The Development and Evolution of Floral Symmetry in the Zingiberales and Interactive Tools for Teaching Evolution (ArborEd)

By

Riva Anne Bruenn

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Chelsea D. Specht, Chair
Professor Jennifer Fletcher
Professor Sarah Hake
Professor Nicole King

Summer 2017

Abstract

The Development and Evolution of Floral Symmetry in the Zingiberales and Interactive Tools for Teaching Evolution (ArborEd)

by

Riva Anne Bruenn

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Chelsea D. Specht, Chair

Floral symmetry is a key innovation in the evolution of flowering plants. Zygomorphy, or single-planed symmetry, is associated with the diversification of many flowering plant lineages. The model system for floral symmetry is the snapdragon (*Antirrhinum majus*). In *A. majus* flowers, a set of TCP and MYB-related transcription factors form a core gene regulatory network necessary for zygomorphy. The genes involved in this network have been implicated in several independent transitions to zygomorphy from actinomorphy (many-planed symmetry). Although the TCP components of the symmetry network have been investigated across flowering plants, MYB-related transcription factors remain largely unstudied outside of the Asterid group containing *A. majus* and close relatives. Here we investigate the evolution of MYB-related genes *DIVARICATA-like (DIV-like), RADIALIS-like (RAD-like),* and *DIVARICATA and RADIALIS INTERACTING FACTOR-like (DRIF-like)* across flowering plants, and their expression patterns in the developing flowers of two zygomorphic species of the monocot order Zingiberales.

We found that *RAD-like* and *DIV-like* are sister MYB-related genes which diverged before the diversification of flowering plants. Each gene contains one MYB-like domain that has been closely conserved throughout flowering plant evolution. Furthermore, we identified candidate homologs to *A. majus RAD* and *DIV* in several monocot taxa, with at least three copies of each in the Zingiberales.

In the Zingiberales, *RAD-like* and *DIV-like* genes are expressed in *Costus spicatus* (Costaceae) and *Musa basjoo* (Musaceae) in patterns consistent with roles in floral symmetry. Using Reverse Transcription PCR and *in situ* hybridization we recovered asymmetric expression patterns for some *RAD-like* genes across the dorsal/ventral plane of developing flowers, and universal expression of *DIV-like* genes, consistent with the model known from *Antirrhinum majus.*

We identified *DRIF-like* genes across flowering plants, recovering a previously undescribed duplication in eudicot *DRIF* Group 1 genes. Furthermore, we recovered candidate *DRIF-like* genes in *Musa basjoo* (Musaceae: Zingiberales) with expression patterns similar to those described in *A. majus DRIF1* and *DRIF2*.

Finally, we developed a tutorial for high school and college students to investigate a coevolutionary hypothesis in sharpshooters and their bacterial endosymbionts. This tool will help students understand how comparative evolutionary research is performed, and give them hands-on experience performing common analyses.

Acknowledgements

There are many people who have made it possible for me to come to Berkeley and to make it through the years I've spent here. I will begin with my dissertation advisor, Dr. Chelsea Specht. Chelsea, you have always encouraged me to take on opportunities and responsibilities I didn't think I was capable of. Thank you for guiding me to the upper reaches of my limits. Thank you also for doing everything you can to make your field welcoming and inclusive to everyone. You inspire me to pay attention, to speak up, and to push for change. To the whole lab – thank you for building the supportive and challenging environment that helped me grow into the scholar I am today, and that has built the foundation for the educator I want to be tomorrow. Two people I'd like to acknowledge specially – thank you Susan for helping me let go of a future in research and thank you Joyce for challenging me to be my best self and forgiving me when I haven't been.

I would like to acknowledge the incredible guidance my committee has given me throughout this process. Thank you to Drs. Jennifer Fletcher, Sarah Hake, and Nicole King. You challenged me with big and small picture questions, gave me valuable advice, inspired me to improve as a speaker, and encouraged me in my career goals.

The resources and advisors at UC Berkeley made my PhD possible. Thank you Rocio Sanchez for helping me without question even when my problems were entirely my fault, and finding solutions even when I was out of options and out of hope. Thank you Dana Jantz, you keep this department going! Thank you to Drs. Denise Schichnes and Steve Ruzin at the Bioimaging Facility for providing advice, training, and resources for all my imaging and histology work. Thank you Drs. Linda von Hoene and Sabrina Soracco at the GSI Teaching and Resource Center for providing the training, knowledge, and practice that has informed my development as a teacher and helped me build confidence in the classroom.

Finally I would like to thank my family and out-of-lab friends. Thank you Carrie Gratiot for helping me be brave enough to move across the country. Thank you Judy Brown for urging me to take care of myself and be mindful of my every day. Thank you Jeremy Bruenn for making your lab some of my earliest memories and for always making me answer my own questions. Thank you to my cohort, class of 2012. You gave me the community and support I needed to feel like I belonged and consistently re-inspired my joy of discovery. I hope we'll keep our camping trips going. Thank you especially to Carine, Kate, Becky, Anna, and Ben. More than anyone the five of you helped me find the science outreach opportunities, seminars, books, dance clubs, national parks, bars, and coffee shops that made this whole process worth while. Carine – our semesters teaching Plant Morphology together convinced me I could be a teacher and our weeks of dissertating together convinced me I could graduate.

Finally, I want to thank my wife. Thank you so much Emily Anne Moon for listening to me practice all my talks, for holding me while I cried about all my real and ridiculous fears and failures, and for always believing in me. It's really something to know that no matter how badly the day goes you are at the end of it.

I hope that those who have helped me along feel how much I value them and that I can give back as much as I've taken.

Table of Contents

Introduction

The Evolutionary Context of Floral Symmetry

Floral symmetry is a key innovation that has likely impacted the diversification of flowering plants over the course of their evolution. Transitions to zygomorphy (one plane of symmetry, Figure 1) from actinomorphy (many planes) have occurred at least 70 times independently throughout the evolution of flowering plants, with a recent estimate of at least 130 times (Citerne et al. 2010; Reyes, Sauquet, and Nadot 2016). Zygomorphy has been linked to reduction in pollen production, more exclusive plant/pollinator relationships, and greater rates of speciation (Peter K Endress 1999a; Giurfa, Dafni, and Neal 1999; Cubas 2004; Sargent 2004; Ushimaru and Hyodo 2005; Hu et al. 2008; Peter K Endress and Doyle 2009; Ushimaru et al. 2009; Waser et al. 2010; Meara et al. 2016).

Many morphological and developmental factors contribute to zygomorphy across flowering plant lineages. Structural zygomorphy arises by the differential repression, elaboration, or reduction of organs while presentation zygomorphy arises through differential organ coloration, placement, or expansion (Rudall and Bateman 2004; Peter K. Endress 2012). Zygomorphy can also arise at different points of development. An initially actinomorphic floral meristem can become zygomorphic at the point of the first organs initiating, which may be placed or sized asymmetrically, as organs expand differentially, fuse, or organ abortion occurs, or during the final stages of development when organ elaboration, coloration, and additional fusion may occur (reviewed in Reyes, Sauquet, and Nadot 2016).

Despite the diversity of morphological patterns giving rise to zygomorphy, floral symmetry seems to be fairly labile. In addition to the large number of transitions from actinomorphy to zygomorphy, reversions are not uncommon (Hileman 2014b; Reyes, Sauquet, and Nadot 2016). The first actinomorphic mutant of a normally zygomorphic species was found to be the result an epigenetic mutation in a single gene, *CYCLOIDEA (CYC)* (Luo et al. 1996). Changes in copy number and expression pattern of *CYC-like* genes correlates with reversions to actinomorphy in lineages across flowering plants (reviewed in Hileman 2014b). Small numbers of loci can have large pleiotropic effects on suites of related floral characters (Smith 2016). In some cases, the same genes have been recruited independently in the convergent evolution of traits such as the elimination of a vernalization requirement for germination or precursors to C_4 photosynthesis (Stern 2013). Alternatively, mutations in different genes within the same regulatory pathway can result in convergence, as in anthocyanin production in Solanaceae (Smith and Rausher 2011). Despite the complexity and diversity of floral characteristics, the same sets of genes may have been independently coopted in zygomorphic lineages across flowering plant evolution.

The Floral Symmetry Gene Network in Antirrhinum majus

Antirrhinum majus, the snapdragon, is the model organism for floral symmetry. The *Antirrhinum majus* flower is zygomorphic with two dorsal petals and two lateral petals that are each internally asymmetrical and a single symmetrical ventral petal. The dorsal stamen is arrested during development, leaving four stamens at anthesis. The two sets are different in length, making the androecium (stamens) zygomorphic as well as the corolla (petals). (Luo et al. 1996).

In *Antirrhinum majus,* the restriction of *CYCLOIDEA (CYC)* and *DICHOTOMA (DICH)* expression, and later *RADIALIS (RAD)* to the dorsal portion of the flower contributes to several developmental processes, from dorsal stamen abortion to differential growth of dorsal petals. In *A. majus*, flowers are nearly actinomorphic in early development, but the retardation of the dorsal stamen and eventually the differential growth of petals and stamens leads to zygomorphy at anthesis, or floral maturity. Early dorsal *CYC* expression leads to stamen abortion and growth reduction in dorsal regions of the floral meristem, while continued expression throughout late development results in increases in petal lobe growth in dorsal regions (Luo et al. 1996; Luo et al. 1999). Initial repression of organ growth, which is never reversed in the dorsal stamen, is likely the result of down-regulation of cell cycling genes (Gaudin et al. 2000; Clark and Coen 2002). The asymmetric organ growth that produces a larger, internally symmetric ventral petal and smaller internally asymmetric lateral and dorsal petals is due to the competitive actions of *RAD* and *DIVARICATA (DIV)* (Almeida, Rocheta, and Galego 1997; Galego and Almeida 2002; Corley et al. 2005; Costa et al. 2005; Perez-Rodriguez et al. 2005). The two proteins compete for binding partners *RAD and DIV INTERACTING FACTORs (DRIFs).* Where RAD protein is absent, DIV and DRIF complexes activate downstream transcription targets specifying ventral petal identity (Sobral 2010; Raimundo et al. 2013). *DIV* expression constricts throughout development, initially being present throughout the developing ventral petal and overlapping regions of lateral petals and eventually retreating to a furrow of cells demarking the boundary between the corolla tube and the lobes (Galego and Almeida 2002). Due to this restriction, in initial development DIV protein induces the expansion of the ventral petal and adjacent portions of lateral petals, but later contributes to the lobe vs. tube boundary.

The Floral Symmetry Gene Network

The core network of genes known from *A. majus* may have been repeatedly coopted in transitions from actinomorphy to zygomorphy. *CYC-like, RAD-like,* and *DIV-like* genes have all been implicated in floral symmetry in clades with independently derived zygomorphy. Developmental genetics research in orders such as Asterales (*Gerbera*), Fabales (*Pisum*), and Poales (*Zea*, *Oryza*) supports a role for *CYC-like* transcription factors in floral symmetry through stamen abortion and differential dorsal-ventral (dorsiventral) petal morphology (Zheng Wang et al. 2008; Preston 2010; Zhang, Kramer, and Davis 2010; Busch et al. 2012; Hoshino et al. 2014). Dorsiventral asymmetry in *CYC-like* expression has been found to be consistent in many additional lineages based on RT-PCR and *in situ* hybridization (reviewed in Hileman 2014a; Hileman 2014b). Interestingly, *CYC-like* genes are expressed dorsally in zygomorphic eudicots, but appear to be expressed ventrally in zygomorphic monocots based on studies of taxa in the Zingiberales and its sister order, the Commelinales (Bartlett and Specht 2011; Preston and Hileman 2012). To date, *RAD-like* and *DIV-like* genes have been studied in zygomorphic lineages within the Asterids, evolutionarily distant from the monocots, and recently in orchids (Valoroso et al. 2017). In the Dipsacales and in *Orchis italica*, at least one copy each of *RAD-like* and *DIV-like* genes are expressed similarly to *RAD* and *DIV* in *A. majus*, consistent with roles in dorsiventral asymmetry. Additionally, these taxa maintain the same ratio of *RAD-like* and *DIV-like* gene copies (Howarth and Donoghue 2009; Boyden, Donoghue, and Howarth 2014; Valoroso et al. 2017).

The Zingiberales as a Study System

The tropical ginger order Zingiberales is a model system in which to study monocot floral symmetry. The species in the order are zygomorphic with very few exceptions. Zygomorphy in the order is independently derived, with the ancestors of the Zingiberales and the sister order Commelinales thought to be actinomorphic (Rudall and Bateman 2004). Across the eight families in the Zingiberales, zygomorphy occurs due to multiple morphological factors including organ reduction, repression and elaboration in the perianth (sepals and petals) and the androecium (stamens) (Figure 2). In the ginger families Marantaceae, Cannaceae, Zingiberaceae, and Costaceae, zygomorphy is largely the result of dorsivenral asymmetry in the androecium. In the banana families Heliconiaceae, Strelitziaceae, Lowiaceae, and Musaceae, dorsiventral asymmetry of the perianth contributes most to overall zygomorphy. The dorsal petal and dorsal stamens of flowers in the order are differentiated from the rest of the corolla and androecium by delays in development (Bartlett and Specht 2011). In the ginger families, the androecium is larger and more elaborate than the banana families, and makes up the bulk of the floral display. The ginger families also experienced reduction in the number of fertile stamens before their diversification, with all but the dorsal stamen sterile and laminar (staminodes). In the Costaceae and Zingiberaceae, these staminodes fuse to form an elaborate staminodial labellum (B. K. Kirchoff 1988; B. K. Kirchoff 1991; Bartlett and Specht 2010).

The presence of novel organs (like the staminodial labellum) as well as the transition from perianth to androecium floral display and zygomorphy present an interesting test case for the adaptation of the floral symmetry gene network to a variety of floral morphologies. Changes in copy number, expression pattern, or up and downstream regulation may be responsible for changes in floral morphology across the order. Additionally, the evolutionary distance between the Zingiberales and the Lamiales (including *A. majus*) allows us to test the extent of conservation of the gene network across flowering plants.

Based on previous studies, *CYC-like* expression in the Zingiberales appears to be consistent with a role in zygomorphy. Three copies of *CYC-like* genes were found in the order (Bartlett and Specht 2011), and expression was characterized for two of these copies based on *in situ* hybridization in a ginger family species *Costus spicatus* (Costaceae) and a banana family species *Heliconia stricta* (Heliconiaceae). *zinTBL1a* expression was found in the petaloid sterile stamens (staminodes) of both species and was additionally found throughout the ventral perianth and dorsal fertile thecae (pollen producing structures) of *Costus spicatus*. *zinTBL2* expression was found in the ventral sepals of *Heliconia stricta* and the dorsal fertile thecae of *Costus spicatus*. *zinTBL1a* may be involved in stamen abortion in the Zingiberales, though the presence of expression in the fertile stamen of *C. spicatus* may indicate multiple functions. *zinTBL2* in *H. stricta* and *zinTBL1a* in *C. spicatus* may restrict growth in the ventral regions of the flowers during early development, resulting in the characteristically larger dorsal perianth members. The reversed expression patterns of *TBL1a* and *TBL2* in the two taxa correlates with the divergent symmetry patterns, and may reflect different regulation or different function in the two groups (Bartlett and Specht 2011).

Using these data and the evolution and expression of other symmetry network components, we tested the hypothesis that the symmetry network known from *A. majus* has been coopted in a

transition to zygomorphy in an evolutionarily distant order of monocots. Given the induction of *RAD* expression caused by *CYC* and *DICH* in *A. majus*, we expected to find *RAD-like* expression patterns that overlapped with the characterized *TBL* patterns. Furthermore, given the different patterns of expression found in Heliconiaceae vs. Costaceae, we expected *RAD-like* expression patterns to be altered between the banana and ginger families, correlated with shifts in symmetry from the perianth to the androecium. Given the universal expression of *DIV* and the *DRIFs* in *A. majus* flowers, we expected universal expression of orthologs to these genes in Zingiberales flowers. We recovered overlapping but not matching expression patterns for two copies of *TBL* and *RAD-like* genes and universal expression of *DIV-like* and *DRIF-like* genes. Although the relationships between copies are unclear, the asymmetric expression patterns of *TBL* and *RAD-like* genes and the universal expression of *DIV-like* and *DRIF-like* are consistent with roles in zygomorphy in these taxa.

Using ArborEd to Teach Evolutionary Concepts

Arbor is an online platform useful to evolutionary biologists interested in performing comparative analyses. Arbor provides a fast and simple way for researchers to use and combine many common analyses without needing to download, compile, and learn to use software on their own. Arbor is also a platform on which to write and combine novel analyses and pipelines. ArborEd is a new instance of Arbor designed to help students learn how evolutionary biologists perform research through hands-on experience with real data. We have used the Arbor platform to design a coevolution tutorial for high school and college students using a published example of mutualism between xylem-feeding insects and their bacterial endosymbionts.

