhinum majus*. *Development* 121, 2143–2154.
- Worden, A. Z., Lee, J. H., Mock, T., Rouzé, P., Simmons, M. P., Aerts, A. L. et al. (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* 324, 268–272. doi: 10.1126/science.1167222
- Yamada, T., Ito, M., and Kato, M. (2003). Expression pattern of INNER NO OUTER homologue in *Nymphaea* (water lily family, Nymphaeaceae). *Dev. Genes Evol.* 213, 510–513. doi: 10.1007/s00427-003-0350-8
- Yamada, T., Ito, M., and Kato, M. (2004). YABBY2-homologue expression in lateral organs of *Amborella trichopoda* (Amborellaceae). *Int. J. Plant Sci.* 165, 917–924. doi: 10.1086/423793
- Yamada, T., Yokota, S., Hirayama, Y., Imaichi, R., Kato, M., and Gasser, C. S. (2011). Ancestral expression patterns and evolutionary diversification of YABBY genes in angiosperms. *Plant J.* 67, 26–36. doi: 10.1111/j.1365-313X.2011.04570.x
- Yamaguchi, T., Nagasawa, N., Kawasaki, S., Matsuoaka, M., Nagato, Y., and Hirano, H. Y. (2004). The YABBY gene DROOPING LEAF regulates carpel specification and midrib development in *Oryza sativa*. *Plant Cell* 16, 500–509. doi: 10.1105/tpc.018044
- Yamaguchi, T., Nukazuka, A., and Tsukaya, H. (2012). Leaf adaxial-abaxial polarity specification and lamina outgrowth: evolution and development. *Plant Cell Physiol.* 53, 1180–1194. doi: 10.1093/pcp/pcs074
- Yang, Z. (2007). PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yockteng, R., Almeida, A. M. R., Yee, S., Andre, T., Hill, C., and Specht, C. D. (2013). A method for extracting high-quality RNA from diverse plants for next-generation sequencing and gene expression analyses. *Appl. Plant Sci.* 1:1300070. doi: 10.3732/apps.1300070

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Morioka, Yockteng, Almeida and Specht. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Flower Development and Perianth Identity Candidate Genes in the Basal Angiosperm *Aristolochia fimbriata* (Piperales: Aristolochiaceae)

Natalia Pabón-Mora^{1,2*}, Harold Suárez-Baron¹, Barbara A. Ambrose² and Favio González³

¹ Instituto de Biología, Universidad de Antioquia, Medellín, Colombia, ² The New York Botanical Garden, Bronx, NY, USA,

³ Instituto de Ciencias Naturales, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

OPEN ACCESS

Edited by:

Jocelyn Hall,
University of Alberta, Canada

Reviewed by:

Annette Becker,
Justus Liebig University Giessen,
Germany
David Smyth,
Monash University, Australia

*Correspondence:

Natalia Pabón-Mora
lucia.pabon@udea.edu.co

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 15 September 2015

Accepted: 22 November 2015

Published: 11 December 2015

Citation:

Pabón-Mora N, Suárez-Baron H,
Ambrose BA and González F (2015)
Flower Development and Perianth
Identity Candidate Genes in the Basal
Angiosperm *Aristolochia fimbriata*
(Piperales: Aristolochiaceae).
Front. Plant Sci. 6:1095.
doi: 10.3389/fpls.2015.01095

Aristolochia fimbriata (Aristolochiaceae: Piperales) exhibits highly synorganized flowers with a single convoluted structure forming a petaloid perianth that surrounds the gynostemium, putatively formed by the congenital fusion between stamens and the upper portion of the carpels. Here we present the flower development and morphology of *A. fimbriata*, together with the expression of the key regulatory genes that participate in flower development, particularly those likely controlling perianth identity. *A. fimbriata* is a member of the magnoliids, and thus gene expression detected for all ABCE MADS-box genes in this taxon, can also help to elucidate patterns of gene expression prior the independent duplications of these genes in eudicots and monocots. Using both floral development and anatomy in combination with the isolation of MADS-box gene homologs, gene phylogenetic analyses and expression studies (both by reverse transcription PCR and *in situ* hybridization), we present hypotheses on floral organ identity genes involved in the formation of this bizarre flower. We found that most MADS-box genes were expressed in vegetative and reproductive tissues with the exception of *AfimSEP2*, *AfimAGL6*, and *AfimSTK* transcripts that are only found in flowers and capsules but are not detected in leaves. Two genes show ubiquitous expression; *AfimFUL* that is found in all floral organs at all developmental stages as well as in leaves and capsules, and *AfimAG* that has low expression in leaves and is found in all floral organs at all stages with a considerable reduction of expression in the limb of anthetic flowers. Our results indicate that expression of *AfimFUL* is indicative of pleiotropic roles and not of a perianth identity specific function. On the other hand, expression of B-class genes, *AfimAP3* and *AfimPI*, suggests their conserved role in stamen identity and corroborates that the perianth is sepal and not petal-derived. Our data also postulates an *AGL6* ortholog as a candidate gene for sepal identity in the Aristolochiaceae and provides testable hypothesis for a modified ABCE model in synorganized magnoliid flowers.

Keywords: AGAMOUS-like6, APETALA3, *Aristolochia fimbriata*, FRUITFULL, magnoliids, MADS-box genes, perianth, PISTILLATA

INTRODUCTION

With approximately 550 species, *Aristolochia* is the largest genus in the Aristolochiaceae, one of the families of the monophyletic order Piperales (APG, 2009). Flowers in the genus differ from those of the other members of the family (*Asarum* L., *Thottea* Rottb., *Lactoris* Phil., *Hydnora* Thunb., and *Saruma* Oliver) in that its perianth is monosymmetric, and formed by a trimerous whorl of petaloid sepals fused to form a variously convoluted and tubular structure, differentiated in three regions: a basal, inflated portion called the utricle; a narrow, tubular portion called the tube; and an expanded, laminar portion called the limb (**Figure 1**). All species of *Aristolochia* also possess a gynostemium, that is, a crown-like structure found inside the utricle, above the five or six carpellate, syncarpic, inferior ovary. Ontogenetic studies have shown that the gynostemium is formed by the congenital fusion of stamens and stigmas (González and Stevenson, 2000a,b). This suite of floral characters is unique among early diverging flowering plants, and strongly differ from the predominant floral construction among magnoliids based on a large number of spirally arranged organs and a predominantly apocarpic and superior ovary (Endress and Doyle, 2007). Thus, the floral construction in *Aristolochia* poses a number of questions regarding the origin, development, and genetic identity of whorls in early diverging synorganized flowers.

The genetic basis of floral organ identity was established three decades ago, based on homeotic mutants of *Arabidopsis thaliana* and *Antirrhinum majus* (Schwarz-Sommer et al., 1990; Coen and Meyerowitz, 1991). This resulted in a model that explained how a combinatorial activity of three classes of MIKC MADS-box transcription factors and an *ERF/APETALA2* gene (collectively named A, B, and C class genes) confer a specific identity to each floral whorl. According to the ABC model, A-class genes (*APETALA1* and *APETALA2*) establish the identity of floral meristem and sepals, A and B-class genes (*APETALA3* and *PISTILLATA*) together control petal identity, B and C-class genes (*AGAMOUS*) regulate stamen identity, and finally, C-class genes alone regulate carpel identity, as well as the termination of the floral meristem (Bowman et al., 1989, 1991; Yanofsky et al., 1990). Although the model points to a mutually exclusive role of A and C class genes to control sterile and fertile portions of the flower, there is little evidence that the direct or indirect repression of AG via *AP1* or *AP2* orthologs, shown to occur in *Arabidopsis thaliana*, occurs in the same fashion in other flowering plants (Litt, 2007; Dihn et al., 2012). This model was modified when additional transcription factors were found to be required for the initiation of all floral organs, which has resulted in the inclusion of the E-class genes (the *Arabidopsis* *SEPALLATA1*, *SEP2*, *SEP3*, and *SEP4*) in the model (Pelaz et al., 2000). In addition, D-class ovule-specific transcription factors were also identified in *Petunia* and since some authors have proposed an expanded ABCDE model (Pinyopich et al., 2003). However, functional data in other angiosperms, has pointed to little generality of these findings favoring the use of the ABCE model, which we will refer hereafter. MADS-box gene duplications have occurred coinciding with the diversification of the Brassicaceae, the eudicots and the monocots, often linked

with whole genome duplication (WGD) events (Jiao et al., 2011). Studies addressed to investigate the genetic control of the diverse floral ground plans in non-model angiosperms have revealed that gene duplications complicate the extrapolation of model to non-model plants, and that functional evolution of different gene lineages involved functions that were not considered in the original model (Litt and Kramer, 2010). Comparative functional studies have demonstrated that, for the most part, stamen and carpel identity are controlled by B and C class genes, in early diverging angiosperms, monocots, magnoliids, and eudicots (reviewed in Litt and Kramer, 2010). In the same way, extra whorls formed outside or inside the stamens, such as the controversial corona in Passifloraceae and *Narcissus* (Hemingway et al., 2011; Waters et al., 2013) are also controlled by the same B + C class gene combination, suggesting that these extra whorls share the same genetic regulation as the stamens themselves.

The genetic basis of perianth identity is, by comparison, more complicated. Petal identity relies on the AP3-PI heterologous interaction with AP1 and SEP proteins in *Arabidopsis thaliana* (Honma and Goto, 2001). Gene duplications in the AP3 gene lineage have often resulted in the specialization of gene copies in unique structures. For instance, AP3 gene duplications in core eudicots, resulting in the AP3 and TM6 gene clades, have led to sub-functionalization with AP3 paralogs contributing to the perianth, and TM6 paralogs functioning in stamen identity (Jack et al., 1994; Liu et al., 2004; Vandenbussche et al., 2004; deMartino et al., 2006; Rijpkema et al., 2006). Independent gene duplication in Ranunculales, have also led to sub-functionalization of AP3-III homologs exclusively providing petal identity, and their paralogs AP3-I and AP3-II controlling stamen identity (Drea et al., 2007; Kramer et al., 2007; Sharma et al., 2011; Sharma and Kramer, 2012; Zhang et al., 2013). Gene expression patterns and functional analyses of AP3 homologs have also been used in an attempt to assess homology of atypical floral structures occurring in the second whorl, which are likely modified petals. These include for example the reduced lodicules in grasses (Ambrose et al., 2000; Whipple et al., 2007) and the labellum in orchids (Mondragón-Palomino and Theissen, 2011). On the other hand, petaloidy outside of petals can occur with or without B-class expression. Petaloid sepals in non-grass monocots, like tulips, seem to occur as a result of the ectopic expression of B-class genes in the first floral whorl (Kanno et al., 2007), whereas sepaloid tepals in *Lacandonia* lack B-class gene expression (Alvarez Buylia et al., 2010). Conversely, the identity of petaloid sepals in *Aristolochia manshuriensis* (magnoliids) does not seem to correlate with AP3 expression, as *ArmAP3* copies are only turned on late in development, exclusively in the adaxial region of the perianth (Jaramillo and Kramer, 2004). Thus, petaloidy is a complex, likely homoplasious feature, both morphologically and genetically and seems to be the result of different gene combinations recruited in a case-by-case scenario (see also Feng et al., 2012; Almeida et al., 2013).

The A-function is by far the most contentious function in the ABCE model. Increasing evidence suggests that *AP1* and *AP2* orthologs rarely control perianth identity outside *Arabidopsis thaliana* (Litt, 2007; Causier et al., 2010). Like the AP3 and the AG



FIGURE 1 | *Aristolochia fimbriata*. Anthetic flower in face (A) and back (B) view. Close-up of the gynostemium in female (C) and male (D) phases. Black arrowheads indicate anthers; asterisks (*) indicate the upper portion of a single carpel alternating with each anther; l, limb; o, ovary; t, tube; u, utricle. Scale bar: 1 cm.

gene lineages, the *API* gene lineage has undergone duplication events coincident with the diversification of monocots and eudicots, as well as the Brassicaceae (Litt and Irish, 2003). Functional studies suggest that *API/FUL* genes are pleiotropic as they are involved in leaf morphogenesis, inflorescence architecture, floral transition, floral meristem identity, and fruit development (Huijser et al., 1992; Gu et al., 1998; Immink et al., 1999; Ferrándiz et al., 2000; Berbel et al., 2001, 2012; Müller et al., 2001; Vrebalov et al., 2002; Murai et al., 2003; Benlloch et al., 2006; Pabón-Mora et al., 2012, 2013). *API/FUL* homologs frequently play roles in the identity of the floral meristem and the sepals (Huijser et al., 1992; Berbel et al., 2001; Benlloch et al., 2006) but their contribution to petal identity is unclear (Mandel et al., 1992; Yu et al., 2004; Castillejo et al., 2005; Pabón-Mora et al., 2012). Sepal identity is tightly coupled with the acquisition of floral meristem identity. Potential candidate gene lineages that function in the transition from inflorescence to floral meristem identity, and turn on floral organ identity genes, include the *SEP* and the *AGL6* gene lineages. Functional analyses in model core eudicots and monocots have shown that they function in floral meristem fate and the identity of all floral organs (Pelaz et al., 2000; Ferrario et al., 2003; Ohmori et al., 2009; Rijpkema et al., 2009; Pan et al., 2014). *AGL6* expression has been detected in sepals or paleas (putative first whorl organs) in grasses (Reinheimer and Kellogg, 2009; Viaene et al., 2010) suggesting that in addition to the ABCE genes, the *AGL6* homologs may be controlling fundamental processes in floral organ identity,

specifically in floral meristem and sepal identity. This is likely to be the case only in early diverging angiosperms and monocots, as *AGL6* is redundant with *SEP* genes in many core eudicots in the specification of floral organ identity (Rijpkema et al., 2009) or have been co-opted to function in inflorescence architecture and flowering, as is the case in *Arabidopsis* (Koo et al., 2010). In addition, *AGL6* genes are also expressed in ovules (Reinheimer and Kellogg, 2009; Rijpkema et al., 2009) and are the only genes from the *API/SEP/AGL6* lineage found in both gymnosperms and angiosperms (Zahn et al., 2005).

The present research aims to assess the genetic basis responsible for the petaloid perianth and the gynostemium identity in *Aristolochia*, having the ABCE model as a reference point (Figure 1). In order to do so we have selected *Aristolochia fimbriata* Cham., as this species has recently been proposed as a candidate magnoliid for evolutionary developmental studies for being a self-compatible herb with continuous flowering, and having high rate of seed germination, small genome size, and low ($2n = 14$) chromosome number (Bliss et al., 2013). The species is native to temperate South America but has been widely spread as an ornamental. It has distichous, kidney-shape, variegated, and glabrous leaves, solitary, axillary flowers, and acropetally dehiscent capsules that contain many heart-shaped seeds (González et al., 2015). Here, we present a detailed study of flower development, provide key developmental stages and their associated features, evaluate expression profiles and *in situ* hybridization of the key regulatory genes that participate in

floral development, and specifically highlight those that may be responsible for the sepal-derived perianth identity. In particular, we would like to identify perianth expressed genes, as previous studies have found that B-class genes do not play a role in determining the petaloid perianth identity in other species of *Aristolochia* (Jaramillo and Kramer, 2004; Horn et al., 2014). Finally, as *A. fimbriata* is a member of the magnoliids, gene expression detected for all ABCE MADS-box genes can be used to elucidate ancestral patterns of expression for the gene lineages that duplicated during the diversification of eudicots and monocots independently.

MATERIALS AND METHODS

Scanning Electron Microscopy

Floral buds at several stages of development of *A. fimbriata* (Voucher N. Pabón-Mora 242, NY) were collected from the living collections at the Nolen greenhouses (NYBG) and at the Universidad de Antioquia (UdeA), fixed in 70% ethanol, and dissected in 90% ethanol. The samples were then dehydrated in a series of 100% ethanol, 50:50% ethanol–acetone, and 100% acetone, critical point-dried using a Samdri 790 CPD (Rockville, MD, USA), coated with gold and palladium using a Hummer 6.2 (Anatech, Springfield, VA, USA) sputter coater, and examined and photographed at 10 kV in a Jeol JSM-5410 LV scanning electron microscope.

Light Microscopy

For light microscopy, floral buds and mature flowers were also prepared by conventional dehydration with ethanol and toluene using a standard series in a Leica TP-1020 automatic tissue processor and embedded in paraplast X-tra using an AP-280 Microm tissue embedding center; the samples were sectioned at 12 μ m with an AO Spencer 820 rotary microtome. Sections were stained in safranin and astra blue, mounted in permount and examined using a Zeiss Compound microscope equipped with a Nikon DXM1200C digital camera with ACT-1 software.

Isolation and Phylogenetic Analyses of MADS-Box Genes from *A. fimbriata*

Fresh inflorescence and floral tissue from cultivated plants was ground using liquid nitrogen and further total RNA extracted using TRIZOL reagent (Invitrogen). The RNA-seq experiment was conducted using the truseq mRNA library construction kit (Illumina) and sequenced in a HiSeq2000 instrument reading 100 bases paired end reads. A total of 85,608,833 raw read pairs were obtained. Read cleaning was performed with PRINSEQ-LITE with a quality threshold of Q35 and contig assembly was computed using Trinity package following default settings. Contig metrics are as follows: Total assembled bases: 85,608,833; total number of contigs (>101 bp): 118941; average contig length: 719 bp; largest contig: 16972 bp; contig N50: 1823 bp; contig GC%: 42.71%; number of Ns: 0. Orthologous gene search was performed using BLASTN (Altschul et al., 1990) using the *Arabidopsis* sequences as a query to identify

a first batch of homologs in the *A. fimbriata* transcriptome. Sequences in the transcriptome were compiled using BioEdit¹, where they were cleaned to keep exclusively the open reading frame. Ingroup sequences included also MADS-box genes from *Amborella trichopoda* (Amborellaceae), *Saruma henryi* (Aristolochiaceae), *Aquilegia coerulea* (Ranunculaceae), *Mimulus guttatus* (Phrymaceae), and *Arabidopsis thaliana* (Brassicaceae) with the purpose of including at least one species of each major group of eudicots and the ANA grade. Nucleotide sequences were then aligned using the online version of MAFFT² (Katoh et al., 2002), with a gap open penalty of 3.0, an offset value of 1.0, and all other default settings. The alignment was then refined by hand using BioEdit taking into account the conserved MADS protein domain. Outgroup sequences include *SUPPRESSOR OF CONSTANS 1* (*SOC1*) copies from basal angiosperms, basal eudicots and core eudicots (Becker and Theissen, 2003; Carlsbecker et al., 2013). Maximum likelihood (ML) phylogenetic analyses using the full nucleotide coding sequences were performed in RaxML-HP2 BlackBox (Stamatakis et al., 2008) on the CIPRES Science Gateway (Miller et al., 2009). The best performing evolutionary model was obtained by the Akaike information criterion (AIC; Akaike, 1974) using the program jModelTest v.0.1.1 (Posada and Crandall, 1998). Bootstrapping was performed according to the default criteria in RAXML where bootstrapping stopped after 200–600 replicates when the criteria were met. Trees were observed and edited using FigTree v1.4.0 (Rambaut, 2014). Uninformative characters were determined using Winclada Asado 1.62 (Nixon, 2002). Accession numbers for all MADS box genes here identified correspond to KT957081–KT957088.

Reverse Transcription – PCR (RT-PCR)

Expression of MADS-box homologs was assayed using RT-PCR on RNA extracted from floral buds at four different stages in preanthesis (1.5, 2.5, 3.5, and 4.5 cm) and anthesis. Total RNA was prepared from dissected organs; in this case we separated the perianth from the gynostemium and the ovary and further separated the sepals from the most proximal portions to the most distal ones, into the utricle, the tube and the limb. In addition total RNA was prepared from young leaf and a young fruit (3 cm long); given that fruits in *A. fimbriata* are septicidal capsules with acropetal dehiscence, we chose a green fruit before sclerenchyma accumulation in the endocarp. Total RNA was prepared using TRIzol (Invitrogen, Carlsbad, CA, USA), and DNaseI (Roche, Switzerland) treated to remove genomic DNA contamination. 2 μ g were used as template for cDNA synthesis with SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA). The resulting cDNA was used for PCR-amplification using locus specific primers (Supplementary Table 1), with a thermal cycling regime consisting of one initial step at 94°C for 3 min, 28 cycles at 94°C for 40 s, 55°C for 45 s, and 72°C for 1 min, and a final extension step at 72°C for 10 min. All reactions were carried out in a MultiGeneTM OptiMax thermocycler (Labnet International, Edison, NJ, USA). PCR was run on a 1% agarose gel with 1X TAE,

¹<http://www.mbio.ncsu.edu/bioedit/bioedit.html>

²<http://mafft.cbrc.jp/alignment/server/>

stained with ethidium bromide and digitally photographed using a Whatman Biometra BioDocAnalyzer (Gottingen, Germany).

In Situ Hybridization

Developing shoot apical meristems in reproductive stages were collected from wild type plants of *A. fimbriata* growing in the Nolen greenhouses at NYBG or at the Universidad de Antioquia (UdeA), and fixed under vacuum in freshly prepared FAA (50% ethanol, 3.7% formaldehyde, and 5% glacial acetic acid). After an incubation of 4 h, samples were dehydrated in an ethanol series. They were then transferred to toluene and infiltrated with Paraplast X-tra tissue embedding medium (Fisher, Waltham, MA, USA) in a Leica TP1020 automatic tissue processor. Samples were then embedded in fresh Paraplast using a Microm AP280 tissue embedding center and stored at 4°C until use. Samples were prepared and sectioned at 10 µm according to standard methods on a Microm HM3555 rotary microtome. DNA templates for RNA probe synthesis were obtained by PCR amplification of 350–500 bp fragments. To ensure specificity, the probe templates included a portion of the 3' UTR and the C-terminal portion of the proteins that is specific to each MADS-box gene. Fragments were cleaned using the QIAquick PCR purification kit (Qiagen, Valencia, CA, USA). Digoxigenin labeled RNA probes were prepared using T7 polymerase (Roche, Switzerland), murine RNase inhibitor (New England Biolabs, Ipswich, MA, USA), and RNA labeling mix (Roche, Switzerland) according to each manufacturers protocol. RNA *in situ* hybridization was performed according to Ambrose et al. (2000) and Ferrándiz et al. (2000), optimized to hybridize overnight at 55°C. Probe concentration was identical for all the experiments including the sense control hybridizations. RNase-treated control slides were not used, as expression patterns observed for the target genes were not indicative of artificial signal due to “stickiness” in any particular tissue. *In situ* hybridized sections were subsequently dehydrated and permanently mounted in Permount (Fisher, Waltham, MA, USA). All sections were digitally photographed using a Zeiss Axioplan microscope equipped with a Nikon DXM1200C digital camera.

RESULTS

Flower Development

The flowering shoots of *A. fimbriata* are indeterminate. A solitary flower is formed axillary to each leaf. An accessory bud is often produced per node, in an adaxial position with respect to the floral bud (Figures 2A,B). Floral primordia are monosymmetric, transversally oblong (Figure 2C). Flower development can be readily divided into nine different stages prior to anthesis (Table 1). Stage 1 (S1) in flower development is here defined by the initiation of two lateral and one median sepal primordia (Figure 2D). Next, at stage 2 (S2) the individual sepal primordia become fused in a dome-shaped perianth, which begins to undergo asymmetric intercalary growth, as the abaxial region corresponding to the median sepal grows noticeably faster (Figures 2E,F). The rapid elongation of the abaxial region of

the growing perianth is likely due to faster cell division rather than cell expansion, as cell size is the same in all flanks of the perianth (Figures 3A,B). Stage 3 (S3) is defined by the formation of the furrow between the flanks of the two lateral sepals, and the initiation of the perianth curvature, the anthers and the ovary (Figures 2G and 3A,B). Perianth curvature occurs as a result of a stronger elongation of the abaxial flank. At S3 initiation of trichomes in the apex of the growing perianth occurs (Figure 2G), and the six anther primordia become evident at the bottom of the perianth above the forming inferior ovary (Figures 2G and 3A,B). Stage 4 (S4) can be distinguished by the emergence of hooked trichomes in the outer epidermis of the ovary (Figure 2H), the growth of the six anthers and the thecae differentiation (Figures 3C,D). By Stage 5 (S5), the future abscission zone between the perianth and the inferior ovary is formed, and a proper utricle, tube and limb can be distinguished (Figure 2I); simultaneously, the six stigmatic lobes gradually elongate fused to the inner side of the anthers, overtopping them, thus forming the gynostemium (Figures 3E–H). At this stage, the anthers become tetrasporangiate; the filaments never differentiate (Figures 3E–H). At Stage 6 (S6), perianth growth and elongation continues and the furrow left by the two lateral sepals closes by a tight interlocking of the marginal epidermis (Figures 2J,K); at this stage the total length of the perianth reaches ca. 0.8 mm and the three parts of it (the basal *utricle*, the narrow *tube*, and the *limb*) are already apparent (Figures 2J,K and 6A).

Prior to anthesis, the young flower undergoes drastic changes that include the growth of the six stigmatic lobes above the level of the anthers, a feature that marks the beginning of Stage 7 (S7; Figures 3I–K and 6B). The six anthers become totally fused with the stigmatic lobes. The commissural origin of each stigmatic lobe is supported by the presence of two bulges toward their apices, whereas the apex of each carpel remains subterminal and alternate with the stamens (Figures 3I–L). In the ovary, the protrusion of the six placentae is simultaneous with the differentiation of two rows of ovule primordia, each formed at the margin of each carpel (Figures 3M,N). Stage 8 (S8) is here defined by the repositioning of the flower through resupination, that is, the future floral entrance turns away from the shoot axis as the result of the gradual torsion of the peduncle (Figures 1A and 6C). Additionally, at this stage the inner epidermis of the perianth acquires a dull-purple pigmentation in the tube and the limb, while the outer epidermis changes from green to pale yellow (Figures 4A and 6C). At Stage 9 (S9; Figure 6D) the limb remains closed, the stigmatic lobes are already expanded and wet, and the anatropous, bitegmic, and crassinucellar ovules are fully developed (Figures 3O,P).

Anthesis occurs acropetally with one flower opening at a time (Figures 1A and 6E). During anthesis, the limb unfolds and expands, displaying the fimbriae and allowing insects to reach the utricle through the tube. At day 1 of anthesis, the gynostemium remains at its “female stage” as wet stigmas are fully expanded and receptive, whereas the anthers are still indehiscent; at day 2–3, the gynostemium enters its “male” stage, recognized by the closure of stigmatic lobes and the dehiscence of the anthers; at this stage the trapped insects are exposed to the pollen (Figures 1C,D). By

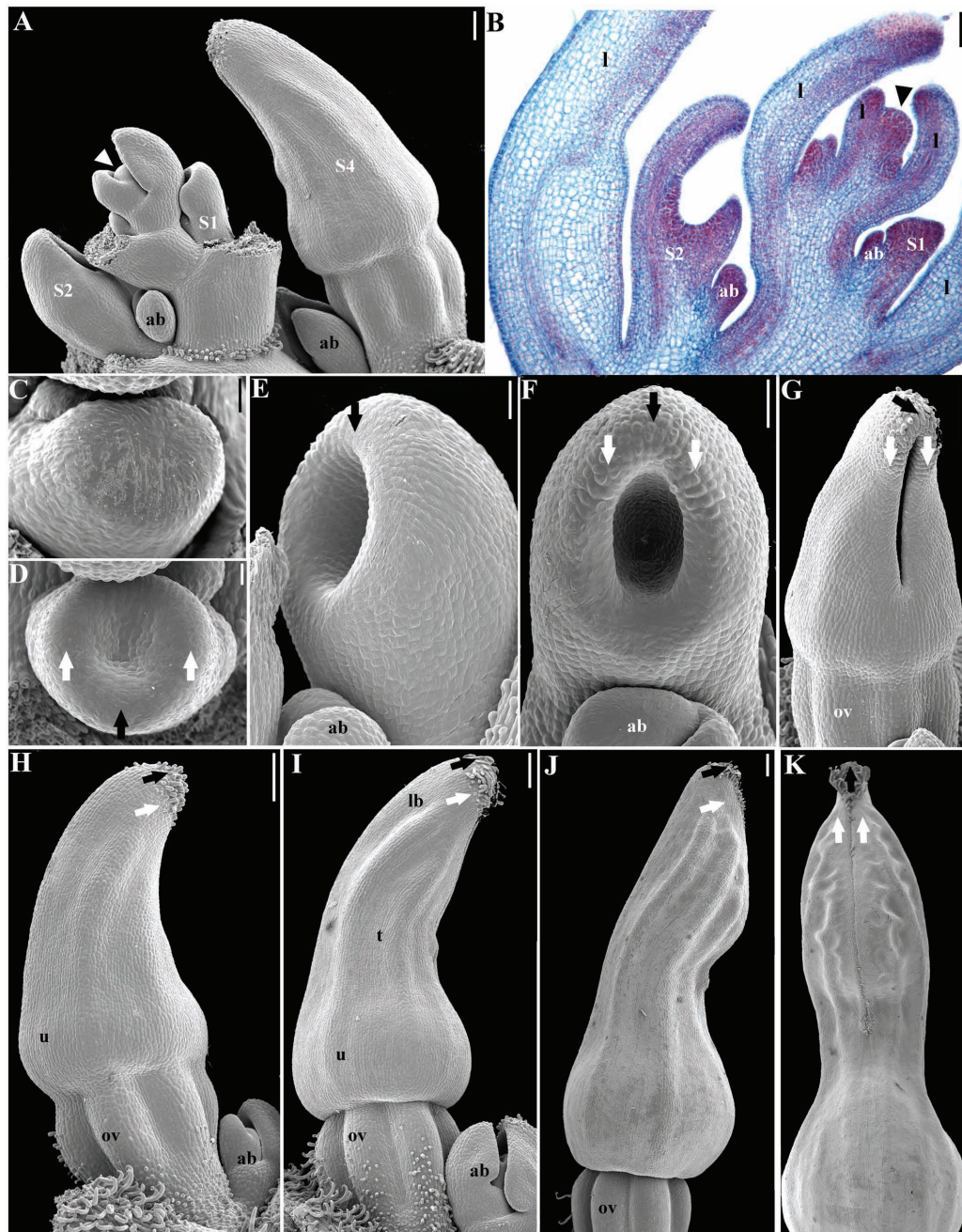


FIGURE 2 | *Aristolochia fimbriata*, perianth development. (A,B) Shoot apex with floral stages S1 to S4. **(C,D)** Initiation of sepal primordia (stage S1). **(E,F)** Elongation of sepal primordia (S2). **(G)** Perianth closure and ovary (ov) differentiation (S3). **(H)** Initiation of hooked trichomes (S4). **(I)** Differentiation of utricle (u), tube (t), and limb (lb) during S5. **(J,K)** Perianth closure (S6). Black arrows indicate the median sepal; white arrows indicate lateral sepals; arrowheads indicate the shoot apical meristem; ab, accessory bud. Scale bars: **(A,B), (G–K):** 100 μm ; **(C,D):** 20 μm ; **(E,F):** 40 μm .

late anthesis, perianth withers and allows the insects to escape (González and Pabón-Mora, 2015).

Epidermis and Trichomes of the Perianth

The outer epidermis of the utricle, the tube and the limb in *A. fimbriata* is homogeneously formed by flat epidermal cells with interspersed stomata, and lacks trichomes (Figures 1B,C

and 4E,J,N). This contrasts with the much more elaborated inner epidermis, as at least four different types of trichomes develop (Figures 4A,G–I,K–M,O–Q). Conical trichomes in the tube develop first at the beginning of S5 (Figures 4K–M); these trichomes are secretory during late preanthesis (S8 and S9), but during anthesis they function as the guard trichomes that keep pollinators temporarily trapped. Next, epidermal elaboration

TABLE 1 | Developmental landmarks for each stage identified during flower development.

STAGE	Developmental landmarks
S1	• Initiation of the sepal primordia
S2	• Sepal fusion • Sepal growth lead by the median sepal
S3	• Furrow formation between the lateral sepals • Initiation of perianth curvature • Inception of the anther primordia • Ovary differentiation
S4	• Differentiation of hooked trichomes on the ovary • Thecae differentiation
S5	• Abscission zone formation between the perianth and the inferior ovary • Perianth differentiation into utricle, tube and limb • Differentiation of the upper portion of the carpels • Fusion between the differentiating carpelar portion and the inner surface of the anthers • Differentiation of the tetrasporangiate anthers
S6	• Closure of the furrow between the two lateral sepals with interlocking epidermis
S7	• Growth of the six stigmatic lobes above the anthers • Ovule initiation (nucellus apparent)
S8	• Resupination of the flower by torsion of the peduncle • Pigmentation initiation in the tube and limb
S9	• Differentiation of the two integuments in the ovules
Anthesis	• Expansion of the limb

in the utricle occurs at S6, as a carpet of long, multicellular, filamentous, nectarial trichomes develop and surround small patches of osmophores and nectarioles (Figures 4O–Q). The marginal fimbriae begin to form at S6 (Figures 4B,C), and by late preanthesis (S8 and S9; Figures 4D,E) they reach their final size but remain folded. They are vascularized, and possess secretory tips with osmophores, and hooked trichomes scattered along their proximal half (Figures 4B–E). Most of the inner epidermis of the limb is formed by osmophores, accompanied by scattered conical, and hooked trichomes (Figures 4G–I).

Isolation and Expression of MADS-Box Genes

In order to identify orthologs of the ABCE genes involved in organ identity as well as *SEEDSTICK* and *AGAMOUS-like6*, we searched the generated transcriptome using as a query orthologous genes previously identified from other basal angiosperms (Kim et al., 2005; Yoo et al., 2010). We were able to obtain hits for gene members representing major gene lineages and were able to identify one *AP1/FUL* gene (named *AfimFUL*), two *LOFSEP/SEP3* genes (named *AfimSEP1* and *AfimSEP2*), one *AGL6* gene (named *AfimAGL6*), one *AP3/DEF* gene (named *AfimAP3*), one *PI/GLO* gene (named *AfimPI*), one *AG* gene (named *AfimAG*), and one *STK* gene (named *AfimSTK*; Figure 5).

In order to investigate the expression patterns across developmental stages of all genes involved in the ABCE model of flower development as well as *STK* and *AGL6* we did an expression screening using reverse transcription (RT)-PCR

(Figure 6). We tested the expression of all copies at different developmental stages from S6 through anthesis. In addition, we tested gene expression in leaves and capsules. Our results show that most genes are expressed in vegetative and reproductive tissues with the exception of *AfimSEP2*, *AfimAGL6*, and *AfimSTK* transcripts that are only found in flowers and capsules but are not detected in leaves (Figure 6). Two genes show ubiquitous expression. *AfimFUL* (the putative “A-class” gene, ortholog of *AP1* and *FUL*) that is found in all floral organs at all developmental stages as well as in leaves and capsules, and *AfimAG* (a “C-class” gene ortholog of *AG* and *SHPI/2*) that has low expression in leaves and is found in all floral organs at all stages with a considerable reduction of expression in the limb of anthetic flowers (Figure 6). The two E-class gene copies *AfimSEP1* and *AfimSEP2* show different expression patterns; *AfimSEP1*, (ortholog of *SEP1/2/4*) is expressed in all floral organs at early stages S6, S7, and S8 except in the ovary and turned off specifically in the utricle of the perianth at S9 and anthesis, stages at which is present in the ovary; in contrast, *AfimSEP2* (ortholog of *SEP3*) is found in all floral organs throughout development and its expression is only reduced in the ovary of the flowers at anthesis (Figure 6). *AfimAGL6* is detected in the perianth at all stages. *AfimAGL6* is never expressed in the gynostemium. In addition, the expression of *AfimAGL6* in the ovary is reduced from S9 to anthesis (Figure 6). The B-class gene copies *AfimAP3* and *AfimPI* are both widely expressed in all floral organs at stages S7–S9, however, at anthesis *AfimAP3* becomes restricted to the limb and the tube in the perianth and to the gynostemium, whereas *AfimPI* is only expressed in the tube at very low levels (Figure 6). Out of all genes evaluated, both *AfimAP3* and *AfimPI* are the only two genes that are not expressed in the capsules (Figure 6). Finally the D-class gene *AfimSTK* is only found in the gynostemium and the ovary during all stages of flower development and its expression is persistent in the capsule (Figure 6).

Detailed Expression Analyses of Perianth Identity Candidate Genes

With the purpose of identifying putative perianth identity candidate genes, in this case sepal-derived, we decided to investigate the expression of *AfimFUL* using *in situ* hybridization experiments to evaluate its contribution to the initiation and development of the perianth parts, as it is a member of the *AP1/FUL* gene lineage to which sepal and petal identity has been attributed in *Arabidopsis* as well as in more early diverging lineages such as *Papaver* and *Eschscholzia* (Papaveraceae, basal eudicots; Bowman et al., 1991; Kempin et al., 1995; Pabón-Mora et al., 2012). Our results are consistent with the RT-PCR and show that expression of *AfimFUL* is very broad and can be detected in the shoot apex, young and old leaves, floral meristems, as well as accessory buds and their respective bracts (Figures 7A,B). However, *AfimFUL* expression shifts from a homogeneous expression in young leaves to a localized adaxial expression in older leaves (Figure 7A). *AfimFUL* is detected throughout flower development between stages S1–S6, starting with a broad expression during the differentiation of sepal primordia (S1; Figure 7C) that is maintained during their

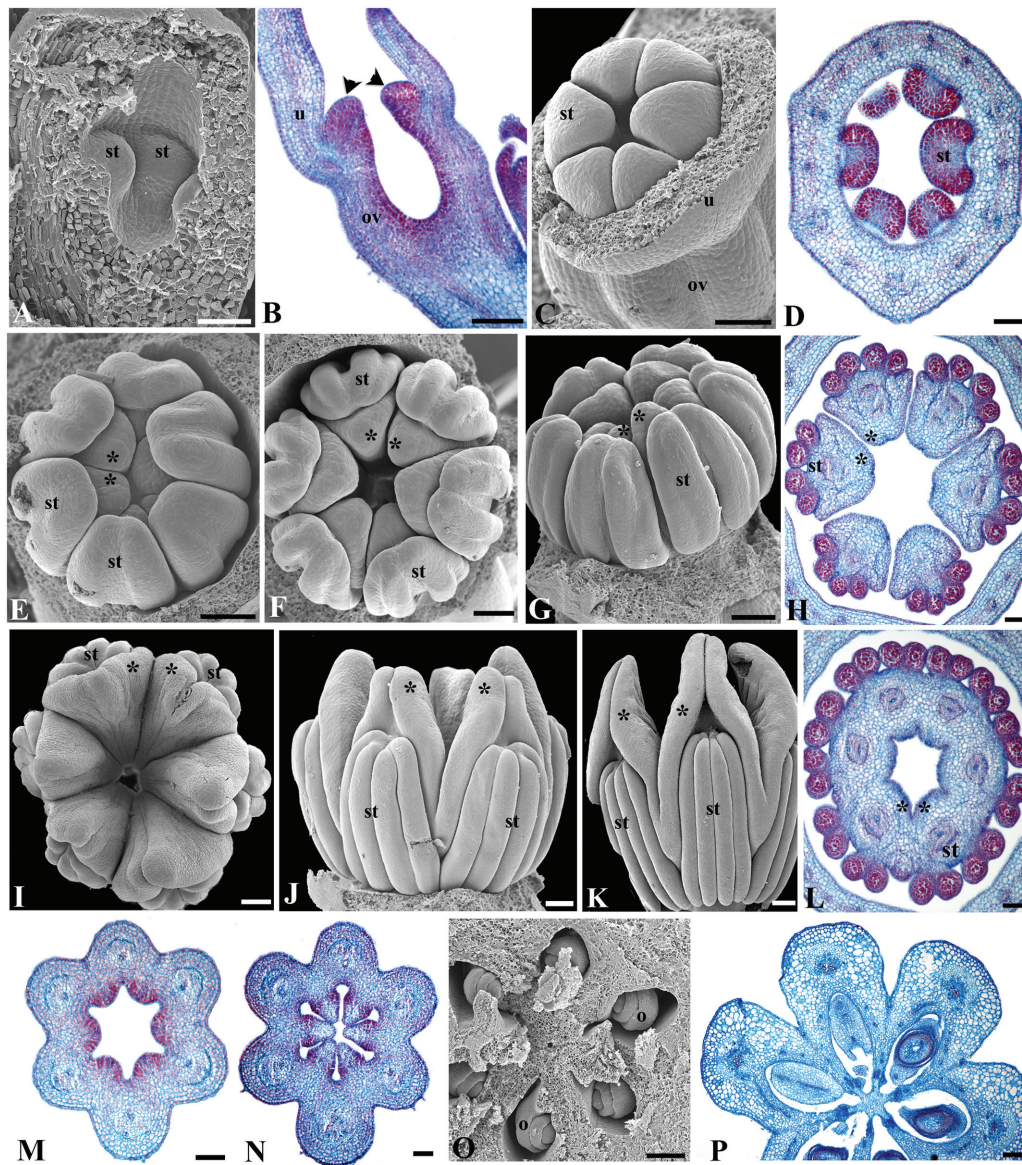


FIGURE 3 | *Aristolochia fimbriata*, gynostemium and ovary development. (A,B) Stamen initiation above the forming inferior ovary at S3. **(C,D)** Anther growth and formation of the sporogenous tissue at S4. **(E–G)** Two successive stages of gynostemium development at S5 **(E)** and S6 **(F,G)**. **(H)** Transverse section of the gynostemium at S6. **(I–K)** Top **(I)** and lateral **(J,K)** views of the “female” stage of gynostemium with stigmatic lobes growing above the staminal tissue at S7. **(L)** Transverse section of the gynostemium at S7 with carpellary tissue growing between the anthers. **(M–O)** Successive stages of ovary development at S6 **(M,N)** and S7 **(O)**. **(P)** Ovary during transition to fruit **(P)**. o, ovule; ov, ovary; st, stamen; u, utricle; black arrowheads indicate anthers; asterisks (*) indicate stigmas. Scale Bars **(A,B,L)**: 60 μm ; **(C,E–G,J,O,P)**: 100 μm ; **(D,H,M)**: 50 μm ; **(I)**, 200 μm ; **(K)**, 300 μm ; **(N)**, 80 μm .

asynchronous growth (S2, 3; **Figures 7E–G**), with a higher expression in the adaxial perianth region between the upper or medial sepal and the two lower sepal lobes (**Figures 7D–F**). *AfimFUL* transcripts are also present in the stamen primordia since their initiation and until pollen differentiation (S3–S6; **Figures 7E–H**). Despite this broad expression, *AfimFUL* was not detected in the stigmatic portion of the carpels at the gynostemium (S5, S6; **Figures 7G,H**). At S6 (the oldest stage analyzed using *in situ* hybridization) *AfimFUL* is expressed in the perianth, in particular in the inner and outer epidermis

and the vasculature, the stamens, the ovary and the ovule primordia.