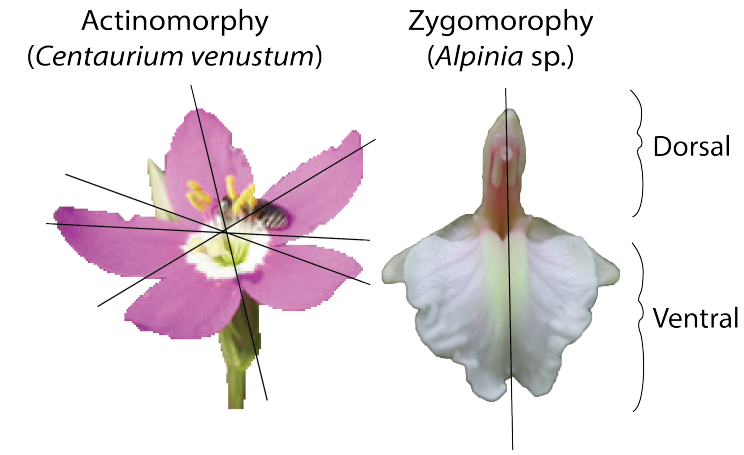


Figure 1: Actinomorphy and zygomorphy. Lines represent planes of symmetry. Photo credit: Suberite&ryquail@flikr, Alan Carle

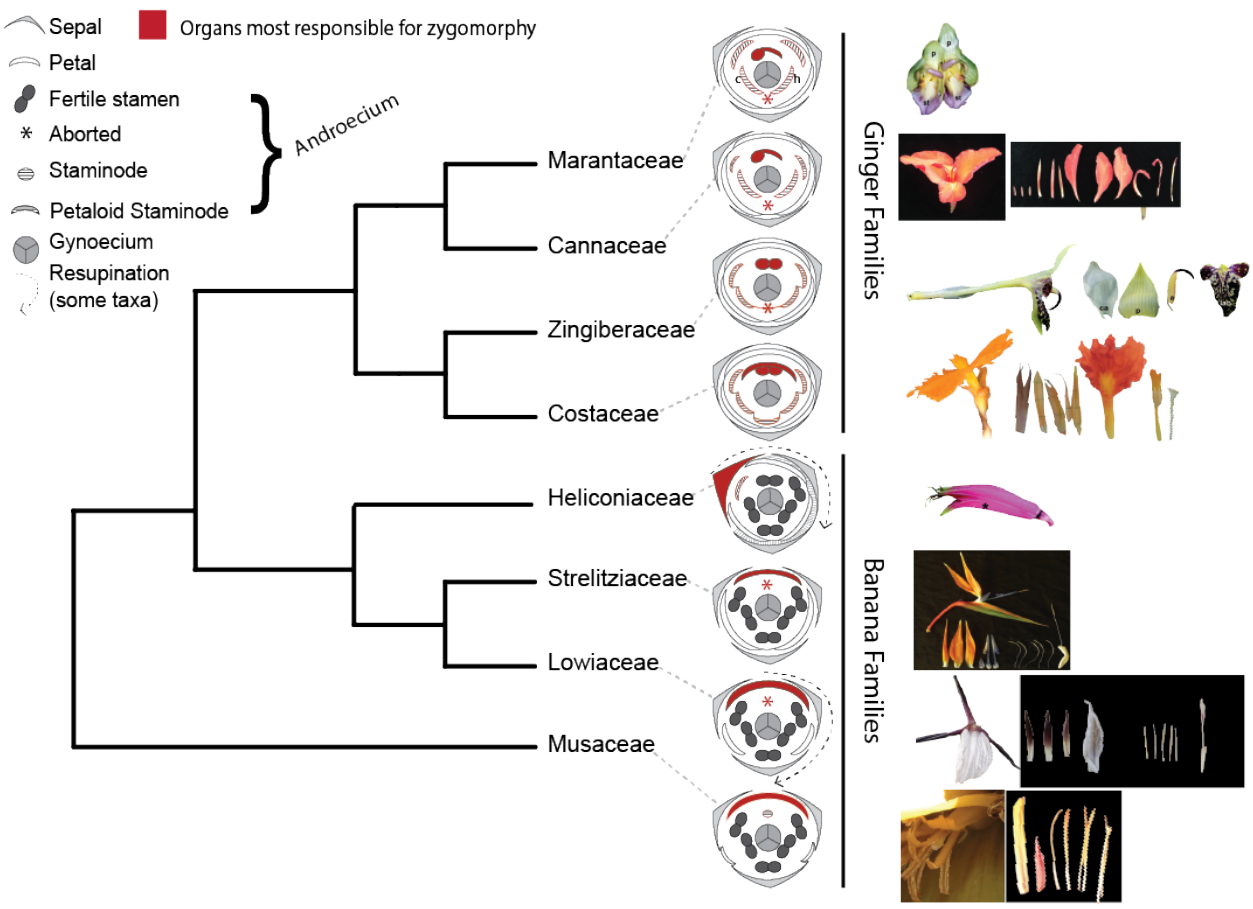


Figure 2: A phylogeny of the Zingiberales (Sass et al. 2016) with floral diagrams and images associated with representative species belonging to each family, adapted from (Bartlett 2010). Dotted lines with arrows indicate reorientation of the flower during development. Red highlighted regions are the organs which contribute most to zygomorphy in that family.

Chapter 1: Evolution of *RADIALIS and DIVARICATA*

Abstract*:*

RADIALIS (RAD) and *DIVARICATA (DIV)* are MYB-related genes with one and two MYB-like repeats respectively. MYB and MYB-related genes form a large group of transcription factors present in species from across the eukaryotes (Feller et al. 2011). Although the evolutionary histories of many MYB genes are known, the MYB-related group is diverse and largely uncharacterized, with hypothesized multiple independent origins in plants (Rosinski and Atchley 1998). Here we identify *RAD-like* and *DIV-like* sequences from 71 taxa across flowering plants and gymnosperms, with sequences from 17 Zingiberales taxa. We present phylogenetic trees for use in identifying *RAD-like* and *DIV-like* genes in a variety of taxa of interest, including taxa across the monocots. We recovered one *DIV-like* gene present across monocots, with lineage-specific duplications. Within the *RAD-like* genes we recovered a duplication prior to the diversification of monocots, followed by additional lineage-specific duplications. Both *DIV-like* and *RAD-like* genes have at least three copies in the Zingiberales, consistent with a 1-to-1 relationship. Our results support the homology of the first MYB-like domain of *DIV* with the MYB-like domain of *RAD*, and the homology of the second MYB-like domain of *DIV* with the MYB-like domain of *CCA1/LHY*. We propose that MYB-related genes be characterized by gene and/or domain homology, rather than by the number of MYB-like repeats.

Introduction:

RADIALIS (RAD), DIVARICATA (DIV), and *DIV and RAD INTERACTING FACTORS (DRIFs)* are MYB-related genes; transcription factors characterized as having 1-4 imperfect tandem repeats of about 50 amino acids. Each MYB-like repeat forms a helix-turn-helix thought to bind the major groove of a target DNA sequence (Rosinski and Atchley 1998; Dubos 2013). The MYB protein family is present in all eukaryotic organisms studied thus far (Feller et al. 2011) and took its acronym from v-MYB, a component of avian myeloblastosis retrovirus (Klempnauer, Gonda, and Michael Bishop 1982). MYB genes have diverse functions, from roles in anthocyanin production to conical cell formation to roles in the circadian clock. They are quite numerous, with hundreds of representatives in *Arabidopsis* and *Oryza* (Feller et al. 2011). The typical MYB motif has 3 regularly spaced tryptophans (Rosinski and Atchley 1998) which together form a hydrophobic core necessary for DNA binding (Dubos 2013). MYB family members are characterized by the number of MYB repeats they possess (Dubos 2013). The most common in plants are 2-domain MYB plant proteins, which are called "R2R3" proteins because their domains are most similar to the second and third MYB repeat of the animal MYB gene c-MYB (Jin and Martin 1999). Although similar in many respects, the MYB gene "family" is thought to have polyphyletic origins, meaning that MYB genes may not have a shared common ancestor (Rosinski and Atchley 1998; Jin and Martin 1999).

Some MYB-related genes, including *DIV, RAD,* and the *DRIFs,* are further differentiated from other MYB genes by having noticeably altered MYB repeats, lacking one of the characteristic 3 tryptophans (Du et al. 2013). *DIV* has 2 MYB-like repeats (Galego and Almeida 2002), *RAD* has 1 (Corley et al. 2005; Costa et al. 2005), and the *DRIFs* have 1 MYB-like domain and an additional domain of unknown function (Raimundo et al. 2013). The first domain of *DIV* is

alignable to the single domain of *RAD*, but the domains of the *DRIFs* are not alignable to either *RAD* or *DIV*. Because the *DRIF* domains are not homologous to those in *RAD* or *DIV,* the remainder of this chapter will focus on *RAD* and *DIV*.

RAD and *DIV* were first functionally characterized in *Antirrhinum majus* (Galego and Almeida 2002; Corley et al. 2005; Costa et al. 2005). RAD and DIV proteins share a protein-binding domain, but only DIV has the DNA binding domain necessary to induce expression of downstream targets (Stevenson et al. 2006). DIV acts in concert with protein binding partners DIV AND RAD INTERACTING FACTORs 1 and 2 (DRIFs). In the presence of RAD, RAD binds to the DRIFs and restricts them to the cytoplasm, while in the absence of RAD, DRIFs travel to the nucleus where they are bound by DIV, together regulating transcription (Sobral 2010; Raimundo et al. 2013). Thus, where RAD is present, DIV/DRIFs cannot activate genes necessary for ventral petal development.

To date, the roles of *RAD-like* and *DIV-like* genes in symmetry have only been studied in zygomorphic lineages within the Asterids, evolutionarily distant from the monocots, and recently in the monocot family Orchidaceae. In *Bournea* (Gesneriaceae: Lamiales), *RAD-like* gene expression is asymmetric across the dorsal/ventral plane but becomes symmetric later in development, correlating with the flower's transition to actinomorphy (Zhou et al. 2008). In the Dipsacales, at least one copy each of *RAD-like* and *DIV-like* genes are expressed similarly to the single copy of *RAD* and *DIV* in *A. majus*, consistent with influencing dorsiventral asymmetry. Additionally, these taxa maintain the same copy number of *RAD-like* and *DIV-like* genes, supporting the one-on-one relationship between *RAD* and *DIV* present in *A. majus* (Howarth and Donoghue 2009; Boyden, Donoghue, and Howarth 2014). Recently published research in orchids likewise uncovered asymmetric expression of a *RAD-like* ortholog (Valoroso et al. 2017). Evidence presented thus far is consistent with *RAD* and *DIV* homologs playing roles in floral symmetry outside the model system of *Antirrhinum*.

Considering the hundreds of MYB and MYB-related genes present in plants, characterization of these genes and their functions is a long way off. Reliable identification of genes homologous to those of interest in model systems remains a challenge (Feuermann et al. 2016). Although BLAST algorithms (Altschup et al. 1990) may reveal the closest homologs of the query genes, this is often not the case. Phylogenetic analysis is the best estimate of homology (Koski and Golding 2001), and functional characterization can be informed through a homology-based approach (Feuermann et al. 2016). While such functional prediction methods - such as annotation by BLAST to characterized homologs or motif finding algorithms - may provide reasonable estimates, there is no substitute for detailed genetic studies. With the rising number of sequenced genomes and the advent of CRISPR-Cas9 technologies, these genetic studies may be faster and less expensive than in the past. The identification of candidate symmetry genes is bottleneck in floral symmetry research.

Phylogenies of MYB-related genes have been published in two separate studies (Yanhui et al. 2006; Du et al. 2013), though both publications would be substantially improved through wider and more balanced taxonomic sampling as well as more reliable methods of phylogenetic inference. Yanhui *et al.* (Yanhui et al. 2006) included 64 MYB-related gene sequences from *Arabidopsis thaliana* and *Oryza sativa* subsp. *japonica.* With only two distantly related taxa

being sampled, statistical assumptions of the models used are likely to be violated (Yang 1994). Furthermore, the phylogeny generated in this study was unrooted, meaning that the authors did not use an outgroup to polarize their tree or test for monophyly of the included sequences. Although the authors categorized MYB-related genes by clade, without an outgroup the individual clades within a tree cannot be reliably interpreted as each individual sequence included in the analysis has the potential to be the least related to any of the remaining sequences. Thus, any potential clade could be broken up by rooting the tree on any of the sequences within it, meaning that no clades were rigorously tested for monophyly.

A more recent study by Du *et al.* (Du et al. 2013) included 16 angiosperm species compared to the 2 included in Yanhui *et al.* (Yanhui et al. 2006) and also included sequences from a moss (*Physcomitrella patens*), a clubmoss (*Selaginella*), and an alga (*Chlorella variabilis*). Although this taxon sampling is broader than that of Yanhui *et al.,* the inclusion of highly divergent outgroups without a large spread of ingroup taxa may have skewed the resulting phylogeny (Lyons-weiler, Hoelzer, and Tausch 1998; Qiu et al. 2001). Additionally, though the authors included hundreds of genes spanning the proposed MYB gene superfamily, it was impossible to reconstruct an adequate gene alignment using their published sequences and stated alignment methods (data not shown). Alignments were not included in the supplementary files and authors did not respond to email requests for data.

In phylogenetic analysis, especially when using one or a few genes, a well-made sequence alignment is vital to the reliability of the result. By inputting a sequence alignment into a phylogenetic analysis program, one is labeling each column of nucleotides or amino acids in that alignment as homologous characters. It is important to note that an alignment program will not 'fail' in the case of non-homologous characters, and therefore does not provide a test of orthology. In the case of extremely divergent gene sequences, due to either non-homologous genes or highly divergent taxa, alignment programs will generate unreliable alignments. Authors must often make quality assessments by eye or by estimating divergence among taxa with known evolutionary distances (as estimated from other sources of information). It is routine practice to trim parts of the alignment which lack sufficient sequence similarity including the beginnings and ends of genes as well as non-conserved areas of the gene if they are prone to insertions and/or deletions (Talavera and Castresana 2007). It is possible that the authors were unaware of these practices, or performed this alignment along with many others as part of a high-throughput analysis without such manual checks. With many dozens of genes, inaccurate homology becomes noise overcome by strong phylogenetic signal from more conserved regions. However, this is a single gene analysis in which there are not sufficient characters to overwhelm such background noise.

To improve the reliability of the MYB-related phylogeny and to increase its utility across flowering plants, we sampled genes from fifty-four published genomes, including a spread of angiosperms, gymnosperms, and the moss *Physcomitrella patens.* We sampled five Zingiberales species across the order using unpublished RNA-seq data, and an additional twelve Zingiberales species in a sequence capture described below.

Methods:

Sequence Retrieval and Alignment

Genes were recovered from fifty-four published genomes, the NCBI database, cleaned raw reads from five Zingiberales and one Commelinales (Angiosperms: Monocots) floral transcriptomes, and twelve Zingiberales sequence capture targets for a final total of 837 sequences from 71 taxa. Initial sequences were retrieved using *RADIALIS (RAD)* and *DIVARICATA (DIV)* sequences from *A. majus* as queries in an National Center for Biotechnology (NCBI) BLAST search (Altschup et al. 1990). Top hits were aligned to *RAD* and *DIV* in Geneious version 6.6.1 (Kearse et al. 2012) using MAFFT (Katoh et al. 2002) and MUSCLE (Edgar 2004) alignment algorithms, then edited manually to correct for inappropriately spaced insertions and deletions. The resulting alignment was used as a template to build a Hidden Markov Model (HMM). Genomes were downloaded from the relevant databases and searched using the generated nucleotide HMM in HMMER (Finn, Clements, and Eddy 2011). Resulting sequences were aligned as above. Because HMMER utilizes an alignment of genes from evolutionarily distant taxa, the resulting sequences include many false hits. In order to filter these hits, sequences that did not align either automatically or by eye were removed. A suitable alternative would be to make a BLAST database from the returned hits, and include genes within a given e-value cutoff after performing a BLAST search.

Zingiberales sequences were retrieved by using *Musa acuminata* genomic sequences as a query in a BLAST search against databases of cleaned raw reads from transcriptomes. These transcriptomes were generated for 5 of the 8 families of the Zingiberales. RNA libraries were made using the illumina® TruSeq® kit and protocol. Libraries were assembled using Trinity (Haas et al. 2013). Perhaps due to insufficient coverage and/or low expression, *DIV-like* and *RAD-like* sequences could not be recovered from assembled transcriptomes. Due to our failure to recover genes from the assemblies, we generated a pipeline to assemble the sequences from unassembled reads, with poor quality and highly repetitive sequences removed and adaptors trimmed. These cleaned raw reads were used to build nucleotide BLAST databases, then queried using commands optimized for short reads (blastn-short). Through aligning the resulting sequences, we determined that even those with the highest e-values were often viable hits. Knowing this, we incorporated all BLAST returns in downstream steps. BLAST results were clustered using cd-hit (Li and Godzik 2006) with a 95% identity clustering threshold, a word size of 8, and set to place each sequence in the best matching cluster rather than the first to meet the alignment threshold. The sequences in each cluster were aligned to *Musa acuminata* references in Geneious version 6.6.1 as above. The sequences that aligned successfully were put through two rounds of de novo assembly in Geneious, using low sensitivity/fastest default settings with exceptions of 10% maximum mismatches per read, 80% minimum overlap identity, and a maximum gap length of 1. Resulting contiguous sequences or "contigs" supported with at least 10X coverage were included in the main alignment. Support for these contigs was checked using BOWTIE (Langmead et al. 2009) to align original transcriptome raw reads to all contigs. BOWTIE, a less stringent alignment tool than the one outlined above, was able to yield higher coverage of the regions represented by all contigs. This process ensured that no contigs with 10X or greater coverage were mistakenly filtered out. *Costus spicatus* and *Musa basjoo* sequences

used as templates for RT-PCR were verified by Sanger sequencing at the Museum of Vertebrate Zoology facility at the University of California at Berkeley.

Additional Zingiberales sequences were recovered by sequence capture and subsequent Illumina sequencing following established lab protocols (Sass et al. 2016). Zingiberales *RAD-like* and *DIV-like* sequences generated from transcriptome data and separated at intron/exon boundaries as predicted by the *Musa acuminata* genome were included as baits to capture sequences from 199 Zingiberales samples. Sequence capture results from a selection of 12 taxa were made into nucleotide BLAST databases and original bait sequences used as queries to retrieve BLAST hits. Separately, BOWTIE was used to align sequence capture results to the baits most closely related to targets (ie *Musa* baits used to retrieve Musaceae sequences). Matching sequences from BLAST and BOWTIE were grouped, duplicate hits were removed, and resulting sequences were put through two rounds of de Novo assembly as above. Resulting contigs with at least 10X coverage were included.

All resulting sequences were aligned by codon using MAFFT and MUSCLE algorithms in Geneious version 6.6.1. Resulting alignments were manually edited in Mesquite (Maddison and Maddison 2014). The beginning and end of the alignments were trimmed, as well as intronic regions in order to improve the accuracy of the resulting trees (Talavera and Castresana 2007). Final nucleotide alignments were translated to amino acids in Mesquite for use in sequence analysis.

Phylogenetic Tree Building

The final nucleotide alignments were used to produce phylogenies in RAxML-HPC v.8 on XSEDE (Stamatakis 2014) with 100 bootstrap replicates in all cases but the full domain 1 phylogeny which was produced using MrBayes (Ronquist and Huelsenbeck 2003) on the CIPRES server (Miller, Pfeiffer, and Schwartz 2010) for 5 million generations, with a burn-in fraction of 0.25. All trees were produced under a GTR + G model as selected by model testing in PHYML (Guindon et al. 2010).

Given that the evolutionary history of *RAD-like* and *DIV-like* genes is unknown, there is no known related gene that can be used as an outgroup to directly test the hypothesis that the recovered genes are homologous to *RAD* and *DIV*. Instead, we made the assumption that genes previously annotated as *RAD-like* and *DIV-like,* including those functionally characterized in *Antirrhinum majus* (Almeida, Rocheta, and Galego 1997; Galego and Almeida 2002; Costa et al. 2005) and *Solanum lycopersicum* (MacHemer et al. 2011) as well as those characterized by expression in *Heptaconium miconioides* (Howarth and Donoghue 2009) and *Bournea* (Zhou et al. 2008) belong to monophyletic groups. Using this assumption, we built preliminary phylogenies (data not shown) in RAxML on CIPRES using trimmed alignments of the first domains of all sequences, the full trimmed alignment including single domain and double domain genes, and a trimmed alignment including only double domain genes. Using the results of these trees, we found groups of sequences that, when used to root each tree, produced monophyletic groups for each gene. These sequences were used as outgroups to root the full taxon phylogenies here. For the *DIV-like* and *RAD-like* phylogenies, gymnosperm sequences

within the relevant clade in the full trees were used as outgroups to root angiosperm sequences. For the monocot focused phylogenies, Brassicales sequences were chosen as outgroups.

Results:

Phylogenetic Trees

Using our complete dataset of 71 taxa across gymnosperms, non-core eudicots, asterids, rosids, and monocots, we built a phylogeny including a total of 837 *DIV-like* and *RAD-like* gene sequences to visualize the broad relationships among these genes (Figure 1). The *RAD-like* + *DIV-like* clade has Bayesian probability of about 70%, the *RAD-like* clade 99%, and the *DIV-like* clade 58%. Within the *DIV-like* clade, the clade excluding gymnosperms and Amborella has 94% support. Sequences falling into the *RAD-like* and *DIV-like* clades match both the BLAST-based annotations of many gene sequences as well as previous results (Du et al. 2013). This phylogeny supports the hypothesis that *RAD-like* and *DIV-like* genes are closely related, as well as the homology of the first MYB repeat from *DIV-like* genes with the MYB domain of *RAD-like.*

Using the relationships reconstructed in our full phylogeny, we trimmed our alignment to generate a phylogeny useful in studying only the relationships between *DIV-like* genes (Figure 2). Characterized genes from *Antirrhinum majus, Bournea leiophylla,* and *Heptaconium miconioides* are highlighted in green. The monocot and eudicot clades both have reasonable to weak bootstrap support (84 and 64 respectively). Support within each clade is weaker, with some exceptions. Previously found relationships among *Bournea, Antirrhinum,* and *Solanum DIV* and *DIV-like* genes are supported, though *Arabidopsis thaliana* genes previously described as belonging to two separate eudicot copies are intermixed in our phylogeny (Howarth and Donoghue 2009; Zhou et al. 2008). Although we recovered several sequences from most taxa, without high support few conclusions can be drawn as to the evolutionary history of duplications on the scale of flowering plant evolution. The presence of several "DIV-like" annotations in sequences recovered in the *DIV* clade demonstrates the concordance between BLAST-based annotations and phylogenetic methods, though fine tuned differentiation among the many paralogous copies requires well sampled family or order level phylogenies.

To focus on *RAD-like* gene evolution, we trimmed our full alignment to include only sequences reconstructed as *RAD-like* in the full phylogeny. We used this trimmed alignment to generate a phylogeny (Figure 3). Similar to the full phylogeny (Figure 1), bootstrap support for clades of *RAD-like* genes is weak. This is understandable given the very few informative characters for this gene, which is also visible on the scale bar, showing that branch length is based on relatively few predicted substitutions per site in this gene group. Although the divergence events close to the backbone of the tree have poor support, reconstructed evolutionary relationships nearer the tips fare better. There may have been two eudicot-specific duplications in the history of *RAD-like* genes, with the second duplication leading to the lineage containing *A. majus RAD* and annotated *RAD-like* genes. The *Arabidopsis RAD-like* sequences recovered do not show previously recovered relationships, though again *Bournea, Antirrhinum,* and *Solanum RAD* and *RAD-like* genes form a clade as previously described (Zhou et al. 2008; Du et al. 2013; Boyden, Donoghue, and Howarth 2014).

We further explored the evolution of *DIV-like* and *RAD-like* genes using taxonomically focused phylogenetic analyses. Using the Brassicales as outgroups to root the tree, we generated a phylogeny for all recovered monocot *DIV-like* genes (Figure 4). Relationships reconstructed in this phylogeny show one monocot *DIV-like* copy, with a Poales specific duplication. Deep relationships among the Arecales and commelinids (Zingiberales and Commelinales) are not well supported, but show two moderately supported Zingiberales groups: DIV 1, with 63 bootstrap support and DIV 2 with 58. DIV 3 is a paraphyletic grade, containing all other non-grass *DIV-like* sequences. Within each Zingiberales DIV group, *Musa* sequences fall into two separate groups. *Calathea* sequences also fall out into separate clades within DIV3 and DIV1. This may indicate family-specific duplications, but is more likely a failure to recover all copies of these genes from Zingiberales transcriptomes and sequence capture. Within the Arecales and Commelinales there may be lineage-specific duplications, though more taxon sampling is necessary to elucidate copy number and history of duplication. This phylogeny supports a history of lineage-specific duplication rather than ancient duplicates conserved across monocots.

As in the whole taxon *RAD-like* phylogeny (Figure 3), core relationships among monocot *RAD-like* genes are poorly supported (Figure 5). There may have been one *RAD-like* duplication before the diversification of monocots (boxed in blue, low bootstrap support, 13) followed by lineage-specific duplications. Zingiberales *RAD-like* sequences are labeled RAD1, RAD2, and RAD3 groups, with group specific duplicates labeled in terms of genes amplified by RT-PCR (Chapter 2), highlighted in blue. Within RAD2, there appears to be at least one duplication, supported by the presence of *Heliconia* and *Musa* species each in three distinct clades. Poales appears to have undergone a duplication prior to diversification, resulting in RAD2a and RAD2b, though *Oryza* sequences were all recovered in RAD2b.

Sequence Analysis

The MYB-like domains of *A. majus RAD* and *DIV* contain an altered pattern relative to typical MYB domains. Instead of the classic three evenly spaced tryptophans (W), the third tryptophan is replaced with a tyrosine (Y) (Table 1, arrows Figures 6 and 7). Motifs found in MEME (Bailey and Elkan 1994) match those found previously for *I-box-like* and *R-R-like* genes (Table 1) (Du et al. 2013). The first MYB domain matches that found for *I-box-like* and the second matches that found in *RR-like* genes including the '*SHAQK(Y/F)*' motif additionally found in *CCA1/LHY-like* genes (Figure 8) (Schaffer et al. 1998; Zhi-yong Wang and Tobin 1998; Yanhui et al. 2006; Du et al. 2013). The presence of the '*SHAQKY'* motif in the second domain of *DIV-like* sequences confirms the relationship of these genes with *DIV*, as this motif has previously been described as a hallmark of *DIV-like* genes (Rose, Meier, and Wienand 1999; Galego and Almeida 2002). The *'SHAQK(Y/F)'* motif in *CCA1/LHY-like* genes is thought to be important for their roles as transcriptional regulators involved in maintaining circadian rhythms, and may play a similar role in the transcriptional activity of *DIV-like* genes (Schaffer et al. 1998; Zhi-yong Wang and Tobin 1998). The strong e-values for the motifs found in MEME despite the broad taxonomic range and the presence of multiple genes speaks to the strong conservation of the MYB-like domain throughout evolutionary time.

Discussion:

The lack of support for the full taxon trees is to be expected, given the broad taxonomic range of species represented as well as the relatively small number of informative characters in the genes. Nevertheless, these phylogenies are a tool researchers can use to identify *RAD-like* and *DIV-like* genes of interest. Researchers can sub-sample the alignments for relevant taxa, with additions as more genomes and transcriptomes are published, and generate their own trees to get finer-scale information in their groups of interest. These phylogenies will contribute to the ongoing effort to characterize the MYB-related gene group. Within monocots, we have now identified several Zingiberales sequences that are likely candidates for future study in floral development.

As a result of this detailed study, we propose that MYB-related proteins should be characterized by their evolutionary relationships with one another rather than their number of MYB repeats. Typically, MYB and MYB-related proteins have been characterized by their number of MYB repeats, and the relationship of these repeats with the three MYB repeats of the human c-MYB protein (Feller et al. 2011; Dubos 2013). Based on the number of MYB repeats, *RAD-like* and other single MYB repeat genes have been annotated R1 or R-MYB proteins and *DIV-like* as R2, R-R, or 2R-MYB. The largest group of MYB genes, the R2R3 genes are so named due to their similarity with the second and third domains of c-MYB. Likely because of the two MYB repeats, *DIV* and *DIV-like* genes have been grouped with R2R3 genes and annotated as such in the past (Galego and Almeida 2002; Du et al. 2009; Feller et al. 2011; Dubos 2013). However, positional comparisons to the c-MYB domain are invalid for *RAD* and *DIV,* as the common MYB domain is not alignable to any of the three MYB repeats in the human c-MYB gene. Furthermore, R2R3 genes in plants are thought to have evolved from common ancestors of animal R1R2R3 genes followed by the loss of the R1 repeat (Du et al. 2009), making the annotation of *RAD-like* genes 'R1' particularly misleading. Alternatively, *RAD-like* and *DIV-like* genes have been annotated by similarity to other MYB-related genes determined by clustering algorithms. These algorithms have placed *RAD-like* with *I-box* related genes, and *DIV-like* with *CCA1/R-R-like* genes. When the domains of *R-R-like* genes were separately analyzed, the first domains clustered with *I-box-like* genes, supported by our finding that these domains are related. For the sake of consistency and incorporation of known relationships, we propose that the genes included in our study be referred to as *RAD-like* and *DIV-like* genes determined by their clade in our full phylogeny (Figure 1). If MYB domain annotation is necessary, the domain of *RAD-like* and the first domain of *DIV-like* should be referred to as *I-box* like and the second domain of *DIV-like* as *CCA1-like* in accordance with previously published results (Du et al. 2013).

The fact that many sequences consistently fell outside of the core *DIV-like* and *RAD-like* clades (Figure 1) shows that even alignable sequences retrieved with BLAST and HMMER may have had separate evolutionary histories. It has been hypothesized that single MYB domain proteins like *RAD* evolved from multiple MYB domain proteins by the loss of a domain (Stevenson et al. 2006). Single domain MYB-related proteins have been shown to block transcriptional roles of double domain proteins in *Arabidopsis* trichome development (Wester et al. 2009). This could be a common evolutionary mechanism to establish transcriptional switches, with the truncated proteins blocking the activity of non-truncated paralogs. Alternatively, some have argued that tandem repeats followed by whole gene duplications is the more parsimonious explanation for R2R3 MYB genes in plants and R1R2R3 in animals (Du et al. 2009). As MYB and MYB-related

proteins contain staggering numbers and diversity of members in plants, it is likely that a combination of these evolutionary trajectories led to the diversity of proteins we see today. In the case of *RAD-like*, *DIV-like,* and *CCA1/LHY-like* genes, it seems likely that a double MYB repeat ancestor gave rise to three lineages – one of which maintained only the first MYB repeat, a second of which maintained both, and the third maintaining only the second MYB repeat. The lack of sequence similarity between these MYB-related genes and others, such as the *DRIF-like* genes and the R2R3 genes, may indicate a more ancient inheritance of one or multiple MYB repeats which have since diverged and diversified, gaining and losing MYB repeats over the course of plant evolution.

Future Directions:

Extensive taxonomic sampling between actinomorphic and zygomorphic sister lineages, especially between the Zingiberales and the Commelinales, could further identify likely candidates for recruitment to zygomorphic floral development. Additionally, studies of synonymous and nonsynonymous mutations across the sister lineages would provide insight into any common patterns of MYB-related gene recruitment and evolution in zygomorphic lineages.

To further understand broader scale evolutionary patterns, additional protein characters such as hydrophobicity and additional sequence characters such as number and location of indels and introns, and sequence length, could be coded to provide additional signal to large-scale phylogenies. Phylogenetic coalescence based algorithms like ASTRAL (Mirarab et al. 2014) may be useful in the future to combine lineage-specific MYB-related gene trees in producing a broader understanding of relationships among these genes.

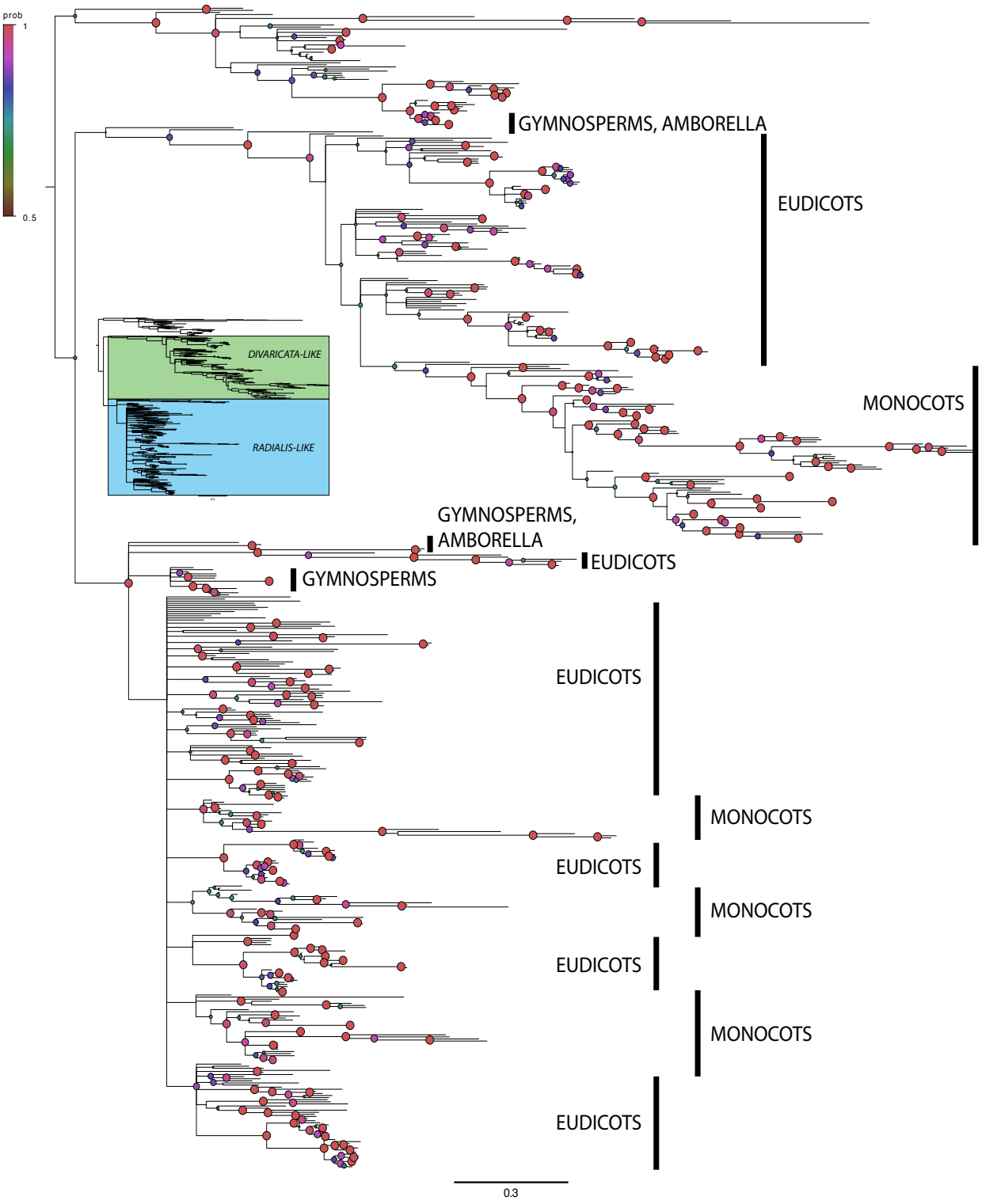


Figure 1: Consensus Bayesian phylogeny of *RAD-like* and the first domain of *DIV-like* genes. Clades supported by less than 50% probability are collapsed, all other Bayesian probability scores shown as symbols on nodes sized proportional to % probability, colored as shown in legend (upper left). Inset shows the broad relationship between the two genes. Scale indicates predicted substitutions per site. Taxon names removed for legibility.

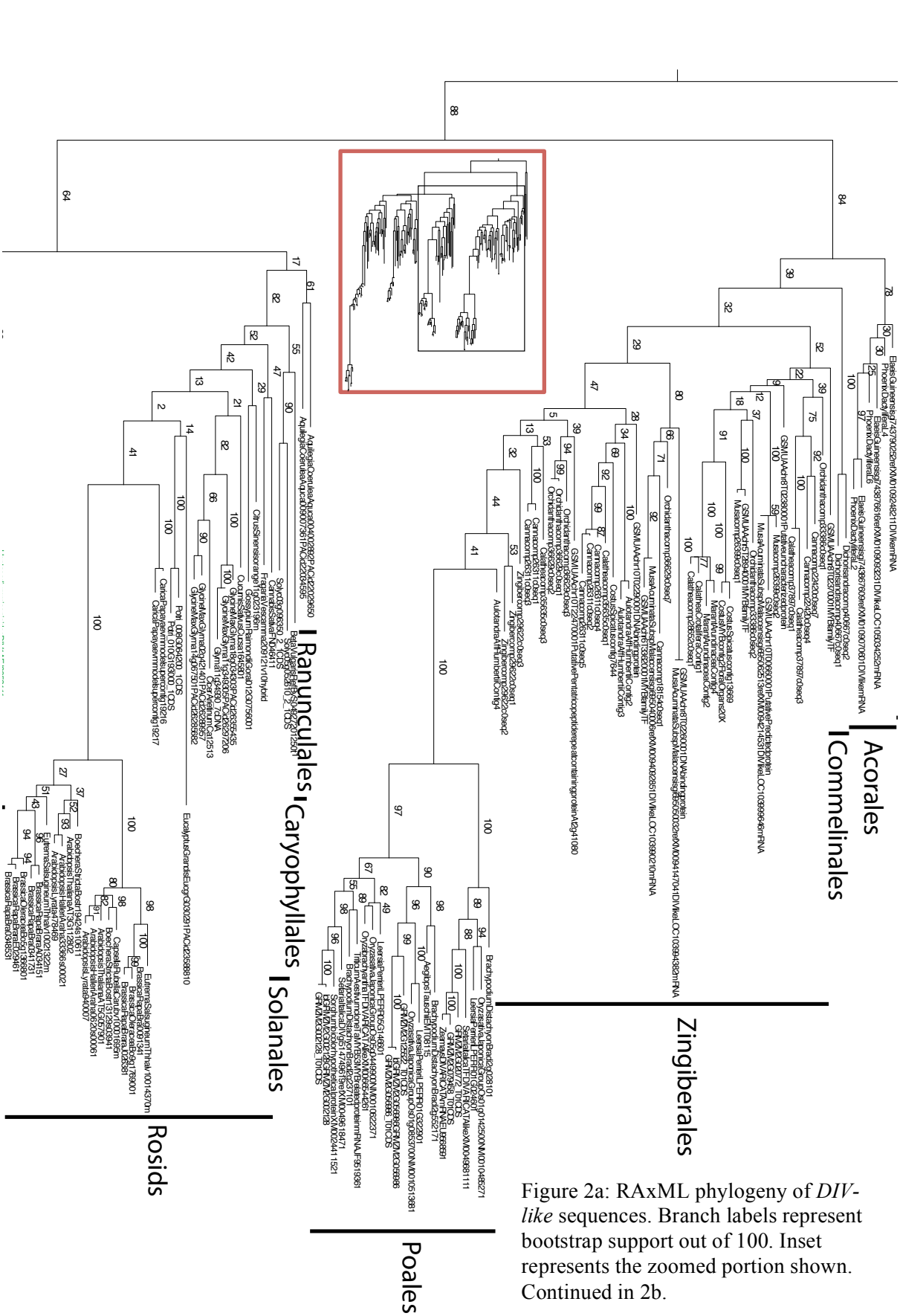


Figure 2a: RAxML phylogeny of *DIV-like* sequences. Branch labels represent bootstrap support out of 100. Inset represents the zoomed portion shown. Continued in 2b.