Next, we decided to analyze *AfimAGL6* detailed expression, which according to our RT-PCR is detected in the perianth during flower development but it is never detected in the gynostemium. Contrary to *AfimFUL*, the expression of *AfimAGL6* is localized to the perianth early on during the initiation of the three sepal primordia (S1; **Figures 8A,B**) and throughout flower development (S2–S6; **Figures 8C–G**). *AfimAGL6* is not detected in the shoot apex, the leaves or the axillary dormant buds

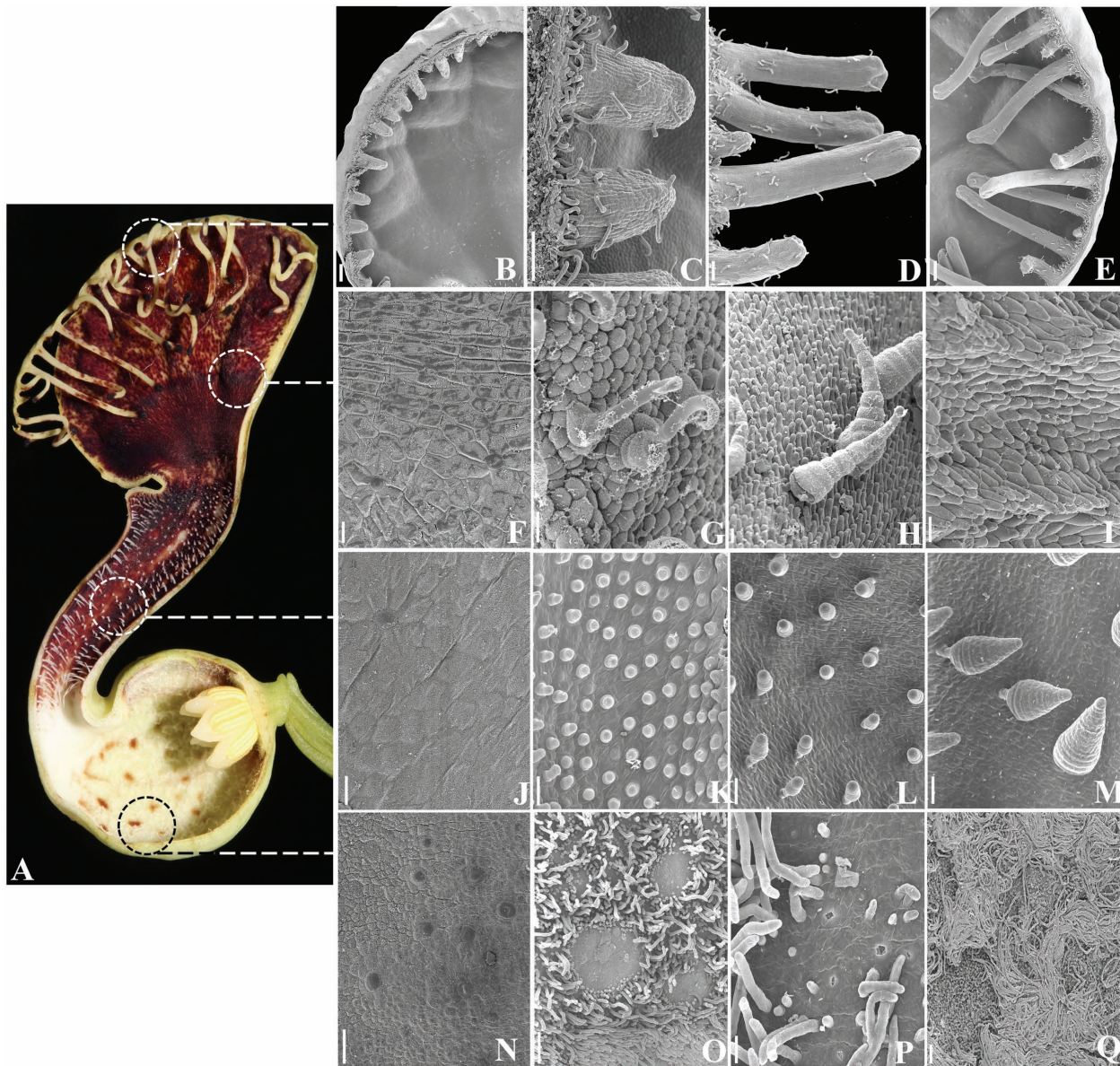


FIGURE 4 | *Aristolochia fimbriata*, epidermal differentiation of the perianth. (A) Preanthetic flower, longitudinal section. **(B–E)** Fimbriae development and ornamentation. **(F–I)** Ornamentation of the abaxial **(F)** and adaxial **(G–I)** epidermis of the limb. **(J–M)** Ornamentation of the abaxial **(J)** and adaxial **(K–M)** epidermis of the tube. **(N–Q)** Ornamentation of the abaxial **(N)** and adaxial **(O–Q)** epidermis of the utricle. Scale bars: **(B,E)**: 200 μm ; **(C,D,O)**: 100 μm ; **(F–N,P)**: 20 μm ; **(P)**: 50 μm .

(Figures 8A,B). Its expression in the perianth is continuous from the utricle to the limb (Figures 8C–G). *AfimAGL6* is not detected in the stamen primordia early on (S3, S4; Figures 8C,D) or in the young gynostemium (S5, S6; Figure 8E) nor in the staminal or the stigmatic tissue. *AfimAGL6* is turned on again in the pedicel at the level of the ovary since S3 and then later on at S6 in the ovule primordia (Figures 8C–E,H). Late expression of *AfimAGL6* is detected in the outer and the inner integuments of the ovule (Figures 8I,J).

Finally, we decided to investigate the contribution of the B-class genes to perianth identity in *A. fimbriata*. The expression

of *AfimAP3* detected through *in situ* hybridization is also very broad, which is consistent with the RT-PCR results, and the transcripts are present at low levels in the shoot apex, leaves and floral meristems (Figures 9A,B). Expression of *AfimAP3* increases at S3 and S4, when it is localized to the stamen primordia (Figures 9A,B). Expression of *AfimAP3* continues to be localized in staminal tissue in the gynostemium (S5, S6; Figures 9C–E), but it is not detected in the stigmatic portion of the gynostemium or in the ovary (Figures 9C–E).

On the other hand, *AfimPI* is broadly expressed in the perianth, the gynostemium and the ovary according to the

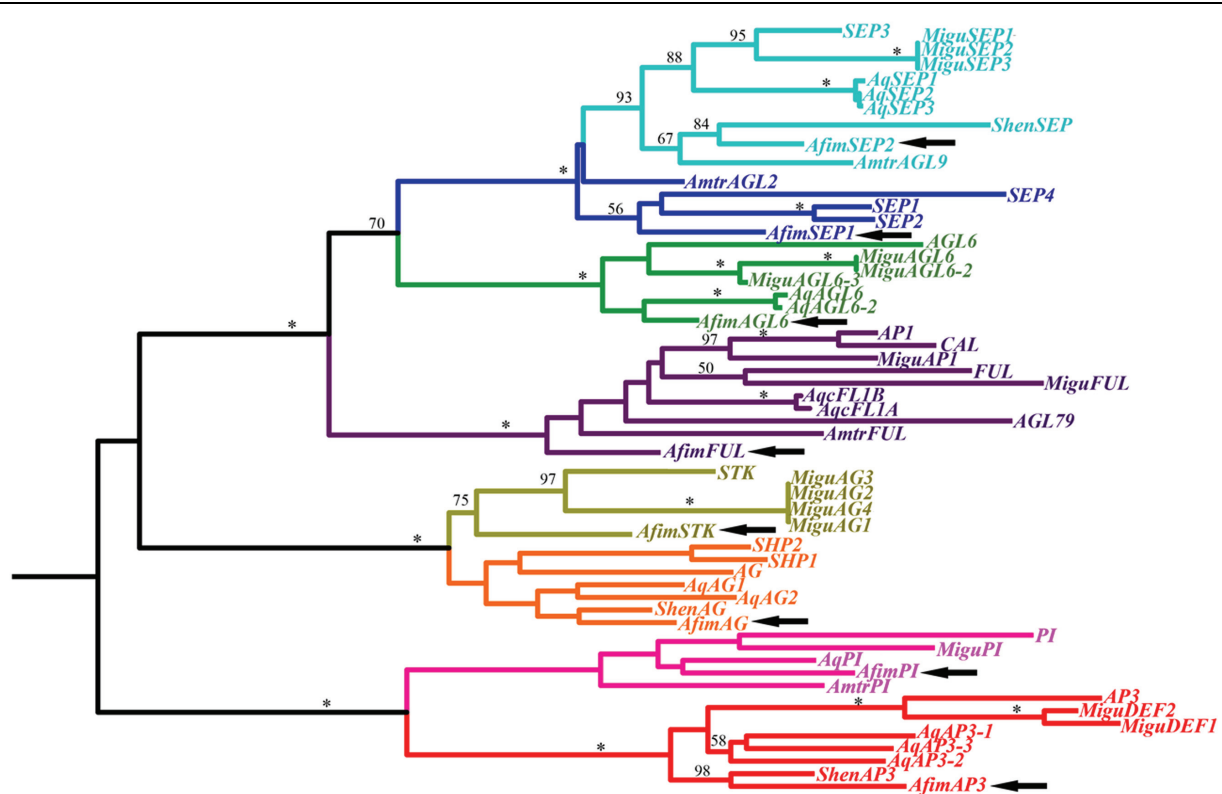


FIGURE 5 | Maximum likelihood tree of MADS-box genes identified in the transcriptome of *A. fimbriata* (arrows). Branch colors denote gene clades, as follows: blue, *SEPALLATA*; green, *AGAMOUS-Like 6*; purple, *APETALA1/FRUITFULL*; ochre, *SEEDSTICK*; orange, *AGAMOUS/SHATTERPROOF*; pink, *PISTILLATA*; red, *APETALA3*. BS values above 50% are placed at nodes. Asterisks indicate bootstrap (BS) of 100. Besides *A. fimbriata*, ingroup sequences include those of *Amborella trichopoda* (Amborellaceae), *Saruma henryi* (Aristolochiaceae), *Aquilegia coerulea* (Ranunculaceae), *Mimulus guttatus* (Phrymaceae), and *Arabidopsis thaliana* (Brassicaceae), representing the ANA grade, the magnoliids, the basal eudicots, the asterids and the rosids, respectively.

RT-PCR (Figure 5). However, its expression turns off almost completely in all floral organs as the flower enters anthesis (Figure 5). *In situ* hybridization results confirm that *AfimPI* is turned on at S3 in stamen primordia and maintained there until S6–S7 (Figures 10A–C). In addition, during S3 *AfimPI* is also expressed in the emerging gynostemium lobes and in the inner cell layers of the perianth starting in the utricle, and expanding toward the tube and limb (Figures 10C–G). By S5–S6, expression of *AfimPI* expands to outer staminal portion of the gynostemium and the inner layers of the base of the perianth, and later to the distal portion of the utricle that limits with the tube (Figures 10H,I). By S6 *AfimPI* becomes restricted to the fertile portions of the stamens and is turned on in the ovules (Figure 10J).

DISCUSSION

The Aristolochiaceae encompasses an enormous variation of life forms and floral morphologies including flowers with radial (*Saruma*, *Asarum*, and *Thottea*) as well as bilateral symmetry (*Aristolochia*), with (*Saruma*), or without petals (*Asarum*, *Thottea*, and *Aristolochia*), and having partial (*Saruma*, *Asarum*, *Thottea*) or total congenital fusion between stamens and carpels

to form a gynostemium (*Aristolochia*). Following such character combination, *Aristolochia* species are unique in possessing deep, tubular, sepal-derived perianths with strong monosymmetric, highly synorganized flowers (González and Stevenson, 2000a,b; González and Pabón-Mora, 2015). *A. fimbriata* produces 6–7 floral plastochrones from the shoot apex that can include all early stages (from S1–S6) in less than a full centimeter (Figures 2,3). *Aristolochia fimbriata*, like most *Aristolochia* species, undergoes early fusion of the sepals followed by asymmetrical growth of the perianth (S2–S4), almost simultaneously with the fusion between anther primordia and the upper portion of the carpels (S3–S4). In addition, *A. fimbriata* differentiates the utricle, tube and limb with distinct unique epidermal specializations (S5–S6). Thus, *A. fimbriata*, serves as a unique reference point for studying complex floral features like organ fusion, and extreme synorganization and transfer of function from the suppressed petals to a sepal-derived perianth.

Aristolochia fimbriata has a Floral MADS Box Gene Toolkit Similar to Other Early Divergent Angiosperms

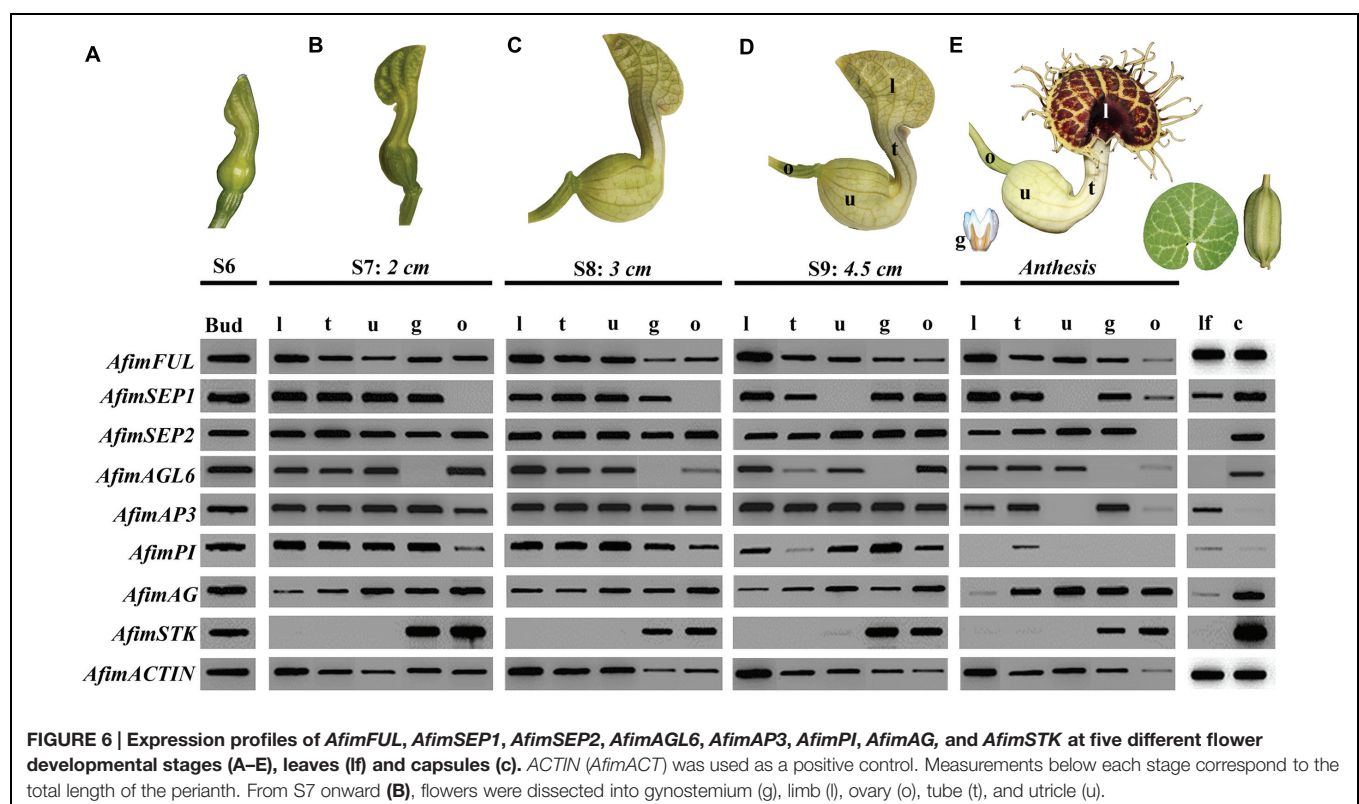
According to the APG (2009), the magnoliids are more closely related to the earliest diverging angiosperms Amborellales,

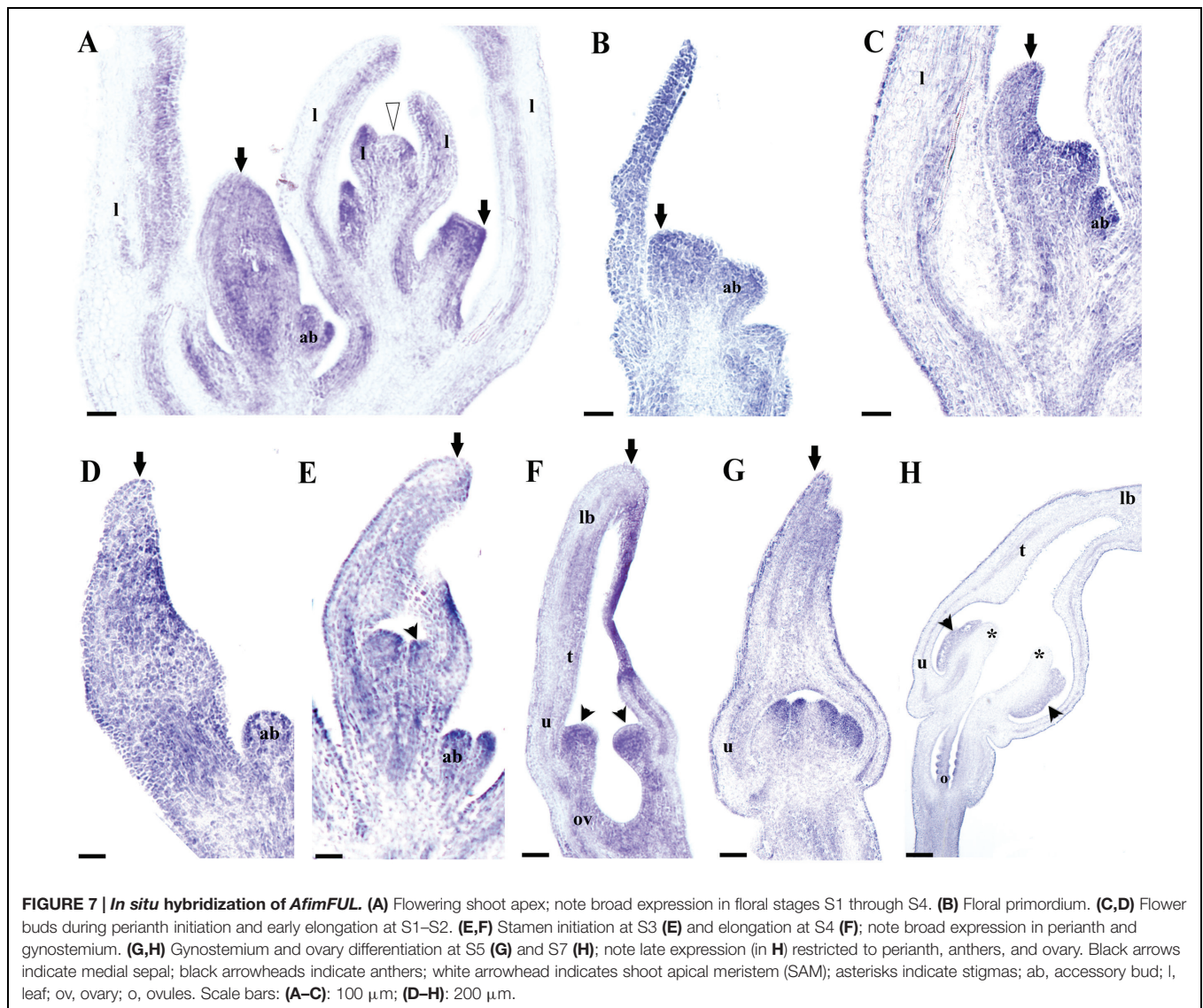
Nymphaeales, and Austrobaileyales and altogether these lineages were thought to have evolved prior to the diversification of monocots and eudicots. More recently, the phylogenomic approach using whole transcriptome data (oneKP project), has resulted in an alternative topology, with the magnoliids as sister to the eudicots only, and the monocots evolving independently (Wickett et al., 2014; Zeng et al., 2014). WGD events have been proposed to occur at different times during plant evolution, once, prior to the diversification of angiosperms (ϵ), two times in the monocots (ρ , σ), once before the diversification of the eudicots (γ), and twice in the Brassicaceae (α , β ; Jiao et al., 2011). In addition to these large-scale duplications, Ks distributions of paralogs have been used to propose several additional basal-angiosperm – specific genome duplications, one shared between Laurales and Magnoliales and another within Piperales only (Cui et al., 2006). These studies raise the question regarding the genetic complement present in basal angiosperms, and in particular in Piperales with respect to the earliest diverging angiosperms on one side and the specious monocot and eudicot clades, on the other. Our results show that *A. fimbriata* has a similar floral genetic toolkit to that found in the earliest diverging ANA members, *Amborella trichopoda* (Amborellaceae) and *Nuphar pumila* (Nymphaeaceae; Amborella Genome Project, 2013; Li et al., 2015). The *A. fimbriata* flower and fruit mixed transcriptome allowed us to find expression of a single copy of each A, C, D-class, and AGL6 MADS-box clade, and two paralogs for the B and the E-class genes, which are known to have duplicated prior to the diversification of angiosperms likely in the ϵ WGD event (Figure 5; Kramer et al., 1998;

Becker and Theissen, 2003; Zahn et al., 2005). These results suggest that no additional duplications or losses have occurred in *A. fimbriata* when compared to *Amborella trichopoda*, the earliest diverging angiosperm. Nevertheless copy number can only be confirmed with genome sequencing and broader phylogenetic samplings will have to be done to assess whether duplications have occurred in other magnoliids independently. This approach is critical, given their phylogenetic affinities with the eudicots under the most updated plant classification system, and considering that other analyses have shown taxa specific duplications for *Persea americana* (Laurales) and *Liriodendron tulipifera* (Magnoliales; Cui et al., 2006; Wickett et al., 2014; Zeng et al., 2014).

Floral Meristem and Perianth Identity are Likely Determined by *FUL*-like and *AGL6* Genes in *A. fimbriata*

Our evaluation of the expression of A-class and *AGL6* genes during *A. fimbriata* flower development indicates that *AfimFUL* and *AfimAGL6* have homogeneous overlapping expression in the sepal primordia (S1) and during perianth fusion and elongation (S2–S4). Moreover, at later stages of flower development, *AfimFUL* expression expands to all other floral organs, including ovules, whereas *AfimAGL6* is expressed only in carpels and ovules (Figures 7 and 8). *AfimFUL* is also expressed in the shoot apical meristem and leaves during development (Figure 7). Our data are consistent with qRT-PCR expression data shown for *FUL* and *AGL6* homologs in other magnoliids and basal angiosperms (Kim

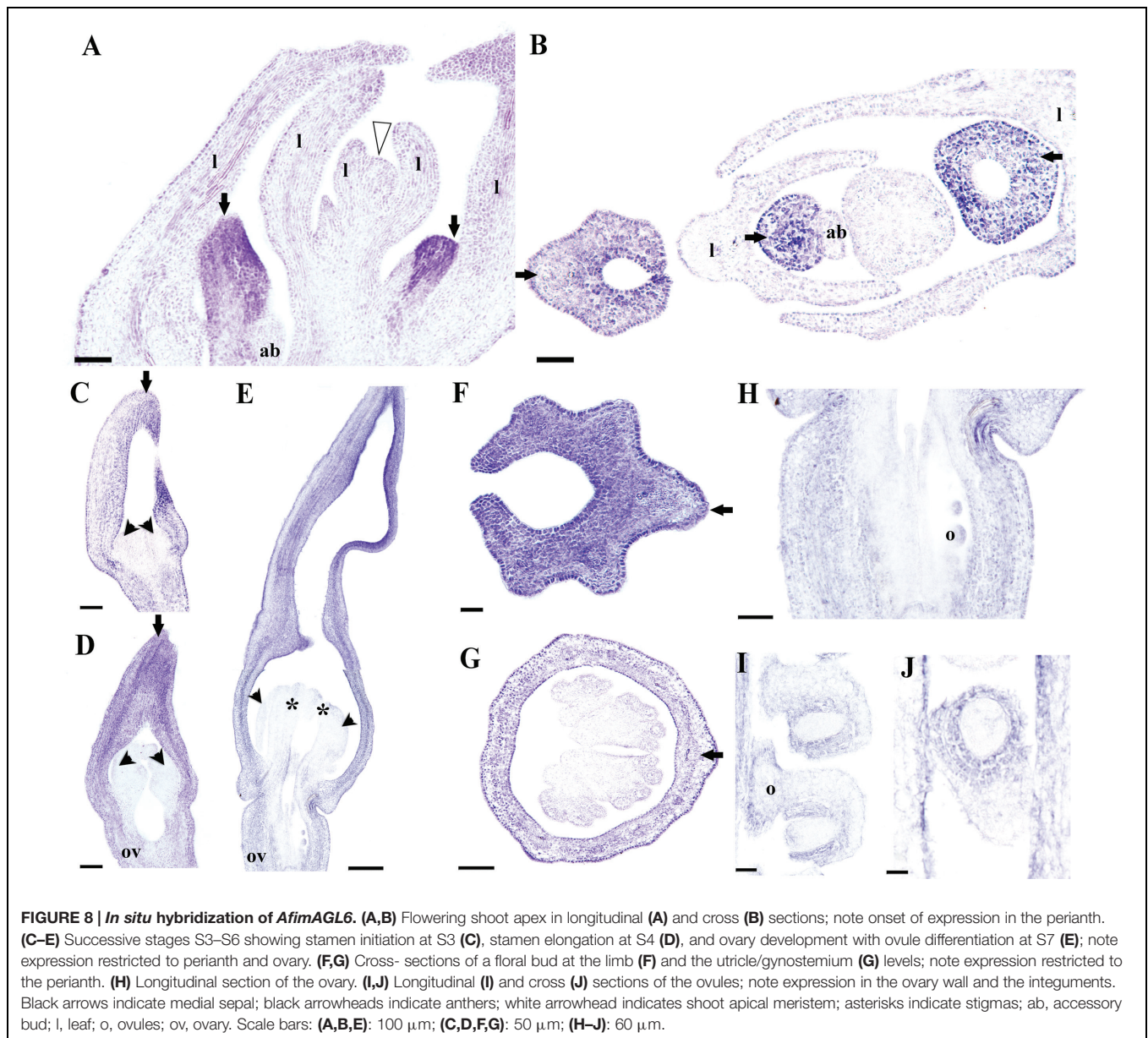




et al., 2005; Yoo et al., 2010). In addition, our *in situ* hybridization results provide a better assessment of spatio-temporal expression patterns and more accurate predictions of putative functions associated with the activation of these transcription factors.

AfimFUL is part of the *API/FUL* gene lineage, while *AfimAGL6* is part of the *AGL6* gene lineage, the former is angiosperm specific, whereas the latter, is present in all seed plants (Litt and Irish, 2003; Viaene et al., 2010; Kim et al., 2013). The original ABCE model in *Arabidopsis* established that *API* (in the *API/FUL* gene lineage) was responsible for floral meristem and perianth identity (Coen and Meyerowitz, 1991). Gene evolution analyses, coupled with expression and functional studies of *API/FUL* homologs across angiosperms has revealed a complex scenario of two rounds of gene duplication, resulting in the *euFULI*, *euFULII*, and *euAPI* clades in core eudicots, accompanied by subfunctionalization (Litt and Irish, 2003; Pabón-Mora et al., 2012). Specifically, *euAPI* and *euFUL* core-eudicot paralogs, have non-overlapping expression patterns

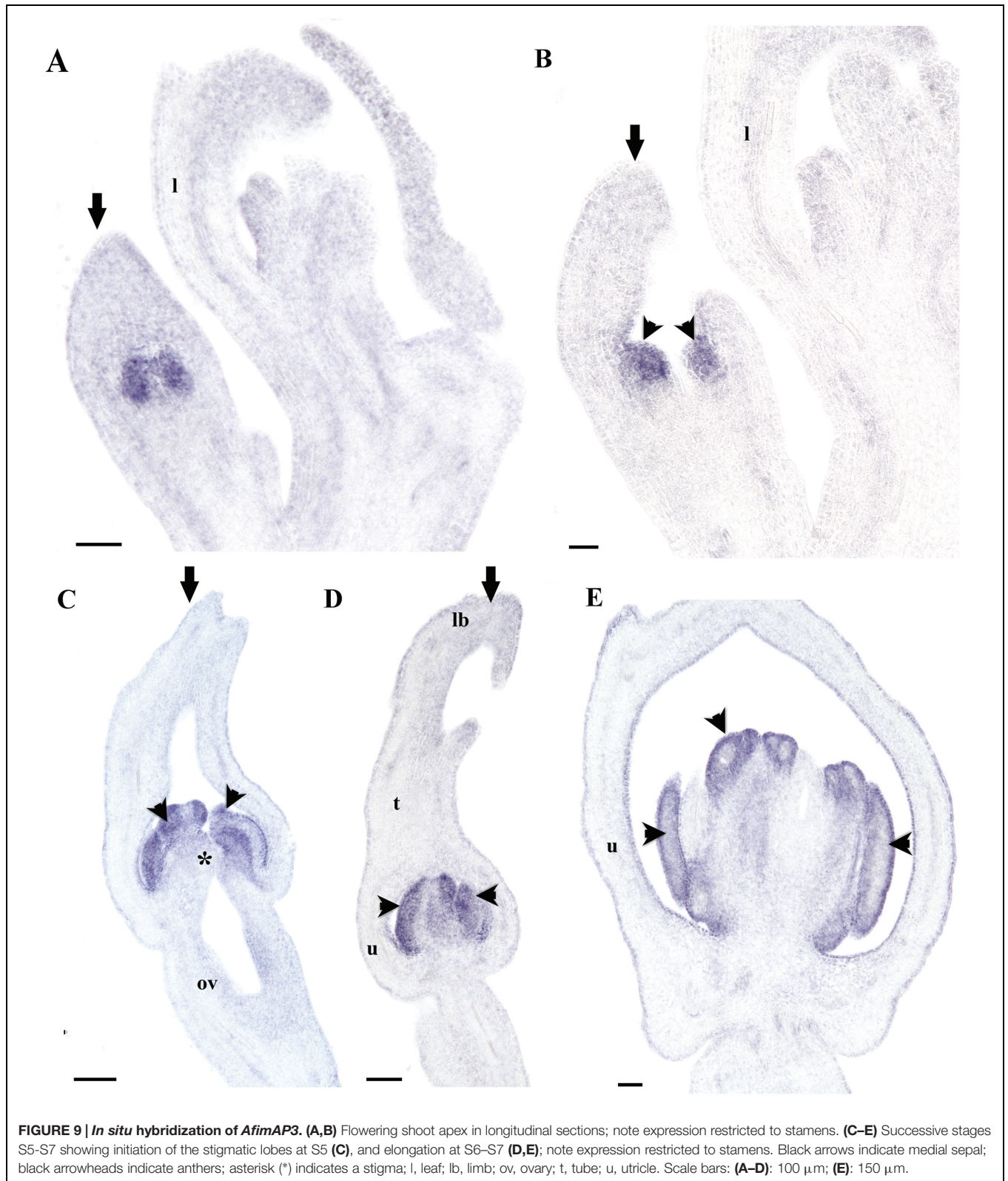
consistent with their unique roles in plant development. While *euAPI* genes including the canonical *API*, are turned on in floral meristems and perianth, and function determining sepal (and sometimes petal) identity, *euFUL* genes are expressed during the transition to inflorescence meristem and later on in the carpel and fruit and control the transition from inflorescence to flower as well as proper fruit wall development (Huijser et al., 1992; Kempin et al., 1995; Gu et al., 1998; Ferrándiz et al., 2000). The pre-duplication *FUL-like* genes have also undergone local duplications (Litt and Irish, 2003). Functionally, *FUL-like* genes have been characterized in grasses (monocots), where they play roles in the transition to reproductive meristems (Murai et al., 2003), and the Ranunculales (basal eudicots), where they function in flowering time, patterning inflorescence architecture, leaf morphogenesis, floral meristem and sepal identity, late petal epidermal differentiation and fruit development (Pabón-Mora et al., 2012, 2013; Sun et al., 2014). The broad expression patterns of *AfimFUL* suggest that the



pleiotropic roles of *FUL-like* genes occur in *Aristolochia*, and thus would predate at least the diversification of basal eudicots and magnoliids.

Gene lineage evolution, together with expression and functional analyses studies for the *AGL6/AGL13* genes support a major duplication event in the Brassicaceae resulting in the *AGL6* and the *AGL13* clades, as well as distinct roles in flower development in monocots, eudicots, and the Brassicaceae (Viaene et al., 2010). In *Arabidopsis*, both copies are expressed in the ovules, while *AGL6* is restricted to the endothelium, *AGL13* is found in the chalaza, however, *agl6* and *agl13* single null mutants exhibit wild type phenotypes suggesting redundancy (Schauer et al., 2009). In other core-eudicots, like *Petunia*, *PhAGL6* is expressed in developing petals, carpels and ovules; the *phagl6* mutant does not exhibit noticeable abnormal phenotypes, but

it shows a role in petal identity in combination with the two *SEPALLATA* copies *FBP2/5*, as the *fbp2/5 phagl6* triple mutant shows enhanced sepaloid petals (Rijkema et al., 2009). In grasses, both *osmads6* and *bde* mutants, in rice and maize, respectively, show altered palea identity, abnormal carpels with multiple stigmas and ovules with protruding nucelli (Thompson et al., 2009; Li et al., 2010). These data suggest that *AGL6* grass homologs control palea identity, as well as carpel and ovule development. More recently, studies in orchids have shown that *AGL6* paralogs, together with *AP3* orchid-specific copies have specialized in determining identity of sepal and lateral petals on one side, *versus* identity of the lip on the other. Thus, *AGL6* controls perianth identity and while the PPI *AGL61–AGL61–AP31–PI* is responsible for the identity of sepals and petals, the combination *AGL62–AGL62–AP32–PI* controls lip identity



(Hsu et al., 2015). The expression pattern here observed for *AfimAGL6*, suggests an early role in establishing floral meristem and sepal identity, together with *AfimFUL* and *AfimSEP* genes,

as well as a late role in ovule development. If we consider that the ancestral roles of *AGL6* genes in early diverging angiosperms include both perianth identity ovule development, the lack of

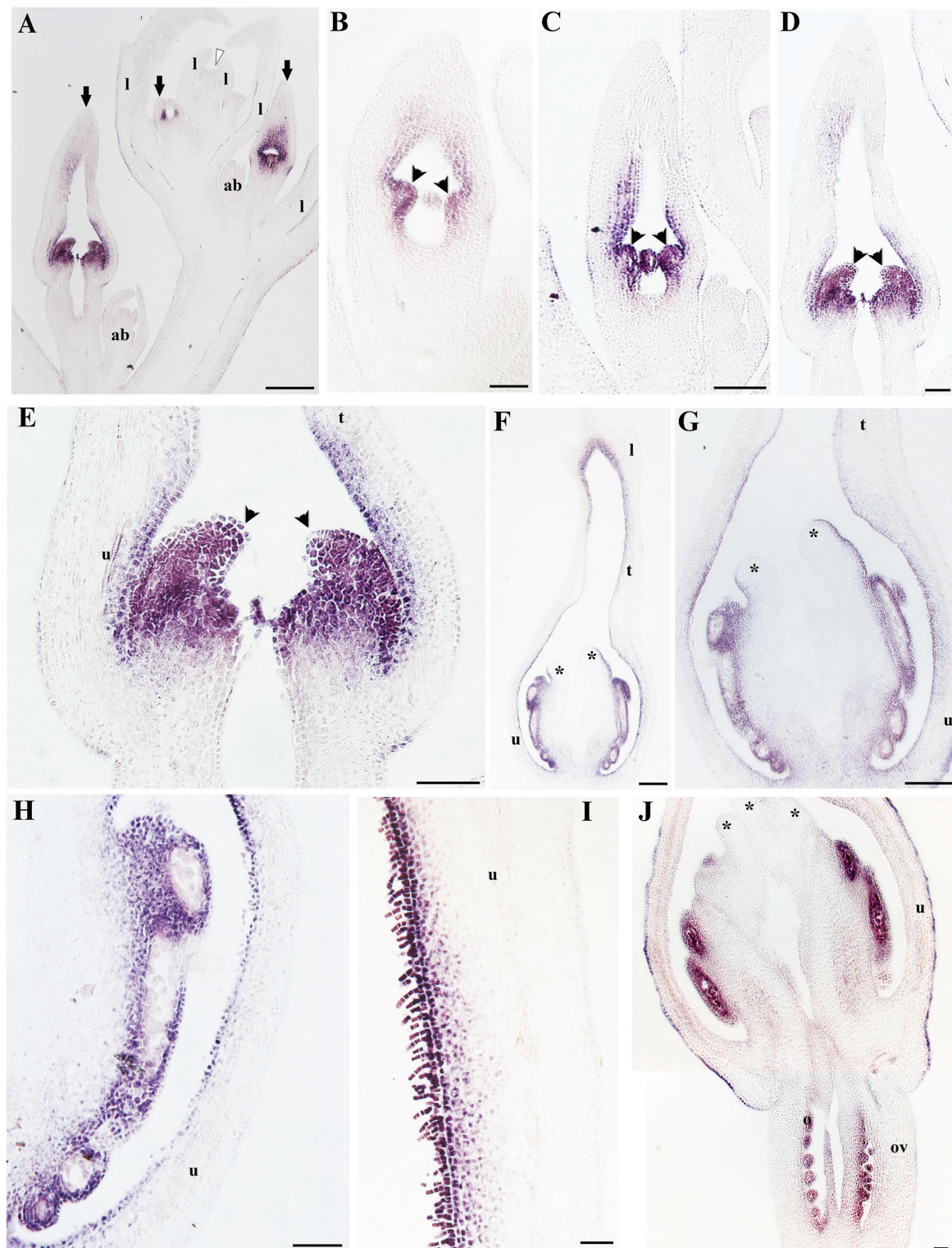


FIGURE 10 | *In situ* hybridization of *AfimPI*. (A) Flowering shoot apex in longitudinal section; note expression in stamens and surrounding perianth tissue. (B) Stamen initiation at S3. (C) Stamen growth at S4. (D,E) Utricle, tube and limb differentiation together with stigmatic tissue initiation at S5; note expression in the young gynostemium. (F–J) Longitudinal sections and details of a floral bud at the utricle/gynostemium level at S7 (H) and the distal-most adaxial portion of the utricle (I); note expression in the adaxial epidermis, trichomes and hypodermis. Black arrows indicate medial sepal; black arrowheads indicate anthers; white arrowhead indicates shoot apical meristem (SAM); asterisks indicate stigmas; ab, accessory bud; l, leaf; o, ovules; ov, ovary. Scale bars: (A–D,F,G): 100 μm ; (E): 150 μm ; (H,I,J): 50 μm .

mutant phenotypes in ovules in other angiosperms suggest redundancy with other MADS-box ovule-specific transcription factors and the absence of mutant phenotypes in the perianth in the core eudicots suggest redundancy with *SEP* genes for determining floral meristem and perianth identity (Pelaz et al., 2000; Schauer et al., 2009; Rijpkema et al., 2009). The role of *AGL6* genes in controlling perianth identity in early diverging angiosperms and in magnoliids, will have to be explored across a number of species having unipartite sepaloid perianths and bipartite perianths to better assess whether *AGL6* plays a general role in perianth identity, like in *Aristolochia* or a distinct role in sepals *versus* petals or extremely modified petals, like in the case of orchids. As the phylogenetic position of the magnoliids is still uncertain (APG, 2009; Wickett et al., 2014), the perianth-specific identity role here proposed for *AGL6* will have to be assessed studying spatio-temporal expression of the *AGL6* orthologs in the ANA grade before extrapolating it to the early radiating angiosperms. In addition, testing protein interactions occurring *in vitro* and *in planta* will be fundamental to identify putative partners of MADS-box floral organ identity proteins with overlapping expression patterns in *Aristolochia*, to propose functional quartets active in floral meristem and perianth identity and to identify protein homo- and heterodimers that are common for eudicots and those that may be unique to the magnoliids.