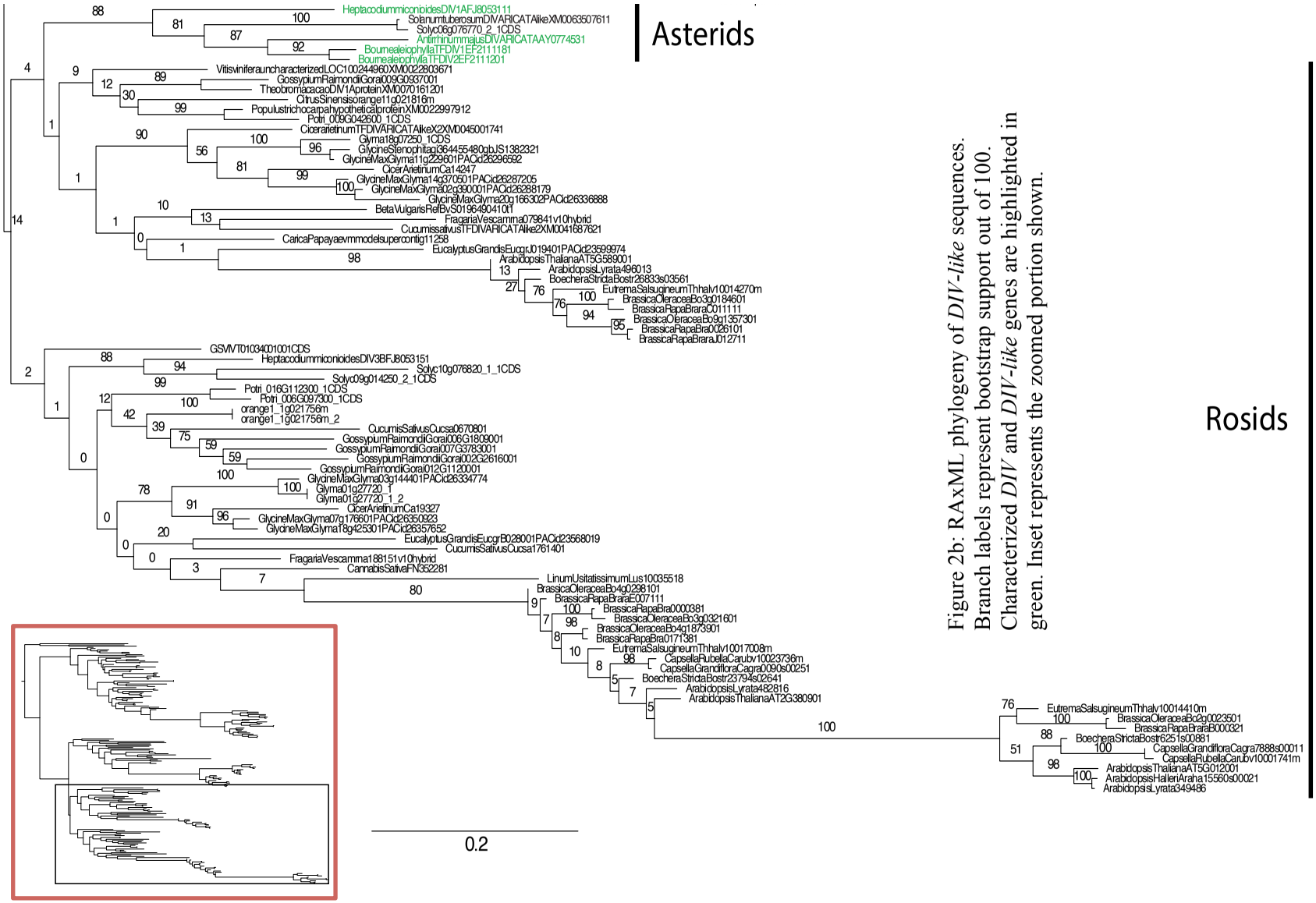


Figure 2b: RAxML phylogeny of *DIV-like* sequences. Branch labels represent bootstrap support out of 100. Characterized *DIV* and *DIV-like* genes are highlighted in green. Inset represents the zoomed portion shown.

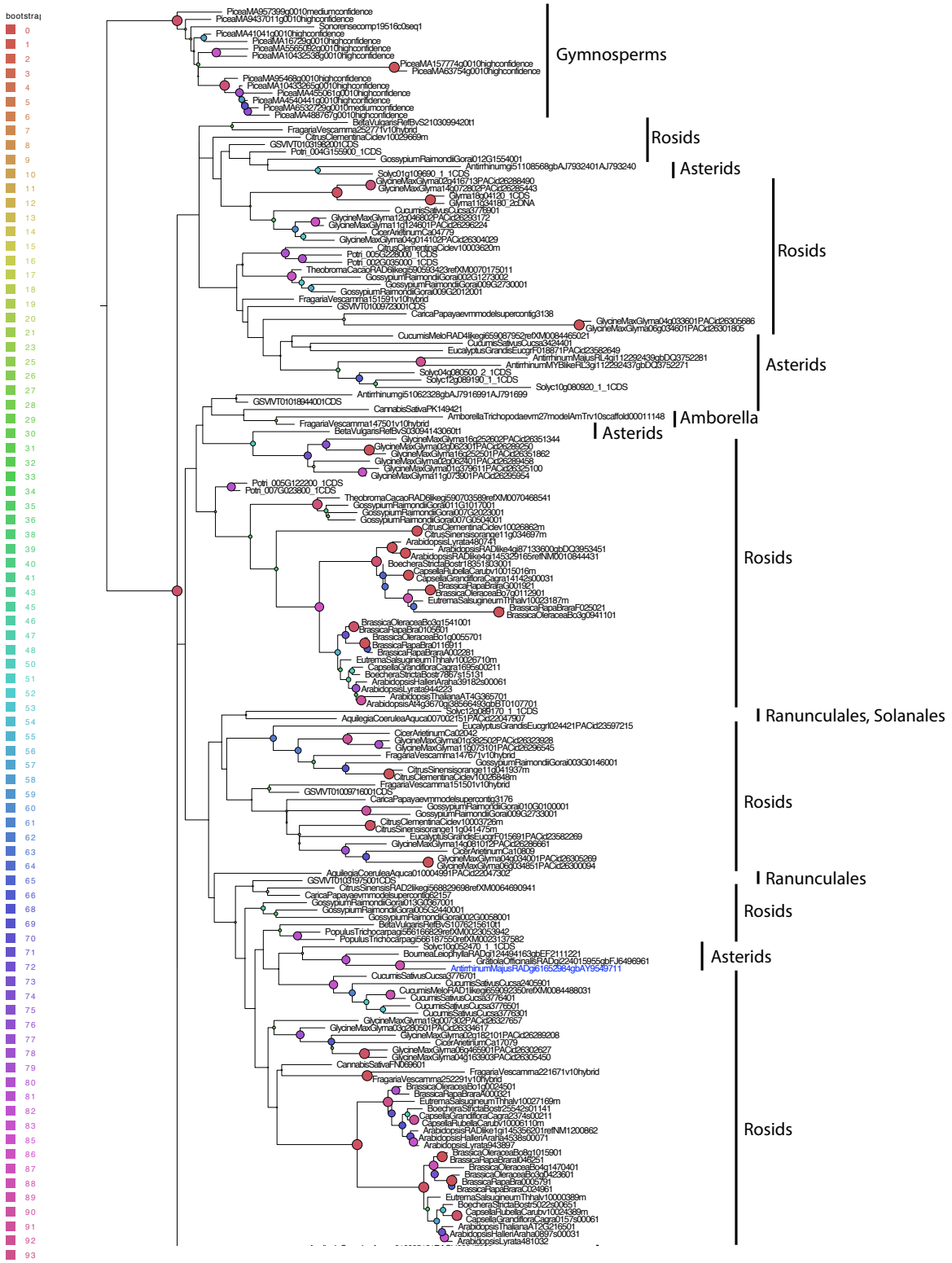


Figure 3a: RAxML phylogeny of *RAD-like* genes. Node symbols represent bootstrap support out of 100, colored according to the legend on the left, sized proportional to % out of 100. Characterized *RAD* and *RAD-like* genes are highlighted in blue. Continued in 3b.

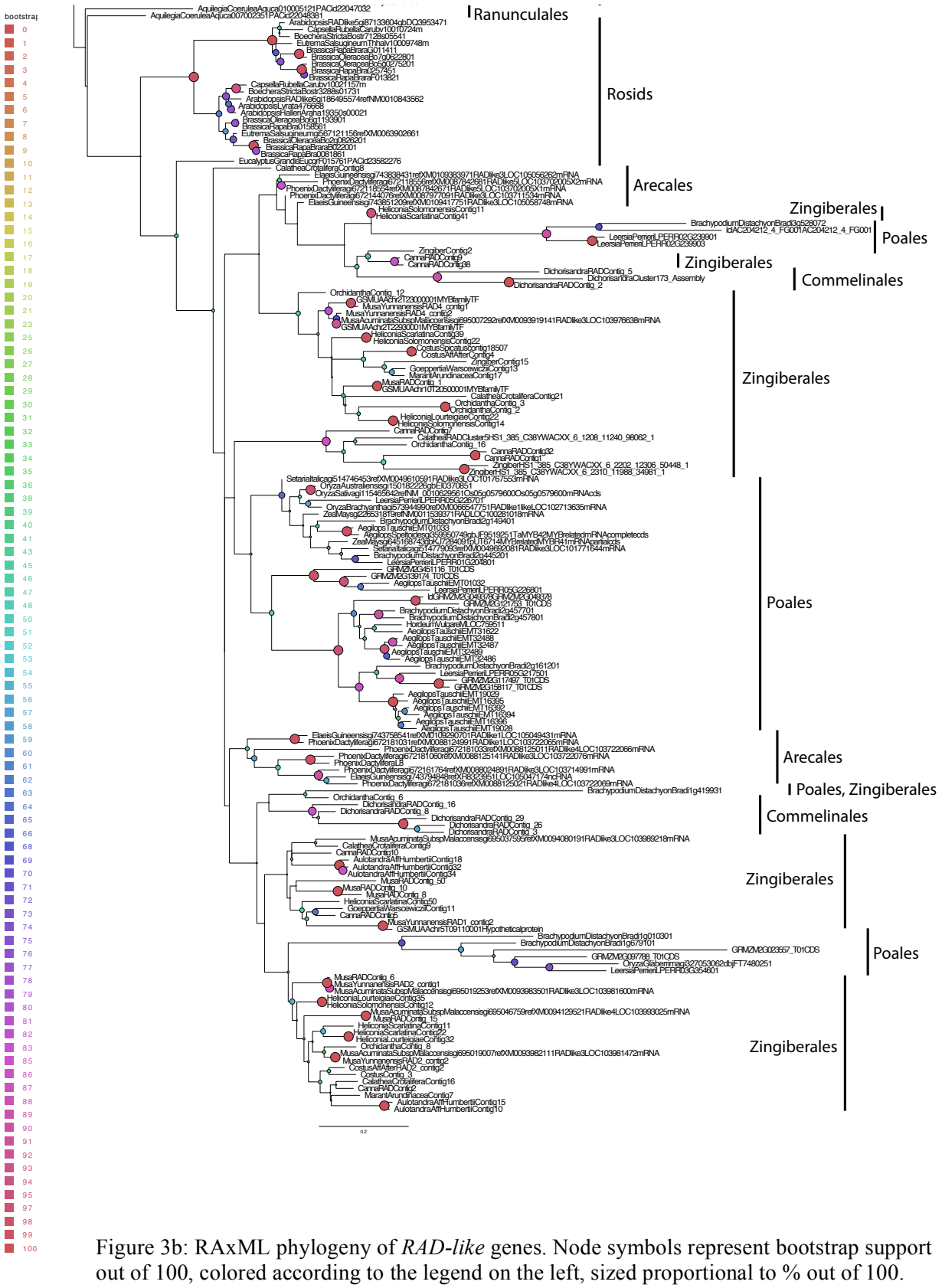


Figure 3b: RAxML phylogeny of *RAD-like* genes. Node symbols represent bootstrap support out of 100, colored according to the legend on the left, sized proportional to % out of 100.

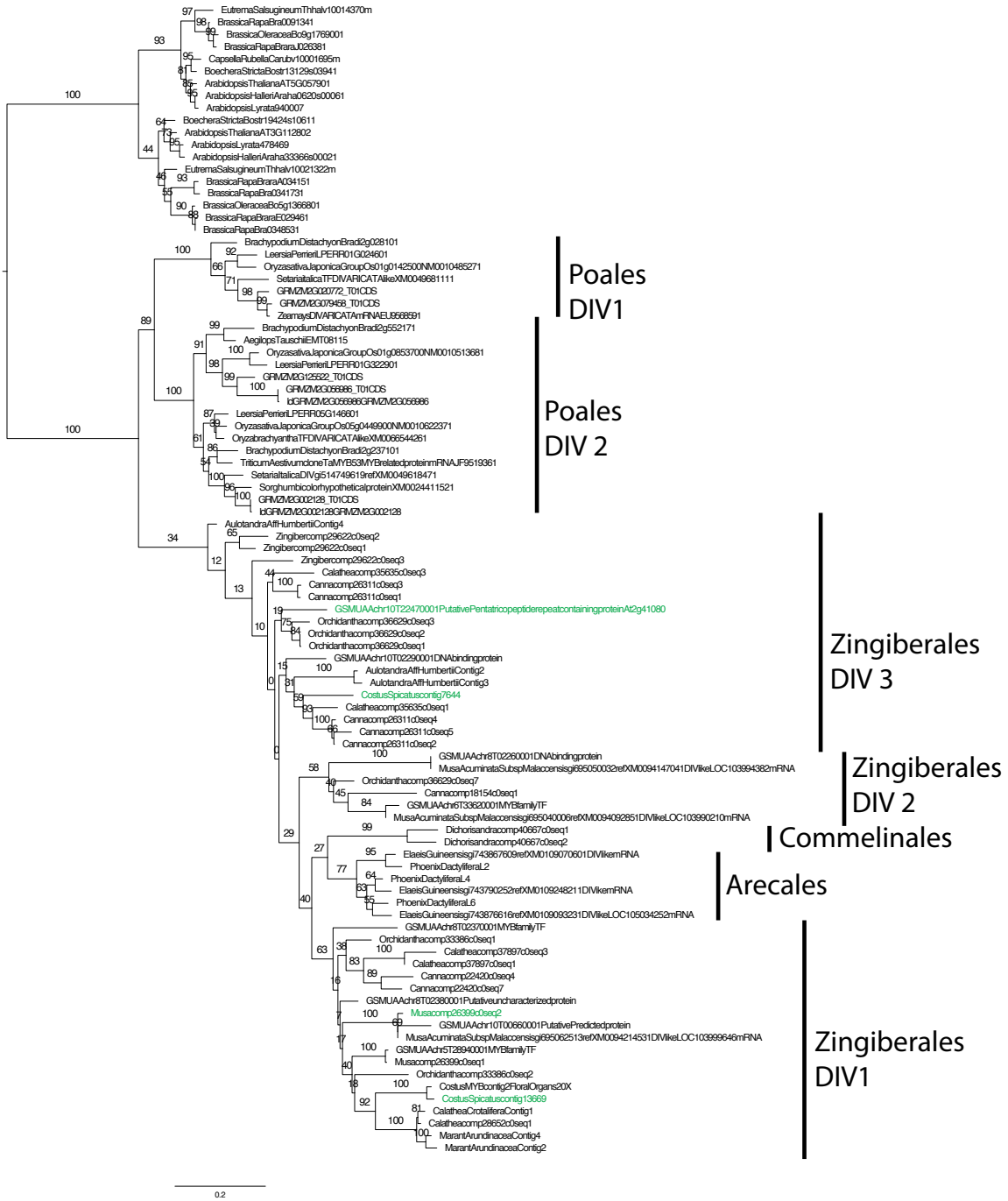


Figure 4: RAxML phylogeny of monocot *DIV-like* genes, with Brassicales sequences used as an outgroup. Branch labels represent bootstrap support out of 100. *DIV-like* genes amplified by RT-PCR are highlighted in green. Scale bar indicates predicted substitutions per site.

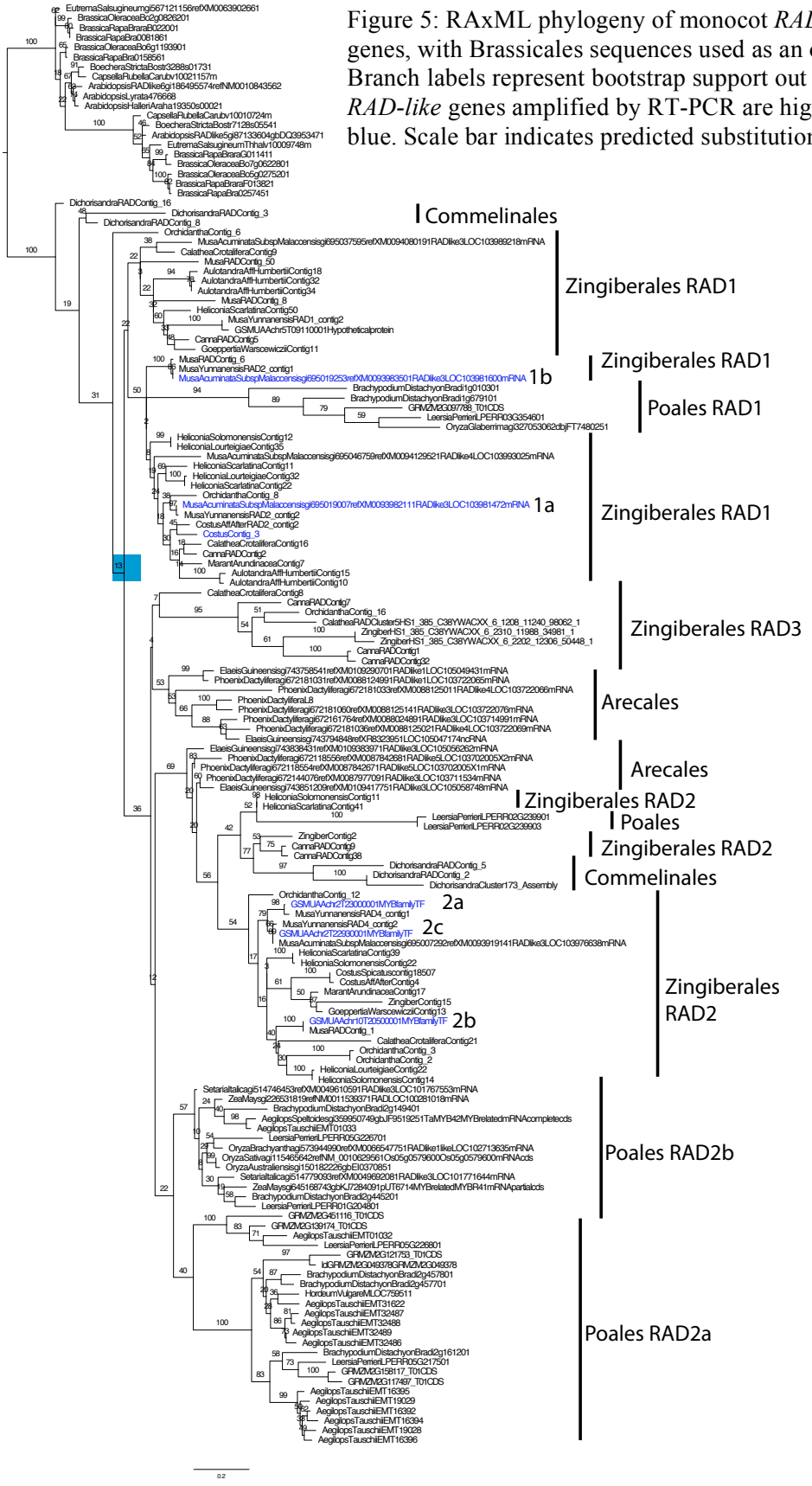
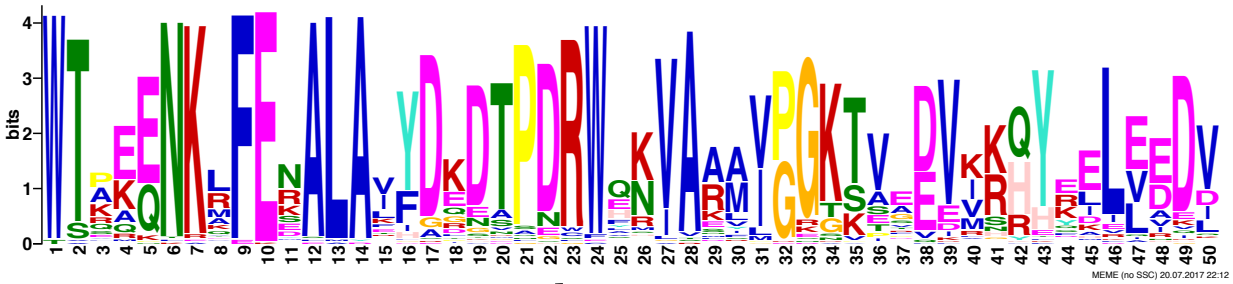
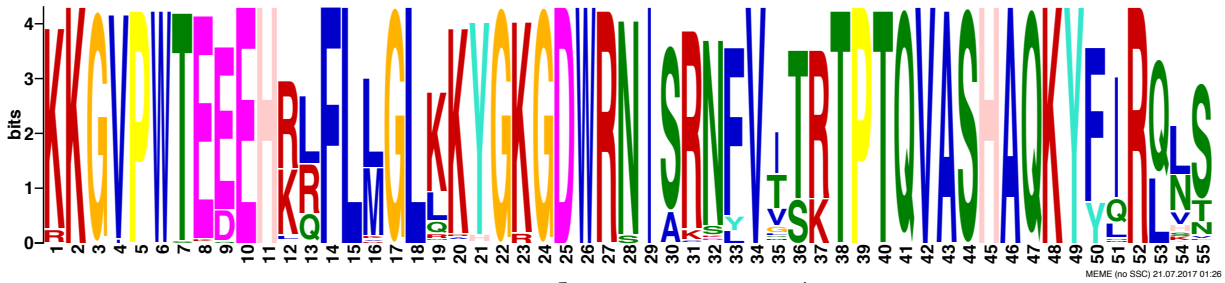


Figure 5: RAxML phylogeny of monocot *RAD-like* genes, with Brassicales sequences used as an outgroup. Branch labels represent bootstrap support out of 100. *RAD-like* genes amplified by RT-PCR are highlighted in blue. Scale bar indicates predicted substitutions per site.

A: e-value 7.5e-2457



B: e-value 1.4e-8107



C: e-value 3.9e-050

Table 1: MEME output from http://meme-suite.org/ Shown are motifs discovered in the first domain of amino acid sequences from across the *RAD-like* and *DIV-like* phylogenies (A), and in both domains of *DIV-like* genes (B,C) with the probability that a random sequence would contain as close a match to the motif as the sequences queried. Positions 1-43 in A, 6-49 in B, and 1-43 in C contain the MYB-like motifs. In B positions 43-49 contain the SHAQKY motif.

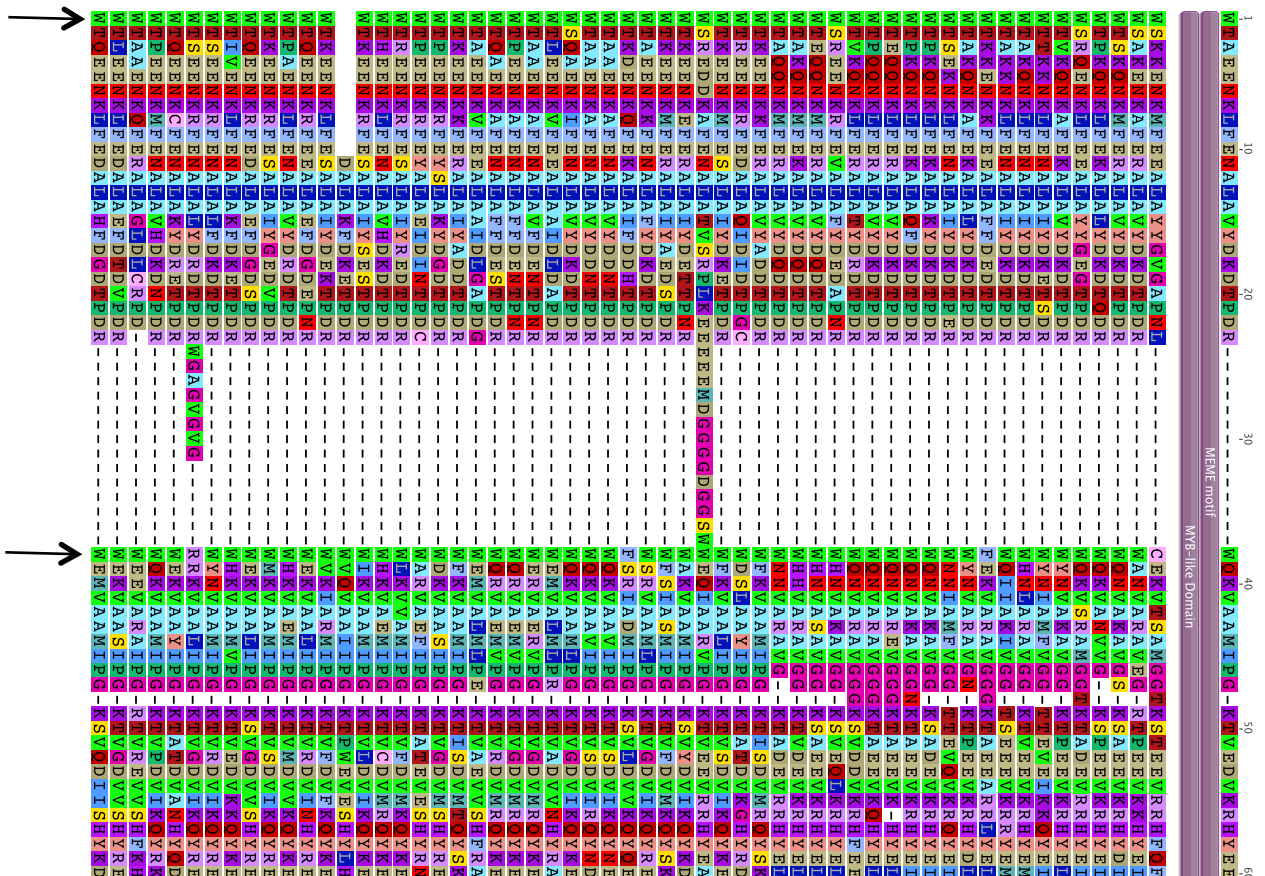


Figure 6: Amino acid alignment of representative sequences of *RAD-like* and the first domain of *DIV-like,* annotated with motif found in MEME (Bailey and Elkan 1994) and the MYB-related domain previously described (Klempnauer, Gonda, and Michael Bishop 1982). Taxa included are a representative subset of *RAD-like* (1-33) and *DIV-like* (34-57) genes. Arrows show the two regularly spaced tryptophans (W) and the third, which is replaced in most taxa with a tyrosine (Y).

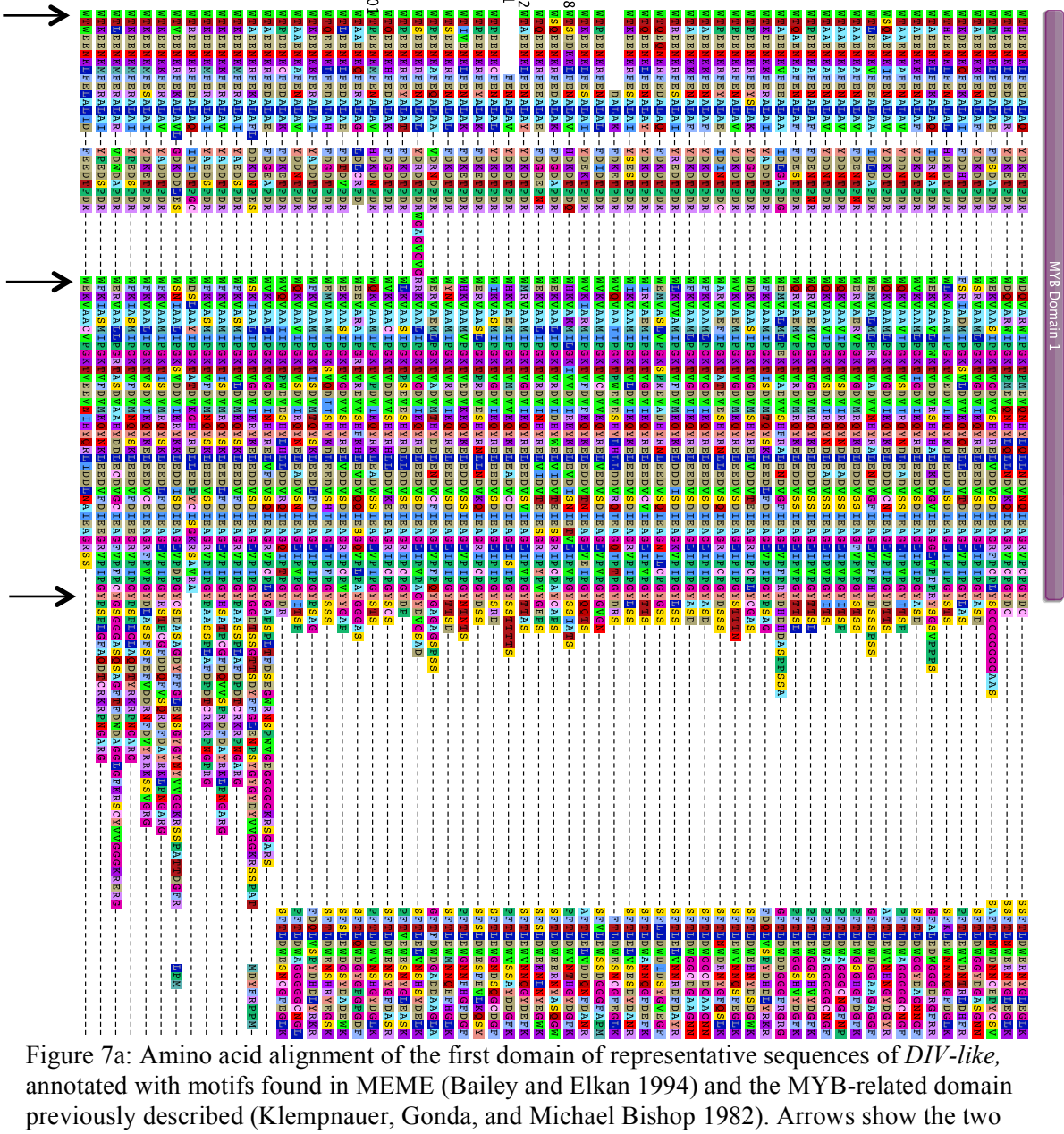


Figure 7a: Amino acid alignment of the first domain of representative sequences of *DIV-like,* annotated with motifs found in MEME (Bailey and Elkan 1994) and the MYB-related domain previously described (Klempnauer, Gonda, and Michael Bishop 1982). Arrows show the two regularly spaced tryptophans (W) and the third, which is replaced in most taxa with a tyrosine (Y). Second domain on the following page.

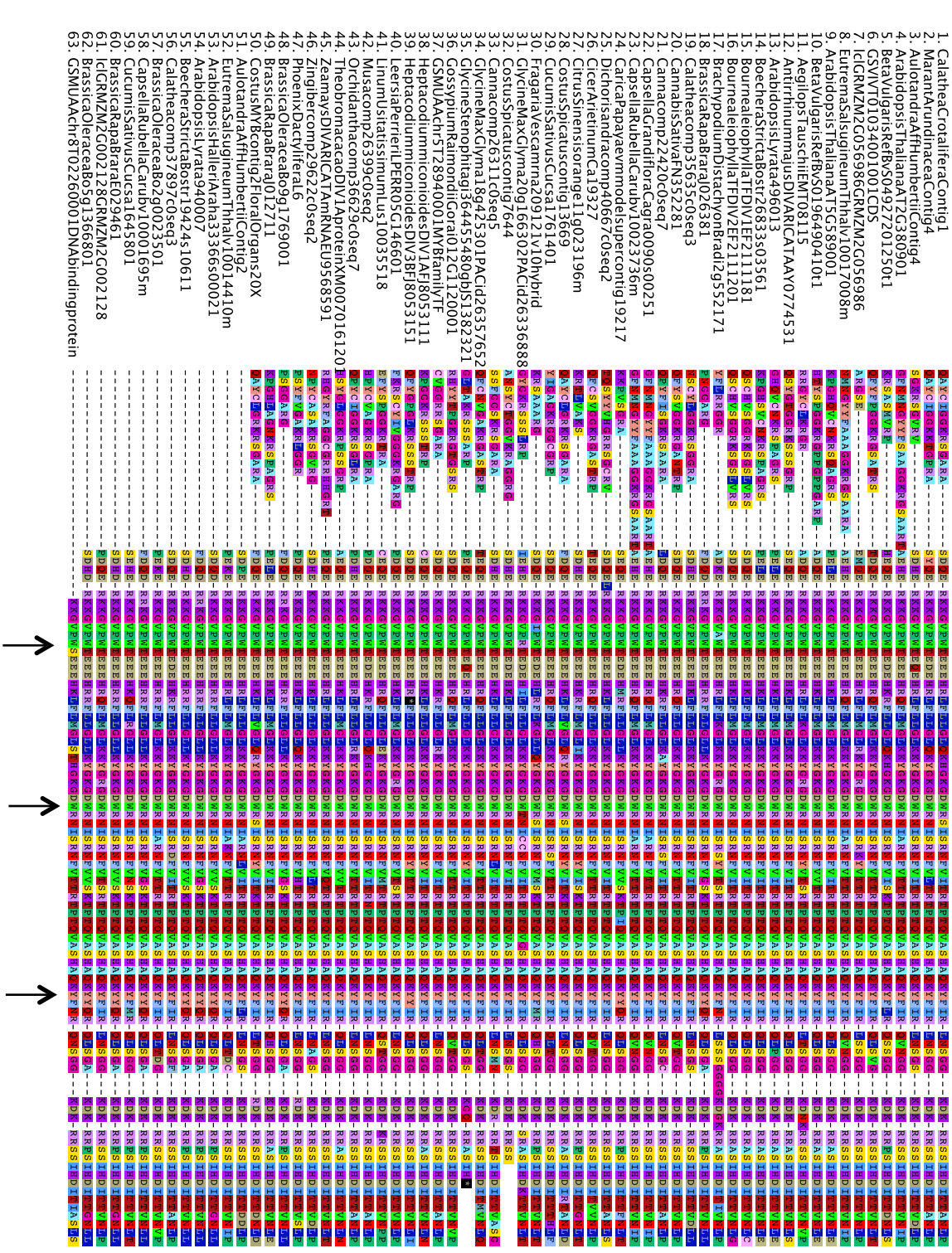


Figure 7b: Amino acid alignment of the second domain of representative sequences of *DIV-like,* annotated with motifs found in MEME (Bailey and Elkan 1994) and the MYB-related domain previously described (Klempnauer, Gonda, and Michael Bishop 1982). Arrows show the two regularly spaced tryptophans (W) and the third, which is replaced in most taxa with a tyrosine (Y).

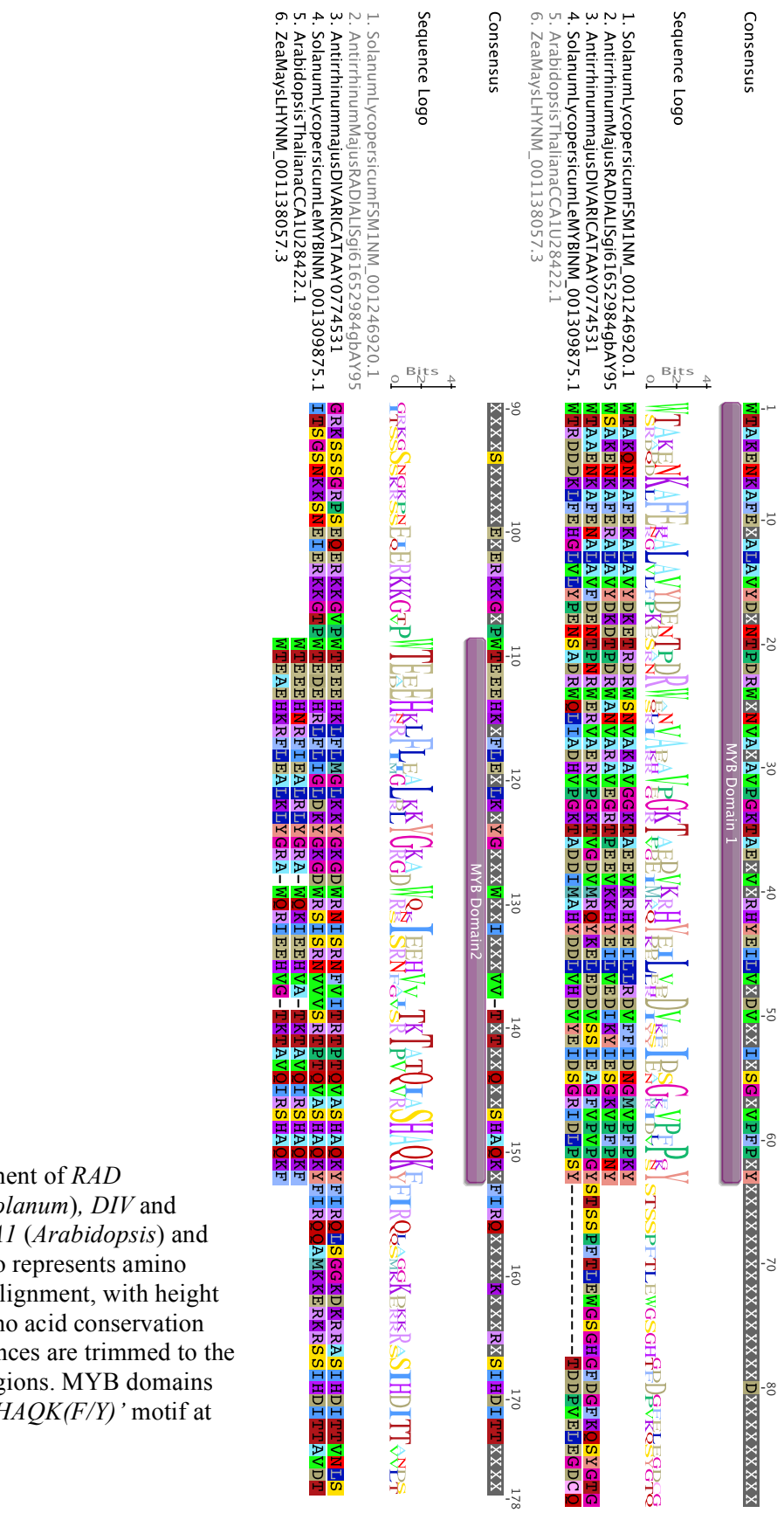


Figure 8: Amino acid alignment of *RAD* (*Antirrhinum*) and *FSM1* (*Solanum*), *DIV* and *LeMYB* (*Solanum*), and *CCA1* (*Arabidopsis*) and *LHY* (Maize). Sequence logo represents amino acids at that site across the alignment, with height proportional to average amino acid conservation across the alignment. Sequences are trimmed to the MYB motifs or alignable regions. MYB domains are annotated. Notice the '*SHAQK(F/Y)'* motif at the end of MYB domain 2.