Expression Patterns of B-class Homologs in *A. fimbriata* Suggest that they do not Contribute to Perianth Identity

The B-class transcription factors originally included in the model as responsible for petal and stamen identity include the *Arabidopsis* *APETALA3* and *PISTILLATA* genes (Bowman et al., 1989; Jack et al., 1992) and the *Antirrhinum* orthologs *DEFICIENS* and *GLOBOSA* (Sommer et al., 1990; Tröbner et al., 1992). Both genes in each species have been shown to function as obligate heterodimers to bind DNA and turn on petal specific genes that include a number of transcription factors involved in conical cell differentiation, multicellular trichome formation often accompanied by pigment accumulation, while turning off photosynthetic genes (Goto and Meyerowitz, 1994; Jack et al., 1994; Zachgo et al., 1995; Martin et al., 2002; Mara and Irish, 2008). The heterodimerization and autorregulation of the AP3-PI dimer detected in core eudicot model species have shown to be conserved in monocots, basal eudicots and basal angiosperms, suggesting that the AP3-PI interaction is a central hub in the petal-stamen identity programs in flowering plants (Kanno et al., 2003; Melzer et al., 2014; Hsu et al., 2015). In addition to this interaction, homodimerization has been detected independently for AP3/DEF proteins as well as for PI/GLO proteins; the occurrence of such interactions in gymnosperms has been proposed as a prerequisite to the heterodimerization in angiosperms; however, its biological significance is still poorly understood (Winter et al., 2002a,b; Kanno et al., 2003; Whipple et al., 2004; Melzer et al., 2014).

Although the expression of AP3 and PI as well as the activation of the AP3-PI heterodimer is for the most part restricted to petals and stamens, it occurs also exceptionally in the first floral whorl in a number of monocot species having petaloid sepals, such as in *Agapanthus*, *Lilium*, and *Tulipa* (van Tunen et al., 1993; Tzeng and Yang, 2001; Kanno et al., 2003; Nakamura et al., 2005). The ectopic expression of the petal/stamen genetic module has been used to explain the petaloid nature of the first whorl in such species in what has been named the sliding borders model, a modified ABC model of flower development (Kanno et al., 2003, 2007). However, as petaloid sepals also occur in the absence of AP3-PI coordinated expression in other monocots, like *Asparagus* (Asparagaceae; Park et al., 2003) and *Habenaria* (Orchidaceae; Kim et al., 2007), in basal eudicots like *Aquilegia* (Ranunculaceae; Kramer et al., 2007; Sharma et al., 2011; Sharma and Kramer, 2012), and in a number of core eudicots like *Gerbera* (Asteraceae), *Impatiens* (Balsaminaceae) and *Rhodochiton* (Scrophulariaceae; Geuten et al., 2006; Broholm et al., 2010; Landis et al., 2012), it is likely that petaloid features including pigment and structural color as well as epidermal modifications can be turned on independently of AP3-PI (Landis et al., 2012; Weiss, 2000).

In Aristolochiaceae, AP3 and PI expression are restricted to petals and stamens in *Saruma* and by comparison, in *A. fimbriata*, as well as in *A. manshuriensis* and *A. arborea*, AP3 and PI only overlap in expression patterns in outer portion of the gynostemium, suggesting that the AP3-PI interaction is critical in stamen identity, but does not play any role in perianth identity (Figures 9 and 10; Jaramillo and Kramer, 2004; Horn et al., 2014). This is in accordance with the interpretation of the gynostemium as a fused structure between stamens and the upper portion of the carpels provided by González and Stevenson (2000a). Nevertheless, the genetic bases of the proximal-distal differentiation of the gynoeceum-derived tissue forming the different portions of the gynostemium will require full examination of the stamen-carpel identity C-class genes as well as the carpel zonation genes responsible for stigmatic and transmitting tissue that include *SPATULA*, *HECATE3*, *CRABS CLAW*, and *NGATHA* (Alvarez and Smyth, 2002; Gremski et al., 2007; Fourquin and Ferrándiz, 2014; Schuster et al., 2015).

Interestingly, in *Aristolochia manshuriensis*, *A. arborea*, and *A. fimbriata* PI expression correlates with the occurrence of conical cellular differentiation and pigment accumulation in the inner epidermis of the perianth (Figures 4 and 9; Jaramillo and Kramer, 2004; Horn et al., 2014). Such epidermal specialization that often includes the formation of multicellular trichomes and osmophores is a unique trait of the diploid species of subgenus *Aristolochia*, to which *A. fimbriata* belongs to, and is lacking from the polyploid species prevalent in subgenus *Siphisia*, to which *A. arborea* and *A. manshuriensis* belong (González and Stevenson, 2000b). These observations generate testable hypothesis about a putative role of PI homodimers that together with other MADS-box perianth expressed genes like *FUL* and *AGL6* are able to effectively activate the genetic circuitry responsible for cellular specialization, pigment accumulation, and nectarial secretion in the inner sepal epidermis.

AUTHOR CONTRIBUTIONS

NP-M and FG designed the study, NP-M, HS-B, BAA, and FG acquired, analyzed, and interpreted the data, NP-M, BAA, and FG wrote the manuscript, and all authors revised and approved the final version.

ACKNOWLEDGMENTS

We thank J.F. Alzate (Centro Nacional de Secuenciación de Genómica, SIU, Universidad de Antioquia, Medellín, Antioquia) for the assembly and storage of the *A. fimbriata* transcriptome.

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control* 19, 716–723. doi: 10.1109/TAC.1974.1100705
- Almeida, A. M. R., Brown, A., and Specht, C. D. (2013). Tracking the development of the petaloid fertile stamen in *Canna indica*: insights into the origin of androecial petaloidy in the Zingiberales. *AoB Plants* 5:pl009. doi: 10.1093/aobpla/pl009
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Alvarez, J., and Smyth, D. R. (2002). CRABS CLAW and SPATULA genes regulate growth and pattern formation during gynoecium development in *Arabidopsis thaliana*. *Int. J. Plant Sci.* 163, 17–41. doi: 10.1086/324178
- Alvarez Buylia, E. R., Ambrose, B. A., Flores-Sandoval, E., Englund, M., Garay-Arroyo, A., García-Ponce, B., et al. (2010). B-function expression in the flower center underlies the homeotic phenotype of *Lacandonia schismatica* (Triuridaceae). *Plant Cell* 22, 3543–3559. doi: 10.1105/tpc.109.069153
- Amborella Genome Project (2013). The *Amborella* genome and the evolution of flowering plants. *Science* 342, 1467–1477. doi: 10.1126/science.1241089
- Ambrose, B. A., Lerner, D. R., Ciceri, P., Padilla, C. M., Yanofsky, M. F., and Schmidt, R. J. (2000). Molecular and genetic analyses of the *silky1* gene reveal conservation in floral organ specification between eudicots and monocots. *Mol. Cell* 5, 569–579. doi: 10.1016/S1097-2765(00)80450-5
- APG (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants, APG III. *Bot. J. Linnean Soc.* 161, 105–121. doi: 10.1111/j.1095-8339.2009.00996.x
- Becker, A., and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* 29, 464–489. doi: 10.1016/S1055-7903(03)00207-0
- Benlloch, R., d'Erfurth, I., Ferrandiz, C., Cosson, V., Beltrán, J. P., Cañas, L. A., et al. (2006). Isolation of *mtpim* proves *Tnt1* a useful reverse genetics tool in *Medicago truncatula* and uncovers new aspects of AP1-like functions in legumes. *Plant Physiol.* 142, 972–983. doi: 10.1104/pp.106.083543
- Berbel, A., Ferrandiz, C., Hecht, V., Dalmais, M., Lund, O. S., Sussmilch, F. C., et al. (2012). VEGETATIVE1 is essential for development of the compound inflorescence in pea. *Nat. Commun.* 3:797. doi: 10.1038/ncomms1801
- Berbel, A., Navarro, C., Ferrandiz, C., Cañas, L. A., Madueño, F., and Beltrán, J. P. (2001). Analysis of PEAM4, the pea AP1 functional homologue, supports a model for AP1-like genes controlling both floral meristem and floral organ identity in different plant species. *Plant J.* 25, 441–451. doi: 10.1046/j.1365-313x.2001.00974.x
- Bliss, B. J., Wanke, S., Bakarar, A., Ayyampalayam, S., Wickett, N., Wall, P. K., et al. (2013). Characterization of the basal angiosperm *Aristolochia fimbriata*: a potential experimental system for genetic studies. *BMC Plant Biol.* 13:13. doi: 10.1186/1471-2229-13-13
- Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1989). Genes directing flower development in *Arabidopsis*. *Plant Cell* 1, 37–52. doi: 10.1105/tpc.1.1.37

We thank D.W. Stevenson and L.M. Campbell for allowing us to use the Structural Botany laboratory at the New York Botanical Garden. This work was funded by the Committee for Research Development (CODI), Convocatoria Programática Ciencias Exactas y Naturales 2013–2014, and the Estrategia de Sostenibilidad 2013–2014 at the Universidad de Antioquia (Medellín, Colombia).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.01095>

- Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1991). Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* 112, 1–20.
- Broholm, S. K., Pollanen, E., Ruokolinen, S., Tahtiharju, S., Kotilainen, M., Albert, V. A., et al. (2010). Functional characterization of B-class MADS-box transcription in *Gerbera hybrida*. *J. Exp. Bot.* 61, 75–85. doi: 10.1093/jxb/erp279
- Carlsbecker, A., Sundstrom, J. F., Englund, M., Uddenberg, D., Izquierdo, L., Kvarnheden, A., et al. (2013). Molecular control of normal and acrocona mutant seed cone development in Norway spruce (*Picea abies*) and the evolution of conifer ovule-bearing organs. *New Phytol.* 200, 261–275. doi: 10.1111/nph.12360
- Castillejo, C., Romera-Branchat, M., and Pelaz, S. (2005). A new role of the *Arabidopsis* SEPALLATA3 gene revealed by its constitutive expression. *Plant J.* 43, 586–596. doi: 10.1111/j.1365-313X.2005.02476.x
- Causier, B., Schwarz-Sommer, Z., and Davies, B. (2010). Floral organ identity: 20 years of ABCs. *Sem. Cell Dev. Biol.* 21, 73–79. doi: 10.1016/j.semdb.2009.10.005
- Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0
- Cui, L., Wall, P. K., Leebens-Mack, J., Lindsay, B. G., Soltis, D. E., and Doyle, J. J. (2006). Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16, 738–749. doi: 10.1101/gr.4825606
- deMartino, G., Pan, I., Emmanuel, E., Levy, A., and Irish, V. F. (2006). Functional analyses of two tomato APETALA3 genes demonstrate diversification in their roles in regulating floral development. *Plant Cell* 18, 1833–1845. doi: 10.1105/tpc.106.042978
- Dihn, T. T., Girke, T., Liu, X., Yant, L., Schmid, M., and Chen, X. (2012). The floral homeotic protein APETALA2 recognizes and acts through an AT-rich sequence element. *Development* 139, 1978–1986. doi: 10.1242/dev.077073
- Drea, S., Hileman, L. C., de Martino, G., and Irish, V. F. (2007). Functional analyses of genetic pathways controlling petal specification in poppy. *Development* 134, 4157–4166. doi: 10.1016/j.jpb.2006.11.007
- Endress, P. K., and Doyle, J. A. (2007). Floral phyllotaxis in basal angiosperms: development and evolution. *Curr. Opin. Plant Biol.* 10, 52–57. doi: 10.1111/j.1469-8137.2012.04255.x
- Feng, C.-M., Liu, X., Yu, Y., Xie, D., Franks, R., and Xiang, Q.-Y. (2012). Evolution of bract development and B-class MADS box gene expression in petaloid bracts of *Cornus* s.l. (*Cornaceae*). *New Phytol.* 196, 631–643.
- Ferrándiz, C., Gu, Q., Martienssen, R., and Yanofsky, M. F. (2000). Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development* 127, 725–734. doi: 10.1105/tpc.010280
- Ferrario, S., Immink, R. G. H., Shchennikova, A., Busscher-Lange, J., and Angenent, G. C. (2003). The MADS-box gene FBP2 is required for SEPALLATA function in *Petunia*. *Plant Cell* 15, 914–925. doi: 10.1111/nph.12703
- Fourquin, C., and Ferrándiz, C. (2014). The essential role of NGATHA genes in style and stigma specification is widely conserved across eudicots. *New Phytol.* 202, 1001–1013. doi: 10.1111/j.1365-313X.2006.02800.x

- Geuten, K., Becker, A., Kaufmann, K., Caris, P., Janssens S., Viaene, T., et al. (2006). Petaloidy and petal identity MADS-box genes in the balsaminoid genera *Impatiens* and *Marcgravia*. *Plant J.* 47, 501–518.
- González, F., Ospina, J. C., and Zanotti, C. (2015). Sinopsis y novedades taxonómicas de la familia Aristolochiaceae para la Argentina. *Darwiniana* 3, 38–64. doi: 10.1111/nph.13328
- González, F., and Pabón-Mora, N. (2015). Trickery flowers: the extraordinary chemical mimicry of *Aristolochia* to accomplish deception to its pollinators. *New Phytol.* 206, 10–13.
- González, F., and Stevenson, D. W. (2000a). Gynostemium development in *Aristolochia* (Aristolochiaceae). *Bot. Jahrb. Syst.* 122, 249–291.
- González, F., and Stevenson, D. W. (2000b). Perianth development and systematics of *Aristolochia*. *Flora* 195, 370–391. doi: 10.1101/gad.8.13.1548
- Goto, K., and Meyerowitz, E. M. (1994). Function and regulation of the *Arabidopsis* floral homeotic gene PISTILLATA. *Genes Dev.* 8, 1548–1560. doi: 10.1242/dev.011510
- Gremski, K., Ditta, G., and Yanofsky, M. F. (2007). The HECATE genes regulate female reproductive tract development in *Arabidopsis thaliana*. *Development* 134, 3593–3601.
- Gu, Q., Ferrandiz, C., Yanofsky, M. F., and Martienssen, R. (1998). The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125, 1509–1517.
- Hemgingway, C. A., Christensen, A. R., and Malcomber, S. T. (2011). B- and C- class gene expression during corona development of the blue passionflower (*Passiflora caerulea*, Passifloraceae). *Am. J. Bot.* 98, 923–934. doi: 10.3732/ajb.1100026
- Honma, T., and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525–529. doi: 10.1038/35054083
- Horn, S., Pabón-Mora, N., Theuß, V. S., Busch, A., and Zachgo, S. (2014). Analysis of the CYC/TB1 class of TCP transcription factors in basal angiosperms and magnoliids. *Plant J.* 81, 559–571. doi: 10.1111/tpj.12750
- Hsu, H.-F., Hsu, W.-H., Lee, Y.-I., Mao, W.-T., Yang, J.-Y., Li, J.-Y., et al. (2015). Model for perianth formation in orchids. *Nat. Plants* 1:15046. doi: 10.1038/nplants.2015.46
- Huijser, P., Klein, J., Lönig, W. E., Meijer, H., Saedler, H., and Sommer, H. (1992). Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene squamosa in *Antirrhinum majus*. *EMBO J.* 11, 1239–1249.
- Immink, R. G. H., Hannapel, D. J., Ferrario, S., Busscher, M., Franken, J., Lookeren Campagne, M. M., et al. (1999). A petunia MADS box gene involved in the transition from vegetative to reproductive development. *Development* 126, 5117–5126.
- Jack, T., Brockman, L. L., and Meyerowitz, E. M. (1992). The homeotic gene APETALA3 of *Arabidopsis thaliana* encodes a MADS-box and is expressed in petals and stamens. *Cell* 68, 683–697. doi: 10.1016/0092-8674(92)90144-2
- Jack, T., Fox, G. L., and Meyerowitz, E. M. (1994). *Arabidopsis* homeotic gene APETALA3 ectopic expression: transcriptional and posttranscriptional regulation determine floral organ identity. *Cell* 76, 703–716. doi: 10.1016/0092-8674(94)90509-6
- Jaramillo, M. A., and Kramer, E. M. (2004). APETALA3 and PISTILLATA homologs exhibit novel expression patterns in the unique perianth of *Aristolochia* (Aristolochiaceae). *Evol. Dev.* 6, 449–458. doi: 10.1111/j.1525-142X.2004.04053.x
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916
- Kanno, A., Nakada, M., Akita, Y., and Hirai, M. (2007). Class B gene expression and the modified ABC model in nongrass monocots. *Sci. World J.* 7, 268–279. doi: 10.1100/tsw.2007.86
- Kanno, A., Saeki, H., Kameya, T., Saedler, H., and Theissen, G. (2003). Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). *Plant Mol. Biol.* 52, 831–841. doi: 10.1023/A:1025070827979
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kempin, S. A., Savidge, B., and Yanofsky, M. F. (1995). Molecular basis of the cauliflower phenotype in *Arabidopsis*. *Science* 267, 522–525. doi: 10.1126/science.7824951
- Kim, S., Koh, J., Yoo, M.-J., Kong, H., Hu, Y., Ma, H., et al. (2005). Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. *Plant J.* 43, 724–744. doi: 10.1111/j.1365-313X.2005.02487.x
- Kim, S., Soltis, P. S., and Soltis, D. E. (2013). AGL6-like MADS-box genes are sister to AGL2-like MADS-box genes. *J. Plant Biol.* 56, 315–325. doi: 10.1007/s12374-013-0147-x
- Kim, S., Yun, P., Fukuda, T., Ochiai, T., Yokoyama, J., Cameya, T., et al. (2007). Expression of a DEFICIENS-like gene correlates with the differentiation between sepal and petal in the orchid *Habenaria radiata* (Orchidaceae). *Plant Sci.* 172, 319–326. doi: 10.1016/j.plantsci.2006.09.009
- Koo, S. C., Bracko, O., Park, M. S., Schwab, R., Chun, H. J., Park, K. M., et al. (2010). Control of lateral organ development and flowering time by the *Arabidopsis thaliana* MADS-box gene AGAMOUS-like6. *Plant J.* 62, 807–816. doi: 10.1111/j.1365-313X.2010.04192.x
- Kramer, E. M., Dorit, R. L., and Irish, V. F. (1998). Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages. *Genetics* 149, 765–783.
- Kramer, E. M., Holappa, L., Gould, B., Jaramillo, M. A., Setnikov, D., and Santiago, P. (2007). Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia* (Ranunculaceae). *Plant Cell* 19, 750–766. doi: 10.1105/tpc.107.050385
- Landis, J. B., Barnett, L. L., and Hileman, L. C. (2012). Evolution of petaloid sepals independent of shifts in B-class MADS-box gene expression. *Dev. Genes Evol.* 222, 19–28. doi: 10.1007/s00427-011-0385-1
- Li, H., Liang, W., Jia, R., Yin, C., Zong, J., Kong, H., et al. (2010). The AGL6-like gene OsMADS6 regulates floral organ and meristem identities in rice. *Cell Res.* 20, 299–313. doi: 10.1038/cr.2009.143
- Li, L., Yu, X.-X., Guo, C.-C., Duan, X.-S., Shan, H.-Y., Zhang, R., et al. (2015). Interactions among proteins of floral MADS-box genes in *Nuphar pumila* (Nymphaeaceae) and the most recent common ancestor of extant angiosperms help understand the underlying mechanisms of the origin of the flower. *J. Sys. Evol.* 53, 285–296. doi: 10.1111/jse.12148
- Litt, A. (2007). An evaluation of A-function: evidence from the APETALA1 and APETALA2 gene lineages. *Int. J. Plant Sci.* 168, 73–91. doi: 10.1086/509662
- Litt, A., and Irish, V. F. (2003). Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* 165, 821–833.
- Litt, A., and Kramer, E. M. (2010). The ABC model and the diversification of floral organ identity. *Sem. Cell Dev. Biol.* 21, 129–137. doi: 10.1016/j.semcdb.2009.11.019
- Liu, Y., Nakayama, N., Schiff, M., Litt, A., Irish, V. F., and Dinesh-Kumar, S. P. (2004). Virus induced gene silencing of a DEFICIENS ortholog in *Nicotiana benthamiana*. *Plant Mol. Biol.* 54, 701–711. doi: 10.1023/B:PLAN.0000040899.53378.83
- Mandel, M. A., Gustafson-Brown, C., Savidge, B., and Yanofsky, M. F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene APETALA1. *Nature* 360, 273–277. doi: 10.1038/360273a0
- Mara, C. D., and Irish, V. F. (2008). Two GATA transcription factors are downstream effectors of floral homeotic gene action in *Arabidopsis*. *Plant Physiol.* 147, 707–718. doi: 10.1104/pp.107.115634
- Martin, C., Bhatt, K., Baumann, K., Jin, H., Zachgo, S., Roberts, K., et al. (2002). The mechanics of cell fate determination in petals. *Philos. Trans. R Soc. Lond. B.* 357, 809–813. doi: 10.1098/rstb.2002.1089
- Melzer, R., Härter, A., Rümpler, F., Kim, S., Soltis, P. S., Soltis, D. E., et al. (2014). DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperms evolution. *Ann. Bot.* 114, 1431–1443. doi: 10.1093/aob/mcu094
- Miller, M. A., Holder, H. T., Vos, R., Midford, P. E., Liebowitz, T., Chan, L., et al. (2009). *The CIPRES Portals*. Available at: <http://www.phylo.org>
- Mondragón-Palomino, M., and Theissen, G. (2011). Conserved differential expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the flowers of Orchidaceae: refining the “orchid code”. *Plant J.* 66, 1008–1019. doi: 10.1111/j.1365-313X.2011.04560.x
- Müller, B. M., Saedler, H., and Zachgo, S. (2001). The MADS-box gene DEFH28 from *Antirrhinum* is involved in the regulation of floral meristem identity and fruit development. *Plant J.* 28, 169–179. doi: 10.1046/j.1365-313X.2001.01139.x

- Murai, K., Miyamae, M., Kato, H., Takumi, S., and Ogihara, Y. (2003). WAP1, a wheat APETALA1 homolog, plays a central role in the phase transition from vegetative to reproductive growth. *Plant Cell Physiol.* 44, 1255–1265. doi: 10.1093/pcp/pcg171
- Nakamura, T., Fukuda, T., Nakano, M., Hasebe, M., Cameya, T., and Kanno, A. (2005). The modified ABC model explains the development of the petaloid perianth of *Agapanthus praecox* ssp. *orientalis* (Agapanthaceae) flowers. *Plant Mol. Biol.* 58, 435–445. doi: 10.1007/s11103-005-5218-z
- Nixon, K. C. (2002). *Win Clada*. Ithaca, NY: Nixon Publisher.
- Ohmori, S., Kimizu, M., Sugita, M., Miyao, A., Hirochika, H., Uchida, E., et al. (2009). MOSAIC FLORAL ORGANS1, an AGL6-like MADS box gene, regulates floral organ identity and meristem fate in rice. *Plant Cell* 21, 3008–3025. doi: 10.1105/tpc.109.068742
- Pabón-Mora, N., Ambrose, B. A., and Litt, A. (2012). Poppy APETALA1/FRUITFULL orthologs control flowering time, branching, perianth identity, and fruit development. *Plant Physiol.* 158, 1685–1704. doi: 10.1104/pp.111.192104
- Pabón-Mora, N., Sharma, B., Holappa, L., Kramer, E. M., and Litt, A. (2013). The *Aquilegia* FRUITFULL-like genes play key roles in leaf morphogenesis and inflorescence development. *Plant J.* 74, 197–212. doi: 10.1111/tjp.12113
- Pan, Z. J., Chen, Y.-Y., Du, J.-S., Chen, Y.-Y., Chung, M.-C., Tsai, W.-C., et al. (2014). Flower development of *Phalaenopsis* orchid involves functionally divergent SEPALLATA-like genes. *New Phytol.* 202, 1024–1042. doi: 10.1111/nph.12723
- Park, J., Ishikawa, Y., Kanno, A., and Kameya, T. (2003). Expression of AoDEF, a B-functional MADS-box gene in stamens and inner tepals of the dioecious species *Asparagus officinalis* L. *Plant Mol. Biol.* 51, 867–875. doi: 10.1023/A:1023097202885
- Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405, 200–203. doi: 10.1038/35012103
- Pinyopich, A., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. (2003). Unraveling the redundant roles of MADS-box genes during carpel and fruit development. *Nature* 424, 85–88. doi: 10.1038/nature01741
- Posada, D., and Crandall, K. A. (1998). MODELTEST, testing the model of DNA substitution. *Bioinformatics* 14, 817–818. doi: 10.1093/bioinformatics/14.9.817
- Rambaut, A. (2014). *FigTree: Tree Figure Drawing Tool, Version 1.4.2*. Institute of Evolutionary Biology, University of Edinburgh. Available at: <http://tree.bio.ed.ac.uk/software/figtree/> [accessed April 1, 2015].
- Reinheimer, R., and Kellogg, E. A. (2009). Evolution of AGL6-like MADS-box genes in grasses (Poaceae): ovule expression is ancient and palea expression is new. *Plant Cell* 21, 2529–2605. doi: 10.1105/tpc.109.068239
- Rijkema, A. S., Royaert, S., Zethof, J., van der Weerden, G., Gerats, T., and Vandenbussche, M. (2006). Functional divergence within the DEF/AP3 lineage: an analysis of PHTM6 in *Petunia hybrida*. *Plant Cell* 18, 1819–1832. doi: 10.1105/tpc.106.042937
- Rijkema, A. S., Zethof, J., Gerats, T., and Vandenbussche, M. (2009). The petunia AGL6 gene has a SEPALLATA-like function in floral patterning. *Plant J.* 60, 1–9. doi: 10.1111/j.1365-313X.2009.03917.x
- Schauer, S. E., Schlüter, P. M., Baskar, R., Gheyselinck, J., Bolaños, A., Curtis, M. D., et al. (2009). Intronic regulatory elements determine the divergent expression patterns of AGAMOUS-like6 subfamily members in *Arabidopsis*. *Plant J.* 59, 987–1000. doi: 10.1111/j.1365-313X.2009.03928.x
- Schuster, C., Gailloch, C., and Lohmann, J. U. (2015). *Arabidopsis* HECATE genes function in phytohormone control during gynoecium development. *Development* 142, 3343–3350. doi: 10.1242/dev.120444
- Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development in *Antirrhinum majus*. *Science* 250, 931–936. doi: 10.1126/science.250.4983.931
- Sharma, B., Guo, C., Kong, H., and Kramer, E. M. (2011). Petal-specific sub-functionalization of an APETALA3 paralog in the Ranunculales and its implications for petal evolution. *New Phytol.* 191, 870–883. doi: 10.1111/j.1469-8137.2011.03744.x
- Sharma, B., and Kramer, E. M. (2012). Sub and neo-functionalization of APETALA3 paralogs have contributed to the evolution of novel floral organ identity in *Aquilegia* (columbine, Ranunculaceae). *New Phytol.* 197, 949–957. doi: 10.1111/nph.12078
- Sommer, H., Beltran, J.-P., Huijser, P., Pape, H., Lönng, W.-E., Saedler, H., et al. (1990). Deficiens, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J.* 9, 605–613.
- Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A fast bootstrapping algorithm for the RAXML web-servers. *Syst. Biol.* 57, 758–771. doi: 10.1080/10635150802429642
- Sun, W., Huang, W., Li, Z., Song, C., Liu, D., Liu, Y., et al. (2014). Functional and evolutionary analysis of the AP1/SEP/AGL6 superclade of MADS-box genes in the basal eudicot *Epimedium sagittatum*. *Ann. Bot.* 113, 653–668. doi: 10.1093/aob/mct301
- Thompson, B. E., Bartling, L., Whipple, C., Hall, D. H., Sakai, H., Schmidt, R., et al. (2009). Bearded-ear encodes a MADS-box transcription factor critical for maize floral development. *Plant Cell* 21, 2578–2590. doi: 10.1105/tpc.109.067751
- Tröbner, W., Ramirez, L., Motte, P., Hue, I., Huijser, P., Lönng, W.-E., et al. (1992). GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of *Antirrhinum* floral organogenesis. *EMBO J.* 11, 4693–4704.
- Tzeng, T.-Y., and Yang C.-H. (2001). A MADS-box gene from lily (*Lilium longiflorum*) is sufficient to generate dominant negative mutation by interacting with PISTILLATA (PI) in *Arabidopsis thaliana*. *Plant Cell Physiol.* 42, 1156–1168. doi: 10.1093/pcp/pce151
- Vandenbussche, M., Zethof, J., Royaert, S., Weterings, K., and Gerats, T. (2004). The duplicated B-class heterodimer model: whorl specific effects and complex genetic interactions in *Petunia hybrida* flower development. *Plant Cell* 16, 741–754. doi: 10.1105/tpc.019166
- van Tunen, A. J., Eikelboom, W., and Angenent, G. C. (1993). Floral organogenesis in *Tulipa*. *FNL* 16, 33–38.
- Viaene, T., Vekemans, D., Becker, A., Melzer, S., and Geuten, K. (2010). Expression divergence of the AGL6 MADS domain transcription factor lineage after a core eudicot duplication suggests functional diversification. *BMC Plant Biol.* 10:148. doi: 10.1186/1471-2229-10-148
- Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., et al. (2002). A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. *Science* 296, 343–346. doi: 10.1126/science.1068181
- Waters, M. T., Tiley, A. M. M., Kramer, E. M., Meerow, A. W., Langdale, J. A., and Scotland, R. W. (2013). The corona of the daffodil *Narcissus bulbocodium* shares stamen-like identity and is distinct from the orthodox floral whorls. *Plant J.* 74, 615–625. doi: 10.1111/tjp.12150
- Weiss, D. (2000). Regulation of flower pigmentation and growth: multiple signaling pathways control anthocyanin synthesis in expanding petals. *Physiol. Plant.* 100, 152–157. doi: 10.1034/j.1399-3054.2000.110202.x
- Whipple, C., Ciceri, P., Padilla, C. M., Ambrose, B. A., Bandong, S. L., and Schmidt, R. J. (2004). Conservation of B-class floral homeotic gene function between maize and *Arabidopsis*. *Development* 131, 6083–6091. doi: 10.1242/dev.01523
- Whipple, C. J., Zanis, M. J., Kellogg, E. A., and Schmidt, R. J. (2007). Conservation of B class gene expression in the second whorl of a basal grass and outgroups links the origin of lodicules and petals. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1081–1086. doi: 10.1073/pnas.0606434104
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., and Carpenter, E. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4859–E4868. doi: 10.1073/pnas.1323926111
- Winter, K. U., Saedler, H., and Theissen, G. (2002a). 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*. *Plant J.* 31, 457–475. doi: 10.1046/j.1365-313X.2002.01375.x
- Winter, K. U., Weiser, C., Kaufmann, K., Bohne, A., Kirchner, C., Kannom, A., et al. (2002b). Evolution of class B floral homeotic proteins: obligate heterodimerization originated from homodimerization. *Mol. Biol. Evol.* 19, 587–596. doi: 10.1093/oxfordjournals.molbev.a004118
- Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A., and Meyerowitz, E. M. (1990). The protein encoded by the *Arabidopsis* homeotic gene AGAMOUS resembles transcription factors. *Nature* 346, 35–39. doi: 10.1038/346035a0
- Yoo, M., Soltis, P. S., and Soltis, D. E. (2010). Expression of floral MADS-box genes in two divergent “water lilies”: Nymphaeales and *Nelumbo*. *Int. J. Plant Sci.* 171, 121–146. doi: 10.1086/648986
- Yu, H., Ito, T., Wellmer, F., and Meyerowitz, E. M. (2004). Repression of AGAMOUS-like24 is a crucial step in promoting flower development. *Nat. Genet.* 36, 157–161. doi: 10.1038/ng1286

- Zachgo, S., Motte, P., Trobner, W., Saedler, H., Saedler, H., and Schwarz-Sommer, Z. (1995). Functional analysis of the *Antirrhinum floral* homeotic *Deficiens* gene in vivo and in vitro by using a temperature-sensitive mutant. *Development* 121, 2861–2875.
- Zahn, L. M., Kong, H., Leebens-Mack, J. H., Kim, S., Soltis, P. S., Landherr, L. L., et al. (2005). The evolution of the SEPALLATA subfamily of MADS-box genes a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* 169, 2209–2223. doi: 10.1534/genetics.104.037770
- Zeng, L., Zhang, Q., Sun, R., Kong, H., Zhang, N., and Ma, H. (2014). Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat. Commun.* 5:4956. doi: 10.1038/ncomms5956
- Zhang, R., Guo, C., Zhang, W., Wang, P., Li, L., Duan, X., et al. (2013). Disruption of the petal identity gene *APETALA3-3* is highly correlated with loss of petals within the buttercup family (Ranunculaceae). *Proc. Natl. Acad. Sci. U.S.A.* 110, 5074–5079. doi: 10.1073/pnas.1219690110
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Pabón-Mora, Suárez-Baron, Ambrose and González. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Role of TDIF Peptide Signaling in Vascular Cell Differentiation is Conserved Among Euphyllophytes

Yuki Hirakawa^{1,2*} and John L. Bowman^{1,3*}

¹ School of Biological Sciences, Monash University, Melbourne, VIC, Australia, ² Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Nagoya, Japan, ³ Section of Plant Biology, University of California, Davis, Davis, CA, USA

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington, USA

Reviewed by:

Mary Byrne,
The University of Sydney, Australia
Barbara Ambrose,
The New York Botanical Garden, USA

*Correspondence:

Yuki Hirakawa
yuki.hirakawa@itbm.nagoya-u.ac.jp;
John L. Bowman
john.bowman@monash.edu

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 03 August 2015

Accepted: 09 November 2015

Published: 26 November 2015

Citation:

Hirakawa Y and Bowman JL (2015) A
Role of TDIF Peptide Signaling in
Vascular Cell Differentiation is
Conserved Among Euphyllophytes.
Front. Plant Sci. 6:1048.
doi: 10.3389/fpls.2015.01048

Peptide signals mediate a variety of cell-to-cell communication crucial for plant growth and development. During *Arabidopsis thaliana* vascular development, a CLE (CLAVATA3/EMBRYO SURROUNDING REGION-related) family peptide hormone, TDIF (tracheary element differentiation inhibitory factor), regulates procambial cell fate by its inhibitory activity on xylem differentiation. To address if this activity is conserved among vascular plants, we performed comparative analyses of TDIF signaling in non-flowering vascular plants (gymnosperms, ferns and lycophytes). We identified orthologs of TDIF/CLE as well as its receptor TDR/PXY (TDIF RECEPTOR/PHLOEM INTERCALATED WITH XYLEM) in *Ginkgo biloba*, *Adiantum aethiopicum*, and *Selaginella kraussiana* by RACE-PCR. The predicted TDIF peptide sequences in seed plants and ferns were identical to that of *A. thaliana* TDIF. We examined the effects of exogenous CLE peptide-motif sequences of TDIF in these species. We found that liquid culturing of dissected leaves or shoots was useful for examining TDIF activity during vascular development. TDIF treatment suppressed xylem/tracheary element differentiation of procambial cells in *G. biloba* and *A. aethiopicum* leaves. In contrast, neither TDIF nor putative endogenous TDIF inhibited xylem differentiation in developing shoots and rhizophores of *S. kraussiana*. These data suggest that activity of TDIF in vascular development is conserved among extant euphyllophytes. In addition to the conserved function, via liquid culturing of its bulbils, we found a novel inhibitory activity on root growth in the fern *Asplenium × lucrosum* suggesting lineage-specific co-option of peptide signaling occurred during the evolution of vascular plant organs.

Keywords: CLE peptides, plant evo-devo, LRR-RLKs, plant vascular development, vascular plants, non-model organism

INTRODUCTION

Recent advances in biochemical, genetic and bioinformatic analyses have unveiled the importance of peptide hormones in plant growth and development (Matsubayashi, 2014). CLE (CLAVATA3/EMBRYO SURROUNDING REGION-related) peptides are a class of peptide hormones involved in an array of plant developmental processes including shoot

apical meristem maintenance, vascular cell differentiation and stem cell maintenance, root meristem maintenance, development of embryo and endosperm, autoregulation of nodulation, lateral root responses to nutrient conditions and pollen viability (Fletcher et al., 1999; Brand et al., 2000; Hirakawa et al., 2008; Okamoto et al., 2009; Kondo et al., 2011; Fiume and Fletcher, 2012; Depuydt et al., 2013; Endo et al., 2013; Araya et al., 2014). Typical CLE proteins contain an N-terminal signal peptide and a CLE peptide motif near the C-terminus which are intervened by non-conserved variable region. The mature signaling peptides are produced from the CLE peptide motif as 12–13 amino acid peptides containing proline hydroxylation and glycosylation via post-translational processing (Ito et al., 2006; Kondo et al., 2006; Ohyama et al., 2009; Ogawa-Ohnishi et al., 2013). Secreted CLE peptides are perceived by receptors residing in target cell membranes to mediate intercellular signaling. Based on the bioactivity and receptor specificity, two major subgroups can be recognized in the CLE peptide family: here we call R-type CLE and H-type CLE (Ito et al., 2006; Strabala et al., 2006; Kinoshita et al., 2007; Whitford et al., 2008; Ohyama et al., 2009). Each has a characteristic amino acid residue (arginine or histidine) at the N-terminus of the peptide. The R-type CLE includes CLV3 (CLAVATA3), which plays a significant role in the maintenance of shoot apical meristem in *Arabidopsis thaliana*, while the H-type CLE includes TDIF, an important regulator of vascular cell differentiation (Brand et al., 2000; Schoof et al., 2000; Ito et al., 2006; Hirakawa et al., 2008). Peptides in each subgroup are perceived through specific receptors of LRR-RLK family, CLV1 (CLAVATA1)/BAM (BARELY ANY MEISTEM) or TDR/PXY (TDIF RECEPTOR/PHLOEM INTERCALATED WITH XYLEM; Clark et al., 1995; Brand et al., 2000; DeYoung et al., 2006; Fisher and Turner, 2007; Hirakawa et al., 2008; Ogawa et al., 2008; Shinohara et al., 2012). The TDIF-TDR pair mediates a phloem-derived signal that inhibits differentiation of procambial cells into xylem cells, which is important during secondary growth of vasculature in *A. thaliana* floral stems (Hirakawa et al., 2008, 2010; Whitford et al., 2008; Etchells and Turner, 2010).

The CLE family is conserved throughout land plants although functional paralogs are not precisely characterized except in angiosperms. Similar to many other gene families of developmental regulators, the number of genes seems lower in early diverging taxa such as the bryophytes and lycophytes (1 and 15 sequences are reported for *Physcomitrella patens* and *Selaginella moellendorffii*, respectively), compared to the number found in flowering plant species such as *A. thaliana*, which possesses 32 CLE genes (Jun et al., 2008; Oelkers et al., 2008; Miwa et al., 2009). Thus, the expansion of the CLE gene family and subsequent neofunctionalization may have played important roles in the evolution of land plant development, particularly in the vascular plant lineages.

In this study, we performed evolutionary and functional comparative analyses of TDIF/H-type CLE peptides among major lineages in vascular plants—angiosperms, gymnosperms, ferns and lycophytes.

MATERIALS AND METHODS

Database Search for Orthologs of TDIF and TDR Genes

Nucleotide or protein sequences corresponding to the CLE peptide motif of *A. thaliana* CLE41/At3g24770 (His⁸⁷ to Asn⁹⁹), the CLE peptide motif of *P. patens* CLE1/CLE170/XM_001752838 (Arg¹³⁶ to Asn¹⁴⁷) and the kinase domain of *A. thaliana* TDR/At5g61480 (Gly⁷²⁶ to Leu⁹⁹⁷) were used as queries for database searches. BLAST searches were performed against the SRA (Sequence Read Archive) and oneKP (one thousand plants, <http://www.onekp.com/>) databases, focusing on EST data for gymnosperms, ferns and lycophytes, as well as Genbank transcript data (de Vries et al., 2015; Vanneste et al., 2015). Each of the obtained sequences was manually validated to determine whether it encodes a complete protein containing an N-terminal signal peptide by SignalP (<http://www.cbs.dtu.dk/services/SignalP/>).

RNA Extraction and cDNA Synthesis

Total RNA was extracted from immature leaves/fronds of *Ginkgo biloba*, *Adiantum aethiopicum*, and *Selaginella kraussiana*, using the RNeasy Plant Mini Kit (Qiagen) with modifications: adding 1% polyethylene glycol into the lysis buffer (RLC buffer) and repeating an extra EtOH buffer (RPE buffer) wash. Reverse-transcription (RT) reactions were performed against the extracted total RNA using either the Super Script III (Life Technologies) or the SMART RACE cDNA Amplification Kit (Clontech) according to the manufacturers' instructions.

Degenerate PCR and Smart-race PCR

Degenerate primers were designed based on the conserved amino acid sequences within the CLE peptide motif for TDIF or the kinase domain for TDR (Table S2). SMART-RACE PCR was performed using SMART RACE cDNA Amplification Kit (Clontech) with primers described in Table S2. Genbank accession numbers for the obtained sequences are KT343281–KT343287 as indicated in Table S2.

Phylogenetic Analysis

The sequences were first aligned in Clustal X. We excluded ambiguously aligned sequence to produce an alignment of 253 amino acid characters. Phylogenetic analyses were performed using MrBayes 3.2.1 (Huelsenbeck and Ronquist, 2001) and analyses were run for 500,000 generations, which was sufficient for convergence of the two simultaneous runs of each analysis. Convergence was assessed by visual inspection of the plot of the log likelihood scores of the two runs calculated by MrBayes (Gelman and Rubin, 1992). Character matrix and command files used to run the Bayesian phylogenetic analysis are provided in Data Sheet S1.

Plant Culture and Peptide Treatment

Immature *G. biloba* leaves, immature *A. aethiopicum* fronds and *S. kraussiana* shoots of 5 mm in length were excised

and surface sterilized in sterilization solution (1% sodium hypochlorite and 0.1% TritonX-100) for 3–5 min, then washed 4 times with water. For *Asplenium × lucrosum* bulbils, all visible leaves were detached and the sterilization was performed for 15 min. All plant samples were cultured in half-strength MS liquid medium containing 1% sucrose and 0.05% MES (pH 5.8) at 22°C under continuous light without shaking. The bulbils were transferred to new liquid culture medium every 3 weeks. In the peptide treatment assays, plant samples of similar size/developmental stage were collected for the replicate of control and peptide-treatment samples. TDIF, (HEVHypSGHypNPISN), SkCLE1 (HSVHypSGHypNPVGN), and SkCLE1L (HSVHypSGHypNPVGNLPG) peptides were chemically synthesized with >95% purity (Operon Biotechnologies). All experiments were replicated at least three times.

Observation of Vasculature

Leaves/fronds were fixed in a 1:3 mixture of acetic acid/ethanol, washed with water and mounted in a mixture of chloral hydrate/glycerol/water (8:1:2). For sectioning, samples were fixed in FAA solution (50% ethanol: 10% formalin: 5% acetic acid in water) and embedded using the JB-4 embedding kit (Polysciences) according to the manufacturer's instructions. Blocks were sectioned at 3 µm thick and the sections were stained with 0.05% toluidine blue and observed with a Zeiss Axioskop microscope.

RESULTS

TDIF Genes in Vascular Plants

TDIF genes in non-flowering vascular plants were identified by searching the Genbank and 1 KP databases using the amino acid sequence of TDIF, HEVPSGPNPISN, as a query. This revealed TDIF-like gene transcripts in many gymnosperms and ferns. For example, CLE peptide motifs identical to TDIF were found in *Picea sitchensis*, *Pseudotsuga menziesii*, *Taxus baccata*, *Sequoia sempervirens*, *Gnetum gnemon*, *G. biloba*, *Equisetum giganteum* (Table S1). In the transcript data for the lycophyte *Huperzia squarrosa*, we found two H-type CLE and an R-type CLE sequences although we could not find CLE peptide motifs identical to TDIF in lycophyte data. These sequences were also different from any of the five H-CLE sequences of *S. moellendorffii*, encoded by SmCLE12–15 (Miwa et al., 2009). In the moss *P. patens*, a CLE gene has been reported and designated as CLE170/PpCLE1 (Oelkers et al., 2008; Miwa et al., 2009). Using the CLE motifs of PpCLE1 in addition to TDIF as queries, we found additional 5 CLE sequences in Genbank transcript database (designated as PpCLE2 to PpCLE6; Table S1). However, all encode R-type CLE genes and no additional H-type CLE gene was detected.

We next isolated TDIF orthologs from cDNA of *G. biloba*, *A. aethiopicum*, and *S. kraussiana* by degenerate PCR and RACE PCR. For *G. biloba* TDIF genes (*GbCLE1* and *GbCLE2*), two partial sequences obtained in the BLAST search were used to design primers for RACE-PCR. *GbCLE1* and *GbCLE2* sequences

exhibit a typical CLE protein organization: an N-terminal signal peptide, a CLE peptide motif near or at the C-terminus and an intervening non-specific region (**Figure 1A**). In *A. aethiopicum* and *S. kraussiana*, amplification of CLE peptide sequences was performed by degenerate SMART-RACE PCR with the primers corresponding to the first several amino acids in the CLE peptide motif and 3'-end universal primers for SMART-RACE PCR (Table S2). We could detect single genes in the two species, namely *AaCLE1* and *SkCLE1*. The *SkCLE1* sequence was highly similar to CLE14 of *S. moellendorffii*. Both *AaCLE1* and *SkCLE1* had the typical CLE protein configuration (**Figure 1A**).

The primary sequences of the CLE peptide motif of *AtCLE41/44*, *GbCLE1*, *GbCLE2*, and *AaCLE1* were identical while *SkCLE1* has a few substitutions relative to the other sequences (**Figure 1A**). As these substituted residues are reported to be not essential for bioactivity in the xylem cell differentiation assay (Ito et al., 2006), *SkCLE1* peptide would be predicted to possess the TDIF-like bioactivity in angiosperms. In *A. thaliana*, exogenous TDIF suppresses xylem differentiation when plants are grown in liquid culture medium (**Figures 1B,C**; Hirakawa et al., 2008), and indeed, *SkCLE1* peptide (H-S-V-Hyp-S-G-Hyp-N-P-V-G-N) exhibited a similar bioactivity (**Figure 1D**). A longer CLE peptide, *SkCLE1L* (H-S-V-Hyp-S-G-Hyp-N-P-V-G-N-S-L-P-G), was also examined since C-terminal cleavage of the *SkCLE1* peptide might occur either at the homologous position (Asn⁸²-Ser⁸³) or between the Gly⁸⁶ and Lys⁸⁷, catalyzed by proteases like the *A. thaliana* SOL1 carboxypeptidase (Tamaki et al., 2013). The *SkCLE1L* peptide showed a similar bioactivity as *SkCLE1* and TDIF peptides in *A. thaliana* (**Figure 1E**).

TDR Genes in Vascular Plants

BLAST searches using the kinase domain of *A. thaliana* TDR/PXY as a query, we found TDR sequences for gymnosperm and fern species from transcript databases. Sequences were obtained from *G. biloba*, *Azolla filiculoides*, *E. giganteum*, *Pteridium aquilinum* (Table S1). In addition, the *Sellaginella moellendorffii* genome contained four sequences highly similar to *AtTDR*, which are designated as *SmTDR1-A,B* and *SmTDR2-A,B* (Table S1; the pairs are two alleles). However, in *P. patens*, we could find no sequence highly similar to *AtTDR*, although orthologs of *AtCLV1*, a CLV3 receptor of *A. thaliana*, are encoded (*PpCLL1* and *PpCLL2* in Table S1; Miwa et al., 2009). CLL genes were also found in *A. filiculoides*, *E. giganteum*, and *S. moellendorffii* (Table S1). In addition, We obtained partial TDR sequences by application of degenerate PCR and RACE PCR to cDNA isolated from *G. biloba*, *A. aethiopicum* and *Sellaginella kraussiana* (Table S1). Kinase domains of the obtained sequences were aligned with the kinase domain sequences of the *ERECTA*, *CLV1/BAM*, *TDR/PXY/PXL* genes of *A. thaliana*. The phylogeny of the genes was reconstructed using a Bayesian method (**Figure 2**). Rooting the tree with the *ERECTA/CLV1/BAM* clade as an outgroup, vascular plant TDR genes form a highly supported monophyletic clade sister to a clade of land plant *CLV1/BAM*. The gene duplication producing the TDR/PXY and PXL clades predated the divergence of ferns from seed plants, with well-supported euphyllophyte clades for each of these gene classes. It seems likely that the

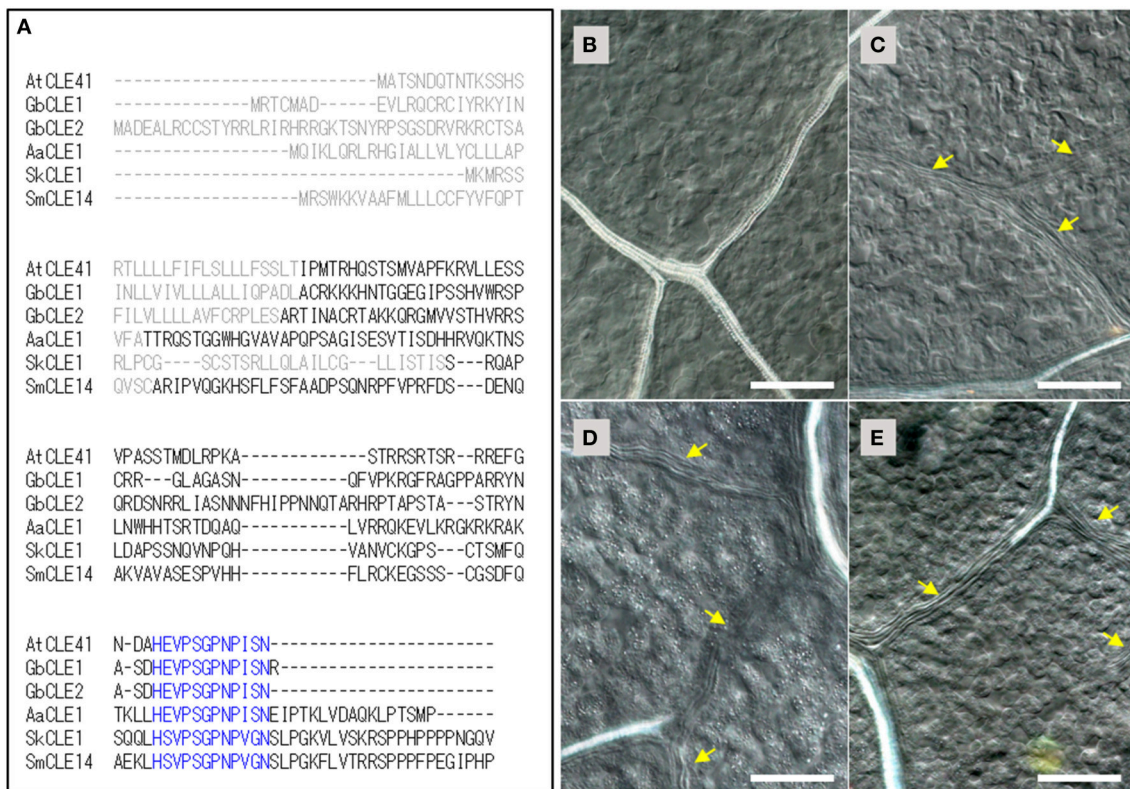


FIGURE 1 | TDIF/H-type CLE genes in land plants and the bioactivities of CLE peptides in *Arabidopsis thaliana*. (A) Alignment of deduced primary sequences for TDIF genes identified in this study (*Ginkgo biloba* CLE1, *Adiantum aethiopicum* CLE1, *Selaginella kraussiana* CLE1) with *A. thaliana* CLE41 and *Selaginella moellendorffii* CLE14. Gray and blue texts indicate signal peptide and the 12 amino-acid CLE peptide motifs, respectively. (B–E) Effects of peptides in *A. thaliana* plants grown for 10 days in liquid medium containing no additional peptide (B), 5 μ M TDIF (C), 5 μ M SkCLE1 peptide (D), or 5 μ M SkCLE1L peptide (E). Yellow arrows indicate veins without visible xylem vessels. Scale bars: 100 μ m.

gene duplication producing TDR and PXL genes occurred prior to the divergence of the lycophytes from the remainder of vascular plants, with *SmTDR2* being an ortholog of PXL and *SmTDR1* an ortholog of TDR, but the ambiguous position of *SmTDR1* precludes a definitive statement. Within the euphyllophyte TDR/PXY clade, phylogenetic relationships of the sequences largely mirror that of accepted euphyllophyte phylogeny. All gymnosperm sequences we identified are TDR orthologs, but broader sampling is required to determine whether the PXL ortholog was lost in gymnosperms. That *P. patens* genes are embedded, with high support, in the CLV1/BAM clade suggest that a TDR homolog was likely lost in the moss lineage.

Effects of TDIF on Vascular Development in Vascular Plants

As demonstrated above, TDIF and TDR orthologs are found throughout vascular plants. To investigate the function of TDIF/H-type CLE in vascular plants, we examined the bioactivity of peptide treatment in species from different taxa—gymnosperms, ferns and lycophytes. In *A. thaliana*, bioactivity of TDIF can be readily observed by liquid culturing of whole

plants, thus we applied a similar approach in other species. Immature leaves on short shoots from a *G. biloba* tree were excised and grown in liquid culture for 10 days. In the vasculature, xylem differentiation occurs near the distal edge of the leaf blade and continuous xylem strands are formed along the veins (Figure 3A). TDIF treatment inhibited xylem differentiation, leading to veins developing without visible xylem tracheids even in the central region of the leaf blade (Figures 3A–D). In the veins without tracheids, elongated procambium-like cells are observed (Figures 3E,F). In cross-section, the loss of tracheids, as determined by secondary wall development (assessed by toluidine blue staining), in TDIF treated leaves was observed while phloem differentiation occurred normally, similar to the effects of exogenous bioactivity of TDIF in *A. thaliana* (Figures 3G,H; Hirakawa et al., 2008). The radius of the leaf blade grew from approximately 1.5–7.5 mm during this period (Figure 3I), and the overall growth was not affected by addition of 10 μ M TDIF in the liquid medium.

TDIF sensitivity of *A. aethiopicum* was examined in similar experiments. Immature unfurled fronds were excised by cutting at the petiole and were grown in liquid medium. TDIF treatment

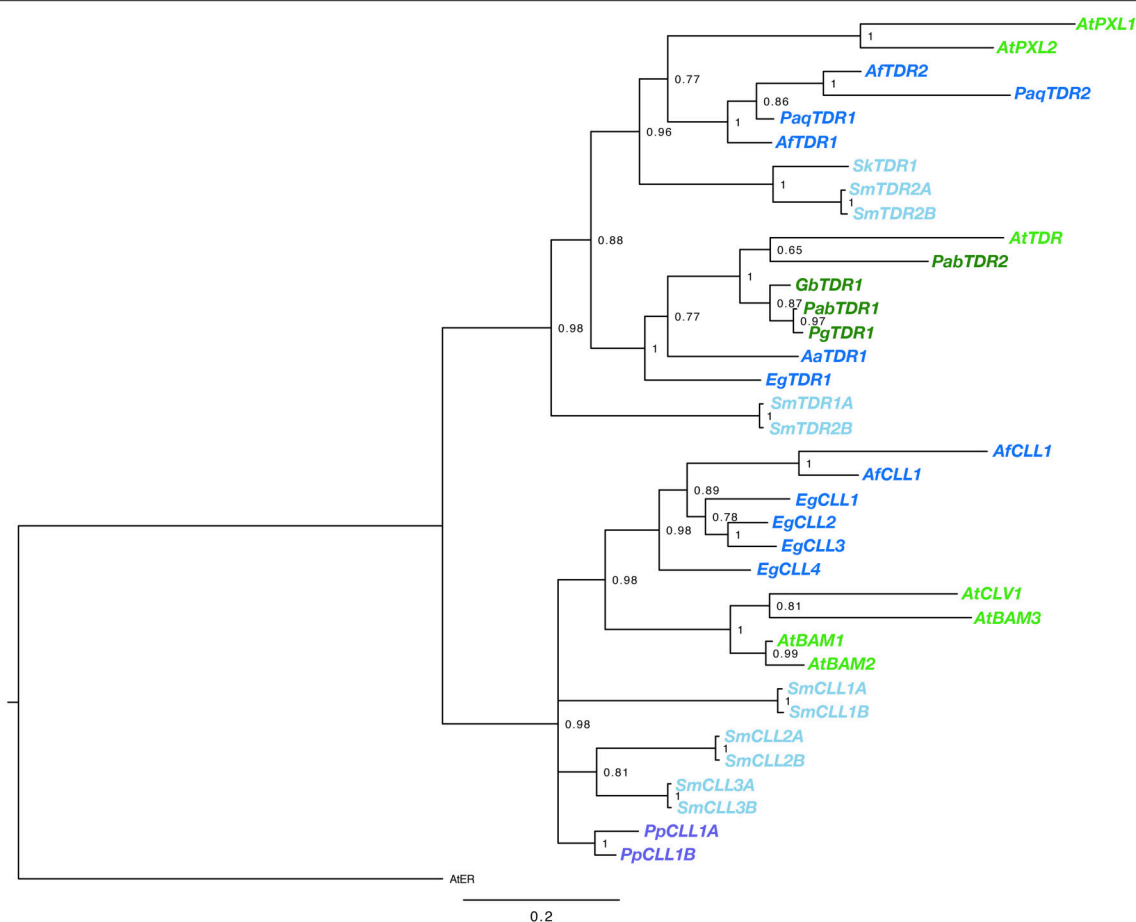


FIGURE 2 | Phylogenetic relationships of vascular plant TDR genes. The phylogram was generated with the kinase domain sequences based on a Bayesian method. The posterior probabilities of trees are shown at the nodes. Using *AtER* as an outgroup, CLV and TDR genes each form monophyletic clades. Genes are color coded by taxon: moss, purple; lycophyte, light blue; ferns, dark blue; gymnosperm, dark green; angiosperm, light green. Species abbreviations are as follows: Pp, *Physcomitrella patens*; Sk, *Selaginella kraussiana*; Sm, *Selaginella moellendorffii*; Aa, *Adiantum aethiopicum*; Af, *Azolla filiculoides*; Eg, *Equisetum giganteum*; Paq, *Pteridium aquilinum*; Gb, *Ginkgo biloba*; Pab, *Picea abies*; Pg, *Picea glauca*; At, *Arabidopsis thaliana*. The paired *S. moellendorffii* sequences (A and B) are likely alleles.

reduced the formation of xylem strands in veins of fronds cultured for 10 days (**Figures 4A,B**). Although the inhibitory effect on xylem formation was not as strong as what was observed in *G. biloba* or *A. thaliana*, the discontinuous xylem formation indicates proper xylem cell differentiation was impeded by TDIF (**Figure 4B**). We further examined the effects of TDIF in *Asplenium × lucrosum* (*A. bulbiferum* × *A. dimorphum*) because this species produces many bulbils on its fronds, and bulbils can be cultured for a long period. All visible fronds were detached from bulbils and they were grown in liquid culture. A few fronds emerged in 17 days, after which the bulbils were transferred to TDIF containing media or control media and were further cultured for 34 days. In 51 day old plants treated with 10 μ M TDIF, leaf veins without visible tracheids were observed (**Figures 4C,D**). Altogether, TDIF inhibits xylem cell differentiation in the two examined fern species.

In addition to the inhibitory activity on xylem strand formation, TDIF had a strong inhibitory activity on overall plant growth in *A. × lucrosum*. After 3 months in culture,

TDIF treatment reduced the growth in a dose dependent manner (**Figures 5A–D**). Although the size and complexity of fronds was reduced in TDIF treated plants, the number of fronds formed was increased. Root growth was also inhibited but the number of the roots formed increased. In addition, while root length was reduced, the thickness of roots increased (**Figure 5E**). In cross-section, roots grown with 1 μ M TDIF had an increased number of cortex cell layers and abnormally shaped epidermal cells (**Figures 5F,G**). The size of the central vascular cylinder was not affected but its cellular organization was altered (**Figures 5H,I**). Under control conditions central vascular tissues are surrounded by one or two layers of pericycle cells. The vascular cylinder contains dipolar protoxylem tracheids, small phloem-like cells at the periphery, and relatively large cells near the center. In the peptide treatment central cylinders had a smaller number of relatively large cells without clear morphological features characteristic of differentiated vascular cells, indicating inhibition of proper cell differentiation and cell division.

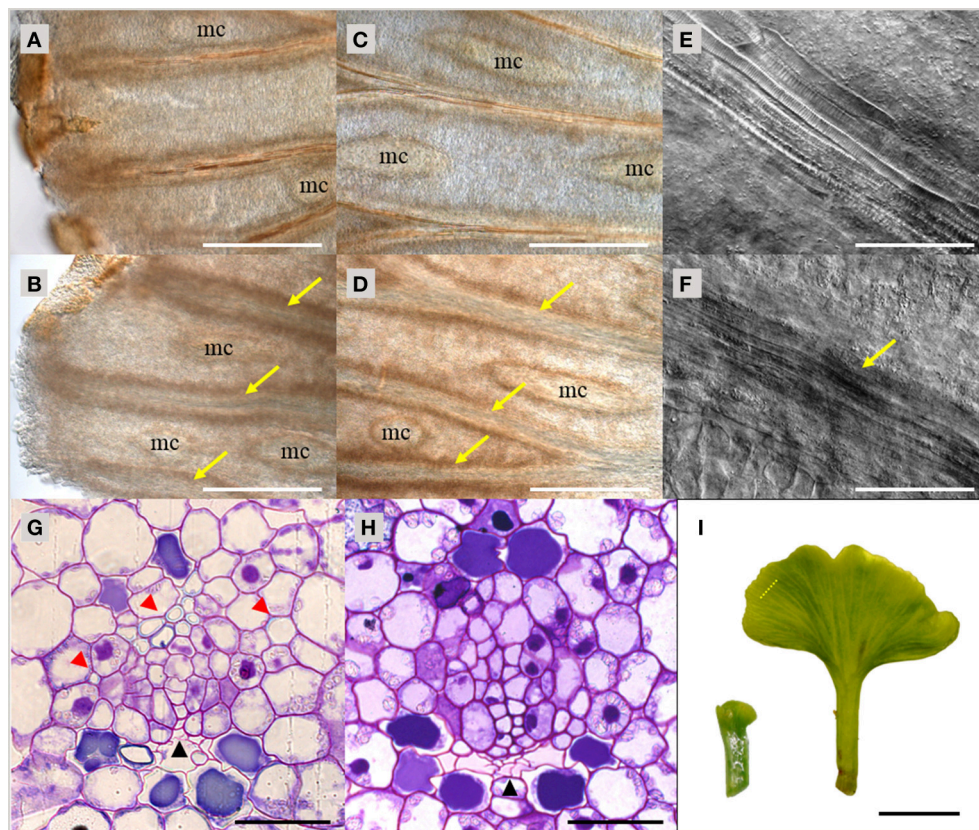


FIGURE 3 | Effects of TDIF treatment in *Ginkgo biloba* leaf. (A–H) Veins of *G. biloba* leaves cultured for 10 days in liquid medium containing no additional peptide (**A, C, E, G**) or 10 μ M TDIF (**B, D, F, H**). (**A–F**) Images near the edge (**A, B**) or middle (**C–F**) of the leaf blade. Yellow arrows indicate veins without visible tracheids. Mucilage canals are indicated as mc. (**G, H**) Cross section of the veins. Black and red arrowheads indicate phloem and xylem, respectively. (**I**) Overall leaf morphology before (left) and after (right) 10-day culture. The dashed line illustrates the approximate position of sectioning plane for (**G, H**). Scale bars: 500 μ m in (**A–D**), 100 μ m in (**E, F**), 50 μ m in (**G, H**), and 0.5 cm in (**I**).

We used *S. kraussiana* as a model to examine TDIF sensitivity in lycophytes. Excised shoots containing a pair of branched shoot tips were cultured 3 weeks in liquid culture with or without peptides. Near the shoot tips of control plants, two rows of xylem strands are formed, which are connected to xylem strands in leaves, following the pattern typical to the vascular development in lycophyte shoot (**Figure 6A**; Steeves and Sussex, 1989). In TDIF-treated shoots, the continuity of xylem strands as well as their relative position was not altered significantly—it was not affected by either SkCLE1 or SkCLE1L peptide (**Figures 6A–D**). Examining tissues other than the vasculature, we could not find any developmental defects by the peptides. We further examined the effect of peptides in rhizophores emerged during liquid culturing, but we did not see significant changes in rhizophore formation and growth due to peptide treatment. In cross-section of the *S. kraussiana* rhizophore the vascular cylinder contains central xylem and surrounding phloem tissues (**Figure 6E**). Tracheid differentiation was not suppressed in plants grown in the presence of 5 μ M of any of the three peptides (**Figures 6E–H**). These data indicate that exogenously applied TDIF or SkCLE1 peptides do not have inhibitory activities on

xylem cell differentiation in *S. kraussiana* although it is still not clear if the *SkCLE1* gene plays no role in xylem differentiation *in planta* because the lack of responses could be due to limitations of method as discussed later.

DISCUSSION

Molecular genetic studies have illustrated the importance of peptide signaling in communication between cells and tissues, which is essential for plant growth and development. An important question is how these signals are integrated into a specific developmental context, such as the formation of vasculature, during plant evolution. Comparative analyses on different plant taxa is one strategy to address this question. However, the availability of molecular genetic techniques is still limited to a small number of species. In this study, we analyzed the evolution of TDIF/CLE genes and the bioactivity of TDIF to examine their function in vascular development in taxa of the major clades of vascular plants including gymnosperms, ferns and lycophytes.

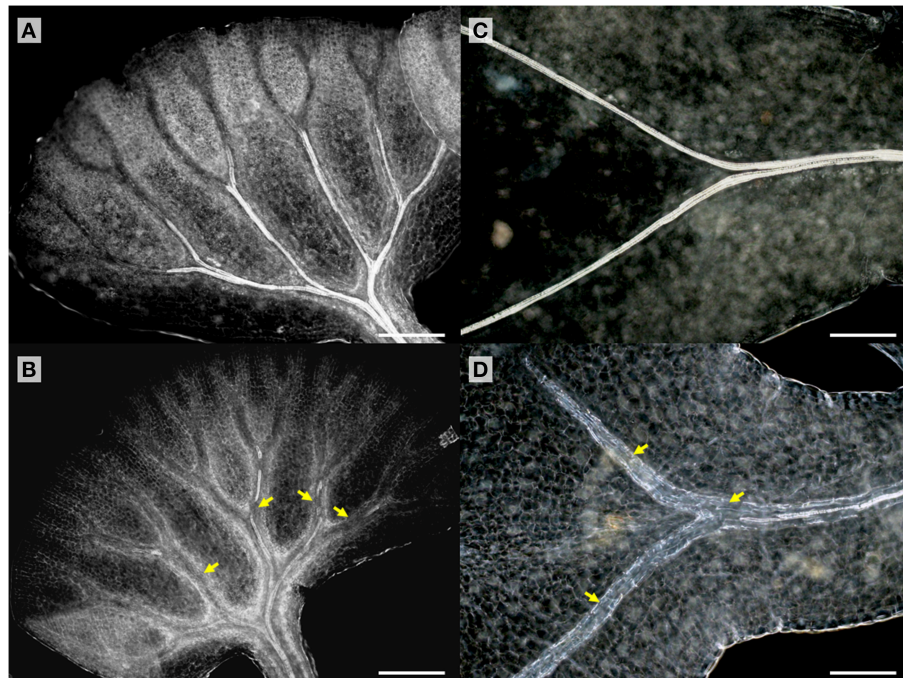


FIGURE 4 | Effects of TDIF treatment in fern fronds. *Adiantum aethiopicum* (A,B) and *Asplenium × lucrosium* (C,D) fronds from plants cultured for 10 or 51 days in liquid medium containing no additional peptide (A,C) or 10 μ M TDIF (B,D). Yellow arrows indicate discontinued xylem strand in (B) and veins without visible tracheids in (D). Scale bars: 200 μ m.

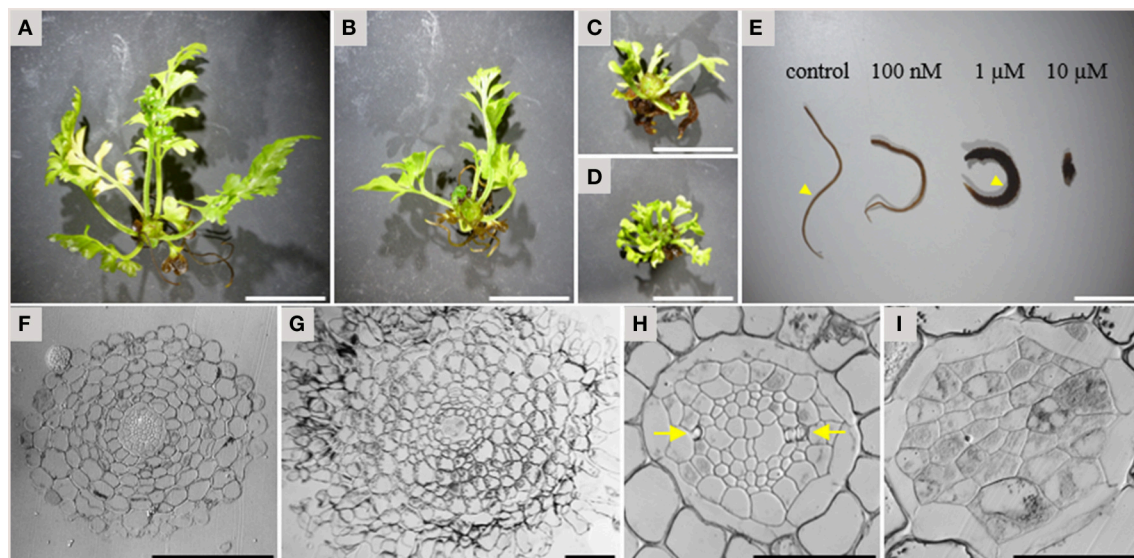


FIGURE 5 | Effects of TDIF on the morphology of *Asplenium × lucrosium*. (A–D) Overall morphology of *A. × lucrosium* plants grown for 3 months in liquid medium containing no additional peptide (A), 100 nM TDIF (B), 1 μ M TDIF (C), or 10 μ M TDIF (D). (E) comparison of root morphology grown for 5 weeks in liquid culture containing different concentration of TDIF peptides as indicated. (F–I) Cross sections at the middle of the roots grown in control (F,H) or 1 μ M TDIF (G,I) medium. Approximate positions for sectioning were illustrated in (E) by arrowheads. The images for (H,I) are magnification of central cylinder in (F,G). Arrows in (H) indicate protoxylem poles. Scale bars: 2 cm in (A–D), 1 cm in (E), 100 μ m in (F,G), and 50 μ m in (H,I).

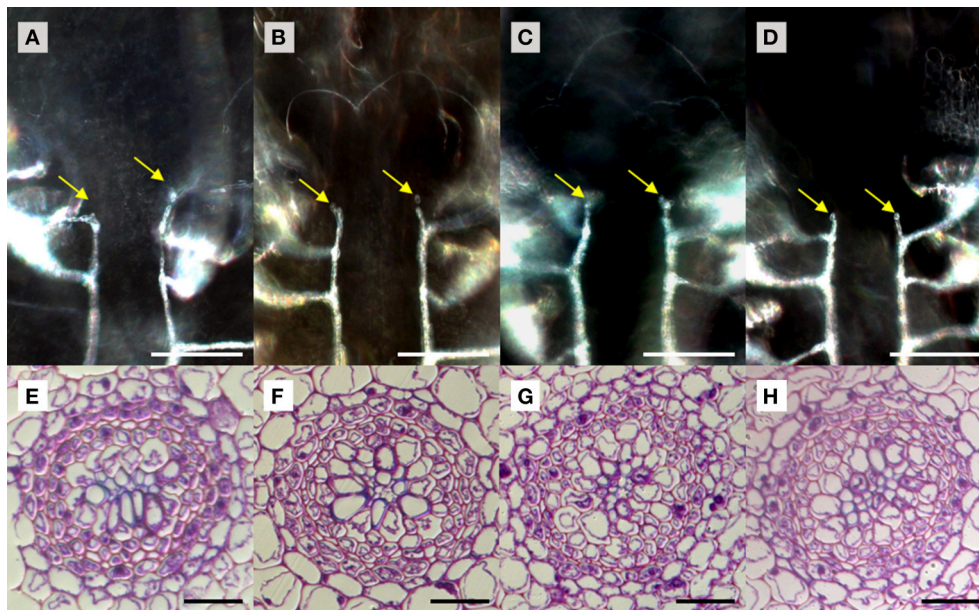


FIGURE 6 | Effects of TDIF in *Selaginella kraussiana*. Cleared whole-mount images of shoot meristem region with xylem strands (**A–D**) and cross sections of rhizophores (**E–H**) of *S. kraussiana* shoots grown for 3 weeks in liquid medium containing no additional peptide (**A,E**), 5 μM TDIF (**B,F**), 5 μM SkCLE1 peptide (**C,G**), or 5 μM SKCLE1L peptide (**D,H**). Yellow arrows in (**A–D**) indicate the termini of xylem strands (white lines in images) just below the shoot apical meristem. Scale bars: 100 μm (**A–D**), 20 μm (**E–H**).

Phylogenetic analyses of sequences obtained in this study indicate that TDIF/H-type CLE genes and TDR receptor genes are conserved among vascular plants, suggesting that this signaling pathway may be active throughout the vascular plant lineage. In contrast, the moss *P. patens* lacks both TDIF/H-type CLE and TDR genes in its genome although it possesses R-type CLE genes. Further analyses on CLE family genes in other bryophytes, as well as charophycean algae, is necessary for understanding the origin and evolution of CLE peptide signaling.

TDIF treatment assays that TDIF is bioactive in shoot vascular tissues of gymnosperms and ferns. Inhibition of xylem strand differentiation in gymnosperm and fern species indicates conservation of the role for TDIF in tracheary element differentiation in euphyllophytes. In *A. thaliana*, TDIF signaling is implicated in the coordination of phloem and xylem differentiation from intervening procambium (Miyashima et al., 2013). As this type of vascular development, secondary xylem and phloem formation, is not active in extant ferns (Gifford and Foster, 1989; Spicer and Groover, 2010), the roles for TDIF in vascular development are not restricted to secondary vascular development. In the lycophyte *S. kraussiana*, we could not detect effects of TDIF on xylem differentiation in either shoots or rhizophores. Based on these observations, we propose that TDIF was integrated into shoot xylem differentiation in the euphyllophyte lineage after divergence from the lycophyte lineage. In this model, lycophytes and euphyllophytes may undergo a different process of xylem differentiation during vascular development. Future comparative analysis on the timing of xylem cell differentiation and the localization of TDIF

signaling among vascular plant lineages is necessary to address if this model is valuable.

The peptide treatment assay in *A. × lucrosum* uncovered a novel developmental role for TDIF/H-type CLE. The strong growth inhibition observed by adding as low as 100 nM of TDIF was not observed in *A. thaliana*, rice or pine (Kinoshita et al., 2007; Strabala et al., 2014). In addition, TDIF did not merely reduce plant growth, but increased the production of lateral organs, increased root cortex cell layers and suppressed cell differentiation in the root vascular cylinder. Thus, TDIF signaling may confer multiple bioactivities in different contexts of tissue/organ development, which might reflect the different origins of lateral organs between seed plants and ferns.

There still exist non-trivial problems in the application of peptide treatment assays. Foremost, the gain-of-function effects caused by peptide treatment provide an idea of the potential for a peptide, which is not necessarily reflected in the role for intrinsic peptide signaling *in planta*. Development of specific agonist and antagonist is a future challenge to overcome this problem. Another problem is the efficacy of peptides. In this study, we did not see any alteration in vascular development of *S. kraussiana* by TDIF treatment, however the peptides might not be delivered to the vasculature, or alternatively, be degraded before reaching cell types expressing an appropriate receptor. In addition to the permeability problem, it is also possible that the native peptides possess side chain modifications, and thus the synthetic peptide does not represent the bioactivity of intrinsic peptides (Okamoto et al., 2013). Finally, a lack of a phenotype might reflect limited expression of the receptor.

To overcome these problems, establishing experimental systems such as transformation techniques or cell culture systems is also important.

ACKNOWLEDGMENTS

We thank the 1 KP initiative for providing information, Sandra Floyd and Stewart Crowley for technical assistance, Bowman lab people for critical discussion. This work was supported by the

Australian Research Council (DP110100070 to JLB). YH was supported by a JSPS Research Fellowship for Research Abroad and an HFSP long-term fellowship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.01048>

REFERENCES

- Araya, T., Miyamoto, M., Wibowo, J., Suzuki, A., Kojima, S., Tsuchiya, Y. N., et al. (2014). CLE-CLAVATA1 peptide-receptor signaling module regulates the expansion of plant root systems in a nitrogen-dependent manner. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2029–2034. doi: 10.1073/pnas.1319953111
- Brand, U., Fletcher, J. C., Hobe, M., Meyerowitz, E. M., and Simon, R. (2000). Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity. *Science* 289, 617–619. doi: 10.1126/science.289.5479.617
- Clark, S. E., Running, M. P., and Meyerowitz, E. M. (1995). CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. *Development* 121, 2057–2067.
- Depuydt, S., Rodriguez-Villalon, A., Santuari, L., Wyser-Rmili, C., Ragni, L., and Hardtke, C. S. (2013). Suppression of Arabidopsis protophloem differentiation and root meristem growth by CLE45 requires the receptor-like kinase BAM3. *Proc. Natl. Acad. Sci. U.S.A.* 110, 7074–7079. doi: 10.1073/pnas.1222314110
- de Vries, J., Fischer, A. M., Roettger, M., Rommel, S., Schluepmann, H., Bräutigam, A., et al. (2015). Cytokinin-induced promotion of root meristem size in the fern *Azolla* supports a shoot-like origin of euphylllophyte roots. *New Phytol.* doi: 10.1111/nph.13630. [Epub ahead of print].
- DeYoung, B. J., Bickle, K. L., Schrage, K. J., Muskett, P., Patel, K., and Clark, S. E. (2006). The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in Arabidopsis. *Plant J.* 45, 1–16. doi: 10.1111/j.1365-3113X.2005.02592.x
- Endo, S., Shinohara, H., Matsubayashi, Y., and Fukuda, H. (2013). A novel pollen-pistil interaction conferring high-temperature tolerance during reproduction via CLE45 signaling. *Curr. Biol.* 23, 1670–1676. doi: 10.1016/j.cub.2013.06.060
- Etchells, J. P., and Turner, S. R. (2010). The PXY-CLE41 receptor ligand pair defines a multifunctional pathway that controls the rate and orientation of vascular cell division. *Development* 137, 767–774. doi: 10.1242/dev.044941
- Fisher, K., and Turner, S. (2007). PXY, a receptor-like kinase essential for maintaining polarity during plant vascular-tissue development. *Curr. Biol.* 17, 1061–1066. doi: 10.1016/j.cub.2007.05.049
- Fiume, E., and Fletcher, J. C. (2012). Regulation of Arabidopsis embryo and endosperm development by the polypeptide signaling molecule CLE8. *Plant Cell* 24, 1000–1012. doi: 10.1105/tpc.111.094839
- Fletcher, J. C., Brand, U., Running, M. P., Simon, R., and Meyerowitz, E. M. (1999). Signaling of cell fate decisions by CLAVATA3 in Arabidopsis shoot meristems. *Science* 283, 1911–1914. doi: 10.1126/science.283.5409.1911
- Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457–511. doi: 10.1214/ss/1177011136
- Gifford, E. M., and Foster, A. S. (1989). *Morphology and Evolution of Vascular Plants, 3rd Edn.* New York, NY: W.H. Freeman and Co.
- Hirakawa, Y., Kondo, Y., and Fukuda, H. (2010). TDIF peptide signaling regulates vascular stem cell proliferation via the WOX4 homeobox gene in Arabidopsis. *Plant Cell* 22, 2618–2629. doi: 10.1105/tpc.110.076083
- Hirakawa, Y., Shinohara, H., Kondo, Y., Inoue, A., Nakanomyo, I., Ogawa, M., et al. (2008). Non-cell-autonomous control of vascular stem cell fate by a CLE peptide/receptor system. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15208–15213. doi: 10.1073/pnas.0808444105
- Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754
- Ito, Y., Nakanomyo, I., Motose, H., Iwamoto, K., Sawa, S., Dohmae, N., et al. (2006). Dodeca-CLE peptides as suppressors of plant stem cell differentiation. *Science* 313, 842–845. doi: 10.1126/science.1128436
- Jun, J. H., Fiume, E., and Fletcher, J. C. (2008). The CLE family of plant polypeptide signaling molecules. *Cell. Mol. Life Sci.* 65, 743–755. doi: 10.1007/s00018-007-7411-5
- Kinoshita, A., Nakamura, Y., Sasaki, E., Kyozuka, J., Fukuda, H., and Sawa, S. (2007). Gain-of-function phenotypes of chemically synthetic CLAVATA3/ESR-related (CLE) peptides in *Arabidopsis thaliana* and *Oryza sativa*. *Plant Cell Physiol.* 48, 1821–1825. doi: 10.1093/pcp/pcm154
- Kondo, T., Sawa, S., Kinoshita, A., Mizuno, S., Kakimoto, T., Fukuda, H., et al. (2006). A plant peptide encoded by CLV3 identified by *in situ* MALDI-TOF MS analysis. *Science* 313, 845–848. doi: 10.1126/science.1128439
- Kondo, Y., Hirakawa, Y., Kieber, J. J., and Fukuda, H. (2011). CLE peptides can negatively regulate protoxylem vessel formation via cytokinin signaling. *Plant Cell Physiol.* 52, 37–48. doi: 10.1093/pcp/pcq129
- Matsubayashi, Y. (2014). Posttranslationally modified small-peptide signals in plants. *Annu. Rev. Plant Biol.* 65, 385–413. doi: 10.1146/annurev-arplant-050312-120122
- Miwa, H., Tamaki, T., Fukuda, H., and Sawa, S. (2009). Evolution of CLE signaling: origins of the CLV1 and SOL2/CRN receptor diversity. *Plant Signal. Behav.* 4, 477–481. doi: 10.4161/psb.4.6.8391
- Miyashima, S., Sebastian, J., Lee, J. Y., and Helariutta, Y. (2013). Stem cell function during plant vascular development. *EMBO J.* 32, 178–193. doi: 10.1038/emboj.2012.301
- Oelkers, K., Goffard, N., Weiller, G. F., Gresshoff, P. M., Mathesius, U., and Frickey, T. (2008). Bioinformatic analysis of the CLE signaling peptide family. *BMC Plant Biol.* 8:1. doi: 10.1186/1471-2229-8-1
- Ogawa, M., Shinohara, H., Sakagami, Y., and Matsubayashi, Y. (2008). Arabidopsis CLV3 peptide directly binds CLV1 ectodomain. *Science* 319, 294. doi: 10.1126/science.1150083
- Ogawa-Ohnishi, M., Matsushita, W., and Matsubayashi, Y. (2013). Identification of three hydroxyproline O-arabinosyltransferases in *Arabidopsis thaliana*. *Nat. Chem. Biol.* 9, 726–730. doi: 10.1038/nchembio.1351
- Ohyama, K., Shinohara, H., Ogawa-Ohnishi, M., and Matsubayashi, Y. (2009). A glycopeptide regulating stem cell fate in *Arabidopsis thaliana*. *Nat. Chem. Biol.* 5, 578–580. doi: 10.1038/nchembio.182
- Okamoto, S., Ohnishi, E., Sato, S., Takahashi, H., Nakazono, M., Tabata, S., et al. (2009). Nod factor/nitrate-induced CLE genes that drive HAR1-mediated systemic regulation of nodulation. *Plant Cell Physiol.* 50, 67–77. doi: 10.1093/pcp/pcn194
- Okamoto, S., Shinohara, H., Mori, T., Matsubayashi, Y., and Kawaguchi, M. (2013). Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nat. Commun.* 4, 2191. doi: 10.1038/ncomm3191
- Schoof, H., Lenhard, M., Haecker, A., Mayer, K. F., Jürgens, G., and Laux, T. (2000). The stem cell population of Arabidopsis shoot meristems is maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. *Cell* 100, 635–644. doi: 10.1016/S0092-8674(00)80700-X
- Shinohara, H., Moriyama, Y., Ohyama, K., and Matsubayashi, Y. (2012). Biochemical mapping of a ligand-binding domain within Arabidopsis BAM1