Chapter 2: Gene expression patterns of *RADIALIS* and *DIVARICATA*

Abstract:

Functional analysis and gene expression studies have indicated a role for *RADIALIS (RAD)* and *DIVARICATA (DIV)* in zygomorphy in *Antirrhinum majus*, Gesneriaceae, *Lonicera x bella,* and *Heptaconium miconioides* (Galego and Almeida 2002; Costa et al. 2005; Zhou et al. 2008; Howarth and Donoghue 2009; Boyden, Donoghue, and Howarth 2014). Here we present the expression patterns of multiple copies of *DIV-like* and *RAD-like* in *Costus spicatus* and *Musa basjoo*, two representative species of the order Zingiberales and compare them with *teosinte branched 1-like (TBL)* gene expression. We find that patterns of expression do not match those described in *Antirrhinum majus*, though expression patterns are consistent with roles in zygomorphy. In *Musa basjoo*, *TBL2, RAD2a,* and *RAD2c* are expressed asymmetrically, with *TBL2* expression concentrated in the ventral region of developing flowers and *RAD2a* and *RAD2c* expression throughout the flower with the exception of the dorsal free petal. In *Costus spicatus, TBL1a* and *1b* are expressed in the ventral portions of the flower, and *RAD1* in the central portion of the flower excluding the dorsal sepals and ventral petals. The divergent expression patterns of *TBL1* and *TBL2* as well as *RAD1* and *RAD2* between the species are consistent with a change in regulation and/or function of *TBL* and *RAD-like* genes between the banana and ginger families. *Musa basjoo TBL2, RAD2a,* and *RAD2c* are good candidates for future functional studies.

Introduction:

The canonical floral symmetry network, first described for the snapdragon (*Antirrhinum majus*), is made up of four transcription factors (reviewed in Citerne et al. 2010). *CYCLOIDEA (CYC)* and *DICHOTOMA (DICH),* two TCP transcription factors, form the core of the network. The remaining genes in the network are MYB-related transcription factors and include *DIVARICATA (DIV), RADIALIS (RAD),* and *DIVARICATA and RADIALIS INTERACTING FACTORS (DRIFs).* In *A. majus,* CYC and DICH proteins together promote the expression of *RAD* in the dorsal petals, resulting in the spread of RAD protein throughout the dorsal and lateral petals, while *DIV* is expressed throughout the developing flower (Figure 1) (Costa et al. 2005). RAD protein interferes with the downstream functions of DIV by sequestering DIV's necessary binding partners, the DRIFs, in the cytoplasm. In the absence of RAD, the DRIFs travel to the nucleus where they are bound by DIV, together acting as a transcriptional regulator promoting ventral petal identity (Sobral 2010; Raimundo et al. 2013). While asymmetric *CYC-like* gene expression has been documented across angiosperms in zygomophic lineages (Hileman 2014a; Hileman 2014b), prior to this year expression patterns in MYB-related genes had only been published in the Asterids (the group containing *A. majus*), with the most divergent lineage being the Dipsacales (Howarth and Donoghue 2009; Boyden, Donoghue, and Howarth 2014). The authors found that in the Dipsacales, zygomorphic lineages had multiple copies of *DIV-like* and *RAD-like,* with some showing dorsiventral expression patterning similar to the *A. majus* model and some showing universal expression. Recently, a study in orchids recovered expression patterns for *DIV-like* and *RAD-like* genes consistent with roles in zygormophy, with *RAD-like* gene expression in the lip and outer tepals of *Orchis italica* and universal *DIV-like* expression

(Valoroso et al. 2017). The authors recovered 8 copies of *DIV-like* and 4 copies of *RAD-like* genes in *O. italica.*

Developmental timing of gene expression plays a critical role in floral symmetry. In *A. majus, DIV* expression, though universal in early stages, is later confined to the corolla and is strongest in ventral petals (Galego and Almeida 2002). In *Bournea leiophylla,* within the Asterid order Lamiales, flowers initiate their development with zygomorphic symmetry due to differing rates of development of petal primordia across the floral meristem, but become nearly actinomorphic at anthesis due to 'ventralization' – or ventral-like petal identity - of all petals (Zhou et al. 2008). The authors found that *BlCYC1, BlRAD,* and *BlDIV* had similar expression patterns to those in *A. majus* until later development when *BlCYC1* and *BlRAD* were downregulated, which the authors hypothesize allows BlDIV protein to function in promoting ventral petal identity throughout the flower.

Although gene expression is not a perfect proxy for protein expression, it is an important first step in establishing hypothetical models of protein network function. Post transcriptional modification, divergent rates of decay or destruction, and differing levels of transcription and translation all contribute to mismatches between gene expression and protein expression. However, gene and protein expression are usually strongly correlated, and gene expression assays remain a cost and time effective proxy for protein expression (Fu et al. 2007).

The Zingiberales is an order of monocotyledonous flowering plants, with flowers that are predominantly zygomorphic with a few species-level exceptions. Species in the 'ginger group' i.e. families Cannaceae, Zingiberaceae, Marantaceae, and Costaceae, have floral displays generated mainly from infertile stamens (staminodes) that are laminar and 'petaloid' in structure. Species in the banana families, including Musaceae, Lowiaceae, Strelitziaceae, and Heliconiaceae, have more traditional floral displays built mainly from petals and sepals (B. K. Kirchoff 2013). Based on previous studies, the monocot ortholog to *CYC, teosinte branched 1-like (TBL)* is expressed in the Zingiberales in a pattern consistent with a role in zygomorphy. Three copies of *TBL* genes were found in the order (Bartlett and Specht 2011), and expression was characterized for two of these copies based on *in situ* hybridization in a ginger group species *Costus spicatus* (Costaceae) and a banana group species *Heliconia stricta* (Heliconiaceae) (Figure 2). In *Costus spicatus*, *TBL1a* is expressed in ventral and lateral sepals and petals, in the anthers of the fertile dorsal stamen, and in the staminodial labellum, while *TBL2* is expressed in the anthers of the fertile dorsal stamen. In *Heliconia stricta*, *TBL1a* is expressed in the aborted dorsal stamen, while *TBL2* is expressed in the fused ventral sepals. These data are consistent with symmetry network conservation across angiosperms and *TBL* gene expression is correlated with shifts in symmetry within the Zingiberales.

By mapping the expression patterns of the *DIV-like* and *RAD-like* Zingiberales orthologs, we can test the hypothesis that the network of genes implicated in core eudicot floral symmetry is likewise operating in producing Zygomorphy in monocot flowers. Floral developmental mechanisms often share core genes but have divergent relationships and functions between eudicots and monocots. A key example is the ABCE model for floral development. In the Zingiberales, the B-class MADS-box genes "*GLOBOSA-like"* (*ZinGLO*) have at least four copies, each with its own expression pattern (Bartlett and Specht 2010). Although *ZinGLO* genes

are likely involved in setting up organ boundaries and contributing to perianth development as in *Arabidopsis thaliana*, the separate Zingiberales copies may be involved in differentiating novel organs or generating novel organ morphologies in the order. Likewise, by mapping the expression patterns of putative floral symmetry genes in representatives of the Zingiberales, we can assess the level of similarity in a floral developmental program across distantly related angiosperms.

For this study, we chose one representative of the ginger families, *Costus spicatus* and one representative of the banana families, *Musa basjoo*. Figure 2 illustrates the *TBL/CYC-like* expression patterns found in (Bartlett and Specht 2011) for *Costus spicatus* and *Heliconia stricta,* another representative of the banana families. If the relationship between *CYC-like, RAD-like* and *DIV-like* genes are comparable in monocots, we expected that at least one *RAD-like* gene copy would match the *TBL*/*CYC-like* expression patterns seen previously. In order to have directly comparable results, we assayed expression via reverse-transcription PCR (RT-PCR) for *TBL, DIV-like,* and *RAD-like* genes as well as *in situ* hybridization for *DIV-like* and *RAD-like* genes for both representative species.

Methods:

Tissue Collection

Floral and vegetative tissue for DNA, RNA, and *in situ* hybridization was collected from clonally propagated individuals at the UC Berkeley Oxford tract greenhouse (*Costus spicatus*) and the UC Berkeley Botanical Garden (*Musa basjoo* – 89.0873). For all DNA extractions, young leaf tissue was harvested and either extracted immediately or preserved in silica gel for later extraction.

For *in situ* hybridization, tissue was fixed in FAA and vacuumed for 10-20 minutes until no rising bubbles were visible. FAA was then changed, and tissue was soaked overnight at 4ºC to complete the fixation (10 hours for *Costus spicatus* flowers and 18 hours for *Musa basjoo* inflorescence). *Musa basjoo* tissue was about 1cm^2 while *Costus spicatus* individual flowers were about 3-5mm^2. Tissue was embedded following the Javelle et al. protocol (Javelle, Marco, and Timmermans 2011) with the exception of fixing in FAA rather than PFA and the exclusion of PBS washes. Tissue embedded in paraplast plus was sliced on a micron retracting rotary microtome in the UC Berkeley Bioimaging Facility in 8 micron sections and mounted on Fisher Scientific probe-on plus slides. Slides were incubated overnight at 42ºC to complete mounting.

For RNA extractions, *Costus spicatus* flowers were collected at various stages of development from a single inflorescence. Tissue from different stages of development was ground and mixed before RNA was extracted. *Musa basjoo* flowers were taken from the first and second unopened bract, those closest to anthesis. Flowers were dissected on site and immediately flash frozen in liquid nitrogen stored on dry ice. Tissue was returned in dry ice to the laboratory and immediately placed in a -80ºC freezer. Tissue dissection occurred as follows:
Costus spicatus : dorsal sepals, ventral sepal, dorsal petal, ventral petals, staminodial labellum, fertile stamen, total flower.
Musa basjoo : Floral tube, free petal, stamens, total flower.

For *in situ* hybridization, entire infloresence stalks were cut, brought to the lab immediately, and dissected. Apices were transferred to ice cold nanopure water on ice to continue the dissection under an ambient light dissection microscope. As much as possible, developing flowers were left intact and subtending bracts were removed. Entire *Musa basjoo* infloresences were fixed and embedded, while individual *Costus spicatus* flowers were fixed and embedded.

For floral developmental images, a *Musa basjoo* infloresence was cut from the UC Berkeley Botanical Garden (*Musa basjoo* – 89.0873) collection, returned to the lab, dissected, and fixed in freshly made FAA. The tissue was then dehydrated stepwise according to the Javelle et al. protocol and stored in 100% ethanol plus nigrosin stain (0.4% nigrosin, 95% ethanol) for three weeks according to the Charlton et al. protocol prior to imaging, with periodic refreshes of stain and ethanol (Charlton et al. 1989; Javelle, Marco, and Timmermans 2011).

Extractions

RNA extractions were performed using the PureLink Plant RNA Reagent (Invitrogen) Cat. No. 12322 following the manufacturers instructions with the following adjustments: tissue was pre-ground with chilled sterile mortars and pestles in liquid nitrogen then ground with Matrix D and a large ceramic bead in a FastPrep™ FP-120, 5PRIME Phase lock gel tubes were used to separate cellular debris from genetic material, and lithium chloride was used to precipitate the RNA. RNA concentration and purity was determined with a 1% agarose gel and an ND-1000 NanoDrop® Spectrophotometer.

DNA extractions were performed using SDS extraction buffer according to a protocol adapted by Dr. Sarah Hake's lab at the Plant Gene Expression Center (USDA: ARS) from Konieczny and Ausubel (Konieczny and Ausubel 1993). Tissue was pre-ground in Matrix A in FastPrep™ FP-120.

cDNA synthesis was performed using the BioRAD iScript Select cDNA synthesis kit, using manufacturer's instructions. Working 1:10 diluted stocks of cDNA were used for downstream protocols.

Reverse Transcription PCR (RT-PCR)

RT-PCR was performed on diluted cDNA samples using Phire taq and 37 cycles as in (Bartlett 2010). RT-PCR reactions were performed in Bio RAD My Cycler thermocyclers. Primers used are in table 1.

in situ *hybridization*

RNA probes were generated as follows. PCR reactions from cDNA were gel extracted using a Qiagen QIAquick gel extraction kit following the spin protocol. Samples were then ligated into a pJET vector using a CloneJET PCR cloning kit from Thermo Scientific™ containing T3 and T7 promoters and an antibacterial resistance marker gene following manufacturer instructions with the exception of ½ volume reactions. Ligated vector was transformed into Mach1™ competent

cells made according to Cold Spring Harbor protocols (Sambrook and Russell 2001) and grown on LB media plus ampicillin. A PCR was performed and sequenced to verify transformation. Successful colonies were grown overnight in liquid LB media, then extracted via the miniprep protocol (Sambrook and Russell 2001). Resulting plasmid DNA was diluted 1:50 and used as a template for PCR using pJET forward primer and gene specific reverse primer (table 1). Resulting PCR reactions were cleaned using a Qiagen QIAquick PCR purification kit following the spin protocol, then purified with a phenol chloroform extraction (Javelle, Marco, and Timmermans 2011). The resulting product was used as a template for *in vitro* transcription using T7 polymerase (Invitrogen) and DIG RNA labeling mix (Roche) as in (Bartlett, Kirchoff, and Specht 2008).

In situ Hybridization was performed in accordance with the Javelle, Marco, and Timmermans protocol (Javelle, Marco, and Timmermans 2011). All Pyrex, glass, and metal materials were wiped down with RNase-away and baked in aluminum foil at 200ºC for at least 4 hours prior to use. Histoclear was used as a clearing agent, Proteinase K to digest proteins, RNase to digest unbound RNA, and acetic anhydride in triethanolamine-HCl to reduce background staining. 0.5-2µl of probe was used for hybridization. Slides were imaged on a Zeiss AxioImager M1/Hamamatsu MicroPublisher color camera at the UC Berkeley Bioimaging Facility. Three rounds of *in situ* hybridization were performed using *Costus spicatus* developing flowers and a *Musa basjoo* inflorescence with two slides of at least 6 *Costus* flowers per slide and at least 15 *Musa* flowers per slide per probe treatment. Sequences amplified are *RAD1 antisense* for *Costus spicatus* and *RAD2a antisense* and *DIV1 antisense* and *sense* for *Musa basjoo* for two rounds, and *RAD1 antisense*, *DIV1 antisense* and *sense* for *Costus spicatus* and *RAD2a antisense* and *DIV1 antisense* and *sense* for *Musa basjoo* for the third round.

Results:

Floral Development

Late stage developing *Costus spicatus* and *Musa basjoo* flowers are shown in Figure 3. In *Costus spicatus*, the sepals have been removed (Figure 3a), revealing the three petals, the fused staminodes forming the staminodial labellum, and the large and fertile dorsal stamen. Although not apparent in this image, the dorsal petal of *Costus* is larger than the ventral counterparts throughout development. In *Musa basjoo* (Figure 3b), the larger dorsal sepals, as well as the ventral petals and ventral sepal are fused into a floral tube, though here the tips of the organs are still differentiated. The dorsal free petal (FP) is the only member of the perianth that remains separate from the floral tube. The five fertile stamens are also visible. At this stage of development all organs have formed in both flowers, though the gynoecium is not visible in either flower, and the aborted dorsal stamen of *Musa basjoo* is likewise not visible.

The stages of *Musa basjoo* floral development were analyzed in order to determine timing of *in situ* expression analyses (Figure 4). Dr. Mohammed Reza Dadpour took the images with a Nikon camera through a Leitz Wetzlar epi-illumination microscope. Images were taken of individual developing flowers from several hands of an infloresence apex. Each successive image shows an additional whorl of floral organ primordia forming. Panel D is the first point at which the actinomorphy of the developing flower is broken by the differential expansion of the dorsal and

ventral sepals. In panel E, the ventral petals have begun to fuse with the sepals, which will result in the floral tube. In panel F, the dorsal stamens have expanded more than the ventral outer stamen, while the ventral petals have expanded more than the dorsal free petal, which will remain independent of the floral tube at anthesis. Panels E and F approximate the stages assayed with *in situ* hybridization, with all organs formed but not yet expanded.

Gene Expression

In many cases, expression patterns determined via *in situ* hybridization and RT-PCR did not match. In comparing published *in situ* results (Bartlett and Specht 2011) for *Heliconia stricta* with our expression results for *Musa basjoo* (Compare Figure 2 to Figures 5 and 6), the disparity may be true differences in the regulation of expression and/or function of the genes assayed. For the same-taxon comparisons of the methods (Figures 5 and 6 vs. Figures 7 and 8), increased sensitivity of RT-PCR in comparison to *in situ* hybridization, lack of specificity in RT-PCR primers, and/or developmental timing differences between the flowers assayed may account for disparities. Flowers assayed with *in situ* hybridization are at early stages of development in which floral organs have differentiated but antheridia are not yet fertile and organs are still in initial expansion stages (Figure 3). Flowers assayed with RT-PCR, in contrast, are just prior to anthesis, with all organs formed, fertile antheridia, and almost complete organ expansion. For the remainder of the chapter, we have interpreted differences between *in situ* hybridization and RT-PCR results as showing differences in expression due to developmental timing.

Gene Expression in Musa basjoo

If *Heliconia stricta in situ* hybridization results (Bartlett and Specht 2011, Figure 2) are comparable to *Musa basjoo TBL* expression, *TBL1a* is expressed in the dorsal staminode and gynoecium of developing flowers (ie Figure 3B) and *TBL2* is expressed in the gynoecium and the ventral sepals. Early *RAD2a* expression overlaps with that of *TBL2*, with expression throughout the stamens and floral tube, strongest in the ventral part of the flower (Figure 7B). *DIV1* expression is similarly universal, with stronger binding in the ventral portion of the flower and weaker overall signal than *RAD2a* (Figure 7F). Binding of the *DIV1sense* probe to the thecae of stamens compared with the no probe control (Figure 7C vs. E) shows that the apparent higher expression of *RAD2a* and *DIV1* in the stamens is an artifact of background signal.

In later developmental stages of *Musa basjoo*, *TBL1a, DIV1, DIV3, RAD1a, RAD1b, RAD2a, RAD2b,* and *RAD2c* expression was assayed via RT-PCR (Figures 5 and 6). Strong *TBL1a* expression was recovered throughout the developing flower, signaling an expansion of expression from earlier developmental stages. *DIV1* expression remained general, but with strongest expression in the free petal. *DIV3* shared a similar pattern, with strongest expression in the floral tube and free petal. *RAD2a* and *RAD2c* expression is similar at this stage of development to *RAD2a* expression assayed via *in situ* hybridization at earlier development. *RAD2a* and *RAD2c* is found expressed everywhere but the free petal, with strong *RAD2a* expression in the stamens. *RAD2b* expression diverges from the other two copies assayed, with universal expression across the developing flower. *RAD1a* and *RAD1b* expression was found in the floral tube and free petal, with *RAD1b* expression stronger than *RAD1a*.

Overall, *TBL1a* expression shifts from dorsal sepals to throughout the flower. *DIV-like* expression appears universal throughout development, though *DIV-like* expression may shift from being strongest in the ventral portion of the flower in early development to stronger in the dorsal free petal at later stages. *RAD-like* genes exhibit 3 patterns, with *RAD2a* showing universal expression (though stronger in the ventral region) at early stages, and *RAD2a* and *RAD2c* losing expression in the free petal at later stages, *RAD1a* and *RAD1b* with perianth expression, and *RAD2b* with universal expression.