- reveals diversified ligand recognition mechanisms of plant LRR-RKs. *Plant J.* 70, 845–854. doi: 10.1111/j.1365-3113X.2012.04934.x
- Spicer, R., and Groover, A. (2010). Evolution of development of vascular cambium and secondary growth. *New Phytol.* 186, 577–592. doi: 10.1111/j.1469-8137.2010.03236.x
- Steeves, T. A., and Sussex, I. M. (1989). *Patterns in Plant Development*, 2nd Edn. Cambridge: Cambridge University Press.
- Strabala, T. J., O'donnell, P. J., Smit, A. M., Ampomah-Dwamena, C., Martin, E. J., Netzler, N., et al. (2006). Gain-of-function phenotypes of many CLAVATA3/ESR genes, including four new family members, correlate with tandem variations in the conserved CLAVATA3/ESR domain. *Plant Physiol.* 140, 1331–1344. doi: 10.1104/pp.105.075515
- Strabala, T. J., Phillips, L., West, M., and Stanbra, L. (2014). Bioinformatic and phylogenetic analysis of the CLAVATA3/EMBRYO-SURROUNDING REGION (CLE) and the CLE-LIKE signal peptide genes in the Pinophyta. *BMC Plant Biol.* 14:47. doi: 10.1186/1471-2229-14-47
- Tamaki, T., Betsuyaku, S., Fujiwara, M., Fukao, Y., Fukuda, H., and Sawa, S. (2013). SUPPRESSOR OF LLP1 1-mediated C-terminal processing is critical for CLE19 peptide activity. *Plant J.* 76, 970–981. doi: 10.1111/tpj.12349
- Vanneste, K., Sterck, L., Myburg, A. A., Van de Peer, Y., and Mizrachi, E. (2015). Horsetails are ancient polyploids: evidence from *Equisetum giganteum*. *Plant Cell* 27, 1567–1578. doi: 10.1105/tpc.15.00157
- Whitford, R., Fernandez, A., De Groot, R., Ortega, E., and Hilson, P. (2008). Plant CLE peptides from two distinct functional classes synergistically induce division of vascular cells. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18625–18630. doi: 10.1073/pnas.0809395105

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Hirakawa and Bowman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Conserved Role for the *NAM/miR164* Developmental Module Reveals a Common Mechanism Underlying Carpel Margin Fusion in Monocarpous and Syncarpous Eurosids

Aurélie C. M. Vialette-Guiraud¹, Aurélie Chauvet¹, Juliana Gutierrez-Mazariegos¹, Alexis Eschstruth², Pascal Ratet² and Charles P. Scutt^{1*}

¹ Laboratoire de Reproduction et Développement des Plantes, UMR 5667, Centre National de la Recherche Scientifique – Institut National de la Recherche Agronomique – Université de Lyon, Ecole Normale Supérieure de Lyon, Lyon, France,

² Institute of Plant Sciences Paris-Saclay, Centre National de la Recherche Scientifique – Institut National de la Recherche Agronomique – Université de Paris Sud, Orsay, France

OPEN ACCESS

Edited by:

Rainer Melzer,
University College Dublin, Ireland

Reviewed by:

Stefan De Folter,
Centro de Investigación y de Estudios
Avanzados del Instituto Politécnico
Nacional, Mexico
Barbara Ambrose,
The New York Botanical Garden, USA

*Correspondence:

Charles P. Scutt
charlie.scutt@ens-lyon.fr

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 03 October 2015

Accepted: 20 December 2015

Published: 13 January 2016

Citation:

Vialette-Guiraud ACM, Chauvet A,
Gutierrez-Mazariegos J, Eschstruth A,
Ratet P and Scutt CP (2016)
A Conserved Role
for the *NAM/miR164* Developmental
Module Reveals a Common
Mechanism Underlying Carpel Margin
Fusion in Monocarpous
and Syncarpous Eurosids.
Front. Plant Sci. 6:1239.
doi: 10.3389/fpls.2015.01239

The majority of angiosperms are syncarpous- their gynoecium is composed of two or more fused carpels. In *Arabidopsis thaliana*, this fusion is regulated through the balance of expression between *CUP SHAPED COTYLEDON* (*CUC*) genes, which are orthologs of the *Petunia hybrida* transcription factor *NO APICAL MERISTEM* (*NAM*), and their post-transcriptional regulator *miR164*. Accordingly, the expression of a *miR164*-insensitive form of *A. thaliana CUC2* causes a radical breakdown of carpel fusion. Here, we investigate the role of the *NAM/miR164* genetic module in carpel closure in monocarpous plants. We show that the disruption of this module in monocarpous flowers of *A. thaliana aux1-22* mutants causes a failure of carpel closure, similar to the failure of carpel fusion observed in the wild-type genetic background. This observation suggested that closely related mechanisms may bring about carpel closure and carpel fusion, at least in *A. thaliana*. We therefore tested whether these mechanisms were conserved in a eurosid species that is monocarpous in its wild-type form. We observed that expression of *MtNAM*, the *NAM* ortholog in the monocarpous eurosid *Medicago truncatula*, decreases during carpel margin fusion, suggesting a role for the *NAM/miR164* module in this process. We transformed *M. truncatula* with a *miR164*-resistant form of *MtNAM* and observed, among other phenotypes, incomplete carpel closure in the resulting transformants. These data confirm the underlying mechanistic similarity between carpel closure and carpel fusion which we observed in *A. thaliana*. Our observations suggest that the role of the *NAM/miR164* module in the fusion of carpel margins has been conserved at least since the most recent common ancestor of the eurosid clade, and open the possibility that a similar mechanism may have been responsible for carpel closure at much earlier stages of angiosperm evolution. We combine our results with studies of early diverging angiosperms to speculate on the role of the *NAM/miR164* module in the origin and further evolution of the angiosperm carpel.

Keywords: *Arabidopsis thaliana*, *Medicago truncatula*, *CUP SHAPED COTYLEDON*, *NO APICAL MERISTEM*, *miR164*, gynoecium, carpel, syncarpy

INTRODUCTION

The female whorl, or gynoecium, of the angiosperm flower consists of one or more carpels which enclose the ovules. In apocarpous gynoecia, the carpels remain separate throughout development, while in syncarpous gynoecia, they are fused together, either from their inception (congenital fusion), or from a later developmental stage (post-genital fusion). If only one carpel is produced per flower, the gynoecium is termed monocarpous. Carpels in apocarpous or monocarpous gynoecia may emerge from the floral meristem with their margins already fused together, in which case they are described as ascidiate (bottle-shaped), or may emerge with unfused margins that subsequently fuse by folding, in which case they are described as plicate.

Syncarpy is believed to confer several selective advantages over apocarpy, including a larger landing platform for pollinating insects, a compitum (a common intersection in the route for pollen tube growth), and larger fruits with more sophisticated mechanisms for seed dispersal. Mapping of character states onto angiosperm phylogeny indicates that syncarpy has arisen at least 17 times in the angiosperms, while the evolution of apocarpy from syncarpy is much less frequent (Armbruster et al., 2002).

The model angiosperm *Arabidopsis thaliana* possesses a syncarpous gynoecium of two congenitally fused carpels. These organs emerge from the center of the floral meristem as a single dome of cells, within which a central slot-like cavity forms as the gynoecium begins to elongate (Smyth et al., 1990). The positions of the carpel margins within the gynoecium wall only become apparent at a later stage, when this structure undergoes differentiation into valve and abaxial replum tissues. Meristematic activity from the abaxial replum then generates the adaxial replum, or septum, which grows inward to divide the ovary into two chambers. Ovule primordia develop from parietal placentae which form along the carpel margins within each chamber of the ovary.

In contrast to *A. thaliana*, the model angiosperm *Medicago truncatula* possesses a plicate, monocarpous gynoecium (Benlloch et al., 2003). At an early stage of *M. truncatula* flower development, the gynoecial primordium becomes crescent-shaped and its margins then fuse together to enclose the single chamber of the ovary. Ovule primordia in *M. truncatula* form from a parietal placenta that develops along the fused carpel margin.

Carpel fusion in *A. thaliana* is regulated by a genetic module, generically termed here the *NAM/miR164* module, which consists of a subset of NAC-family (NAC for *NAM*, *ATAF* and *CUC*; Aida et al., 1997) transcription factors and their post-transcriptional regulator *miR164* (Mallory et al., 2004). In *A. thaliana*, the NAC genes involved in this module are *CUP-SHAPED COTYLEDON1* (*CUC1*) and *CUC2* (Aida et al., 1997), which are orthologs of the single gene *NO APICAL MERISTEM* (*NAM*) from *Petunia hybrida* (Souer et al., 1996). Loss of *miR164* function through mutations to all three *MIR164* paralogs in *A. thaliana* (Sieber et al., 2007), or genetic transformation of *A. thaliana* with a *miR164*-resistant version of *CUC2* (*CUC2g-m4*; Nikovics et al., 2006), results in a breakdown of carpel fusion. Accordingly, in

miR164 triple mutants or *CUC2g-m4* transformants, the two carpels of the *A. thaliana* gynoecium emerge separately and remain unfused and open throughout development.

In addition to their role in carpel development, studies of *NAM* orthologs in eudicots show these factors to be involved in meristem formation and cotyledon development (Souer et al., 1996; Aida et al., 1997, 1999; Takada et al., 2001; Weir et al., 2004), leaf development (Ishida et al., 2000; Nikovics et al., 2006; Blein et al., 2008), ovule development (Galbiati et al., 2013; Kamiuchi et al., 2014) and phyllotaxy (Peaucelle et al., 2007). These transcription factors are expressed at organ margins and tissue boundaries, and their down-regulation by *miR164* consequently facilitates organ outgrowth and/or developmental fusion. The action of the *NAM/miR164* module in the *A. thaliana* leaf margin has been modeled and found to generate, via effects on the auxin efflux carrier PINFORMED1 (PIN1), an alternating series of auxin maxima and minima that, respectively, generate regions of higher and lower marginal growth (Bilborough et al., 2011).

In this work, we hypothesized that the role of the *NAM/miR164* module in syncarpous fusion in *A. thaliana* might reflect a more general role in the fusion of carpel margins in angiosperms. Consequently, we tested the role of this module in the closure of monocarpous gynoecia produced both in *A. thaliana aux1-22* mutants, which are null mutants of the *AUX1* auxin influx transporter (Bennett et al., 1996) and in a wild-type genetic background of the distantly related eurosoid *M. truncatula*. From the results of these experiments, we conclude that the *NAM/miR164* module has conserved a role in carpel margin fusion, at least since the most recent common ancestor (MRCA) of living eurosids. A detailed comparison of gene expression patterns suggests that fine-tuning of the *NAM/miR164* module may regulate species-specific differences in the timing of carpel margin fusion. Accordingly, we discuss the possibility that the activity of the *NAM/miR164* module may be conserved in carpel development throughout the angiosperms, while subtle modulations to this mechanism may determine the distinction between congenital and post-genital carpel margin fusion events in specific angiosperm groups. We further speculate on mechanisms acting upstream of the *NAM/miR164* module that may have contributed to the origin of the carpel in the first flowering plants.

MATERIALS AND METHODS

Plant Cultivation

Arabidopsis thaliana plants were grown from seed on peat-based compost in growth chambers at a daytime temperature of ~21°C and ~55% relative humidity (RH). Plants were initially grown under 8/16 h day/night cycles generated using fluorescent lighting consisting of equal numbers of “cool daylight” (Osram Lumilux L36W/865) and “warm white” (Osram Lumilux L36W/830) lamps, giving a total photon flux at bench level of 170 $\mu\text{mol.m}^{-2}.\text{s}^{-1}$. To induce flowering, plants were transferred to long days (16/8 h day/night cycles) under otherwise similar conditions.

Medicago truncatula plants were grown from seed on peat-based compost in a greenhouse at a daytime temperature of ~22.5°C and 40–60% RH under natural daylight, extended to 16 h daylight using sodium lamps, as necessary.

Vector Construction

MtNAM (MTR_2g078700; Cheng et al., 2012) was initially isolated by radioisotopic screening of an *M. truncatula* bacterial artificial chromosome (BAC) library (Nam et al., 1999). A 1.2-kb fragment containing the *miR164*-binding site of *MtNAM* was released from a sub-cloned BAC DNA fragment by cleavage with *Sst*I and re-ligated into the *pGEM T-Easy* vector. The resulting plasmid was subjected to oligonucleotide-directed site-specific mutagenesis following the method of Kirsch and Joly (1998), using the sense- and antisense-strand oligonucleotides 5'-GAGCACGTGTCCTGTTTtagtACAACATCTACAACATC and 5'-GATGTTGTAGATGTTGTactaAAACAGGACACGTGCTC, respectively. These oligonucleotides generate the same four-base mismatch (shown above in lower case) present in the *miR164*-binding site of *CUC2g-m4* (Nikovics et al., 2006). Mutagenised and wild-type versions of a *MtNAM* genomic sequence of 9883 bp, from an *Eco*RI site situated 6437 bp upstream of the *MtNAM* initiation codon to an *Nco*I site situated 2151 bp downstream of its termination codon, were then inserted by ligation between unique *Eco*RI and *Not*I sites situated between the Left and Right T-DNA borders of the *pGREEN II-NosHyg* plant transformation vector, thereby generating the plasmids *MtNAMg-m4* and *MtNAMg-wt*, respectively.

Plant Transformation

Arabidopsis thaliana aux1-22 mutants (null mutants of *AUX1*; Bennett et al., 1996) were transformed by the “floral dip” method (Clough and Bent, 1998) using the *CUC2g-wt* and *CUC2g-m4* constructs of Nikovics et al. (2006) in *Agrobacterium tumefaciens* strain GV3101 harboring the plasmids *pMP90* (Koncz and Schell, 1986) and *pSOUP* (Hellens et al., 2000). Transformants were selected on plant agar containing 50 µg/mL kanamycin.

MtNAMg-wt and *MtNAMg-m4* constructs were introduced into *A. tumefaciens* GV3101, as described above, and used to transform *M. truncatula* R108 leaf disks by the protocol of Cosson et al. (2015), in which transgenic calli were selected on media containing 30 µg/mL hygromycin.

In Situ Hybridization

Double-stranded cDNAs representing the full-length coding sequences of *A. thaliana CUC1* and *CUC2* and of *M. truncatula MtNAM* were generated by reverse-transcriptase PCR, incorporating a T7-RNA-Polymerase promoter sequence in the reverse primer. Digoxigenin-labeled riboprobes were prepared from these templates using T7 RNA-polymerase and these were then purified and used in *in situ* hybridizations to sections of fixed floral buds embedded in Paraplast Xtra (Leica-Surgipath), as described by Viallette-Guiraud et al. (2011b). Gene expression patterns were observed and photographed under bright field illumination using a Leica Axio Imager M2 inverted microscope fitted with a Leica AxioCam MRC digital camera.

Phenotypic Observations

Flower buds were dissected, observed, and photographed using a Leica MZ12 dissecting microscope fitted with an AxioCam ICc5 digital camera. Carpel anatomy was revealed in transverse sections of fixed flower buds, prepared as for *in situ* hybridization and stained with 0.05% (w/v) Toluidine Blue-0 in 0.1 M sodium phosphate buffer (pH 6.8). Scanning electron microscopy was performed on unfixed material using a Hirox 3000 bench-top environmental scanning electron microscope (SEM).

Character State Mapping

A partial cladogram of angiosperm phylogeny was produced, based on the current consensus view of angiosperm phylogeny given by the Angiosperm Phylogeny Group III (APG III, <http://www.mobot.org/MOBOT/research/APweb/>; Bremer et al., 2009). Carpel fusion character states, obtained from the APG III website and from bibliographic searches, were mapped on this cladogram by maximum parsimony using MacClade4 software.

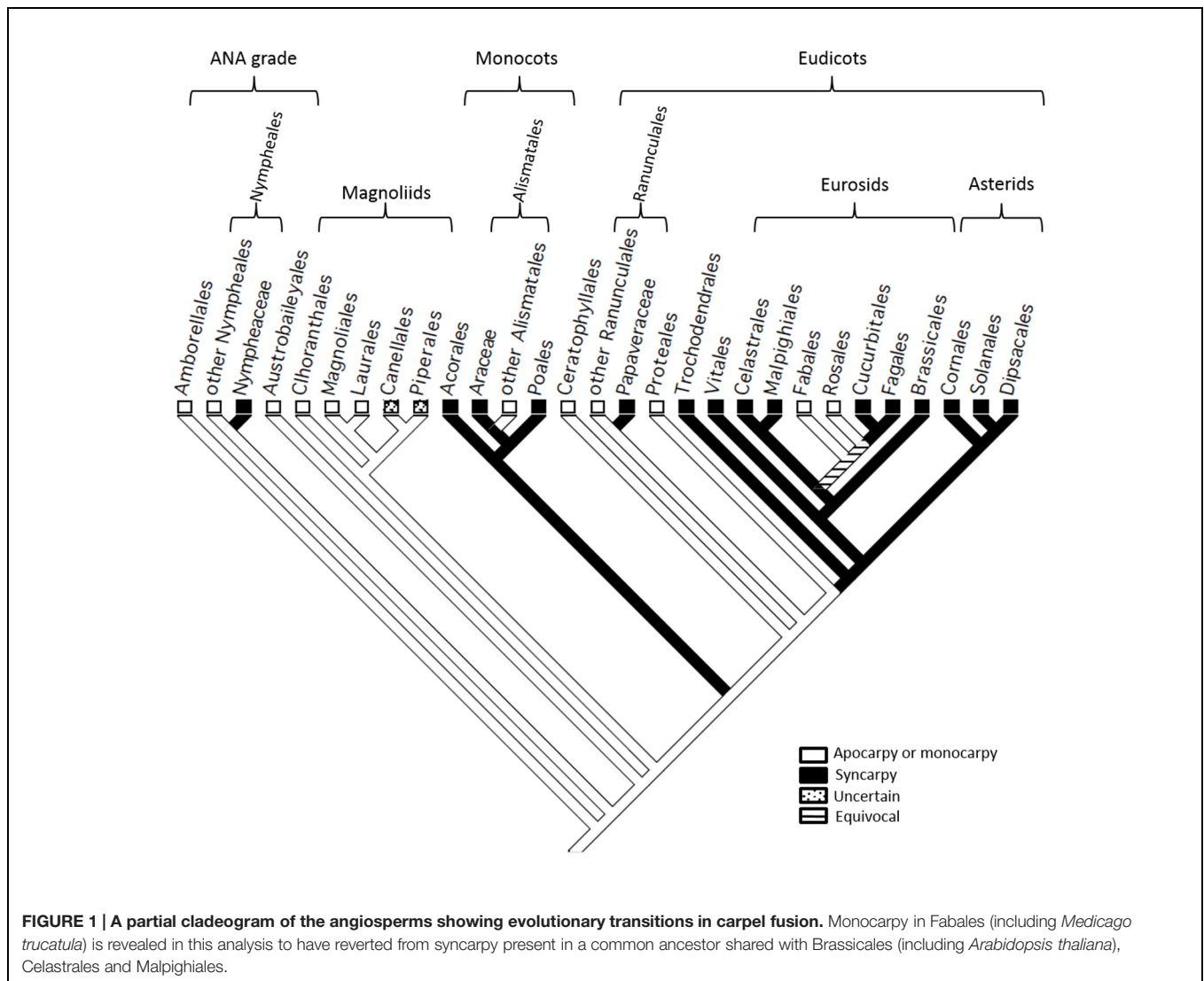
RESULTS

Monocarpy in *Medicago truncatula* Arose by Reversion from Syncarpy in a Common Ancestor Shared with *Arabidopsis thaliana*

To elucidate transitions in carpel fusion in the angiosperms, with emphasis on the model eurosids *M. truncatula* and *A. thaliana*, we mapped this character state onto a cladogram (Figure 1) representing the consensus view of angiosperm phylogeny (Bremer et al., 2009). This analysis confirms the findings of earlier studies (Armbruster et al., 2002) which indicated that the MRCA of living angiosperms was apocarpous, and that syncarpy arose several times independently, including in Nymphaeaceae, monocots, Papaveraceae and a common ancestor of the rosids and asterids. Within the eurosids, our analysis indicates that monocarpy in Fabales (including *M. truncatula*), arose secondarily from syncarpy, which was present in a common ancestor shared with Brassicales (including *A. thaliana*), Celastrales and Malpighiales. By localizing transitions between apocarpy/monocarpy and syncarpy, this analysis provides a phylogenetic framework for the evolutionary interpretation of data on the molecular mechanisms involved in these processes in living angiosperms.

The *NAM/miR164* in *Arabidopsis thaliana* Plays a Role in Both Syncarpy and the Closure of Single Carpels

As the *NAM/miR164* developmental module is necessary for carpel fusion in wild-type, syncarpous *A. thaliana* (Nikovics et al., 2006; Sieber et al., 2007), we aimed to discover whether this mechanism could also contribute to the closure of single carpels in this species. To do this, we tested whether the introduction of a *miR164*-resistant version of *CUC2* (*CUC2g-m4*) could cause a breakdown in the closure of the single carpels that are



produced in *A. thaliana* *aux1-22* mutants (Bennett et al., 1996), as compared to control plants transformed with a wild-type construct (*CUC2g-wt*).

Wild-type Col-0 gynoecia are syncarpous (Figure 2A), as are approximately 50% of gynoecia produced in *aux1-22* mutants (Figure 2B). The ovary wall in these gynoecia contains two valves, alternating with two abaxial repla. The monocarpous gynoecia, which are also produced in *aux1-22* mutants (Figures 2C,D), develop as closed structures whose ovary wall consists of only one valve and one abaxial replum (Figure 2C). These monocarpous gynoecia are not divided by a septum, or adaxial replum. Transformation of *aux1-22* mutants with *CUC2g-wt* produced no apparent change in the morphology of monocarpous gynoecia (Figure 2E). However, transformation of these mutants with *CUC2g-m4* produced a high proportion of monocarpous gynoecia that remained open to maturity (Figures 2F–I). In eight of 20 T1 transformants analyzed, all flowers containing two carpels showed carpel fusion defects, while all monocarpous flowers showed a complete or

partial lack of carpel closure, remaining open over part or all of the valve margin. In these eight plants, carpel fusion/closure defects resulted in an almost complete loss of female fertility. Thus, disruption of the *NAM/miR164* developmental module in monocarpous mutant gynoecia of *A. thaliana* causes the failure of developmental closure in these structures in a similar manner to the disruption of carpel fusion in syncarpous, wild-type gynoecia.

Expression of *NAM* Orthologs is Absent or Reduced During Carpel Margin Fusion in *Arabidopsis thaliana* and *Medicago truncatula*

The observation that the *NAM/miR164* module regulates developmental closure events in the gynoecium in both syncarpous and monocarpous genotypes of *A. thaliana* led us to speculate that this molecular mechanism might be widely conserved within the angiosperms. We chose *M. truncatula*,

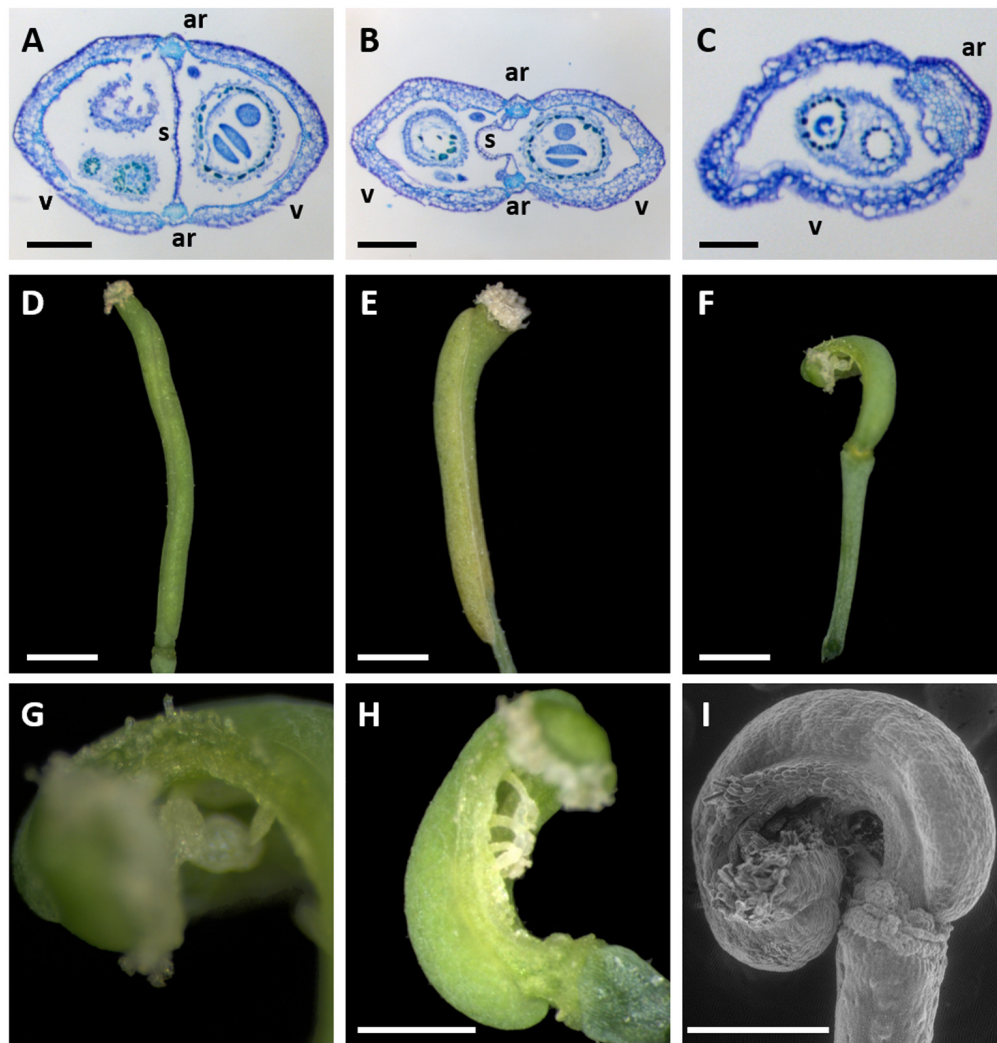


FIGURE 2 | Gynoecium morphology of *A. thaliana* *aux1-22* mutants transformed with *miR164*-resistant (*CUC2g-m4*) or un-mutated control (*CUC2g-wt*) constructs. (A–C) Toluidine blue staining of transverse sections of Col-0 wild-type (A) and *aux1-22* mutant gynoecia composed of two fused carpels (B), and one closed carpel (C), respectively. (D,E) An untransformed *aux1-22* mutant (D) and an *aux1-22* *CUC2g-wt* (control) transformant showing entirely closed monocarpous gynoecia. (F–I) *aux1-22* *CUC2g-m4* transformants showing a breakdown carpel closure in monocarpous gynoecia. (G is a magnification of the apex of the carpel shown in (F). I is a scanning electron microscope image) ar, abaxial replum; s, septum (or adaxial replum); v, valve. Bars = 200 μ m in (A,B), 100 μ m in (C), and 1 mm in (D–F,H,I).

which produces in its wild-type form a single carpel in each flower, as a candidate model species in which to test this hypothesis. The MRCA between *M. truncatula* and *A. thaliana*, which is also the MRCA of the living eurosid clade (comprising Fabidae, or eurosids I and Malvidae, or eurosids II), is estimated to have lived 114–113 million years ago (MYA; Wang et al., 2009). Prior to initiating functional experiments in *M. truncatula*, we used *in situ* hybridization to examine the conservation of expression of NAM orthologs in flower tissues between *A. thaliana* and *M. truncatula* and thereby ascertain the likelihood that the NAM/*miR164* module might function in carpel closure in the latter species.

In situ hybridization in *A. thaliana* flowers at Stage 7 (Smyth et al., 1990), in which a central slot is beginning to form in the gynoecial cylinder, revealed the expression of *CUC2* in the adaxial domain of the gynoecium and in the loculi of the developing anthers (Figure 3A). Recent studies (Galbiati et al., 2013) revealed similar results for *CUC1*. Thus, no expression of either *CUC1* or *CUC2* has been detected in regions of the ovary wall destined to become the abaxial repla, or the fusion zones between these tissues and the valves. At Stage 9–10, both *CUC1* and *CUC2* were expressed in the placentae and at presumptive tissue boundaries within the elongating ovule primordia (Figures 3B,C). At Stage 11, *CUC1* was expressed at the base of the expanding ovule integuments (Figure 3D).

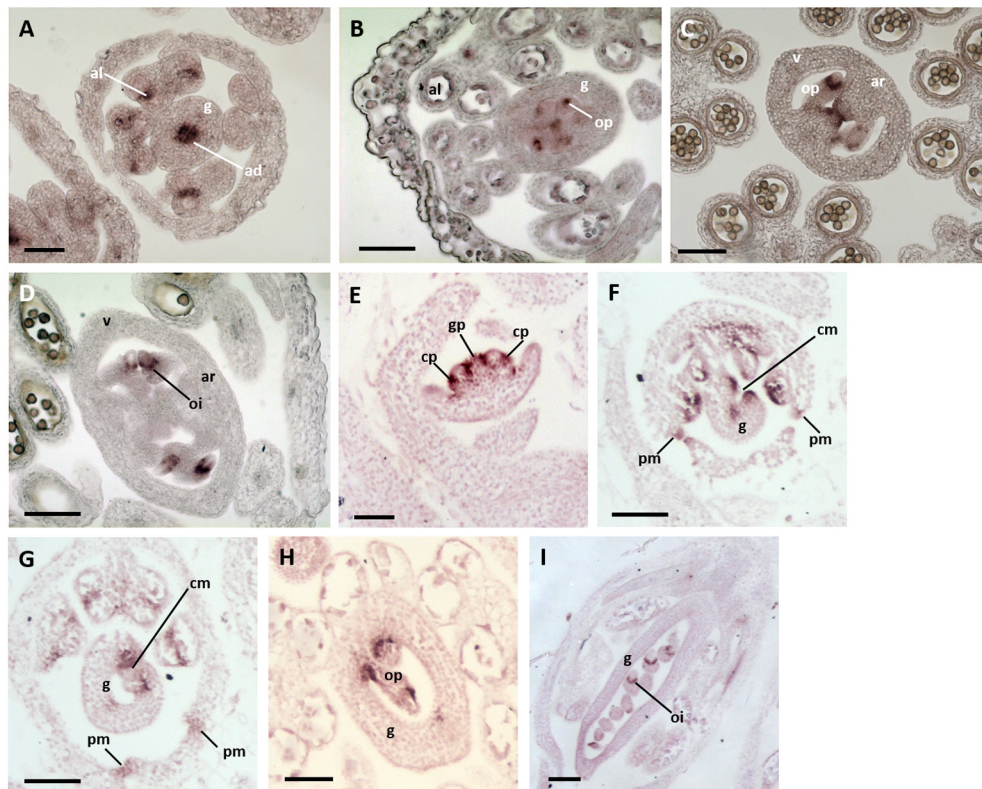


FIGURE 3 | Expression of NAM orthologs in *A. thaliana* and *Medicago truncatula* flower buds. (A–D) *A. thaliana* flower buds hybridized to *CUC1* and *CUC2* probes. **(A)** A Stage-7 bud showing *CUC2* expression in the abaxial domain of the gynoecium and anther loculi. **(B,C)** Buds at Stages 9–10 showing *CUC1* **(B)** and *CUC2* **(C)** expression in the placenta and within ovule primordia. **(D)** A bud at Stage 11 showing *CUC1* expression at the base of the expanding ovule integuments. ad, adaxial zone of the gynoecium; al, anther loculus; ar, abaxial replum; g, gynoecium; oi, ovule integuments; op, ovule primordium. Bars = 50 μ m. **(E–I)** *M. truncatula* flower buds hybridized to an *MtNAM* probe. **(E)** A bud at Stages 3–4 showing *MtNAM* expression between and within organ primordia. **(F)** A Stage-7 bud showing strong *MtNAM* expression in the unfused carpel margins. **(G)** A bud at Stage 8, in which *MtNAM* expression is reduced in the fused carpel margins. **(H,I)** Later stages of flower development in which *MtNAM* expression is absent in the carpel margins, but present first in ovule primordia **(H)** and then at the base of the expanding integuments **(I)**. cm, carpel margin; cp, common primordium; g, gynoecium; gp, gynoecial primordium; oi, ovule integument; op, ovule primordium; pm, petal margin. Bars = 100 μ m.

In situ hybridization in *M. truncatula* at Stages 3–4 of flower development, following the time course defined by Benlloch et al. (2003), showed *MtNAM* expression between the gynoecium primordium and the surrounding common primordia that give rise to both stamens and petals (Figure 3E). Signals were also detected within these common primordia (Figure 3E), marking the boundary between the zones destined to produce petals and stamens. At Stage 7, the gynoecium appeared crescent-shaped in transverse section and *MtNAM* was clearly expressed in the carpel margins, and at the margins of the developing free petals (Figure 3F). By early Stage 8, expression of *MtNAM* was observed to decline in the carpel margins (Figure 3G), which had, by this time, fused together to close the gynoecium. At later developmental stages, *MtNAM* expression is present in presumptive tissue boundaries in the elongating ovule primordia (Figure 3H) and, following this, at the base of the expanding integuments of the ovule (Figure 3I). Similar NAM-ortholog expression patterns in floral organ and ovule primordia were previously shown

in another species of Fabaceae, *Pisum sativum* (Blein et al., 2008).

These expression data reveal several underlying similarities in the expression of *miR164*-regulated NAM orthologs between *A. thaliana* and *M. truncatula*. These orthologs are highly expressed in both species at frontiers between and within floral organs, particularly during ovule development. These data do, however, reveal a difference in NAM expression in the carpel margins- no such expression was detected in the presumptive abaxial repla of the gynoecial tube at early stages of *A. thaliana* flower development, whereas NAM expression was detected in the carpel margins of the early *M. truncatula* gynoecium. This difference may relate to the contrasting modes of congenital and post-genital carpel margin fusion in *A. thaliana* and *M. truncatula*, respectively. Despite the differences observed, we concluded that the presence of *MtNAM* expression in *M. truncatula* carpel margins suggested that the NAM/*miR164* module may be involved in the fusion of these structures, leading us to test this hypothesis experimentally.

Expression of a *miR164*-Resistant form of *MtNAM* Leads to a Breakdown in Carpel Margin Fusion and Other Developmental Fusion Events in *Medicago truncatula* Flowers

To test the role of the *NAM/miR164* developmental module on carpel closure in *M. truncatula*, we produced transgenic plants expressing genomic constructs of *MtNAM* (*MtNAMg-m4* and *MtNAMg-wt*), respectively, with or without four point mutations in their predicted *miR164*-binding sites, identical to those present in the *CUC2g-m4* construct (Nikovics et al., 2006). Three independent transgenic *MtNAMg-m4* calli were generated, two of which were successfully regenerated into fertile adult plants, as was one transgenic callus containing an *MtNAMg-wt* construct (Table 1). Phenotypic observations were made on T2 progeny representative of one of each of these transformed lines, and on untransformed plants for comparison (Table 2; Figure 4).

Toluidene-blue staining was performed to highlight the ovule and the fused region of the carpel margins in the gynoecium of untransformed *M. truncatula* (Figure 4A). Transformation with *MtNAMg-wt* (Figures 4B–E) showed no effects on flower development compared to wild type *M. truncatula*. Accordingly, in *MtNAMg-wt* transformants, as in wild-type, five petals were produced, including two fused “keel” petals, two unfused “wing” petals, and a single “standard” petal (Figures 4B,C). As in the wild-type, all stamen filaments, with the exception of a single stamen positioned adjacent to the standard, were fused into a sheath surrounding the gynoecium (Figure 4D). The carpel margins of *MtNAMg-wt* transformants were also developmentally fused in the mature gynoecium, as in wild-type (Figure 4E).

By contrast, a range of mutant phenotypes were noted in flowers of plants transformed with the *MtNAMg-m4* construct (Table 2; Figures 4F–Q). Two standard petals were produced in some flowers (Figure 4K), while in others, petals with altered morphology and fusion were produced, rendering difficult their identification as standard, wing or keel petals (Figures 4G,O). Unfused stamens were produced in some cases (Table 2; Figures 4H,L), while stamens were absent in others (Table 2; Figure 4Q). The carpel margins remained unfused in many flowers (Table 2; Figures 4I,M), revealing the ovules within these, though a small proportion of flowers did show completely fused carpel margins (Table 2; Figure 4Q).

These data indicate a range of roles of the *NAM/miR164* developmental module in fusion events in the corolla, androecium, and gynoecium of *M. truncatula* flowers. Of

particular interest to the current work, the elimination of post-transcriptional regulation of *MtNAM* in the gynoecium is shown to have a similar effect in *M. truncatula* to that shown on *aux1-22* mutants of *A. thaliana* (Figure 2) by disrupting the fusion of carpel margins.

DISCUSSION

A Role of the *NAM/miR164* Module in the Fusion of Carpel Margins has Been Conserved at Least Since the MRCA of the Eurosids

In this study, we show that a previously characterized developmental module involving the post-transcriptional regulation of *NAM* orthologs by *miR164* is involved not only in carpel fusion in syncarpous *A. thaliana* (Nikovics et al., 2006; Sieber et al., 2007), but also in the closure of the single carpels present in two species whose lineages diverged at the base of the eurosid clade, some 114–113 MYA. The two species concerned are *A. thaliana* itself, as *aux1-22* mutants of *A. thaliana* produce single carpels, and *M. truncatula*, which is monocarpous in its wild-type form. We show that disruption of the *NAM/miR164* module in both *A. thaliana aux1-22* mutants (Figure 2) and a wild-type background of *M. truncatula* (Figure 4) produces single carpels that are no longer completely fused at their margins.

These data indicate that the *NAM/miR164* module has conserved a role in developmental fusion events between carpel margins at least since the MRCA of the eurosids. The mapping of character states onto angiosperm phylogeny (Figure 1) indicates that the MRCA of the eurosids was syncarpous, and we may thus conclude that the *NAM/miR164* module contributed to carpel fusion in that key ancestor, from which some 70 000 extant species are descended (Wang et al., 2009).

The *NAM/miR164* Module Maintained its Role in Carpel Margin Fusion During a Transition from Syncarpy to Monocarpy in an Ancestor of Fabales

Character-state mapping (Figure 1) further indicates that the monocarpy present in Fabales (including *M. truncatula*) is a derived condition that occurred by reversion from syncarpy, present in earlier eurosids. In the present work, we show that the role of the *NAM/miR164* module in carpel margin fusion was conserved during this developmental transition. Thus, our study

TABLE 1 | Results summary for transformation of *Medicago truncatula*.

Constructs	Number of transformation experiments performed	Number of transgenic calli produced (combining all experiments)	Number of plantlets regenerated	Number of T1 plants surviving to reproductive phase
<i>MtNAMg-wt</i>	2	7 (cal 1–7)	7 from cal 1	2 from cal 1
<i>MtNAMg-m4</i>	2	3 (cal 1–3)	18 from cal 1 1 from cal 2 1 from cal 3	12 from cal 1 1 from cal 2

TABLE 2 | Floral phenotypes of five representative T2 *MtNAM/g-m4* transformants.

T1 parent	T2 plant	Total number of flowers dissected	Number of flowers showing abnormalities in the corolla	Number of flowers showing abnormalities in the androecium		Number of flowers showing abnormalities in the gynoecium	
				Unfused stamens	Stamens absent	Slight defects in carpel fusion	Extensive defects in carpel fusion
m4 cal 1	1_1	3	3	0	0	0	0
m4 cal 1	1_2	17	6	2	0	2	7
m4 cal 1	4_1	13	11	3	0	0	4
m4 cal 1	5_1	14	5	0	0	3	2
m4 cal 1	6_3	12	12	4	5	2	8

strongly suggests that the *NAM/miR164* module provides an underlying mechanism that is necessary for fusion events at the carpel margins of both syncarpous and monocarpous eurosids.

It is interesting to note that the *aux1-22* mutation in *A. thaliana* causes a transition from a congenitally fused gynoecium of two carpels to a closed, monocarpous gynoecium. Thus, a single loss-of-function mutation in a gene involved in auxin signaling can bring about, in *A. thaliana*, a similar type of morphological transition to that which led to monocarpy in Fabales. The genetic simplicity of this transition suggests that reversions from syncarpy to monocarpy might occur frequently in natural populations. The general trend in the angiosperms, however, is for evolutionary transitions toward syncarpous gynoecia, which are believed to confer numerous selective advantages (Armbruster et al., 2002). Thus, while the loss of syncarpy may be a genetically “easy” transition to make, the fixation of this trait in populations by natural selection may occur much less frequently.

A Possible Role for the *NAM/miR164* Module in the Timing of Carpel Fusion

In *A. thaliana*, the gynoecium forms as a radially symmetrical cylinder that later differentiates to show the positions of the carpel margins. By contrast, the single carpel of the *M. truncatula* gynoecium is plicate, and closes post-genitally by the fusion of preexisting carpel margins. *In situ* hybridization in this work (Figure 3) and other studies (Galbiati et al., 2013) failed to detect any expression of *CUC1* or *CUC2* in the carpel margins of *A. thaliana*. However, *CUC2* is known to be highly expressed in the carpel margins of the unfused gynoecium at Stage 9 of flower development in *mir164abc* triple mutants (Sieber et al., 2007). Comparison of these data strongly suggests that the *NAM/miR164* expression balance in *A. thaliana* lies heavily in favor of *miR164* from the earliest stages of gynoecium development. By contrast, detectable levels of *MtNAM* were present in margin tissues at early stages of *M. truncatula* carpel development, and these levels were observed to decline at subsequent stages, as the margins fused (Figure 3). Thus, the different balances of *NAM* and *miR164* expression observed at very early stages of *A. thaliana* and *M. truncatula* carpel development (Figure 3; Galbiati et al., 2013) correlate closely with the different timings of carpel closure observed in these species (Smyth et al., 1990; Benlloch et al., 2003).

Given the role of the *NAM/miR164* module in carpel closure in both *A. thaliana* and *M. truncatula* (Figures 2 and 4), and the gene expression differences we have noted between the congenitally and post-genitally fused carpel margins of these two respective species, it would be interesting to compare the expression of *NAM* orthologs in a range of Fabales that show different spatial and temporal patterns of carpel closure. Candidate species for this analysis include *Acacia celastrifolia* and *Inga bella* (Paulino et al., 2014), in which the carpels include both congenitally fused (ascidiate) and later-fusing (plicate) zones, and *Amberstia nobilis* and *Caesalpina* spp. (Tucker and Kantz, 2001), in which the carpel margins remain unfused until after ovule initiation, much later than in most other Fabales.

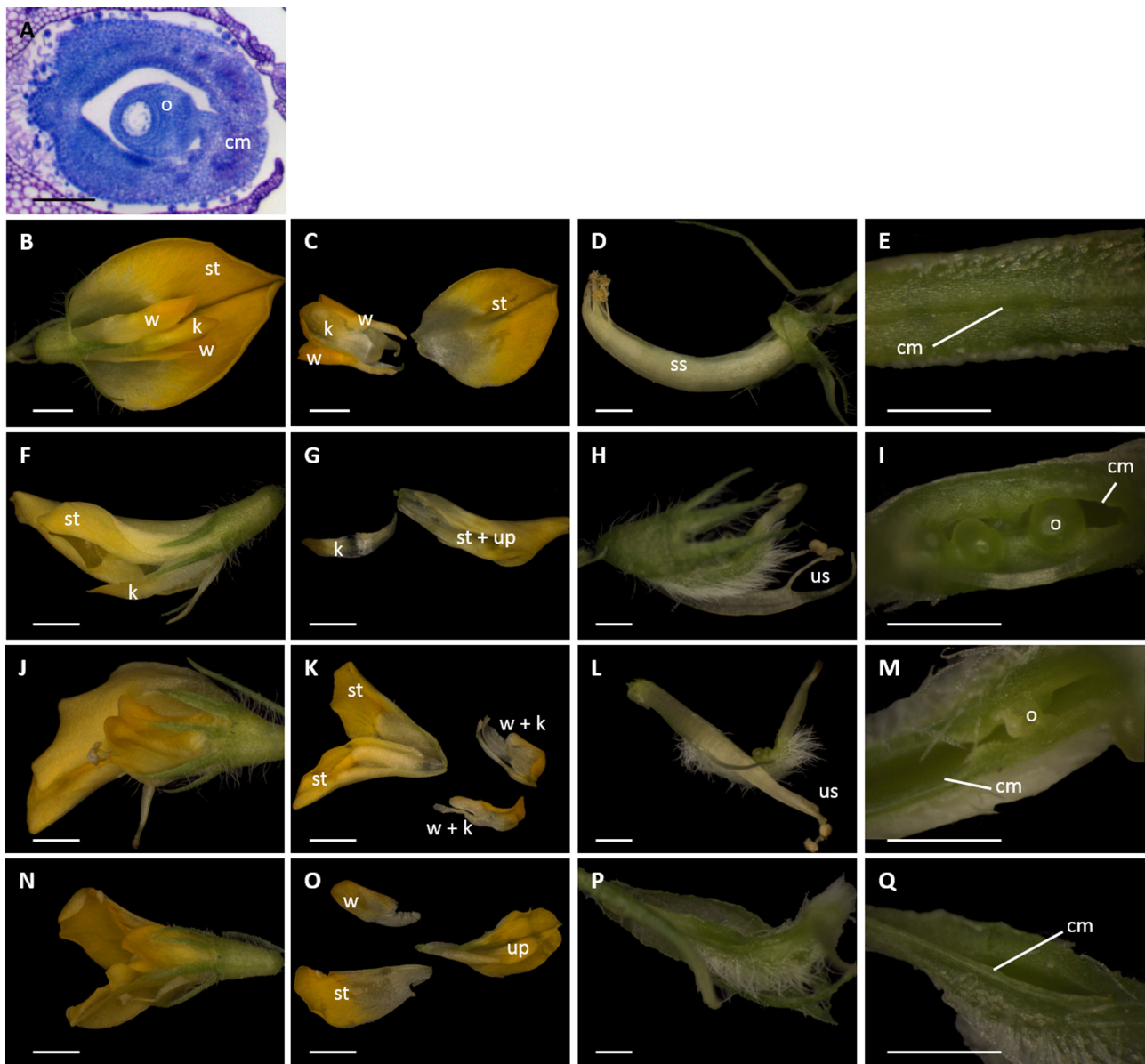


FIGURE 4 | Dissections of *M. truncatula* flowers transformed with *miR164*-resistant (*MtNAMg-m4*) or wild-type control (*MtNAMg-wt*) constructs. (A) Transverse section of wild-type *M. truncatula* gynoecium stained with toluidine blue. **(B–E)** A typical flower of an *MtNAMg-wt* transformant showing **(B)** the intact flower, **(C)** petal morphology, **(D)** the sheath of anther filaments surrounding the gynoecium, and **(E)** the carpel margins. All structures in **(B–E)** appear identical to wild-type. **(F–I)**, **(J–M)**, and **(N–Q)** Three representative flowers from *MtNAMg-m4* transformants showing **(F,J,N)** the intact flower, **(G,K,O)** petal morphology, **(H,L,P)** after removal of the perianth, and **(I,M,Q)** the carpel margins. Defects in the corolla, androecium and gynoecium are apparent, including a marked breakdown in carpel margin fusion in most flower buds (e.g., **I,M**). cm, carpel margin; k, keel petal(s); o, ovule; ss, stamen sheath; st, standard petal; up, unidentified petal(s); us, unfused stamens; w, wing petal. Bars = 100 μ m in **(A)**, 1 mm in **(B,C,F,G,J,K,N,O)**, and 0.5 mm in **(D,E,H,I,L,M,P,Q)**.

Such experiments, in quite closely related species showing marked differences in gynoecium anatomy, could provide strong correlative evidence of a role for the subtle modulation of gynoecium development by changes to the balance of the *NAM/miR164* module. Notably, the *NAM/miR164* expression balance at very early stages of carpel development may be important in determining whether carpel margins will fuse congenitally or postgenitally.

The Role of the *NAM/miR164* Module in Carpel Evolution

As its genetic components are present in both gymnosperms and angiosperms (Axtell and Bartel, 2005; Larsson et al., 2012), the *NAM/miR164* genetic module is clearly of ancient origin in seed plants. This module is involved in leaf, carpel, and ovule development in model angiosperms (Nikovics et al., 2006; Blein et al., 2008; Galbiati et al., 2013; Goncalves et al., 2015),

while expression studies in *Amborella trichopoda*, the only living representative of Amborellales (see **Figure 1**), and hence the likely sister to all other living angiosperms, suggest that its role in ovule development, at least, has been conserved from the earliest stages of angiosperm evolution (Viallette-Guiraud et al., 2011a).

Like most basally diverging angiosperms, *A. trichopoda* is apocarpous and has ascidiate carpels. Thus, in both *A. trichopoda* and *A. thaliana*, the carpel margins are congenitally fused from the earliest stages of gynoecium development, albeit in the different contexts of apocarpy and syncarpy, respectively. No expression of the *NAM* ortholog from *A. trichopoda*, *AtrNAM*, was observed in the early carpel wall (Viallette-Guiraud et al., 2011a), as is the case for *CUC1* and *CUC2* in *A. thaliana* (Galbiati et al., 2013; **Figure 3**). Thus, it appears reasonable to postulate that the *NAM/miR164* module operates in favor of the expression of *miR164*, and against that of *NAM* orthologs, from the earliest stages of gynoecium development in *A. trichopoda*, as it does in *A. thaliana*.

From the above observations, we hypothesize that the *NAM/miR164* module may have played a role in the fusion of carpel margins in the MRCA of the living angiosperms, as it does in present-day model angiosperms. An important test of this hypothesis will depend on the development of plant transformation strategies in basally diverging angiosperms, which would allow, for example, the transformation of *A. trichopoda* with a *miR164*-resistant form of *AtrNAM*. Comparison of early diverging angiosperm lineages strongly suggests that the first flowering plants possessed ascidiate, rather than plicate carpels (Endress and Igersheim, 2000). Accordingly, we furthermore hypothesize, based on our gene expression analyses (**Figure 3**), that the origin of plicate carpels in various later-emerging angiosperm lineages may have depended on subtle modifications to the *NAM/miR164* module that allowed a limited level of early expression of *NAM* orthologs in the carpel margins, as occurs in present-day *M. truncatula*.

Interestingly, it is known that in *A. thaliana*, the role of *CUC2* in the closure of the gynoecium apex is under indirect negative control by the bHLH transcription factor SPATULA (SPT; Nahar et al., 2012). In addition, SPT is known to play a role in carpel fusion along the entire length of the gynoecium, which is revealed in double-mutant combinations with the YABBY transcription

factor *CRABS CLAW* (*crc-1 spt-2*; Alvarez and Smyth, 1999). Like the *NAM/miR164* module, it seems that SPT may have conserved its function in carpel development from the earliest stages of angiosperm evolution (Reymond et al., 2012). Thus, the establishment of negative regulation by SPT of a *miR164*-regulated *NAM* gene in a common ancestor of the angiosperms may have been a crucial step in the evolution of the closed carpel. Analysis of the pathway linking SPT, and its cofactors such as the HECATE transcription factors (Schuster et al., 2015), with the *NAM/miR164* module in model angiosperms could provide insights into this possibility, and thus potentially indicate a molecular mechanism for the enclosure of the ovule with the carpel in the first angiosperms.

AUTHOR CONTRIBUTIONS

AV-G performed all of the work except Medicago transformation, prepared the figures and collaborated with CS to plan and write the paper. AC assisted with *in situ* hybridizations of *Arabidopsis*. JG-M assisted with *in situ* hybridizations of Medicago. AE performed Medicago transformations. PR supervised Medicago transformations. CS supervised all work except Medicago transformation and collaborated with AV-G to plan and write the paper.

FUNDING

This work was supported by research grants ANR-13-BSV2-0009 “ORANGE” to CS and ANR-11-BSV2-0005 “Charmful” to PR. AV-G is funded through an ENS-Lyon research and teaching position.

ACKNOWLEDGMENTS

We thank Téva Vernoux for supplying *aux1-22* seed and Patrick Laufs for making available the *CUC2g-m4* and *CUC2g-wt* constructs. We are grateful to Patrick Laufs and Mike Frohlich for helpful discussions.

REFERENCES

- Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., and Tasaka, M. (1997). Genes involved in organ separation in *Arabidopsis*: an analysis of the cup-shaped cotyledon mutant. *Plant Cell* 9, 841–857. doi: 10.1105/tpc.9.6.841
- Aida, M., Ishida, T., and Tasaka, M. (1999). Shoot apical meristem and cotyledon formation during *Arabidopsis* embryogenesis: interaction among the CUP-SHAPED COTYLEDON and SHOOT MERISTEMLESS genes. *Development* 126, 1563–1570.
- Alvarez, J., and Smyth, D. R. (1999). *CRABS CLAW* and *SPATULA*, two *Arabidopsis* genes that control carpel development in parallel with *AGAMOUS*. *Development* 126, 2377–2386.
- Armbruster, W. S., Debevec, E. M., and Willson, M. F. (2002). Evolution of syncarpy in angiosperms: theoretical and phylogenetic analyses of the effects of carpel fusion on offspring quantity and quality. *J. Evol. Biol.* 15, 657–672. doi: 10.1046/j.1420-9101.2002.00414.x
- Axtell, M. J., and Bartel, D. P. (2005). Antiquity of microRNAs and their targets in land plants. *Plant Cell* 17, 1658–1673. doi: 10.1105/tpc.105.032185
- Benlloch, R., Navarro, C., and Beltran, J. P. (2003). Floral development of the model legume *Medicago truncatula*: ontogeny studies as a tool to better characterize homeotic mutations. *Sex. Plant Reprod.* 15, 231–241. doi: 10.1007/s00487-002-0157-1
- Bennett, M. J., Marchant, A., Green, H. G., May, S. T., Ward, S. P., Millner, P. A., et al. (1996). *Arabidopsis* AUX1 gene: a permease-like regulator of root gravitropism. *Science* 273, 948–950. doi: 10.1126/science.273.5277.948
- Bilsborough, G. D., Runions, A., Barkoulas, M., Jenkins, H. W., Hasson, A., Galinha, C., et al. (2011). Model for the regulation of *Arabidopsis thaliana* leaf margin development. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3424–3429. doi: 10.1073/pnas.1015162108
- Blein, T., Pulido, A., Viallette-Guiraud, A., Nikovics, K., Morin, H., Hay, A., et al. (2008). A conserved molecular framework for compound leaf development. *Science* 322, 1835–1839. doi: 10.1126/science.1166168

- Bremer, B., Bremer, K., Chase, M. W., Fay, M. F., Reveal, J. L., Soltis, D. E., et al. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi: 10.1016/j.jep.2015.05.035
- Cheng, X., Peng, J., Ma, J., Tang, Y., Chen, R., Mysore, K. S., et al. (2012). NO APICAL MERISTEM (MtNAM) regulates floral organ identity and lateral organ separation in *Medicago truncatula*. *New Phytol.* 195, 71–84. doi: 10.1111/j.1469-8137.2012.04147.x
- Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x
- Cosson, V., Eschstruth, A., and Ratet, P. (2015). *Medicago truncatula* transformation using leaf explants. *Methods Mol. Biol.* 1223, 43–56. doi: 10.1007/978-1-4939-1695-5_4
- Endress, P. K., and Igersheim, A. (2000). Gynoecium structure and evolution in basal angiosperms. *Int. J. Plant Sci.* 161, S211–S223. doi: 10.1086/317572
- Galbiati, F., Sinha Roy, D., Simonini, S., Cucinotta, M., Ceccato, L., Cuesta, C., et al. (2013). An integrative model of the control of ovule primordia formation. *Plant J.* 76, 446–455. doi: 10.1111/tpj.12309
- Goncalves, B., Hasson, A., Belcram, K., Cortizo, M., Morin, H., Nikovics, K., et al. (2015). A conserved role for CUP-SHAPED COTYLEDON genes during ovule development. *Plant J.* 83, 732–742. doi: 10.1111/tpj.12923
- Hellens, R., Mullineaux, P., and Klee, H. (2000). Technical focus: a guide to *Agrobacterium* binary Ti vectors. *Trends Plant Sci.* 5, 446–451. doi: 10.1016/S1360-1385(00)01740-4
- Ishida, T., Aida, M., Takada, S., and Tasaka, M. (2000). Involvement of CUP-SHAPED COTYLEDON genes in gynoecium and ovule development in *Arabidopsis thaliana*. *Plant Cell Physiol.* 41, 60–67. doi: 10.1093/pcp/41.1.60
- Kamiuchi, Y., Yamamoto, K., Furutani, M., Tasaka, M., and Aida, M. (2014). The CUC1 and CUC2 genes promote carpel margin meristem formation during *Arabidopsis* gynoecium development. *Front. Plant Sci.* 5:165. doi: 10.3389/fpls.2014.00165
- Kirsch, R. D., and Joly, E. (1998). An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res.* 26, 1848–1850. doi: 10.1093/nar/26.7.1848
- Koncz, C., and Schell, J. (1986). The promoter of TL-DNA Gene 5 controls the tissue-specific expression of chimeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* 204, 383–396. doi: 10.1007/BF00331014
- Larsson, E., Sundstrom, J. F., Sitbon, F., and von Arnold, S. (2012). Expression of PaNAC01, a Picea abies CUP-SHAPED COTYLEDON orthologue, is regulated by polar auxin transport and associated with differentiation of the shoot apical meristem and formation of separated cotyledons. *Ann. Bot.* 110, 923–934. doi: 10.1093/aob/mcs151
- Mallory, A. C., Dugas, D. V., Bartel, D. P., and Bartel, B. (2004). MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. *Curr. Biol.* 14, 1035–1046. doi: 10.1016/j.cub.2004.06.022
- Nahar, M. A.-U., Ishida, T., Smyth, D. R., Tasaka, M., and Aida, M. (2012). Interactions of CUP-SHAPED COTYLEDON and SPATULA genes control carpel margin development in *Arabidopsis thaliana*. *Plant Cell Physiol.* 53, 1134–1143. doi: 10.1093/pcp/pcs057
- Nam, Y.-W., Penmetts, R. V., Endre, G., Uribe, P., Kim, D., and Cook, D. R. (1999). Construction of a bacterial artificial chromosome library of *Medicago truncatula* and identification of clones containing ethylene-response genes. *Theor. Appl. Genet.* 98, 638–646. doi: 10.1007/s001220051115
- Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., et al. (2006). The balance between the MIR164A and CUC2 genes controls leaf margin serration in *Arabidopsis*. *Plant Cell* 18, 2929–2945. doi: 10.1105/tpc.106.045617
- Paulino, J. V., Prenner, G., Mansano, V. F., and Teixeira, S. P. (2014). Comparative development of rare cases of a polycarpellate gynoecium in an otherwise monocarpellate family, Leguminosae. *A. J. Bot.* 101, 572–586. doi: 10.3732/ajb.1300355
- Peaucelle, A., Morin, H., Traas, J., and Laufs, P. (2007). Plants expressing a miR164-resistant CUC2 gene reveal the importance of post-meristematic maintenance of phyllotaxy in *Arabidopsis*. *Development* 134, 1045–1050. doi: 10.1242/dev.02774
- Reymond, M. C., Brunoud, G., Chauvet, A., Martinez-Garcia, J. F., Martin-Magniette, M.-L., Moneger, F., et al. (2012). A light-regulated genetic module was recruited to carpel development in *Arabidopsis* following a structural change to SPATULA. *Plant Cell* 24, 2812–2825. doi: 10.1105/tpc.112.097915
- Schuster, C., Gaillochet, C., and Lohmann, J. (2015). *Arabidopsis* HECATE genes function in phytohormone control during gynoecium development. *Development* 142, 3343–3350. doi: 10.1242/dev.120444
- Sieber, P., Wellmer, F., Gheyselinck, J., Riechmann, J. L., and Meyerowitz, E. M. (2007). Redundancy and specialization among plant microRNAs: role of the MIR164 family in developmental robustness. *Development* 134, 1051–1060. doi: 10.1242/dev.02817
- Smyth, D., Bowman, J., and Meyerowitz, E. (1990). Early flower development in *Arabidopsis*. *Plant Cell* 2, 755–767. doi: 10.2307/3869174
- Souer, E., vanHouwelingen, A., Kloos, D., Mol, J., and Koes, R. (1996). The NO APICAL MERISTEM gene of petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell* 85, 159–170. doi: 10.1016/S0092-8674(00)81093-4
- Takada, S., Hibara, K., Ishida, T., and Tasaka, M. (2001). The CUP-SHAPED COTYLEDON1 gene of *Arabidopsis* regulates shoot apical meristem formation. *Development* 128, 1127–1135.
- Tucker, S. C., and Kantz, K. E. (2001). Open carpels with ovules in Fabaceae. *Int. J. Plant Sci.* 162, 1065–1073. doi: 10.1086/321923
- Viallette-Guiraud, A. C. M., Adam, H., Finet, C., Jasinski, S., Jouannic, S., and Scutt, C. P. (2011a). Insights from ANA-grade angiosperms into the early evolution of CUP-SHAPED COTYLEDON genes. *Ann. Bot.* 107, 1511–1519. doi: 10.1093/aob/mcr024
- Viallette-Guiraud, A. C. M., Alaux, M., Legeai, F., Finet, C., Chambrier, P., Brown, S. C., et al. (2011b). Cabomba as a model for studies of early angiosperm evolution. *Ann. Bot.* 108, 589–598. doi: 10.1093/aob/mcr088
- Wang, H., Moore, M. J., Soltis, P. S., Bell, C. D., Brockington, S. F., Alexandre, R., et al. (2009). Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3853–3858. doi: 10.1073/pnas.0813376106
- Weir, I., Lu, J. P., Cook, H., Causier, B., Schwarz-Sommer, Z., and Davies, B. (2004). CUPULIFORMIS establishes lateral organ boundaries in *Antirrhinum*. *Development* 131, 915–922. doi: 10.1242/dev.00993

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Viallette-Guiraud, Chauvet, Gutierrez-Mazariegos, Eschstruth, Ratet and Scutt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Prevalent Exon-Intron Structural Changes in the *APETALA1/FRUITFULL*, *SEPALLATA*, *AGAMOUS-LIKE6*, and *FLOWERING LOCUS C* MADS-Box Gene Subfamilies Provide New Insights into Their Evolution

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington, USA

Reviewed by:

Stefan Gleissberg,
Gleissberg.org, USA
Amy Litt,
University of California, Riverside, USA
Ji Yang,
Fudan University, China

*Correspondence:

Guixia Xu
xuguixia1982@ibcas.ac.cn;
Hongyan Shan
shanhongyan@ibcas.ac.cn

†These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 01 October 2015

Accepted: 18 April 2016

Published: 02 May 2016

Citation:

Yu X, Duan X, Zhang R, Fu X, Ye L,
Kong H, Xu G and Shan H (2016)
Prevalent Exon-Intron Structural
Changes in the
APETALA1/FRUITFULL, *SEPALLATA*,
AGAMOUS-LIKE6, and *FLOWERING*
LOCUS C MADS-Box Gene
Subfamilies Provide New Insights into
Their Evolution.
Front. Plant Sci. 7:598.
doi: 10.3389/fpls.2016.00598

Xianxian Yu^{1,2†}, Xiaoshan Duan^{1,2†}, Rui Zhang¹, Xuehao Fu^{1,2}, Lingling Ye^{1,2},
Hongzhi Kong¹, Guixia Xu^{1*} and Hongyan Shan^{1*}

¹ State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China, ² University of Chinese Academy of Sciences, Beijing, China

AP1/FUL, *SEP*, *AGL6*, and *FLC* subfamily genes play important roles in flower development. The phylogenetic relationships among them, however, have been controversial, which impedes our understanding of the origin and functional divergence of these genes. One possible reason for the controversy may be the problems caused by changes in the exon-intron structure of genes, which, according to recent studies, may generate non-homologous sites and hamper the homology-based sequence alignment. In this study, we first performed exon-by-exon alignments of these and three outgroup subfamilies (*SOC1*, *AG*, and *STK*). Phylogenetic trees reconstructed based on these matrices show improved resolution and better congruence with species phylogeny. In the context of these phylogenies, we traced evolutionary changes of exon-intron structures in each subfamily. We found that structural changes have occurred frequently following gene duplication and speciation events. Notably, exons 7 and 8 (if present) suffered more structural changes than others. With the knowledge of exon-intron structural changes, we generated more reasonable alignments containing all the focal subfamilies. The resulting trees showed that the *SEP* subfamily is sister to the monophyletic group formed by *AP1/FUL* and *FLC* subfamily genes and that the *AGL6* subfamily forms a sister group to the three abovementioned subfamilies. Based on this topology, we inferred the evolutionary history of exon-intron structural changes among different subfamilies. Particularly, we found that the eighth exon originated before the divergence of *AP1/FUL*, *FLC*, *SEP*, and *AGL6* subfamilies and degenerated in the ancestral *FLC*-like gene. These results provide new insights into the origin and evolution of the *AP1/FUL*, *FLC*, *SEP*, and *AGL6* subfamilies.

Keywords: *APETALA1/FRUITFULL*, *SEPALLATA*, *AGAMOUS-LIKE6*, *FLOWERING LOCUS C*, exon-intron structural change

INTRODUCTION

MADS-box genes encode a family of transcription factors that have been found in plants, animals, and fungi (Theissen et al., 2000; Becker and Theissen, 2003; Ferrario et al., 2004; Causier et al., 2010; Rijpkema et al., 2010). In plants, the best-studied MADS-box genes are those involved in the specification of floral meristem and floral organ identities. Protein products of these genes are characterized by existence of four regions: the MADS (M) domain, the intervening (I) region, the keratin-like (K) domain, and the C-terminal (C) region (Theissen et al., 1996; Nam et al., 2003). Extensive phylogenetic studies have revealed that these MADS-box genes belong to eight different subfamilies or lineages: *APETALA1* (*API*)/*FRUITFULL* (*FUL*), *APETALA3* (*AP3*), *PISTILLATA* (*PI*), *AGAMOUS* (*AG*), *SEEDSTICK* (*STK*), *SEPALLATA1* (*SEP1*), *SEPALLATA3* (*SEP3*), and *AGAMOUS-LIKE6* (*AGL6*) (reviewed in Theissen et al., 2000; Becker and Theissen, 2003; Nam et al., 2003). Among these, the evolutionary histories of the *AP3*, *PI*, *AG*, and *STK* subfamilies are relatively clear and can be traced back to the most recent common ancestor (MRCA) of extant seed plants (Aoki et al., 2004; Kramer et al., 2004; Dreni and Kater, 2013; Dreni et al., 2013). The relationships among the remainder four subfamilies, however, are still controversial, although the sisterhood of *SEP1* and *SEP3* (collectively called *SEP*) has got consistent support. In some studies, *SEP* was resolved as the sister of *API/FUL* (Carlsbecker et al., 2003; Litt and Irish, 2003; Kim et al., 2005; Futamura et al., 2008; Li et al., 2010), whereas in others, it forms a sister to *AGL6* (Kofuji et al., 2003; Nam et al., 2003; Parenicova, 2003; Zahn et al., 2005; Litt, 2007; *Amborella* Genome Project, 2013; Kim et al., 2013; Ruelens et al., 2013; Ubi et al., 2013; Wong et al., 2013; Yockteng et al., 2013). Interestingly, if the former scenario is correct, then it implies that both *API/FUL* and *SEP* have originated before the diversification of angiosperms; otherwise, it implies that both *API/FUL* and *SEP* have existed in the MRCA of extant seed plants but have been independently lost in the lineage leading to extant gymnosperms. The observation that the *FLOWERING LOCUS C* (*FLC*) may be the real sister of *API/FUL* (Ruelens et al., 2013) further complicated the issue, making it necessary to re-investigate the relationships among the aforementioned gene subfamilies.

Many factors, such as biased sampling, long-branch attraction, and heterogeneous substitution rates, can lead to skewed topology of a phylogenetic tree (Kong et al., 2004; Leebens-Mack et al., 2005). However, the most important factor is the reliability of the alignment used for phylogeny estimation. Since using only conserved regions would reduce resolution, most studies include as many as possible alignable sites. Yet, it has recently been revealed that changes in the exon-intron structure of genes (i.e., structural changes, which may be caused by exon/intron gain/loss, exonization/pseudoexonization, and intraexonic insertion/deletion; Roy and Gilbert, 2005; Xu et al., 2012; Long et al., 2013) may hamper the homology-based alignment because they may lead to the addition of nonhomologous sequence or removal of homologous nucleotide. Since almost all studies only used coding sequences (CDS) or protein sequences to generate their alignment, nonhomologous

sites caused by structural changes could be forced to align together. In the MADS-box gene family, structural changes have been shown to be rather common and can indeed cause shifts of reading frame (Litt and Irish, 2003; Vandenbussche et al., 2003a; Litt, 2007; Shan et al., 2007; Xu and Kong, 2007; Liu et al., 2011; Xu et al., 2012). A good example comes from comparing the three core eudicots lineages of the *API/FUL* subfamily: eu*FUL*, eu*AGL79* (also called core eudicot *FUL*-like), and eu*API* (Litt and Irish, 2003; Litt, 2007; Shan et al., 2007). Proteins encoded by the first two lineages have a paleo*API* motif at the C-terminal region, the first six amino acids of which were also defined as *FUL*-like motif in some studies (Litt and Irish, 2003; Litt, 2007) and show high similarity with part of *AGL6* II and *SEP* II motifs. The eu*API* lineage, however, encodes for a quite different C-terminal region with two different motifs: a transcription activation domain and a eu*API* motif, the final four amino acids of which were also called farnesylation motif (Litt and Irish, 2003; Litt, 2007). Detailed investigation revealed that the novel sequence was generated by a 1-bp deletion in exon 8 of the ancestral eu*API* gene (Litt and Irish, 2003; Vandenbussche et al., 2003a; Litt, 2007; Shan et al., 2007). Similarly, an 8-bp insertion (Vandenbussche et al., 2003a) or a 1-bp deletion (Kramer et al., 2006) in the last exon has likely given rise to a new eu*AP3* motif in the eu*AP3* lineage of the *AP3* subfamily. During phylogenetic reconstruction of the *API/FUL*, *SEP*, *AGL6*, and *FLC* subfamilies, however, none of the previous studies considered exon-intron structural changes when generating the final alignment, which may explain why different studies have obtained slightly different topologies.

In this article, we first investigated structural changes during the evolution of these and related subfamilies such as *SUPPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*), *AG*, and *STK*. We found that structural changes have occurred frequently in these subfamilies and could indeed affect phylogenetic estimation and the understanding of gene evolution. With the knowledge of structural changes, we generated more reasonable alignments containing all the focal subfamilies. All the resulting trees support the sisterhood of *API/FUL* and *FLC*, with *SEP* and *AGL6* being successive sisters to them. In the context of this new topology, we discussed the contribution of structural changes to the origin and functional diversification of different subfamilies.

MATERIALS AND METHODS

Sequence Retrieval and Classification

The protein, coding, and genomic (if available) sequences of focal MADS-box genes were retrieved by BLAST searches against the GenBank (<http://www.ncbi.nlm.nih.gov>), FGP (<http://fgp.bio.psu.edu>), Phytozome (<http://phytozome.jgi.doe.gov>), *Amborella* Genome Database (<http://www.amborella.org>), TAIR (<https://www.arabidopsis.org>), MPOB (<http://genomsawit.mpob.gov.my>), and PlantGDB (<http://www.plantgdb.org>) databases, with multiple sequences being used as queries. The resulting dataset was then trimmed by the following strategies. First, CDSs shorter than 400 bp were excluded, because they are not very informative or accurate. Second, all but one of the multiple highly similar (i.e., >95% identical at the CDS level) sequences from the same species

were eliminated, because they represent alleles of the same gene. Third, for genes with alternative splicing, only the transcript showing the least structural divergence from closely related homologs was adopted. And fourth, poorly annotated sequences from whole-genome sequenced species were excluded. As a result, 792 sequences were retained for further analyses.

To assign the retained sequences into different subfamilies, we built a preliminary phylogenetic tree (using the same methods described below) with shared regions (**Dataset S1**). The matrix for every subfamily has a broad taxonomic coverage, including sequences from early-diverging angiosperms, monocots, magnoliids, basal eudicots, core eudicots, and gymnosperm species (if applicable). Detailed information of genes included in this study was listed in **Table S1**.

Sequence Alignment and Phylogenetic Reconstruction

For each subfamily, protein sequences were initially aligned using ClustalX 1.83 with default options (Thompson et al., 1997), and its corresponding codon-based CDS alignment was generated by the PAL2NAL program (<http://www.bork.embl.de/pal2nal/>). A preliminary tree was constructed with the CDS alignment excluding poorly aligned regions (i.e., columns). The sequences in both protein and CDS alignments were then reordered according to their phylogenetic placements as well as the phylogenetic relationships among species. By comparing closely related sequences, we were able to determine homologous sites and refine the alignments. Considering the effect of structural changes on the reliability of alignment, we marked the exon-intron boundaries for genes with structural annotation (from genome-sequenced species) and carefully checked the alignments of neighboring sequences exon by exon. Special attention was paid to the exons that showed considerable divergence in sequences or lengths, in which structural changes have likely occurred. To improve the alignment quality, a pairwise alignment was performed by using both focal exons and their flanking noncoding sequences. Referring to these results, the CDS alignment can be adjusted with confidence, which were carried out in MEGA 6.0 (Tamura et al., 2013). Since our alignments involved human judgment and might be arbitrary, we also generated an amino acid alignment using Probalign (Roshan and Livesay, 2006) for each subfamily and its corresponding codon-based CDS alignment. Eventually, the CDS alignments excluding nonhomologous and highly divergent regions/sites were used for phylogenetic analyses.

To estimate the phylogenetic relationships among different subfamilies, we generated a combined matrix using the “profile-profile alignment” method in Muscle 3.6 (Edgar, 2004), followed by manual adjustments as described above. To maximize the reliability of our phylogenetic analyses, we created three different alignments (I, II, and III). For alignment I, all the 792 sequences were included. Alignment II contained 498 sequences with the exclusion of genes or gene lineages that experienced structural changes shortly after gene duplications. More stringently, in alignment III, we only included 57 exemplars from basal

angiosperms, basal eudicots and gymnosperms (if applicable), which showed less structural divergence during evolution (for details, see results). Because no *FLC*-like gene has ever been identified from basal angiosperms and basal eudicots (Ruelens et al., 2013; this study), *FLC*-like genes from core eudicot species and *Musa* were used for this subfamily. For all the alignments, only homologous sites and regions were used for phylogenetic analyses (**Dataset S9**).

Phylogenetic relationships of genes within each subfamily were revealed by the maximum-likelihood (ML) method, which was performed on the DNA matrix with PhyML (version 2.4) (Guindon and Gascuel, 2003). The most appropriate molecular evolution model (GTR+I+ Γ) was selected, following the estimate with MODELTEST version 3.06 (Posada and Crandall, 1998). A BIONJ tree was used as a starting point for ML searches (Guindon and Gascuel, 2003), and bootstrap analyses were performed with 100 replicates. In addition to the ML method, we also performed Bayesian inference (BI; Ronquist et al., 2012) for alignments I, II, and III to confirm the phylogenetic relationships among the *API1/FUL*, *SEP*, *AGL6*, and *FLC* subfamilies. We ran four chains, sampling one tree every 1000 generations for 15,000,000 generations using GTR+I model (starting with a random tree). The first 25% trees were considered burn-in and discarded from further analysis.

Determination of Exon-Intron Structural Changes

To understand the history of structural changes, we first determined the causal of each gap in the alignment and then tried to trace the origin of each gap on the phylogenetic tree. Gaps located at one or both sides of an exon could be caused by exonization/pseudoexonization or exon gain/loss events. The former could be inferred when exonic sequence of one gene was alignable with intronic or intergenic sequence of the other gene. The latter is the phenomenon when an entire exon of one gene could not be aligned to any region (including noncoding sequences) of the other. Gaps within an exon are usually caused by intraexonic insertions/deletions. We mapped the occurrence and the causal of each gap on the phylogenetic tree and deduced at which branch they have happened according to the maximum parsimony principle. In addition to the above mechanisms, intron gain/loss is also responsible for structural changes as previously reported (Xu et al., 2012), which was regarded when one exon of a certain gene could be perfectly aligned with two neighboring exons of the other gene. Different from other mechanisms, no gaps could be found in the alignment if intron gain/loss has happened, but it could lead to the difference in exon numbers. Therefore, the evolutionary history of intron gain/loss was also inferred. With the knowledge of these exon-intron structural changes, we estimated the exon-intron structures of the various ancestral genes in the MRCAs of extant core eudicots, Ranunculales, magnoliids, monocots, angiosperms, and gymnosperms (if applicable).

RESULTS

Structural Changes within the *AP1/FUL* Subfamily

A total of 209 genes were used for the structural analysis of *AP1/FUL* subfamily members. By performing exon-by-exon alignment, we generated a dataset consisting of 711 nucleotide sites, among which 607 were phylogenetically informative (**Dataset S1**). The topology of the final phylogenetic tree was largely consistent with previous studies and not sensitive to missing data (Litt and Irish, 2003; Preston and Kellogg, 2006; Shan et al., 2007; Xu and Kong, 2007; Litt and Kramer, 2010; Pabón-Mora et al., 2013). Nonetheless, the resolution was slightly improved and the positions of most genes were better congruent with angiosperm phylogeny. In contrast, the dataset created based on an alignment produced by Probalign only included 696 sites, among which 591 were informative (**Dataset S2**). Moreover, in the resulting phylogenetic tree, the positions of some major plant groups were discordant with angiosperm phylogeny (**Dataset S2**). Similar results were obtained when other MADS-box gene subfamilies were analyzed (**Datasets S3–S8**). This suggests that phylogenetic estimation can indeed be improved when structural changes were taken into consideration during alignment.

In the context of the improved phylogeny, we attempted to trace the evolutionary changes in the exon-intron structure of *AP1/FUL* subfamily members. We found that the *AP1/FUL*-like genes generally consist of eight exons, among which the first six have been highly conserved. In contrast, exons 7 and 8 vary greatly in length (from 77 to 209 bp for exon 7 and 34 to 148 bp for exon 8), suggestive of dramatic structural changes (**Figure S1**). Detailed comparisons revealed that intraexonic insertion/deletion occurred more frequently than exonization/pseudoexonization in this subfamily, and that structural changes were not distributed evenly among branches. For example, an average of 2 insertion/deletion events was detected in the Solanaceae euFUL-like genes (**Figure 1A**), while at least 8 structural change events were observed for each of the *OsMADS15* lineage members (**Figure 1B**).

We also found many structural change events shared by certain plant groups or major gene lineages. For example, in exon 7, a 3-bp deletion was detected in all *OsMADS14/15* members of monocots, and two independent 3-bp insertions were found in the *OsMADS14* and *OsMADS15* lineages of Poaceae (**Figure S1**). In exon 8, one 3-bp insertion near the 5' boundary was shared by all the sampled eudicot members (**Figure S1**), suggestive of an ancient structural change event occurred before the diversification of eudicots. There are also multiple cases where structural changes have caused divergence of duplicate genes. For instance, in the *OsMADS18/20* lineage of monocots, a gene duplication event resulted in the creation of two sublineages in Liliaceae (**Figure 1C**). The ancestor of one sublineage has experienced a 3-bp insertion in exon 7, while that of the other sublineage has undergone three insertions of different lengths in the same exon. Consistent with previous studies (Litt and Irish, 2003; Vandebussche et al., 2003a; Shan et al., 2007), we also detected a 1-bp deletion in exon 8 of all

examined euAPI-like genes, which led to pseudoexonization of the last 8 nucleotides (**Figure 1D**). With the knowledge of these structural changes, we inferred that the *AP1/FUL*-like gene in the MRCA of extant angiosperms is composed of eight exons with the lengths of 185, 79, 65, 100, 42, 42, 113, and 106 bp, respectively.

Structural Changes within the *SEP* Subfamily

We obtained 119 *SEP1*- and 87 *SEP3*-like genes to analyze exon-intron structural changes in the *SEP* subfamily. According to a previous study (Zahn et al., 2005) and this study, *SEP1*-like genes contain three major lineages in both core eudicots (i.e., *SEP1/2*, *FBP9*, and *SEP4*) and grasses (i.e., *OsMADS1*, *OsMADS5*, and *OsMADS34*; **Figure S2**). Except for *SEP1/2*-like genes in Brassicaceae and *EgAGL2-5* in *Elaeis guineensis*, all these genes have eight exons. For the Brassicaceae *SEP1/2*-like genes, the fifth exon (84 bp) could be aligned perfectly to the fifth (42 bp) plus the sixth (42 bp) exon of other genes, suggestive of an intron loss event that occurred before the diversification of Brassicaceae (**Figure 2A**). Like the situation in the *AP1/FUL* subfamily, structural change events were mostly observed in the seventh and eighth exons, but the occurrence frequency was much lower (**Figure 2** and **Figure S2**). A large number of structural changes could be found before the diversification of certain plant groups. For example, one 3-bp insertion in exon 2, one 15-bp insertion in exon 7, and two insertions (3 and 6 bp, respectively) in exon 8 of *SEP1/2*-like genes have likely occurred in the MRCA of Brassicaceae and Cleomaceae (**Figure 2A** and **Figure S2**). The longest insertion (66 bp) was observed in exon 7 of the *SEP4* gene of *Capsella rubella*, adjacent to which was an extra 33-bp insertion that has occurred in the ancestor of this and two other related species (*Brassica rapa* and *Arabidopsis*; **Figure 2B**). There were also evidences showing the contribution of structural changes to the divergence of duplicate genes. For instance, maize has a pair of duplicate genes (*ZmM24* and *ZmM31*) in the *OsMADS34* lineage. A 3-bp deletion happened in exon 8 of *ZmM24*, making the lengths of this exon different between them (**Figure 2C**). In addition to recent duplicates, structural changes in more ancient duplicates were also detected. One 3-bp deletion event in exon 2 of the *OsMADS1* lineage, as well as one 45-bp pseudoexonization event in exon 8 of the *OsMADS5* lineage, has likely taken place before the diversification of grasses (**Figure S2**). Within the *SEP1* clade, no structural change event has likely occurred before the origins of major plant groups (i.e., monocots, magnoliids, and core eudicots; **Figure 2D** and **Figure S2**). Based on this information, we inferred that the *SEP1*-like gene in the MRCA of extant angiosperms contains eight exons, with the lengths of 185, 79, 62, 100, 42, 42, 137, and 85 bp, respectively.