Gene Expression in Costus spicatus

In early developmental stages of *Costus spicatus* (ie Figure 3A), *TBL1a* is expressed in the gynoecium, the ventral petals and sepal, the staminodial labellum, and the thecae of the fertile stamen. *TBL2* is expressed in the thecae of the dorsal fertile stamen. *RAD1* expression is general, with the strongest expression in the dorsal fertile stamen thecae and the staminodial labellum (Figure 7A). *RAD1* expression overlaps with *TBL1a*, though *RAD1* expression is also present in the dorsal perianth. The sepals and bract surrounding the *Costus spicatus* flowers are dappled in the no-probe treatment as well as the *RAD1* antisense (Figure 7A and D), making it difficult to interpret staining in those regions. Additional dark circular spots present in the perianth of both species are vascular bundles, which tend to stain more than surrounding tissues due to rapid cellular division.

Although *TBL1a* could not be amplified with RT-PCR, *TBL1b* expression was universal in later stage *Costus spicatus* flowers, but weakest in the dorsal sepal and fertile stamen (Figures 5B and 6). *TBL2* expressed was recovered everywhere but the staminodial labellum, with strongest expression in the dorsal petal. *RAD1* expression was very weak, and present in the ventral sepal, dorsal petal, and staminodial labellum. *DIV1* and *DIV3* expression was universal, with strong *DIV3* expression throughout the flower and *DIV1* expression weakest in the staminodial labellum.

Overall, *TBL1* and *TBL2* expression is overlapping but partially complementary throughout *Costus spicatus* floral development. *TBL1a* is localized to the ventral portion of the flower in early development, and though *TBL1b* spreads throughout the flower in late development, expression remains weak in the dorsal portion of the flower. *TBL2* expression begins localized to the fertile thecae in early development, and later spreads throughout the flower aside from the staminodial labellum, with the strongest expression in the dorsal petal. *DIV1* and *DIV3* expression is universal in late stage *C. spicatus* flowers. *RAD1* expression is universal in early stages but strongest in the androecium. In later stages, *RAD1* expression is weak and found throughout the central regions of the flower.

Further in situ *Hybridization Results*

In tissues assayed by RT-PCR, the free petal was the organ lacking *RAD2* expression for *Musa basjoo*, but this organ was unfortunately not caught in the sections shown (Figure 7). Importantly, a negative control is lacking in the *Costus spicatus* tissue. A third round of *in situ* hybridization was performed (Figure 8), in which probe binding was less successful and subsequent staining much lighter. These reactions are included to provide the genes and floral

organs missing from Figure 7. In comparing Figures 7 and 8, notice that the *Musa basjoo* slides used in Figure 7 were oblique, missing the dorsal portion of the flower where the aborted stamen and free petal are visible, and where the tips of the organs in the floral tube come together. *RAD-like* and *DIV-like* expression in *Musa basjoo* corroborates the results shown in Figure 7 (Figure 8 B and D vs. Figure 7 B and F), and though staining is weak, it appears that probe bound to the aborted stamen and free petal in these sections. Gene expression in *Costus spicatus* in inconclusive as very light staining appears to have occurred exclusively in the androecium in all three probe treatments (Figure 8 A, C, and E). The protocol should be repeated to verify results.

Discussion:

The Establishment of Symmetry Patterns in Developing Flowers

The symmetry patterns of flowers can change as they develop, from the initiation of sets of organs (in whorled phyllotaxy) to just prior to anthesis (floral maturity and opening). Although there are exceptions, most flowers begin development with actinomorphic symmetry (Peter K Endress 1999b; P K Endress 2001; Reyes, Sauquet, and Nadot 2016). This symmetry can be broken through asymmetric organ initiation, asymmetric organ expansion, organ fusion (connation or adnation), or differential organ elaborations that typically occur in the final stages of floral development (Tucker 1999). Unlike leaf initiation, the location of floral organ primordia is determined prior to any organ's formation and is not positionally determined by previous organ initiation events (Peter K Endress 1999b). For this reason, changes in gene expression throughout development are of interest, as they may play a definitive role in differential development of organs across the flower, contributing to zygomorphy.

The Roles of Symmetry Genes in Antirrhinum majus

In the model *Antirrhinum majus, CYC* and *DICH* expression is concentrated in the dorsal portion of the flower. Early in development these genes are responsible for the abortion of the dorsal stamen and the internal asymmetry of the dorsal petals, while later on they promote the greater growth of dorsal petals in comparison to the lateral and ventral petals (Luo et al. 1996; Luo et al. 1999). *RAD* expression is concentrated in the dorsal portion of the flower, with RAD protein also present in the lateral petals. In these regions, RAD protein prevents DIV transcriptional activity by sequestering the protein binding partners DRIF1 and DRIF2. This competition is lacking in the ventral petal, in which DIV + DRIF protein complexes induce the symmetric growth of the large ventral petal (Almeida, Rocheta, and Galego 1997; Galego and Almeida 2002; Corley et al. 2005; Costa et al. 2005; Perez-Rodriguez et al. 2005). *DIV* expression constricts throughout development, initially being present throughout the developing ventral petal and overlapping regions of lateral petals and eventually retreating to a furrow of cells demarking the boundary between the corolla tube and the lobes (Galego and Almeida 2002). Due to this restriction, in initial development DIV protein induces the expansion of the ventral petal and adjacent portions of lateral petals, but later contributes to the lobe vs. tube boundary.

Floral Development in the Musaceae

In the Musaceae, the sepals and ventral petals are fused to form the floral tube, while the dorsal petal remains free (B. K. Kirchoff 1988). The sepals are fused on their abaxial sides, while the ventral petals are adnate on the abaxial side. The free petal is anatomically similar to the ventral sepals, but does not undergo this fusion (Inta, Traiperm, and Swangpol 2015). A ring meristem, or a circular collection of meristematic cells from which organ primordia form, is thought to be involved in organ formation in the Musaceae (Tucker 1999). From the ring meristem, the petals and outer stamens form, followed by inner stamens. Although the dorsal aborted stamen is not visible in Figures 3 and 4, it can be seen in cross section (Figure 8). Potentially due to physical constraints on developing *Musa* flowers caused by the bract encasing each hand of flowers, the dorsal sepals are larger at initiation and remain larger throughout development until indistinguishable from other members of the floral tube. This type of mechanical constraint leading to differential size at organ initiation is common, though many flowers compensate later in development, as in *A. majus* and *Bournea leiophylla* (Luo et al. 1996; Peter K Endress 1999b; Zhou et al. 2008). The asymmetry of the developing *Musa* flower is thus apparent at the onset of organ initiation, though it is enhanced at the point of inner androecium initiation, at which point the free petal is noticeably set apart from the developing floral tube (Figure 4F).

Predicted Roles of TBL*,* RAD-like*, and* DIV-like *in Musaceae Floral Development*

If *Heliconia stricta* and *Musa basjoo* share the same expression patterns, *TBL2* expression is ventral in early stage *M. basjoo* flowers, while *TBL1a* expression is localized to the dorsal staminode (Figure 2). In later stages of development *TBL1a* is expressed throughout the developing flower. *TBL2* could not be amplified for RT-PCR, and sequences could not be recovered from the *Musa basjoo* floral transcriptome (data not shown, methods as in Chapter 1). This could indicate lack of expression in late development. In early development, *RAD2a* expression overlaps that of *TBL2*, with expression throughout the androecium and floral tube. Although unclear due to oblique sections in Figure 7 and poor staining in Figure 8, expression may be lighter in the dorsal regions of the flower. Later in development, *RAD2a* and *2c* expression contracts to eliminate the free petal, while *RAD2b* is expressed throughout the flower. *RAD1a* and *1b* have a divergent pattern, expressed only in the perianth (figures 5 and 6). As in *A. majus*, *DIV1* is expressed throughout early and late developmental stages and *DIV3* throughout late developmental stages (Figures 5, 6, 7, and 8). If the relationships from *A. majus* hold true in the Zingiberales, *TBL* genes should induce the expression of *RAD-like* genes, resulting in reduced growth compared to tissues lacking *RAD-like* gene expression. We find that *TBL2, RAD2a,* and *RAD2c* adhere to this pattern. Ventral *TBL2* expression is consistent with smaller ventral sepals, and *RAD2a* and *RAD2c* may play roles in reducing the growth of the dorsal petal, resulting in a small petal free from the petal and sepal fusion that forms the floral tube. *DIV-like* expression is universal, matching expression patterns from the snapdragon model. *RAD2* genes may play a role in suppressing *DIV-like* activation of growth and expansion genes where expressed, allowing the dorsal region to outpace the ventral.

Floral Development in the Costaceae

In the Costaceae, the sepals are fused into a synsepalous calyx (B. Kirchoff 1991). At anthesis, all floral organs including the style and stamens form a floral tube. The dorsal stamen remains fertile, though its filament is laminar. The lateral and ventral stamens are infertile and laminar,

and fuse to form a staminodial labellum (B. Kirchoff 1991). In the Costaceae, all organs apart from the calyx form on a ring meristem, or a round ridge of meristematic cells that form successive rounds of organ primordia. After the sepals, the ring primordium produces the petals and the inner androecial members relatively simultaneously and opposite one another, followed by the outer members (second stamina whorl) at their margins. Lastly the gynoecium forms at the center, growing downward into its inferior position. As they develop, the sepals fuse together forming the calyx (B. K. Kirchoff 1987). In the development of Costaceae flowers, sepals form individually in a predictable pattern. For this reason, actinomorphic symmetry is broken as the first floral organs are initiated. However, because these organs are symmetrical and fused at anthesis, we instead consider symmetry broken at the point when the ring primordium (which will give rise to the rest of the floral organs) widens at the dorsal side of the flower to produce the dorsal petal and the large fertile stamen. These organs are larger than their ventral counterparts at the onset, and continue to grow at pace with the rest of the organs, maintaining a larger size throughout floral development through to anthesis. The outer whorl androecial members that will form the staminodial labellum never fully differentiate from the inner members, though distinct primordia apices are visible early in development.

Predicted Roles of TBL*,* RAD-like*, and* DIV-like *in Costaceae Floral Development*

The RT-PCR results presented in Figures 5 and 6 as well as previous work on *TBL* gene expression in *Costus* (Figure 2) (Bartlett and Specht 2011) are consistent with a role for orthologs to *A. majus* symmetry genes in setting up *Costus* zygomorphy, though relationships among gene copies has yet to be determined. In early developmental stages assessed with *in situ* hybridization, *TBL1a* is expressed in the ventral portions of the flower, while *TBL2* is expressed in the anthers of the fertile stamen (Bartlett and Specht 2011, Figure 2). At later developmental stages, assessed with RT-PCR, *TBL1b* has universal expression, while *TBL2* is expressed throughout the flower aside from the staminodial labellum (compare Figures 2 and 5). *DIV1* and *3* are expressed throughout the development of *Costus spicatus* flowers. *RAD1* has an interesting pattern, expressed in the dorsal petal, ventral sepal, and staminodial labellum, a central expression pattern that excludes the fertile stamen. Additionally, expression is very slight, with poor amplification (Figure 5). *RAD1* expression overlaps with that of *TBL1*. If *DIV-like* genes promote organ expansion and *RAD-like* genes inhibit organ expansion by indirectly antagonizing *DIV-like,* the organs found with *RAD-like* expression should be those with inhibited growth at that stage of development.

The dorsal petal and fertile stamen of Costaceae flowers are larger at initiation than their counterparts, and grow faster. In early development, Bartlett and Specht (2011) found that *TBL1* and *2* are not expressed in the laminar filament of the fertile stamen or in the dorsal perianth. *RAD1* expression in early development is general, but like *TBL1,* does not seem to be highly expressed in the dorsal perianth (Figure 7A). This pattern is consistent with *DIV-like* promotion of growth in those organs. In later development, *TBL2* is expressed everywhere but the staminodial labellum, with the strongest expression in the dorsal petal. This pattern is consistent with the rapid expansion of the staminodial labellum late in development to form the elaborate labellum-like portion of the flower that creates the majority of the *Costus* floral display. *TBL1b* is expressed throughout late development *Costus* flowers, with weakest expression in the dorsal sepal and fertile stamen. The lack of *RAD1* expression in the ventral petals later in development

might correspond to late developmental expansion of those organs to compensate for the early expansion of the dorsal region of the flower, though *RAD1* expression does not perfectly match any *TBL* copy thus far assayed. *RAD2* could not be amplified. This is likely the fault of poor primer matching rather than lack of expression. Understanding the expression pattern of *RAD2* may help elucidate how the expression patterns of *TBL* and *RAD-like* genes overlap in the developing *Costus* flowers.

Relationships Between Gene Copies

In the model floral symmetry network (Figure 1), *CYC* expression induces the expression of *RAD*, while *DIV* is expressed everywhere. Given our expression patterns, this direct one-to-one relationship likely does not hold true in the Zingiberales. None of the *TBL* copies show expression exactly matching the *RAD-like* copies assayed in either species studied. The relationships between these genes may exist, but are likely complicated by multiple sets of gene duplications in the order for both *RAD-like* and *DIV-like* genes. In *Musa basjoo TBL2, RAD2a,* and *RAD2c* had asymmetric expression across the dorsal/ventral plane concentrated in the ventral region, and are likely candidates for functional study. In *Costus spicatus, TBL1a* and *RAD1* are the likeliest candidates, with *TBL1a* exhibiting a ventral expression pattern and *RAD1* a central expression pattern.

The ventral expression of *TBL2* matches previously described expression patterns in the Commelinales, the sister order to the Zingiberales (Preston and Hileman 2012). The authors found *TB1a,* an ortholog of Zingiberales *TBL2,* expression was strongest in the ventral portions of zygomorphic *Commelina communis* and *Commelina dianthifolia* flowers. *TB1b,* an ortholog of Zingiberales *TBL1,* was not expressed in floral tissue. Given the reversed patterns of expression of *TBL1* and *TBL2,* and *RAD1* and *RAD2* in the banana species vs. ginger species studied, *TBL1* and *RAD1* genes may have been coopted in the ginger families to produce asymmetry of the androecium. *TBL* and *RAD-like* genes may be differently regulated and/or have different functions between the groups.

Developmental Timing of Gene Expression Results

The gene expression results assayed by RT-PCR (Figures 3 and 4) are at a later developmental stage than shown in the developmental images (Figures 7 and 8). The *in situ* hybridization results (Figures 5 and 6) for *Musa basjoo* are likely at a similar stage to Figure 7 panel F, as all organs are observed, though cross sections were taken deeper within the flower as the gynoecium and fusion of floral tube are observed. Similarly, the *Costus spicatus in situ* results are relatively late stage, as the pollen sacks are visible within the anthers (Figures 5 and 6). At this point the zygomorphy of both species has been established. Although the *in situ* results portray different stages of development, we have not ascertained expression patterns in a developmental series that would illuminate potentially shifting roles of *RAD-like* genes as symmetry patterns are established. Given the RT-PCR results, which allow us to distinguish among organs more than the *in situ* results, *RAD-like* genes may play a role in the suppression of *DIV-like* transcriptional activity, leading to differential growth across the dorsal and ventral regions of developing *Costus spicatus* and *Musa basjoo* flowers.

Future Directions

In order to fully elucidate the overlapping expression patterns of *TBL, DIV-like,* and *RAD-like* genes, each copy should be amplified with RT-PCR, and a developmental series should be captured with *in situ* hybridization, including all relevant organs. This would allow a more complete model of how these genes might interact in floral development.

Although expression data is a reasonably proxy for protein presence, understanding the localization of RAD-like proteins would help us better understand the roles of these proteins in floral development, and paired with an understanding of DRIF-like protein localization can help test the hypothesis that this gene network is acting as in model systems. *Musa basjoo* RAD2a and *Costus spicatus* RAD1 are the most likely proteins to have asymmetric localizations.

Furthermore, Yeast-2 hybrid analyses could help elucidate how proteins interact and whether these interactions are copy-specific in the Zingiberales. Sequence conservation (Chapters 1 and 3) indicates that DRIF-like, DIV-like, and RAD-like genes may have conserved binding sites. Determining binding affinities of these genes in *Musa basjoo*, for which we have a closely related sequenced genome (*Musa acuminata*) could paint a stronger picture of the symmetry network in the Zingiberales. For this purpose, DIV1, 2, and 3, RAD1a, 1b, 2a, 2b, and 2c should be tested for binding affinity to DRIF-like proteins found in Chapter 3.

Finally, the putative floral symmetry genes identified in Chapters 1 and 2 are candidates for future knock-out studies. Using the CRISPR-Cas9 system (O'Connell et al. 2014) in *Musa basjoo*, we could produce single and double mutants of the *RAD2a* and *RAD2c* genes, which are expressed in the floral tube and androecium to the exclusion of the free petal. In these mutants, we could assess not only the effect on the *Musa* floral developmental program, but also localization of DIV-like and DRIF-like proteins in the absence of *RAD2* gene expression. Knock-out experiments would be the most certain evidence of a role for *RAD-like* genes in establishing the zygomorphy of Zingiberales flowers.

Gene Amplified	Primer Name	Primer Sequence (5'-3')
RAD1	RAD1_Costus_F4	GGA GAA CAA GAT GTT CGA GAAG
	RAD1_Costus_R3	GTA AGG CAT TTG GCC CTC CT
	insituCostRADR2antisense	GGA AGA AAG GAG TCC CTT GG
RAD1a	RAD1_Musa_F2	GAC CGC GAA GGA GAA CAA GA
	RAD1_Musa_R2	CCA CGA GCA GGT CGT AGT AGC
RAD1b	RAD1_Musa_F2	GAC CGC GAA GGA GAA CAA GA
	RAD1_Musa_R1	TAG TGG CGC TTC ACT TCC TC
RAD2a	RAD2_Musa_F1	AGC AGA ACA AGC TCT TCG AG
	RAD2_Musa_R1	YTC GAT KCG GTG GAG GTC
	insituMusaRADR1antisense	AAG CTC TTC GAG TGG GCT CT
RAD2b	RAD2_Musa_F1	AGC AGA ACA AGC TCT TCG AG
	RAD2_Musa_R2	GAG CTC GTA GTG TCG CTT GA
RAD2c	RAD2_Musa_F1	AGC AGA ACA AGC TCT TCG AG
	RAD2_Musa_R3	GAC CTC TTC GGC GGT CTT
DIV1	DIV1_Cost_F3	GAT CAG GAG AGG AAG AAA GGA G
	DIV1_Cost_R3	AAT TGG CGG TCC TAA TGT CA
	Div1_Musa_F1	GAA GAA AGG AGT CCC CTG GA
	Div1_Musa_R1	TCC TCT TAT CTT TGC MAC CTG
	insituCostDIVR1antisense	CAT AAT TCC GGG AGA TGC TT
	insituDIV1MusaF1	TCG GAC CAA GAG AGG AAG AA
	insituDIV1MusaR1	GGT CCG AGT GAG CAC AAA AT
	insituDIV1MusaR2antisense	GGA CGG AAG AGG AGC ACA A
DIV3	DIV3_Cost_F2	TGG ACT GAA GAC GAG CAC AA
	DIV3_Cost_R3	CTG CCA GAA TTG AGC CTT ATG
	Div3_Musa_F3	CAC GAG AGG AAG AAA GGA RTC
	Div3_Musa_R1	TCT TGT CTT TGC TGC CAG AG
CYC1a	CYC1a_Musa_F1	AGA TCG TCG RAT ACG GCT CT
	CYC1a_Musa_F2	GAG TTC TTG GAT GGC GTG TT
CYC1b	CYC1b_Cost_F1	CTA CAG GAC ATG CTC GGC TTC
	CYC1b_Cost_R1	TCT TTG ATG GCG GCT TTC
CYC2	CYC2_Cost_F1	GGC WCA GCA AGA TCC ACA C
	CYC2_Cost_R1	GAC TGC TTG AGG AGC CAA TC

Table 1: Primers used to amplify *RAD-like, DIV-like,* and *TBL1-like* genes for RT-PCR and *in situ* hybridization.