The phylogenetic tree of *SEP3*-like genes indicates no major gene duplication event (**Figure S3**). All of the 87 genes have eight exons. For exons 1, 4, 5, and 6, the lengths are largely conserved (185, 100, 42, and 42 bp, respectively) with a few exceptions (**Figure S3**). Exons 2, 3, 7, and 8, in contrast, vary remarkably in length, suggestive of multiple structural changes (**Figure 3** and **Figure S3**). For exon 2, independent exonization events

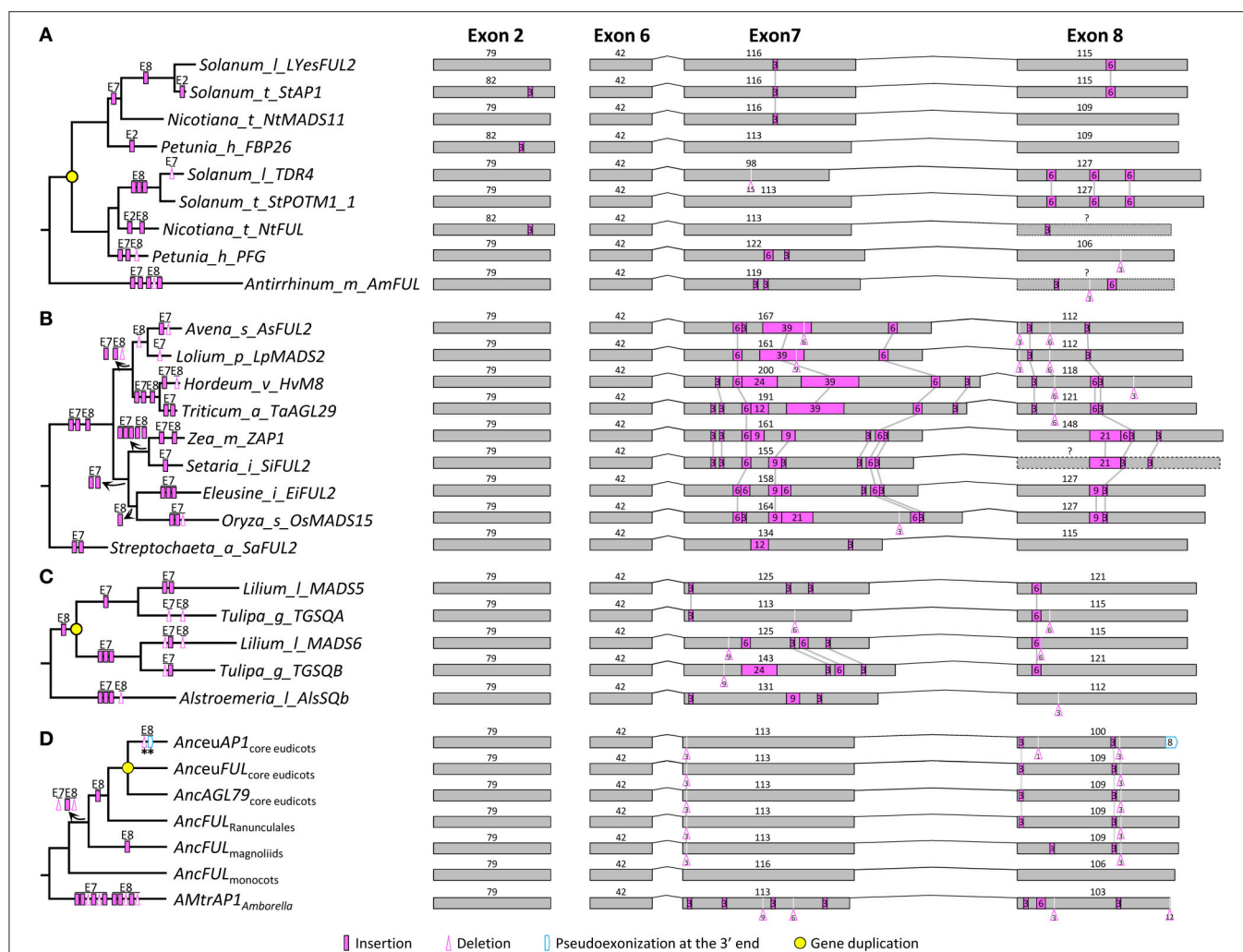
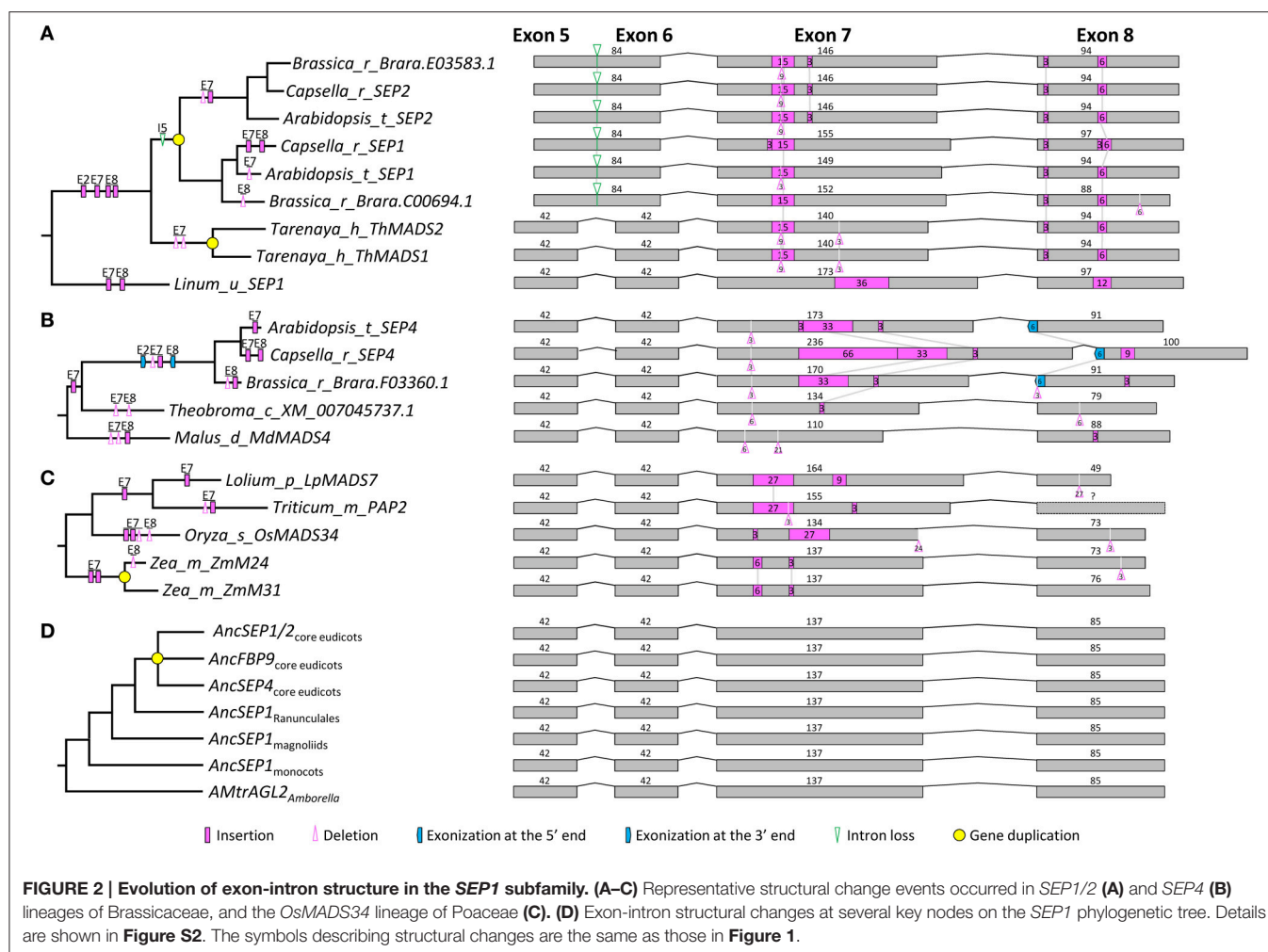


FIGURE 1 | Evolution of exon-intron structure in the AP1/FUL subfamily. (A–C) Representative structural change events occurred in the euFUL lineage of Solanaceae **(A)**, the OsMADS15 lineage of Poaceae **(B)**, and AP1/FUL-like genes of Liliales **(C)**. **(D)** Exon-intron structural changes at several key nodes on the AP1/FUL phylogenetic tree. “Anc” (for Ancestor) is prefixed to the name of each gene lineage. Details are shown in **Figure S1**. Exons and introns are represented by boxes and curved lines, respectively. The length of each exon is shown above the box. Shared structural change events are linked by gray lines. Different mechanisms responsible for structural changes are marked on corresponding branches of the phylogenetic tree. Stars indicate structural changes involving non-triplet sequences.

were observed in several taxa, such as Fabaceae, Brassicaceae, and *Eupomatia*, among others (**Figures 3A,B**, and **Figure S3**). In exon 3, a 9-bp exonization event was detected in members of Asparagales, Commelinales, and Poales, suggestive of an early structural change event during the evolution of monocots. Still in this exon, a more ancient exonization (6 bp) event was found before the divergence of Chloranthaceae (**Figure S3**). In exon 7, the MRCA of eudicots has experienced a 3-bp deletion event, while that of grasses has undergone two independent insertion events (**Figures 3B,C** and **Figure S3**). The earliest structural change event was a 9-bp deletion in exon 8, which happened after the divergence of *Amborella trichopoda* (hereafter called *Amborella*; **Figure 3D**). Taking into account of all the structural change events, we estimated that the *SEP3*-like gene in the MRCA of extant angiosperms contains eight exons, the lengths of which are 185, 79, 62, 100, 42, 42, 140, and 85 bp, respectively.

Structural Changes in the AGL6 Subfamily

Within the *AGL6* subfamily, 119 genes from angiosperms and 13 from gymnosperms were used for structural change analyses. The topology of the *AGL6* gene tree was similar to previous studies (Li et al., 2010; Kim et al., 2013). All the sampled genes except for *ZfAGL6a* in *Zamia fischeri* possess eight exons (**Figure S4**). The lengths of exons 1, 3, 4, and 5 (182, 62, 100, and 42 bp, respectively) are largely the same with exceptions in only five genes. In exon 2, other than a 3-bp deletion event occurred before the diversification of core eudicots, multiple independent insertion events were detected in several taxa, such as Brassicaceae and Ranunculaceae (**Figures 4A,B**, and **Figure S4**). In exon 6, a 21-bp exonization event occurred in the MRCA of asterids (**Figure S4**). Like the situation in the above two subfamilies, exons 7 and 8 were subject to multiple structural change events. In exon 7, major events include a



6-bp insertion in the MRCA of extant gymnosperms, a 3-bp insertion and three independent 3-bp deletions in the MRCA of extant angiosperms, a 3-bp insertion in the MRCA of Ranunculales, a 6-bp insertion in the MRCA of core eudicots, a 3-bp insertion and a 3-bp deletion in the MRCA of rosids, a 3-bp insertion and a 3-bp deletion in the MRCA of Asteraceae, a 6-bp insertion in the MRCA of Brassicaceae, and two 3-bp insertions in the MRCA of Poaceae (**Figures 4A,D** and **Figure S4**). In exon 8, independent insertion/deletion events were observed prior to the origins of eudicots, Asteraceae, and Poaceae, respectively (**Figure S4**). Structural divergence after gene duplication was also not a rare case in this subfamily. For example, *OsMADS6* and *OsMADS17* are two lineages generated by the pre-Poaceae gene duplication event, subsequent to which the former lineage went through two insertions in each of exon 7 and exon 8, while the latter experienced a 3-bp insertion in exon 2 and two 3-bp insertions in exon 8 (**Figure 4C**). Independent insertion/deletion events were also found in the duplicate lineages (Gg1 and Gg2) of gymnosperms (**Figure 4D** and **Figure S4**; Li et al., 2010). Considering all these structural change events, we inferred that the *AGL6*-like gene in the MRCA of extant angiosperms contains eight exons,

with the lengths of 182, 79, 62, 100, 42, 42, 134, and 85 bp, respectively.

Structural Changes within the *FLC* Subfamily

A recent study showed that *FLC*-like genes form a sister group to the *API/FUL* subfamily, and are closely related to the *SEP* and *AGL6* subfamilies (Ruelens et al., 2013). By carefully examining the sequences and deeply mining all available plant genomic data, we found that, as Ruelens et al. (2013) revealed, *FLC*-like genes could only be identified in core eudicots, Poaceae, and *Musa* (Musaceae). These findings suggest that *FLC*-like genes may have been lost independently in several lineages of angiosperms (Ruelens et al., 2013). Our phylogenetic tree showed that the *FLC*-like genes form two clades. One clade contains genes from core eudicots, including *FLC* and *MAF1/2/3/4/5* lineages generated by a pre-Brassicaceae gene duplication event; the other is composed of monocot genes, including *OsMADS51* and *OsMADS37* lineages produced by a pre-Poaceae gene duplication event. Unlike the aforementioned subfamilies, the core eudicot *FLC*-like genes have seven exons and exons 1, 4, 5, and 6 (185, 100, 42, and

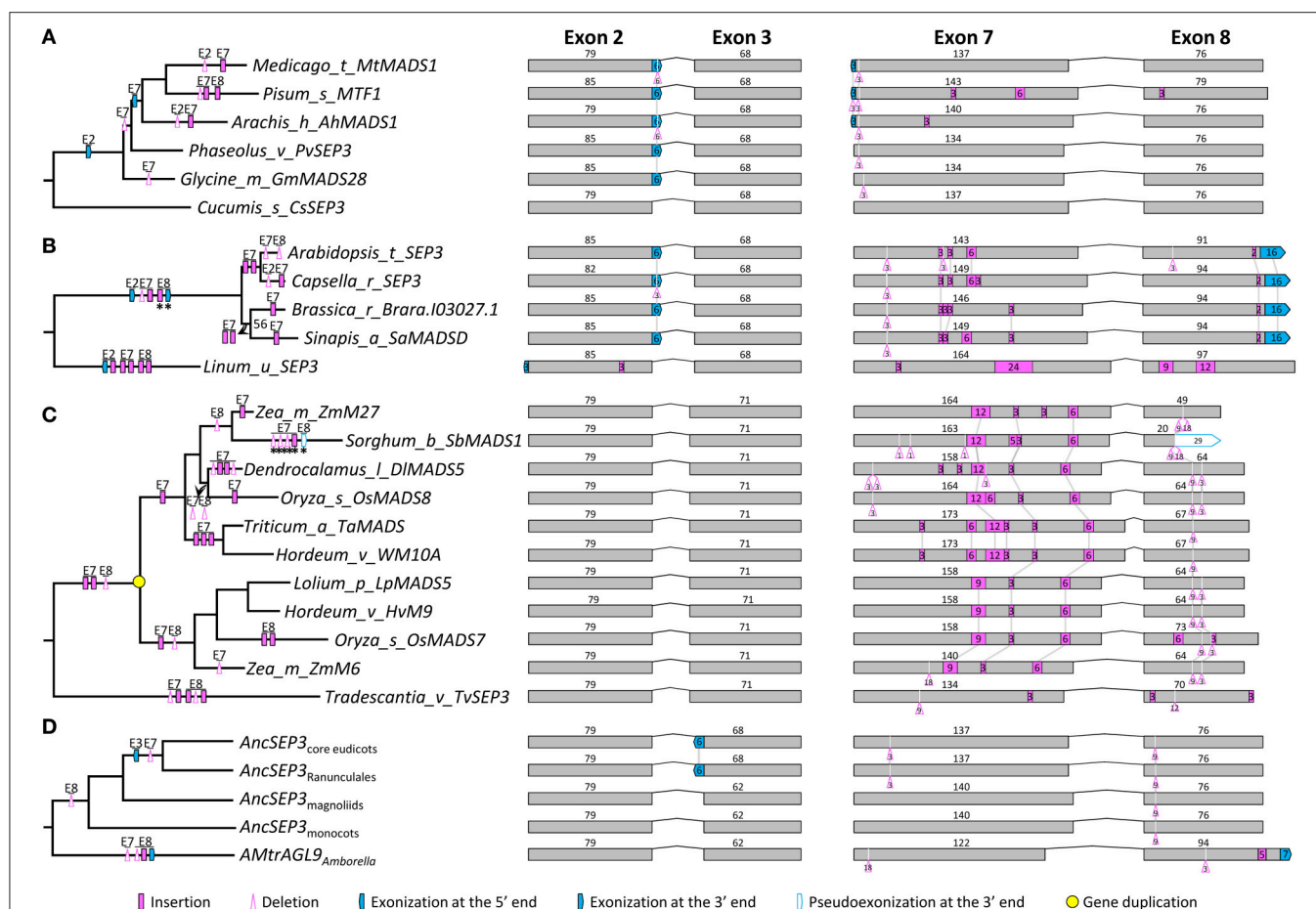


FIGURE 3 | Evolution of exon-intron structure in the SEP3 subfamily. (A–C) Representative structural change events occurred in SEP3-like genes of Fabaceae (A), Brassicaceae (B), and Poaceae (C). (D) Exon-intron structural changes at several key nodes on the SEP3 phylogenetic tree. Details are shown in Figure S3. The symbols describing structural changes are the same as those in Figure 1.

42 bp, respectively) are evolutionarily conserved. In contrast, most monocot genes possess only five exons (Figure S5). Given the dramatic divergence of exon-intron structures of the *OsMADS37*-lineage genes, they were excluded from further analysis.

In the context of the phylogeny, we traced the history of structural changes in this subfamily. We found that some structural change events were shared by core eudicot genes or Brassicaceae genes. In Poaceae, multiple structural change events are likely to have happened in the ancestor of the *OsMADS51* lineage. For example, an intron loss event was detected in exon 5 because it could be aligned to the fifth and sixth exons of core eudicot genes. The last exon, which is the counterpart of the seventh exon in genes from core eudicots, probably has been lost; however, due to rapid sequence evolution of this subfamily, the underlying mechanism is hard to determine. Other relatively trivial structural change events include a 3-bp insertion and a 3-bp deletion in exon 1, a 3-bp insertion and a 15-bp deletion in exon 3, and a 3-bp deletion in exon 4 (Figure S5). Based on these analyses, we inferred that the *FLC*-like gene in the MRCA of extant angiosperms has lost an exon and thus contains seven

exons, with the lengths of 185, 79, 68, 100, 42, 42, and 105 bp, respectively.

Structural Changes within the SOC1, AG, and STK Subfamilies

Structural changes of the outgroup subfamilies (*SOC1*, *AG*, and *STK* subfamilies) were also examined, which show relatively close relationships with the *AP1/FUL*, *SEP*, *AGL6*, and *FLC* subfamilies (Kim et al., 2005, 2013; Amborella Genome Project, 2013; Ruelens et al., 2013). *SOC1* subfamily members are present in both angiosperms and gymnosperms. All genes from monocots form a monophyletic clade with moderate bootstrap support (72%), with Poaceae genes falling into three lineages (*WSOC1*, *TaAGL7*, and *TaAGL23*). Within core eudicots, another three lineages, each containing genes from rosids and asterids, may have been generated by the γ genome triplication event (Tang et al., 2008). Here we named them *euSOC1*, *AGL42/71/72*, and *AGL14/19* after the homologs in *Arabidopsis* (Figure S6). All except for three *SOC1*-like genes (i.e., *Brara.I00679.1* in *Brassica rapa*, *SOC1* in *Linum usitatissimum*, and *CsSOC1B* in *Cucumis sativus*) are composed of seven exons. For the first six exons,

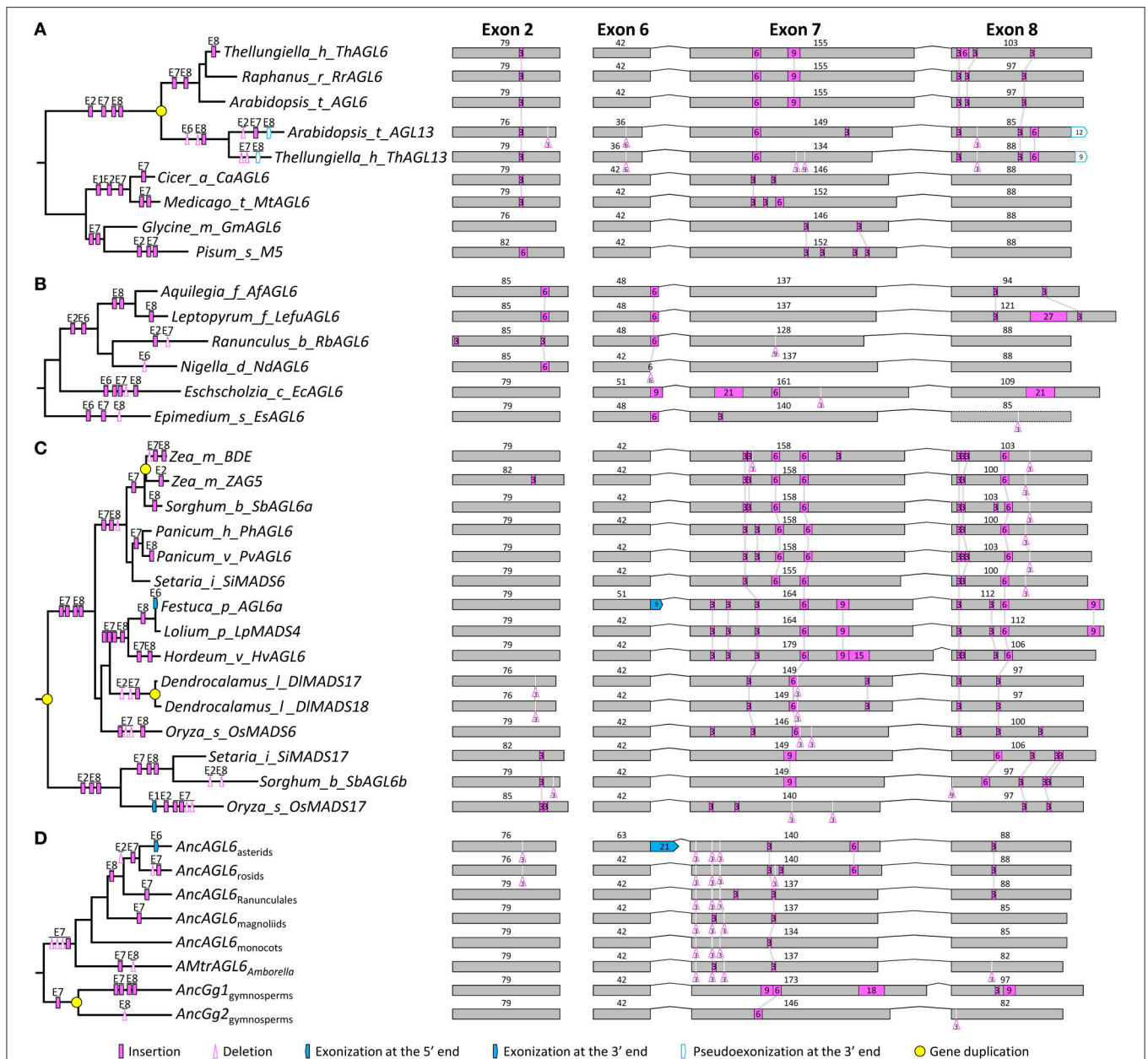


FIGURE 4 | Evolution of exon-intron structure in the AGL6 subfamily. (A–C) Representative structural change events occurred in AGL6-like genes of Brassicaceae (A), Ranunculaceae (B), and Poaceae (C). **(D)** Exon-intron structural changes at several key nodes on the AGL6 phylogenetic tree. Details are shown in Figure S4. The symbols describing structural changes are the same as those in Figure 1.

only a few structural change events were detected, which sparsely distributed across the angiosperm clade. Most structural changes were found in exon 7, including multiple insertion/deletion and exonization/pseudoexonization events (Figure S6). Taken together, we inferred that the *SOC1*-like gene in the MRCA of extant seed plants likely contains seven exons, with the lengths of 182, 82, 62, 100, 42, 42, and 132 bp, respectively.

The phylogenetic relationships of the AG and STK subfamilies were largely consistent with a previous study (Zahn et al., 2006), with the majority of genes containing seven exons (Figure S7).

Structural analyses revealed several major structural changes in the AG subfamily, such as a 3-bp insertion in exon 7 after the divergence of *Amborella* and a 6-bp insertion in exon 7 before the diversification of eudicots. In the STK subfamily, one 3-bp exonization event in exon 3 and two separate insertions in exon 7 have occurred in the MRCA of monocots (Figure S7). Tracing back to the MRCA of extant seed plants, we concluded that the ancestral AG/STK-like gene contains seven exons, with the lengths of 182, 82, 62, 100, 42, 42, and 159 bp, respectively.

Phylogenetic Relationships and Structural Differences among Subfamilies

To resolve the relationships among all focal subfamilies, we constructed phylogenetic trees with three different matrices (alignments I, II, and III) (see Section Materials and Methods; **Dataset S9**). Topologies of all three trees were largely consistent, but the nodal supports at key nodes increased as more structurally diverged sequences were removed (**Figure S8** and **Figure 5**). In the first tree, which was constructed using the matrix composed of all 792 sequences (alignment I), *AP1/FUL* and *FLC* are sisters, with 57% ML bootstrap support (BP) and 0.99 Bayesian posterior probabilities (PP), and *SEP* is the sister to them (50% BP and 0.97 PP). *AGL6* shows a sister relationship with the abovementioned three subfamilies (89% BP and 1.00 PP; **Figure S8A**). Considering that duplicate genes usually show accelerated evolutionary rate and more frequent structural changes that may screw the phylogeny, we next removed duplicated genes that diverged greatly in structure and generated a second matrix (alignment II). The tree built using this matrix gained increased supports for almost all of the abovementioned nodes (**Figure S8B**). To further improve the resolution, we selected genes (alignment III) with more conserved exon-intron structure from the second matrix and constructed the third tree. All focal nodes were strongly supported in both ML and BI trees (**Figure 5**).

Based on our alignment and the topology of the resultant phylogenetic trees, we traced the evolutionary changes of exon-intron structures in these subfamilies. As described earlier, in the MRCA of extant angiosperms or seed plants (if applicable), the *AP1/FUL*, *SEP*, and *AGL6* genes all possess eight exons, while the *FLC*, *AG/STK*, and *SOC1* genes all contain seven exons (**Figure 6**). Unambiguous homologous relationships of exon 1 to exon 6 could be determined based on conservation of the encoded amino acid sequences, i.e., the MADS domain, I region, and K domain. Structural change events were found in exons 1, 2, 3, 7, and 8, some of which were shared by different subfamilies and consistent with their phylogenetic relationships (**Figure 6**). In exon 1, Kim et al. (2013) found a 3-bp gap in all *AGL6*-like genes but not in the *AP1/FUL* and *SEP* subfamilies. Here we found that this gap also appears in genes of *AG/STK* and *SOC1* subfamilies, suggesting a 3-bp insertion in the ancestor of *AP1/FUL*, *FLC*, and *SEP* subfamily genes (**Figure 6**). In exon 2, a 3-bp deletion has likely occurred in the ancestor of *AP1/FUL*, *FLC*, *SEP*, and *AGL6* subfamily genes. The length of exon 3 in all except for the *AP1/FUL* and *FLC* subfamilies is 62 bp. A 3-bp insertion plus an independent 3-bp exonization have resulted in an exon of 65 bp in the ancestor of the *AP1/FUL* subfamily and 68 bp in that of the *FLC* subfamily.

In all these subfamilies, exons 7 and 8 (if present), which encode(s) for the C-terminal region, is highly variable but contains short, relatively conserved, lineage-specific motifs. We found that in exon 7, the AG II motif (Kramer et al., 2004) was alignable to the SEP I motif (Zahn et al., 2005), the FUL motif (Shan et al., 2007), and the AGL6 I motif (Ohmori et al., 2009), and that the last four amino acids (LxxG) are quite conserved. This suggests that the seventh exons of different subfamilies

(**Figure S9**) are homologous. In this exon, two 3-bp insertions and one 21-bp deletion have occurred before the divergence of *AP1/FUL*, *FLC*, *SEP*, and *AGL6* subfamilies. Three deletions with lengths of 3-, 3-, and 9-bp, respectively, as well as a 15-bp insertion were shared by the *AP1/FUL*, *FLC* and *SEP* subfamilies. The ancestor of *AP1/FUL* and *FLC* subfamily genes has likely experienced two deletion events. A 3-bp insertion shared by the *SEP* subfamily genes was also observed (**Figure 6**). These shared structural change events provide further support for the phylogenetic relationships among the four subfamilies.

Exon 8 is specific for the *AP1/FUL*, *SEP*, and *AGL6* subfamilies. Based on our phylogeny, it is highly likely that this exon originated before the divergence of these subfamilies. To figure out the mechanisms responsible for the evolutionary changes of this exon, we further searched putatively homologous sequences of this exon at the downstream 200 kb intergenic region of representative genes from the *FLC*, *SOC1*, *AG*, and *STK* subfamilies. However, due to the relatively long divergence time, we could not find any alignable region. Thus it is hard to determine whether this exon was generated by exonization or exon gain in the ancestor of the four subfamilies. Likewise, it is difficult to determine how this exon was lost in the *FLC*-like genes. More interestingly, we found that the ancestor of the *AP1/FUL* subfamily has experienced an exonization event at the 3' boundary of exon 8. As we mentioned earlier, except for euAP1 proteins, all the other members of this subfamily encode for a paleoAP1 motif (Vandenbussche et al., 2003a; Shan et al., 2007), the first six amino acids of which is defined as FUL-like motif (Litt and Irish, 2003; Litt, 2007) and could be aligned to the C-terminal ends of the *SEP* and *AGL6* proteins (**Figure S10**). To understand the origin of the extra 5 amino acids in the paleoAP1 motif, we tried to align the coding sequence of this region to the 3' untranslated regions of *SEP* and *AGL6* subfamily genes. The resultant alignment (**Figure S10**) suggested that two point mutations (T→C and A→C) may have broken the original stop codon in the ancestor of the *AP1/FUL* subfamily, thereby leading to exonization of the next in-frame 15 bp and thus addition of new amino acids in the protein product (**Figure S10**). Intriguingly, the *Amborella* AMtrAP1 does not contain the extra 5 amino acids. Further investigation showed that this may have been caused by independent insertions and point mutations because the corresponding region in this species does not show much similarity with other *AP1/FUL*-like genes, or with *SEP* or *AGL6* subfamily members.

DISCUSSION

Prevalence and Functional Impacts of Exon-Intron Structural Changes

Although previous studies have reported structural changes in MADS-box genes (Litt and Irish, 2003; Vandenbussche et al., 2003a; Kramer et al., 2006; Shan et al., 2007; Xu and Kong, 2007; Xu et al., 2012; Fourquin et al., 2013), it is ours that first trace the evolution of them in several subfamilies. By conducting such a detailed analysis, we found that: (1) structural changes are highly prevalent during the evolution of MADS-box genes,

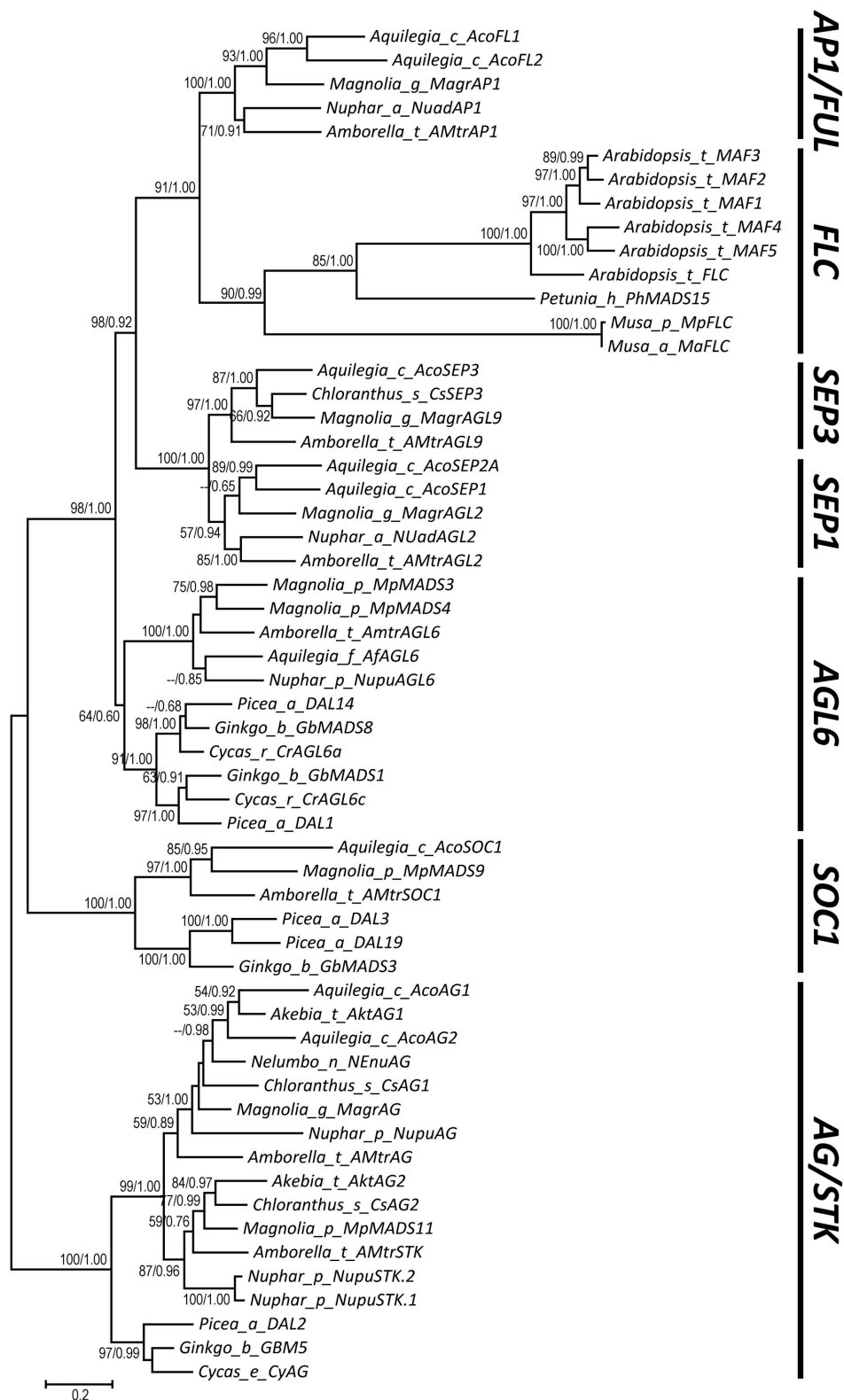


FIGURE 5 | A phylogenetic tree showing relationships of the AP1/FUL, FLC, SEP, and AGL6 subfamilies. The bootstrap values (>50%) obtained from maximum likelihood analysis and the posterior probabilities (>0.5) estimated by Bayesian inference are shown next to the nodes.

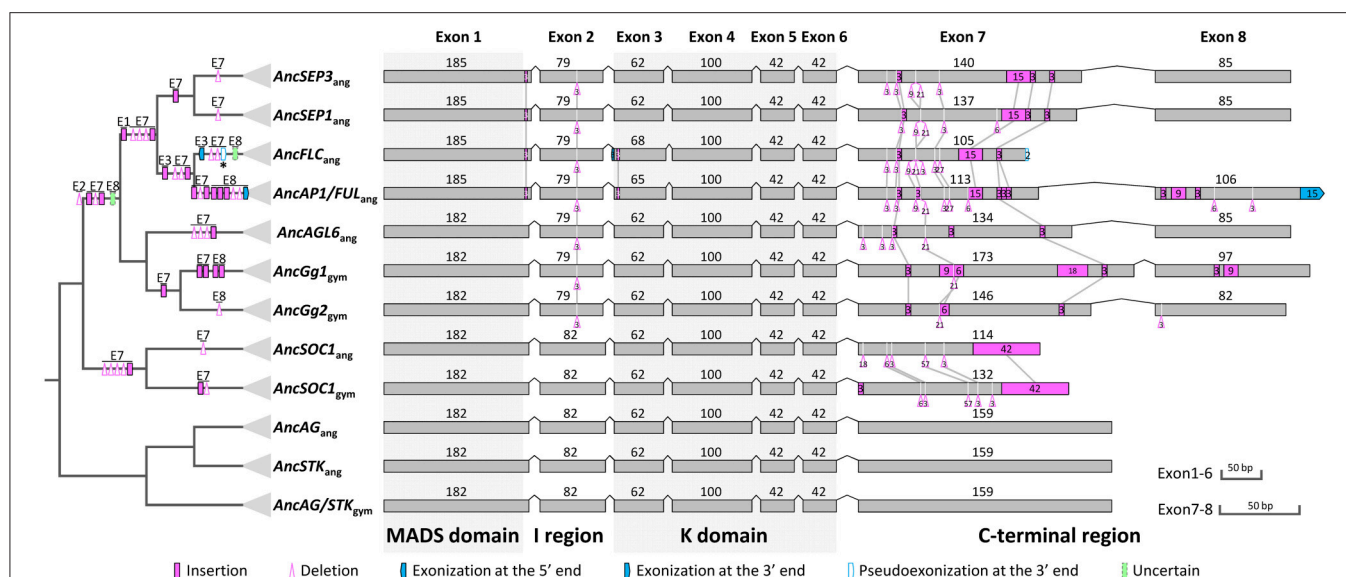


FIGURE 6 | Evolution of exon-intron structures of the AP1/FUL, FLC, SEP, and AGL6 subfamilies. The simplified tree is from Figure 5 and Figure S8. Show here is the ancestral exon-intron structure of each subfamily in the MRCA of extant angiosperms and in the MRCA of extant gymnosperms (if applicable). The MADS domain, I region, K domain, and C-terminal region are indicated below exons, and the MADS and K domains are highlighted with gray boxes. "ang" is the abbreviation for "angiosperms," and "gym" for "gymnosperms." The symbols describing structural changes are the same as those in Figure 1.

which contributed to the divergence of genes within and among subfamilies; (2) as has been shown in previous studies (Xu and Kong, 2007; Xu et al., 2009, 2012; Liu et al., 2011), structural changes could be achieved by three types of mechanisms, i.e., exon/intron gain/loss, exonization/pseudoexonization, and intraexonic insertion/deletion; (3) although structural changes can occur in every exon, most of them took place in exons or the part of an exon that encodes for the I region or the C-terminal region; (4) most structural changes were fixed in a specific gene or species, but some important ones were preserved over long evolutionary time. Clearly, these results provide a comprehensive and updated insight into the significant role that structural changes have played in the diversification of gene families.

The frequent occurrence of structural changes in the C-terminal region is not surprising because it has long been demonstrated that this region varies considerably in length and sequence among MADS-box proteins. However, highly variable as it is, this region contains quite conserved motifs. Structural changes rarely occurred in these motifs, but when they did, they could occasionally cause the formation of new motifs (Litt and Irish, 2003; Vandenbussche et al., 2003a; Kramer et al., 2006; Litt, 2007; Shan et al., 2007). One typical example is the generation of the euAP3 motif by either insertion of eight nucleotides (Vandenbussche et al., 2003a) or deletion of one nucleotide (Vandenbussche et al., 2003a) in an ancestral paleoAP3-motif encoding gene. Another example is the generation of two new motifs in euAP1 proteins by 1-bp deletion (Litt and Irish, 2003; Vandenbussche et al., 2003a; Kramer et al., 2006; Litt, 2007; Shan et al., 2007). The above examples both involve out-of-frame insertions/deletions, which are generally deleterious. However, when occurring in duplicate genes, the presence of a redundant

copy could compensate for the possible loss of function caused by frameshift mutations, enabling these mutations to lead to functional divergence (Raes and Van de Peer, 2005). As a previous study suggested, this might be the main pattern for novel motif generation in transcription factor families (Vandenbussche et al., 2003a). Interestingly, we found that other structural change mechanisms could also contribute to the generation of novel motifs. For example, the paleoAP1 motif was created by degeneration of the original stop codon and exonization of adjacent 15 nucleotides. More dramatically, the eighth exon, part of which encodes for conserved motifs in the AP1/FUL, SEP, and AGL6 subfamilies, was likely generated by an exonization or exon gain event. These new motifs, which have been highly conserved for a remarkably long evolutionary time, are likely of extraordinary importance and could be a good starting point for functional studies.

Currently, there are only limited data on the functions of several C-terminal motifs and the results are conflicting. For example, one study showed that the euAP3 motif endowed euAP3-like proteins with new functions in specifying perianth structures in core eudicots (Lamb and Irish, 2003); whereas two other studies demonstrated that this motif was dispensable for floral organ identity determination (Piwarzyk et al., 2007; Su et al., 2008). The transactivation domain could indeed confer activation capability to euAP1-like proteins of *Arabidopsis*, radish (*Raphanus sativus*), and tobacco (*Nicotiana tabacum* and *Nicotiana sylvestris*; Cho et al., 1999). However, a couple of functional studies showed that euFUL and FUL-like proteins were able to substitute for AP1, indicating that the C-terminal motifs may not be essential for the functions of euAP1-like proteins (Gocal et al., 2001; Jang et al., 2002; Chen et al., 2008).

Also, Krizek and Meyerowitz (1996) presented evidence that the C-terminal domains of AP1 and AG are not necessary for functional specificity. These opposing results may have been caused by different experimental methods, or possible redundancy of these proteins in high-order complexes (Litt and Kramer, 2010). Further investigations are needed in the future to address this question.

Effects of Structural Changes on Alignment and Phylogenetic Relationships among the AP1/FUL, SEP, AGL6, and FLC Subfamilies

A reliable alignment is extremely important for the accuracy of phylogenetic estimation. Sequence similarity is empirically considered as a hint for homology; however, when evolutionary time is too long, it would be quite difficult to draw an unambiguous conclusion. In the present study, we demonstrated that structural changes are common during the evolution of a gene subfamily, and would directly or indirectly disrupt the homology of corresponding sites or regions in a couple of ways. First, insertion/deletion or exonization/pseudoexonization of non-triplet sequences would lead to shifts of reading frame and thus destroy homology of the downstream coding region. Second, independent changes at the same position in different species may be aligned together and thus erroneously produce nonhomologous sites in the matrix. We found quite a few such cases, one of which is several independent exonization events in exon 2 of core eudicot SEP3-like genes (Figure S3). Third, when a certain position is a hot spot for insertion/deletion, it would be hard to determine whether corresponding sites are homologous or not. This phenomenon has been observed frequently in grass genes (Figures S1, S3–S4, S6–S7). Finally, a structural change event may occur within a codon, and thus the homology is interrupted. Multiple cases have been found in this study, such as independent exonization events at the 5' end of exon 8 in some genes of the AP1/FUL and SEP subfamilies (Figures S1–S3). Therefore, with the accessibility of more complete genome sequences, it is feasible to generate a more reasonable alignment by referring to exon-intron structure information.

In this study, with the knowledge of structural changes in each subfamily, we refined our alignments and estimated phylogenetic relationships of the AP1/FUL, FLC, SEP, and AGL6 subfamilies. Our tree showed that SEP is sister to the monophyletic group formed by AP1/FUL and FLC, and that AGL6 is the sister to the three abovementioned subfamilies. The topology is different from the one reported by Ruelens et al. (2013), in which SEP and AGL6 are sister to each other and together they are nested with the lineage formed by AP1/FUL and FLC. Based on their phylogenetic tree and syntenic evidence, Ruelens et al. (2013) proposed that the ancestor of AP1/FUL, FLC, SEP, and AGL6 subfamily genes experienced a tandem duplication event in the MRCA of extant seed plants, creating the ancestor of SEP and AGL6, and the ancestor of AP1/FUL and FLC. Then the former went through a duplication event and generated ancestral SEP and AGL6 genes. The segment containing the ancestral SEP and the ancestor of AP1/FUL and FLC was then lost in the MRCA of extant gymnosperms. However, according to our

phylogenetic tree and taking the syntenic evidence into account, we hypothesize that the ancestor of AGL6, SEP, AP1/FUL, and FLC has experienced a duplication event in the MRCA of extant seed plants, generating the ancestral AGL6 and the ancestor of SEP, AP1/FUL, and FLC. The latter was then lost in the MRCA of extant gymnosperms but went through a tandem duplication event prior to the origin of angiosperms, bringing forth the ancestral SEP and the ancestor of AP1/FUL and FLC. Then the two genes underwent a whole genome duplication event in the MRCA of extant angiosperms and created SEP1 and SEP3, and AP1/FUL and FLC, respectively. Our hypothesis is equally parsimonious with that of Ruelens et al. (2013) and the phylogenetic tree also showed stronger supports at key nodes than previous studies (Carlsbecker et al., 2003; Kim et al., 2005; Futamura et al., 2008; Li et al., 2010). Moreover, structural changes shared by different subfamilies provide extra evidence for our topology (Figure 6). The gradual improvement of nodal supports with successive removal of structurally diverged sequences suggests that structural changes could indeed influence sequence alignment and then phylogenetic estimation, which need to be carefully considered when studying the evolution of a certain gene family.

Structural Diversification Is Associated with Functional Divergence among Subfamilies

Our results showed that structural changes have taken place in all the focal subfamilies but with different extents. The divergence pattern is significantly associated with their functions. For example, SEP-like genes have experienced much less structural changes than the AP1/FUL, FLC, and AGL6 subfamily genes during evolution. Accumulating evidences have shown that the SEP subfamily members play conserved and vital roles in specifying floral organ identities of angiosperms. Silencing or mutation of SEP-like genes in different species, such as *Arabidopsis* SEP1/2/3/4, *petunia* FBP2/FBP5, *tomato* TM5/TM29, *Nigella damascena* NdSEP1/2/3, and rice *OsMADS1/5/7/8*, can lead to the transition of floral organs to sepal-, bract-, or leaf-like organs (Pnueli et al., 1994; Pelaz et al., 2001; Ampomah-Dwamena et al., 2002; Ferrario et al., 2003; Vandenbussche et al., 2003b; Ditta et al., 2004; Cui et al., 2010; Wang et al., 2015). Biochemical data revealed that the SEP-like proteins are able to form quaternary complexes with other floral MADS-box proteins in many species, such as *Arabidopsis*, *petunia*, *Gerbera hybrida*, *Vitis vinifera*, and rice (Honma and Goto, 2001; Ferrario et al., 2003; Ruokolainen et al., 2010; Seok et al., 2010; Smaczniak et al., 2012; Mellway and Lund, 2013). Recently, we reported that heterodimers between the SEP-like proteins and other floral MADS-box proteins can be formed in early diverging angiosperms, such as *Amborella* and *Nuphar pumila* (*Amborella* Genome Project, 2013; Li et al., 2015). Moreover, by conducting yeast two-hybrid assays with resurrected proteins of the MRCA of extant angiosperms, we found that the ancestral SEP-like proteins have broad interactions with other ancestral floral MADS-box proteins (Li et al., 2015). Therefore, it is highly likely that the SEP-like gene in the MRCA of extant angiosperms has obtained the

function of determining floral organ identities and the ability to mediate the formation of floral quartets, which has been retained during the evolution due to their stable gene structures and conserved sequence features.

Unlike *SEP*, the *API/FUL* and *FLC* subfamilies have undergone severe rounds of structural divergence since the duplication of the ancestral gene. In addition to the insertion/deletion events that occurred in the ancestor of *API/FUL* and *FLC*, dramatic exon-intron structural changes, including exon loss, exonization, pseudoexonization, insertions, and deletions, have taken place in the respective ancestors of *FLC* and *API/FUL*. Divergence in gene structure of these two subfamilies resulted in shorter *FLC*-like proteins, but longer *API/FUL*-like proteins. Consistent with this, members of these two subfamilies tend to perform different functions in floral development. As has been reported, some *FLC* subfamily members act as floral repressors responsive to vernalization (Michaels and Amasino, 1999; Sheldon et al., 2006), while the *API/FUL*-like genes mainly function as positive regulators in determining the identities of inflorescences, floral meristems, and floral organs, and controlling the development of compound leaves and fruits (Irish and Sussex, 1990; Huijser et al., 1992; Gu et al., 1998; Pabón-Mora et al., 2012, 2013; Burko et al., 2013). Intriguingly, some *API/FUL* subfamily members are also involved in vernalization, such as *WAP1* in wheat (*Triticum aestivum*; Danyluk et al., 2003; Murai et al., 2003; Trevaskis et al., 2003; Yan et al., 2003; Kim et al., 2009). However, since members of other MADS-box gene subfamilies, such as *STMADS11*-like genes in grasses (Kane et al., 2005), are also identified as vernalization repressors, this type of function may have evolved multiple times independently. Frequent structural changes happened in the *API/FUL* subfamily may also be the cause of functional divergence between *API/FUL* and *SEP* subfamilies. We have recently revealed that the ancestral *API/FUL* protein lost the ability to interact with the AG and STK proteins in the MRCA of extant angiosperms (Li et al., 2015). This suggests that the two gene subfamilies have diverged at the early stage of angiosperm evolution, and that the functions of *API/FUL*-like genes further diversified during evolution due to the accumulation of more gene structural changes.

Different from the *SEP*, *API/FUL*, and *FLC* subfamilies, the *AGL6* subfamily originated before the diversification of extant seed plants, and experienced one round of gene duplication event in the MRCA of extant gymnosperms. In angiosperms, *AGL6*-like genes show various functions. For example, one of the *Arabidopsis* *AGL6*-like genes, *AGL6*, is responsible for the regulation of lateral organ development, flowering time, and circadian clock (Koo et al., 2010; Yoo et al., 2011; Huang et al., 2012, 2013), but the other one, *AGL13*, is involved in male and female gametophyte morphogenesis (Hsu et al., 2014). The *AGL6*-like gene in a basal eudicot species, *Nigella damascena*, acts as an A-function gene to determine the sepal and petal identities (Wang et al., 2015). In Zingiberales (monocot plants), the *AGL6*-like genes may regulate stamen morphology (Yockteng et al., 2013). Interestingly, in several angiosperm species, *AGL6*-like genes, such as *PhAGL6* of petunia (Rijpkema et al., 2009), *BEARDED-EAR (BDE)* of maize (Thompson et al., 2009), and

OsMADS6 of rice (Ohmori et al., 2009), function redundantly with *SEP*-like genes. In this article, we found that frequent structural change events have taken place during the evolution of angiosperm *AGL6*-like genes. Presumably, the unstable gene structures, plus regulatory divergence, have contributed to the functional diversification of angiosperm *AGL6*-like genes. Although some structural divergence events have also been revealed in the ancestor of angiosperm *AGL6*-like genes and the respective ancestors of gymnosperm Gg1 and Gg2 lineages, it seems that these ancestral proteins have similar interaction patterns. For instance, in gymnosperms, the *AGL6*-like proteins of *Gnetum gnemon*, GGM9 and GGM11, can interact with proteins of the *AP3/PI* and *AG/STK* subfamilies, and may have the ability to mediate multimeric protein complex formation (Wang et al., 2010). In the MRCA of extant angiosperms, *AGL6* has relatively high possibility to interact with other floral proteins, similar to *SEP* (Li et al., 2015). Therefore, it is very likely that the quaternary complexes mediated by *AGL6* have existed in the MRCA of extant seed plants (Wang et al., 2010). With the origin of *SEP* and the formation of obligate heterodimers between AP3 and PI in the MRCA of extant angiosperms (Melzer et al., 2014; Li et al., 2015), the multimerization of floral MADS-box proteins becomes equally dependent on *SEP* or *AGL6*. Afterwards, due to quick divergence of ancestral *SEP* and *AGL6* genes in exon-intron structure, together with point mutations and changes in expression regulation, the *SEP*-like proteins become major mediators of floral quartets in extant angiosperms. Overall, the evolution of the *SEP*, *API/FUL*, *FLC*, and *AGL6* subfamilies are complicated; their differences in exon-intron structures are only one aspect of their divergence. More studies are needed to clarify the functional diversification of these genes.

AUTHOR CONTRIBUTIONS

XY, XD, RZ, XF, and LY analyzed data; XY, XD, RZ, GX, and HS wrote the paper; GX, HS, and HK designed the research.

ACKNOWLEDGMENTS

We thank Kong lab members for helpful discussions, and anonymous reviewers for their constructive comments. This work was supported by National Natural Science Foundation of China (Grants 31125005, 31422006, and 31570225) and the Specialized Fund from the CAS Youth Innovation Promotion Association to HS and GX.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00598>

Table S1 | Genes used in this study.

Dataset S1 | The phylogenetic tree used for gene classification.

Dataset S2 | Matrices of the *AP1/FUL* subfamily. (A) The matrix for phylogenetic analysis in consideration of exon-intron structural changes, in which only alignable sites are included. The resulting tree is shown in **Figure S1**. (B) The matrix based on an alignment generated by Probalign and the resulting tree. (C) The matrix for exon-intron structural change analysis.

Dataset S3 | Matrices of the *SEP1* subfamily. (A) The matrix for phylogenetic analysis in consideration of exon-intron structural changes, in which only alignable sites are included. (B) The matrix based on an alignment generated by Probalign and the resulting tree. (C) The matrix for exon-intron structural change analysis.

Dataset S4 | Matrices of the *SEP3* subfamily. (A) The matrix for phylogenetic analysis in consideration of exon-intron structural changes, in which only alignable sites are included. (B) The matrix based on an alignment generated by Probalign and the resulting tree. (C) The matrix for exon-intron structural change analysis.

Dataset S5 | Matrices of the *AGL6* subfamily. (A) The matrix for phylogenetic analysis in consideration of exon-intron structural changes, in which only alignable sites are included. (B) The matrix based on an alignment generated by Probalign and the resulting tree. (C) The matrix for exon-intron structural change analysis.

Dataset S6 | Matrices of the *FLC* subfamily. (A) The matrix for phylogenetic analysis in consideration of exon-intron structural changes, in which only alignable sites are included. (B) The matrix based on an alignment generated by Probalign and the resulting tree. (C) The matrix for exon-intron structural change analysis.

Dataset S7 | Matrices of the *SOC1* subfamily. (A) The matrix for phylogenetic analysis in consideration of exon-intron structural changes, in which only alignable sites are included. (B) The matrix based on an alignment generated by Probalign and the resulting tree. (C) The matrix for exon-intron structural change analysis.

Dataset S8 | Matrices of the *AG/STK* subfamily. (A) The matrix for phylogenetic analysis in consideration of exon-intron structural changes, in which only alignable sites are included. (B) The matrix based on an alignment generated by Probalign and the resulting tree. (C) The matrix for exon-intron structural change analysis.

Dataset S9 | Matrices for phylogenetic construction and exon-intron structural change inference among subfamilies. (A) The matrix of alignment I and the resulting maximum likelihood tree. (B) The matrix of alignment II and the resulting maximum likelihood tree. The simplified trees are shown in **Figure S8**. (C) The matrix of alignment III. The resulting tree is shown in **Figure 5**. (D) The matrix for exon-intron structural change analysis among different subfamilies.

Figure S1 | Evolution of exon-intron structure in the *AP1/FUL* subfamily. (A) A maximum-likelihood tree of the *AP1/FUL* subfamily, with higher-than-50% bootstrap values indicated for each node. Different mechanisms responsible for structural changes are marked on corresponding branches of the phylogenetic tree. Stars indicate structural changes involving non-triplet sequences. (B) Schematic representation of exon-intron structural changes. Exons and introns are represented by boxes and curved lines, respectively. Exon length is shown

above the box, and intron length (if available) is indicated below the curved lines. Shared structural change events are linked by gray lines.

Figure S2 | Evolution of exon-intron structure in the *SEP1* subfamily. (A) A maximum-likelihood tree of the *SEP1* subfamily. (B) Schematic representation of exon-intron structural changes. The symbols describing structural changes are the same as those in **Figure S1**.

Figure S3 | Evolution of exon-intron structure in the *SEP3* subfamily. (A) A maximum-likelihood tree of the *SEP3* subfamily. (B) Schematic representation of exon-intron structural changes. The symbols describing structural changes are the same as those in **Figure S1**.

Figure S4 | Evolution of exon-intron structure in the *AGL6* subfamily. (A) A maximum-likelihood tree of the *AGL6* subfamily. (B) Schematic representation of exon-intron structural changes. The symbols describing structural changes are the same as those in **Figure S1**.

Figure S5 | Evolution of exon-intron structure in the *FLC* subfamily. (A) A maximum-likelihood tree of the *FLC* subfamily. (B) Schematic representation of exon-intron structural changes. Note that due to the dramatic sequence divergence of *OsMADS37*-like genes after gene duplication, the mechanisms underlying structural changes are difficult to determine. For these genes, only the exon-intron structures are shown. The symbols describing structural changes are the same as those in **Figure S1**.

Figure S6 | Evolution of exon-intron structure in the *SOC1* subfamily. (A) A maximum-likelihood tree of the *SOC1* subfamily. (B) Schematic representation of exon-intron structural changes. The symbols describing structural changes are the same as those in **Figure S1**.

Figure S7 | Evolution of exon-intron structure in the *AG/STK* subfamily. (A) A maximum-likelihood tree of the *AG/STK* subfamily. (B) Schematic representation of exon-intron structural changes. The symbols describing structural changes are the same as those in **Figure S1**.

Figure S8 | Simplified phylogenetic trees showing relationships of the *AP1/FUL*, *FLC*, *SEP*, and *AGL6* subfamilies, constructed based on alignments I (A) and II (B). The bootstrap values (>50%) obtained from maximum likelihood analysis and the posterior probabilities (>0.5) estimated by Bayesian inference are shown next to the nodes.

Figure S9 | Alignment of amino acids encoded by exon 7 of representatives of the *AP1/FUL*, *SEP*, *AGL6*, and *AG/STK* subfamilies. Subfamily-specific motifs are highlighted by red boxes.

Figure S10 | Creation of the paleoAP1 motif. Both nucleotide (A) and amino acid alignments (B) of the paleoAP1 motif in the sampled *AP1/FUL*-like genes and its corresponding regions in representatives of *SEP*- and *AGL6*-like genes are shown. On top of the alignments, an asterisk or a number indicates every ten nucleotides or amino acids. In (A), coding sequences and 3' untranslated regions are represented by uppercase and lowercase letters, respectively. In (B), the paleoAP1 motif is boxed. Stars in the amino acid sequence correspond to stop codons.

REFERENCES

- Amborella Genome Project (2013). The *Amborella* genome and the evolution of flowering plants. *Science* 342:1241089. doi: 10.1126/science.1241089
- Ampomah-Dwamena, C., Morris, B. A., Sutherland, P., Veit, B., and Yao, J. L. (2002). Down-regulation of *TM29*, a tomato *SEPALLATA* homolog, causes parthenocarpic fruit development and floral reversion. *Plant Physiol.* 130, 605–617. doi: 10.1104/pp.005223
- Aoki, S., Uehara, K., Imafuku, M., Hasebe, M., and Ito, M. (2004). Phylogeny and divergence of basal angiosperms inferred from *APETALA3*- and *PISTILLATA*-like MADS-box genes. *J. Plant Res.* 117, 229–244. doi: 10.1007/s10265-004-0153-7
- Becker, A., and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* 29, 464–489. doi: 10.1016/S1055-7903(03)00207-0
- Burko, Y., Shleizer-Burko, S., Yanai, O., Shwartz, I., Zelnik, I. D., Jacob-Hirsch, J., et al. (2013). A role for *APETALA1/FRUITFULL* transcription factors in tomato leaf development. *Plant Cell* 25, 2070–2083. doi: 10.1105/tpc.113.1.13035
- Carlsbecker, A., Sundström, J., Tandré, K., Englund, M., Kvarnheden, A., Johanson, U., et al. (2003). The *DAL10* gene from Norway spruce (*Picea abies*) belongs to a potentially gymnosperm-specific subclass of MADS-box genes and is specifically active in seed cones and pollen cones. *Evol. Dev.* 5, 551–561. doi: 10.1046/j.1525-142X.2003.03060.x
- Causier, B., Schwarz-Sommer, Z., and Davies, B. (2010). Floral organ identity: 20 years of ABCs. *Semin. Cell Dev. Biol.* 21, 73–79. doi: 10.1016/j.semcdb.2009.10.005

- Chen, M. K., Lin, I. C., and Yang, C. H. (2008). Functional analysis of three lily (*Lilium longiflorum*) *APETALA1*-like MADS box genes in regulating floral transition and formation. *Plant Cell Physiol.* 49, 704–717. doi: 10.1093/pcp/pcn046
- Cho, S., Jang, S., Chae, S., Chung, K. M., Moon, Y. H., An, G., et al. (1999). Analysis of the C-terminal region of *Arabidopsis thaliana* *APETALA1* as a transcription activation domain. *Plant Mol. Biol.* 40, 419–429. doi: 10.1023/A:1006273127067
- Cui, R., Han, J., Zhao, S., Su, K., Wu, F., Du, X., et al. (2010). Functional conservation and diversification of class E floral homeotic genes in rice (*Oryza sativa*). *Plant J.* 61, 767–781. doi: 10.1111/j.1365-313X.2009.04101.x
- Danyluk, J., Kane, N. A., Breton, G., Limin, A. E., Fowler, D. B., and Sarhan, F. (2003). *TaVRT-1*, a putative transcription factor associated with vegetative to reproductive transition in cereals. *Plant Physiol.* 132, 1849–1860. doi: 10.1104/pp.103.023523
- Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., and Yanofsky, M. F. (2004). The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr. Biol.* 14, 1935–1940. doi: 10.1016/j.cub.2004.10.028
- Dreni, L., and Kater, M. M. (2013). MADS reloaded: evolution of the AGAMOUS subfamily genes. *New Phytol.* 201, 717–732. doi: 10.1111/nph.12555
- Dreni, L., Osnato, M., and Kater, M. M. (2013). The ins and outs of the rice AGAMOUS subfamily. *Mol. Plant* 6, 650–664. doi: 10.1093/mp/sst019
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Ferrario, S., Immink, R. G., and Angenent, G. C. (2004). Conservation and diversity in flower land. *Curr. Opin. Plant Biol.* 7, 84–91. doi: 10.1016/j.pbi.2003.11.003
- Ferrario, S., Immink, R. G., Shchennikova, A., Busscher-Lange, J., and Angenent, G. C. (2003). The MADS-box gene *FBP2* is required for *SEPALLATA* function in petunia. *Plant Cell* 15, 914–925. doi: 10.1105/tpc.010280
- Fourquin, C., del Cerro, C., Victoria, F. C., Vialette-Guiraud, A., de Oliveira, A. C., and Ferrándiz, C. (2013). A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in *Medicago* genus. *Plant Physiol.* 162, 907–917. doi: 10.1104/pp.113.217570
- Futamura, N., Totoki, Y., Toyoda, A., Igasaki, T., Nanjo, T., Seki, M., et al. (2008). Characterization of expressed sequence tags from a full-length enriched cDNA library of *Cryptomeria japonica* male strobili. *BMC Genomics* 9:383. doi: 10.1186/1471-2164-9-383
- Gocal, G. F., King, R. W., Blundell, C. A., Schwartz, O. M., Andersen, C. H., and Weigel, D. (2001). Evolution of floral meristem identity genes. Analysis of *Lolium temulentum* genes related to *APETALA1* and *LEAFY* of *Arabidopsis*. *Plant Physiol.* 125, 1788–1801. doi: 10.1104/pp.125.4.1788
- Gu, Q., Ferrándiz, C., Yanofsky, M. F., and Martienssen, R. (1998). The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125, 1509–1517.
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520
- Honma, T., and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525–529. doi: 10.1038/35054083
- Hsu, W., Yeh, T., Huang, K., Li, J., Chen, H., and Yang, C. (2014). AGAMOUS-LIKE13, a putative ancestor for the E functional genes, specifies male and female gametophyte morphogenesis. *Plant J.* 77, 1–15. doi: 10.1111/tpj.12363
- Huang, X., Ding, J., Effgen, S., Turck, F., and Koornneef, M. (2013). Multiple loci and genetic interactions involving flowering time genes regulate stem branching among natural variants of *Arabidopsis*. *New Phytol.* 199, 843–857. doi: 10.1111/nph.12306
- Huang, X., Effgen, S., Meyer, R. C., Theres, K., and Koornneef, M. (2012). Epistatic natural allelic variation reveals a function of AGAMOUS-LIKE6 in axillary bud formation in *Arabidopsis*. *Plant Cell* 24, 2364–2379. doi: 10.1105/tpc.112.099168
- Huijser, P., Klein, J., Lönig, W., Meijer, H., Saedler, H., and Sommer, H. (1992). Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus*. *EMBO J.* 11, 1239.
- Irish, V. F., and Sussex, I. M. (1990). Function of the *apetala-1* gene during *Arabidopsis* floral development. *Plant Cell* 2, 741–753. doi: 10.1105/tpc.2.8.741
- Jang, S., An, K., Lee, S., and An, G. (2002). Characterization of tobacco MADS-box genes involved in floral initiation. *Plant Cell Physiol.* 43, 230–238. doi: 10.1093/pcp/pcf015
- Kane, N. A., Danyluk, J., Tardif, G., Ouellet, F., Laliberté, J. F., Limin, A. E., et al. (2005). *TaVRT-2*, a member of the *StMADS-11* clade of flowering repressors, is regulated by vernalization and photoperiod in wheat. *Plant Physiol.* 138, 2354–2363. doi: 10.1104/pp.105.061762
- Kim, D. H., Doyle, M. R., Sung, S., and Amasino, R. M. (2009). Vernalization: winter and the timing of flowering in plants. *Annu. Rev. Cell Dev. Biol.* 25, 277–299. doi: 10.1146/annurev.cellbio.042308.113411
- Kim, S., Koh, J., Yoo, M. J., Kong, H., Hu, Y., Ma, H., et al. (2005). Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. *Plant J.* 43, 724–744. doi: 10.1111/j.1365-313X.2005.02487.x
- Kim, S., Soltis, P. S., and Soltis, D. E. (2013). *AGL6*-like MADS-box genes are sister to *AGL2*-like MADS-box genes. *J. Plant Biol.* 56, 315–325. doi: 10.1007/s12374-013-0147-x
- Kofuji, R., Sumikawa, N., Yamasaki, M., Kondo, K., Ueda, K., Ito, M., et al. (2003). Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol. Biol. Evol.* 20, 1963–1977. doi: 10.1093/molbev/msg216
- Kong, H., Leebens-Mack, J., Ni, W., and Ma, H. (2004). Highly heterogeneous rates of evolution in the *SKP1* gene family in plants and animals: functional and evolutionary implications. *Mol. Biol. Evol.* 21, 117–128. doi: 10.1093/molbev/msh001
- Koo, S. C., Bracko, O., Park, M. S., Schwab, R., Chun, H. J., Park, K. M., et al. (2010). Control of lateral organ development and flowering time by the *Arabidopsis thaliana* MADS-box Gene AGAMOUS-LIKE6. *Plant J.* 62, 807–816. doi: 10.1111/j.1365-313X.2010.04192.x
- Kramer, E. M., Jaramillo, M. A., and Di Stilio, V. S. (2004). Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS-box genes in angiosperms. *Genetics* 166, 1011–1023. doi: 10.1534/genetics.166.2.1011
- Kramer, E. M., Su, H. J., Wu, C. C., and Hu, J. M. (2006). A simplified explanation for the frameshift mutation that created a novel C-terminal motif in the *APETALA3* gene lineage. *BMC Evol. Biol.* 6:30. doi: 10.1186/1471-2148-6-30
- Krizek, B. A., and Meyerowitz, E. M. (1996). Mapping the protein regions responsible for the functional specificities of the *Arabidopsis* MADS domain organ-identity proteins. *Proc. Natl. Acad. Sci. U.S.A.* 93, 4063–4070. doi: 10.1073/pnas.93.9.4063
- Lamb, R. S., and Irish, V. F. (2003). Functional divergence within the *APETALA3/PISTILLATA* floral homeotic gene lineages. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6558–6563. doi: 10.1073/pnas.0631708100
- Leebens-Mack, J., Raubeson, L. A., Cui, L., Kuehl, J. V., Fourcade, M. H., Chumley, T. W., et al. (2005). Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Mol. Biol. Evol.* 22, 1948–1963. doi: 10.1093/molbev/msi191
- Li, H., Liang, W., Jia, R., Yin, C., Zong, J., Kong, H., et al. (2010). The *AGL6*-like gene *OsMADS6* regulates floral organ and meristem identities in rice. *Cell Res.* 20, 299–313. doi: 10.1038/cr.2009.143
- Li, L., Yu, X., Guo, C., Duan, X., Shan, H., Zhang, R., et al. (2015). Interactions among proteins of floral MADS-box genes in *Nuphar pumila* (Nymphaeaceae) and the most recent common ancestor of extant angiosperms help understand the underlying mechanisms of the origin of the flower. *J. Syst. Evol.* 53, 285–296. doi: 10.1111/jse.12148
- Litt, A. (2007). An evaluation of A-function: evidence from the *APETALA1* and *APETALA2* gene lineages. *Int. J. Plant Sci.* 168, 73–91. doi: 10.1086/509662
- Litt, A., and Irish, V. F. (2003). Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* 165, 821–833.
- Litt, A., and Kramer, E. M. (2010). The ABC model and the diversification of floral organ identity. *Semin. Cell Dev. Biol.* 21, 129–137. doi: 10.1016/j.semcdb.2009.11.019
- Liu, Y., Guo, C., Xu, G., Shan, H., and Kong, H. (2011). Evolutionary pattern of the regulatory network for flower development: insights gained from a comparison of two *Arabidopsis* species. *J. Syst. Evol.* 49, 528–538. doi: 10.1111/j.1759-6831.2011.00158.x

- Long, M., VanKuren, N. W., Chen, S., and Vrbancan, M. D. (2013). New gene evolution: little did we know. *Annu. Rev. Genet.* 47, 307–333. doi: 10.1146/annurev-genet-111212-133301
- Mellway, R. D., and Lund, S. T. (2013). Interaction analysis of grapevine MIKCC-type MADS transcription factors and heterologous expression of putative véraison regulators in tomato. *J. Plant Physiol.* 170, 1424–1433. doi: 10.1016/j.jplph.2013.05.010
- Melzer, R., Härter, A., Rümpler, F., Kim, S., Soltis, P. S., Soltis, D. E., et al. (2014). DEF and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. *Ann. Bot.* 114, 1431–1443. doi: 10.1093/aob/mcu094
- Michaels, S. D., and Amasino, R. M. (1999). *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11, 949–956. doi: 10.1105/tpc.11.5.949
- Murai, K., Miyamae, M., Kato, H., Takumi, S., and Ogihara, Y. (2003). *WAP1*, a wheat *APETALA1* homolog, plays a central role in the phase transition from vegetative to reproductive growth. *Plant Cell Physiol.* 44, 1255–1265. doi: 10.1093/pcp/pcg171
- Nam, J., dePamphilis, C. W., Ma, H., and Nei, M. (2003). Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol. Biol. Evol.* 20, 1435–1447. doi: 10.1093/molbev/msg152
- Ohmori, S., Kimizu, M., Sugita, M., Miyao, A., Hirochika, H., Uchida, E., et al. (2009). *MOSAIC FLORAL ORGANS1*, an *AGL6*-like MADS-box gene, regulates floral organ identity and meristem fate in rice. *Plant Cell* 21, 3008–3025. doi: 10.1105/tpc.109.068742
- Pabón-Mora, N., Ambrose, B. A., and Litt, A. (2012). Poppy *APETALA1/FRUITFULL* orthologs control flowering time, branching, perianth identity, and fruit development. *Plant Physiol.* 158, 1685–1704. doi: 10.1104/pp.111.192104
- Pabón-Mora, N., Sharma, B., Holappa, L. D., Kramer, E. M., and Litt, A. (2013). The *Aquilegia FRUITFULL*-like genes play key roles in leaf morphogenesis and inflorescence development. *Plant J.* 74, 197–212. doi: 10.1111/tjp.12113
- Parenicova, L. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis* new openings to the MADS world. *Plant Cell* 15, 1538–1551. doi: 10.1105/tpc.011544
- Pelaz, S., Tapia-López, R., Alvarez-Buylla, E. R., and Yanofsky, M. F. (2001). Conversion of leaves into petals in *Arabidopsis*. *Curr. Biol.* 11, 182–184. doi: 10.1016/S0960-9822(01)00024-0
- Piwarzyk, E., Yang, Y., and Jack, T. (2007). Conserved C-terminal motifs of the *Arabidopsis* proteins *APETALA3* and *PISTILLATA* are dispensable for floral organ identity function. *Plant Physiol.* 145, 1495–1505. doi: 10.1104/pp.107.105346
- Pnueli, L., Hareven, D., Broday, L., Hurwitz, C., and Lifschitz, E. (1994). The *TM5* MADS-box gene mediates organ differentiation in the three inner whorls of tomato flowers. *Plant Cell* 6, 175–186. doi: 10.1105/tpc.6.2.175
- Posada, D., and Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818. doi: 10.1093/bioinformatics/14.9.817
- Preston, J. C., and Kellogg, E. A. (2006). Reconstructing the evolutionary history of paralogous *APETALA1/FRUITFULL*-like genes in grasses (Poaceae). *Genetics* 174, 421–437. doi: 10.1534/genetics.106.057125
- Raes, J., and Van de Peer, Y. (2005). Functional divergence of proteins through frameshift mutations. *Trends Genet.* 21, 428–431. doi: 10.1016/j.tig.2005.05.013
- Rijkema, A. S., Vandenbussche, M., Koes, R., Heijmans, K., and Gerats, T. (2010). Variations on a theme: changes in the floral ABCs in angiosperms. *Semin. Cell Dev. Biol.* 21, 100–107. doi: 10.1016/j.semcdb.2009.11.002
- Rijkema, A. S., Zethof, J., Gerats, T., and Vandenbussche, M. (2009). The petunia *AGL6* gene has a *SEPALLATA*-like function in floral patterning. *Plant J.* 60, 1–9. doi: 10.1111/j.1365-3113.2009.03917.x
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Roshan, U., and Livesay, D. R. (2006). Probalalign: multiple sequence alignment using partition function posterior probabilities. *Bioinformatics* 22, 2715–2721. doi: 10.1093/bioinformatics/btl472
- Roy, S. W., and Gilbert, W. (2005). Rates of intron loss and gain: implications for early eukaryotic evolution. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5773–5778. doi: 10.1073/pnas.0500383102
- Ruelens, P., de Maagd, R. A., Proost, S., Theissen, G., Geuten, K., and Kaufmann, K. (2013). *FLOWERING LOCUS C* in monocots and the tandem origin of angiosperm-specific MADS-box genes. *Nat. Commun.* 4:2280. doi: 10.1038/ncomms3280
- Ruokolainen, S., Ng, Y. P., Albert, V. A., Elomaa, P., and Teeri, T. H. (2010). Large scale interaction analysis predicts that the *Gerbera hybrida* floral E function is provided both by general and specialized proteins. *BMC Plant Biol.* 10:129. doi: 10.1186/1471-2229-10-129
- Seok, H. Y., Park, H. Y., Park, J. I., Lee, Y. M., Lee, S. Y., An, G., et al. (2010). Rice ternary MADS protein complexes containing class B MADS heterodimer. *Biochem. Biophys. Res. Commun.* 401, 598–604. doi: 10.1016/j.bbrc.2010.09.108
- Shan, H., Zhang, N., Liu, C., Xu, G., Zhang, J., Chen, Z., et al. (2007). Patterns of gene duplication and functional diversification during the evolution of the *API/SQUA* subfamily of plant MADS-box genes. *Mol. Phylogenet. Evol.* 44, 26–41. doi: 10.1016/j.ympev.2007.02.016
- Sheldon, C. C., Jean Finnegan, E., Dennis, E. S., and James Peacock, W. (2006). Quantitative effects of vernalization on *FLC* and *SOC1* expression. *Plant J.* 45, 871–883. doi: 10.1111/j.1365-3113.2006.02652.x
- Smaczniak, C., Immink, R. G., Muiño, J. M., Blanvillain, R., Busscher, M., Busscher-Lange, J., et al. (2012). Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1560–1565. doi: 10.1073/pnas.1112871109
- Su, K., Zhao, S., Shan, H., Kong, H., Lu, W., Theissen, G., et al. (2008). The MIK region rather than the C-terminal domain of AP3-like class B floral homeotic proteins determines functional specificity in the development and evolution of petals. *New Phytol.* 178, 544–558. doi: 10.1111/j.1469-8137.2008.02382.x
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi: 10.1126/science.1153917
- Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J. T., Munster, T., et al. (2000). A short history of MADS-box genes in plants. *Plant Mol. Biol.* 42, 115–149. doi: 10.1023/A:1006332105728
- Theissen, G., Kim, J. T., and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* 43, 484–516. doi: 10.1007/BF02337521
- Thompson, B. E., Bartling, L., Whipple, C., Hall, D. H., Sakai, H., Schmidt, R., et al. (2009). *bearded-ear* encodes a MADS-box transcription factor critical for maize floral development. *Plant Cell* 21, 2578–2590. doi: 10.1105/tpc.109.067751
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882. doi: 10.1093/nar/25.24.4876
- Trevaskis, B., Bagnall, D. J., Ellis, M. H., Peacock, W. J., and Dennis, E. S. (2003). MADS box genes control vernalization-induced flowering in cereals. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13099–13104. doi: 10.1073/pnas.1635053100
- Ubi, B. E., Saito, T., Bai, S., Nishitani, C., Ban, Y., Ikeda, K., et al. (2013). Characterization of 10 MADS-box genes from *Pyrus pyrifolia* and their differential expression during fruit development and ripening. *Gene* 528, 183–194. doi: 10.1016/j.gene.2013.07.018
- Vandenbussche, M., Theissen, G., Van de Peer, Y., and Gerats, T. (2003a). Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucleic Acids Res.* 31, 4401–4409. doi: 10.1093/nar/gkg642
- Vandenbussche, M., Zethof, J., Souer, E., Koes, R., Tornelli, G. B., Pezzotti, M., et al. (2003b). Toward the analysis of the petunia MADS-box gene family by reverse and forward transposon insertion mutagenesis approaches: B, C, and D floral organ identity functions require *SEPALLATA*-like MADS-box genes in petunia. *Plant Cell* 15, 2680–2693. doi: 10.1105/tpc.017376
- Wang, P., Liao, H., Zhang, W., Yu, X., Zhang, R., Shan, H., et al. (2015). Flexibility in the structure of spiral flowers and its underlying mechanisms. *Nat. Plants* 2, 15188. doi: 10.1038/nplants.2015.188
- Wang, Y., Melzer, R., and Theissen, G. (2010). Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm *Gnetum gnemon*

- provide a clue to the evolutionary origin of 'floral quartets'. *Plant J.* 64, 177–190. doi: 10.1111/j.1365-3113X.2010.04325.x
- Wong, C. E., Singh, M. B., and Bhalla, P. L. (2013). Novel members of the *AGAMOUS LIKE 6* subfamily of MIKCC-type MADS-box genes in soybean. *BMC Plant Biol.* 13:105. doi: 10.1186/1471-2229-13-105
- Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1187–1192. doi: 10.1073/pnas.1109047109
- Xu, G., and Kong, H. (2007). Duplication and divergence of floral MADS-box genes in grasses: evidence for the generation and modification of novel regulators. *J. Integr. Plant Biol.* 49, 927–939. doi: 10.1111/j.1744-7909.2007.00502.x
- Xu, G., Ma, H., Nei, M., and Kong, H. (2009). Evolution of F-box genes in plants: different modes of sequence divergence and their relationships with functional diversification. *Proc. Natl. Acad. Sci. U.S.A.* 106, 835–840. doi: 10.1073/pnas.0812043106
- Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., and Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene *VRN1*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6263–6268. doi: 10.1073/pnas.0937399100
- Yockteng, R., Almeida, A. M., Morioka, K., Alvarez-Buylla, E. R., and Specht, C. D. (2013). Molecular evolution and patterns of duplication in the *SEP/AGL6*-like lineage of the Zingiberales: a proposed mechanism for floral diversification. *Mol. Biol. Evol.* 30, 2401–2422. doi: 10.1093/molbev/mst137
- Yoo, S. K., Hong, S. M., Lee, J. S., and Ahn, J. H. (2011). A genetic screen for leaf movement mutants identifies a potential role for *AGAMOUS-LIKE 6 (AGL6)* in circadian-clock control. *Mol. Cells* 31, 281–287. doi: 10.1007/s10059-011-0035-5
- Zahn, L. M., Kong, H., Leebens-Mack, J. H., Kim, S., Soltis, P. S., Landherr, L. L., et al. (2005). The evolution of the *SEPALLATA* subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* 169, 2209–2223. doi: 10.1534/genetics.104.037770
- Zahn, L. M., Leebens-Mack, J. H., Arrington, J. M., Hu, Y., Landherr, L. L., dePamphilis, C. W., et al. (2006). Conservation and divergence in the *AGAMOUS* subfamily of MADS-box genes: evidence of independent sub- and neofunctionalization events. *Evol. Dev.* 8, 30–45. doi: 10.1111/j.1525-142X.2006.05073.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Yu, Duan, Zhang, Fu, Ye, Kong, Xu and Shan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read,
for greatest visibility



COLLABORATIVE PEER-REVIEW

Designed to be rigorous
– yet also collaborative,
fair and constructive



FAST PUBLICATION

Average 85 days from
submission to publication
(across all journals)



COPYRIGHT TO AUTHORS

No limit to article
distribution and re-use



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



SUPPORT

By our Swiss-based
editorial team



IMPACT METRICS

Advanced metrics
track your article's impact



GLOBAL SPREAD

5'100'000+ monthly
article views
and downloads



LOOP RESEARCH NETWORK

Our network
increases readership
for your article

Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on