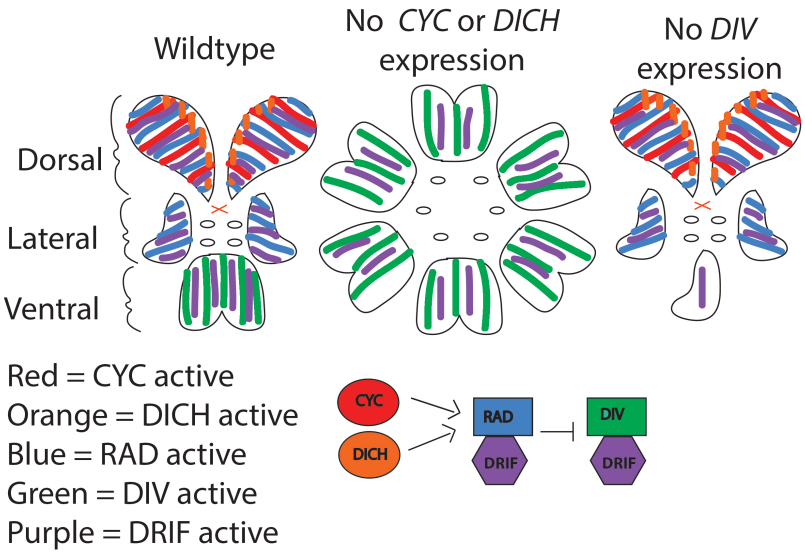


Figure 1: Expression patterns of TCP (*CYC* and *DICH)* and MYB-related (*DIV, RAD,* and *DRIF)* genes. Shading represents regions of protein activity in wildtype and mutant *Antirrhinum majus* flowers. Below is a diagram modeling the interactions between proteins. Arrows indicate protion of expression while perpendicular lines indicate repression of protein activity.

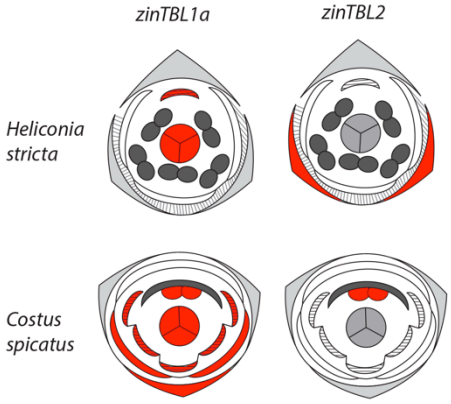
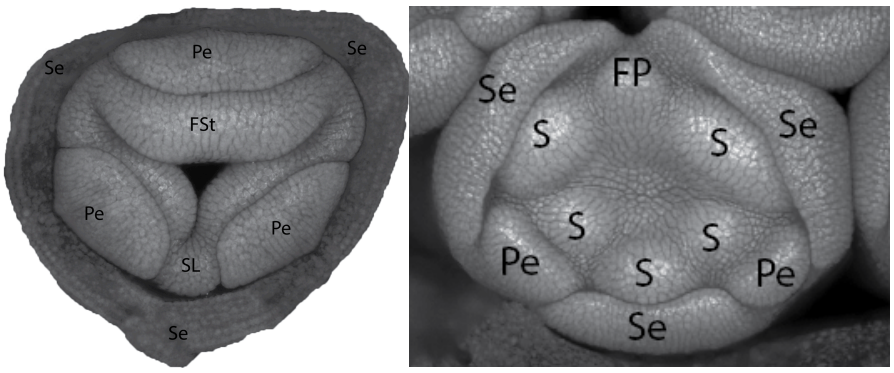


Figure 2: Floral diagrams representing expression (red-shaded regions) of *TBL* (homologs of *CYC*) genes in two Zingiberales species. Expression determined through *in situ* hybridization (Bartlett and Specht 2011).



A B

Figure 3: Developing flowers of *Costus spicatus* (A) and *Musa basjoo* (B). FP = free petal, FSt = fertile stamen, Pe = petal, S = stamen, Se = sepal SL = staminodial labellum. Developing carpels are not visible in these images. Tissue preparation and images by Dr. Mohammed Reza Dadpour.

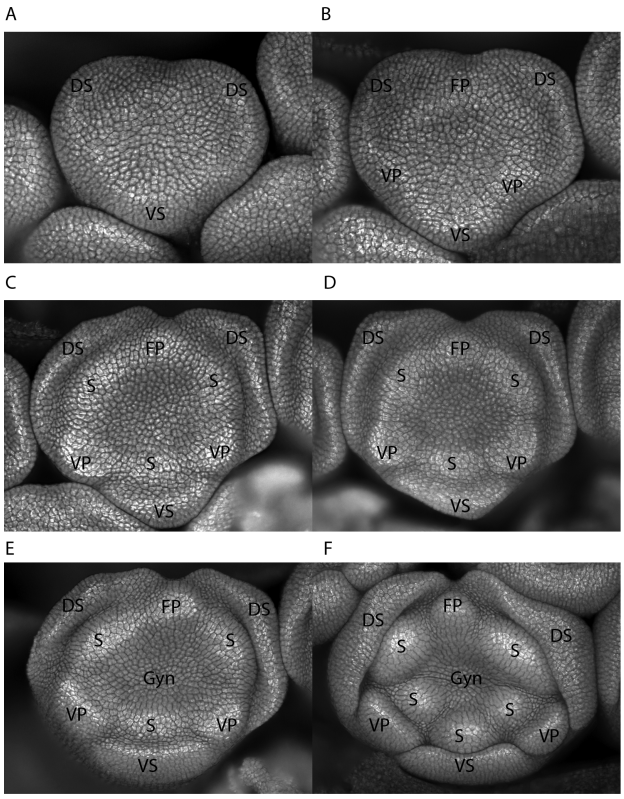


Figure 4: Floral development of *Musa basjoo* in order (A->F). DS = dorsal sepal, VS = ventral sepal, FP = free petal, VP = ventral petal, S = stamen, Gyn = gynoecium.
Tissue preparation and images by Dr. Mohammed Reza Dadpour.

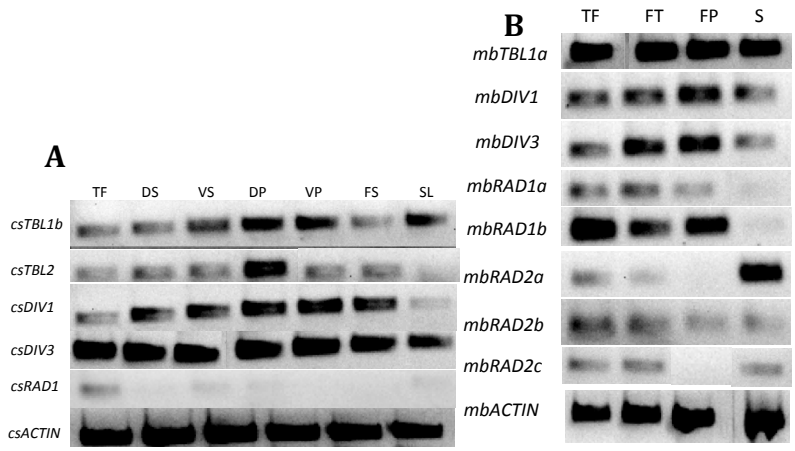


Figure 5: Reverse-transcription PCRs of *CYCLOIDEA/teosinte branched 1-like (TBL)* and *MYB-related (DIV, RAD)* genes for *Costus spicatus* (A) and *Musa basjoo* (B). Key for *Costus spicatus* (A): TF = total flower, DS = dorsal sepal, VS = ventral sepal, DP = dorsal petal, VP = ventral petal, FS = fertile stamen, SL = staminodial labellum. Key for *Musa basjoo* (B): TF = total flower, FT = floral tube, FP = free petal, S = stamens. Full gel images in supplementary figure X.

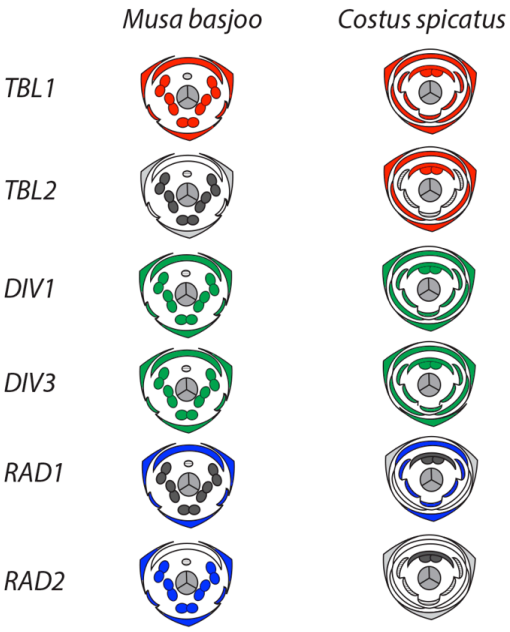


Figure 6: Floral diagrams showing gene expression patterns assayed through RT-PCR. Red: *TBL*, Green = *DIV*, Blue = *RAD*

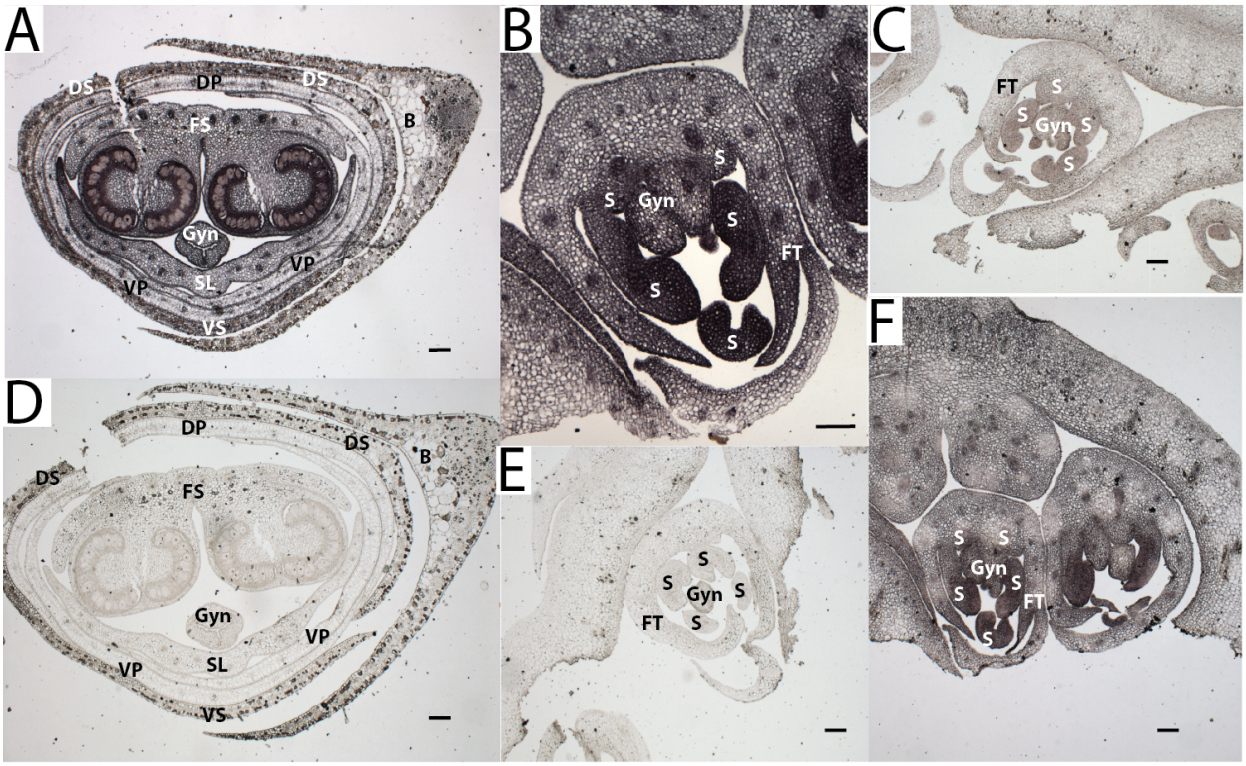


Figure 7: *in situ* hybridization performed on *Costus spicatus* (A,D) and *Musa basjoo* (B,C,E,F) floral tissue with *RAD1 antisense* (A), *RAD2 antisense* (B), *DIV1antisense* (F), *DIV1sense* (C), and without probe (D,E). Scale bars indicate 100µm. Flower are oriented with the dorsal side up.
DP = dorsal petal, DS = dorsal sepal, FS = fertile stamen, FT = floral tube, Gyn = gynoecium, S = stamen, SL = staminodial labellum VP = ventral petal, VS = ventral sepal, B = bract

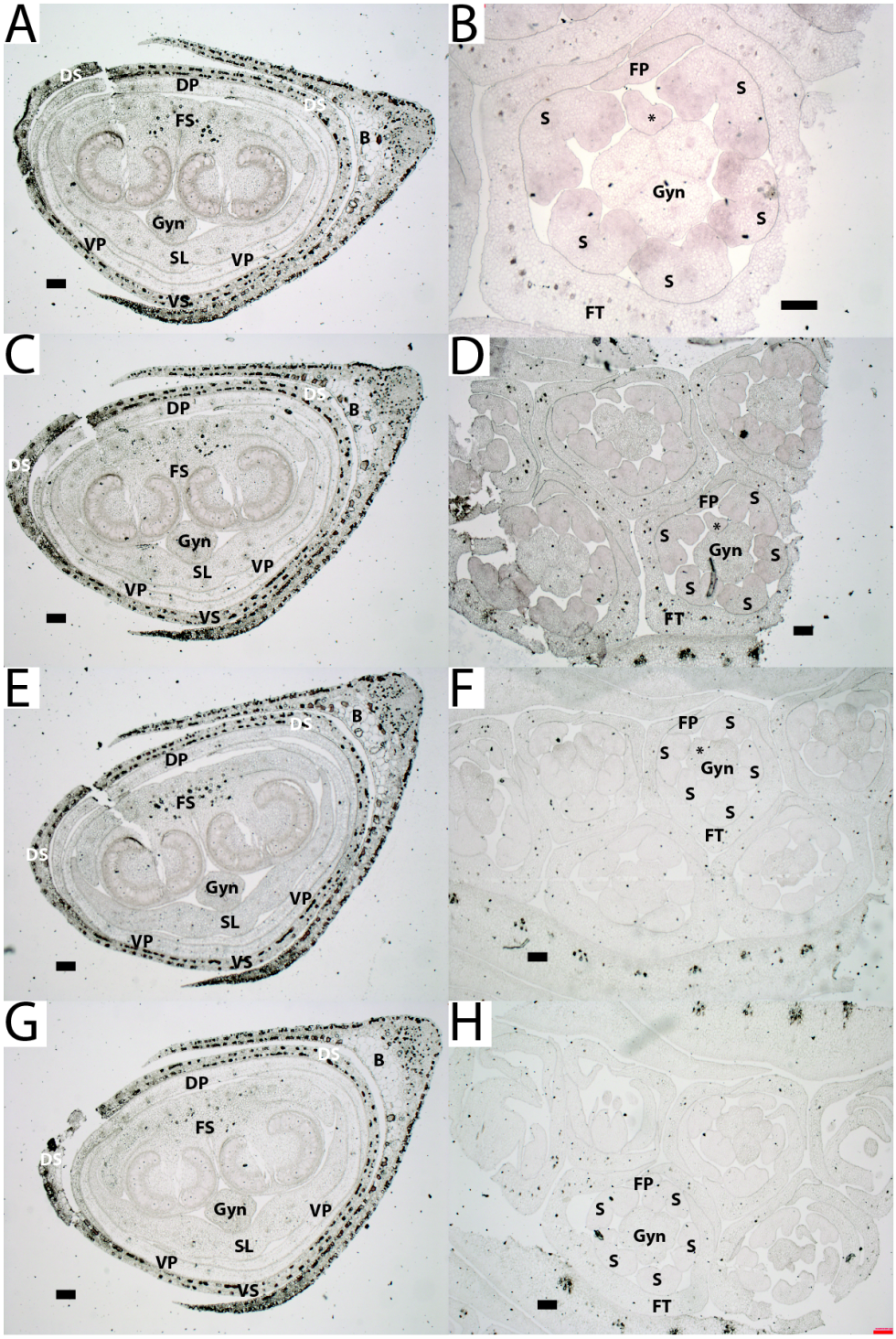


Figure 8: *in situ* hybridization performed on *Costus spicatus* (A,C,E,G) and *Musa basjoo* (B,D,F,H) floral tissue with *RAD1antisense* (A), *RAD2antisence* (B), *DIV1antisense* (C,D), *DIV1sense* (E,F), and without probe (G,H). Scale bars indicate 100µm. Flowers are oriented with the dorsal side up.
DP = dorsal petal, DS = dorsal sepal, FS = fertile stamen, FT = floral tube, Gyn = gynoecium, S = stamen, SL = staminodial labellum VP = ventral petal, VS = ventral sepal, B = bract, * = aborted stamen

Chapter 3: The Roles of *DIVARICATA and RADIALIS INTERACTING FACTORS* in Zingiberales Floral Symmetry

Abstract:

DIVARICATA and RADIALIS INTERACTING FACTORS (DRIFs) influence the floral development of the model plant *Antirrhinum majus* by serving as protein binding partners to the competing proteins DIVARICATA (DIV) and RADIALIS (RAD). This competitive relationship results in the zygomorphy of mature *A. majus* flowers. We present a phylogenetic tree identifying *DRIF-like* orthologs from taxa across flowering plants. We recovered a duplication in eudicot *DRIF* Group 1 genes, resulting in two clades containing *A. majus DRIF1* and *DRIF2* respectively. Two *DRIF* Group 1 Zingiberales clades were recovered, with one containing only sequences from the banana families. We assayed the expression of two *Musa basjoo DRIF* Group 1 sequences with reverse transcription PCR (RT-PCR). The two genes assayed were universally expressed in developing *Musa basjoo* flowers, consistent with a conserved role for *DRIF-like* genes in Zingiberales floral zygomorphy.

Introduction:

DIVARICATA and RADIALIS INTERACTING FACTORS (DRIFs) play an important role in the zygomorphy of the *Antirrhinum majus* flower. DRIF1 and DRIF2 in *A. majus* are protein-binding partners necessary for DIVARICATA (DIV) protein to perform downstream transcriptional tasks. In the presence of RADIALIS (RAD) protein, the DRIFs bind to RAD and localize to the cytoplasm. Only in the absence of RAD do the DRIFs bind to DIV and localize to the nucleus (Raimundo et al. 2013). In this way, despite the general expression of *DIV* throughout the developing flower, RAD protein confines the transcriptional work of DIV protein to the ventral area of the flower.

The *DRIFs* are MYB-related genes like *RAD* and *DIV*. Like *RAD,* the *DRIFs* have a single MYB-like domain, but unlike *RAD* the *DRIFs* have a second domain of unknown function. Raimundo et al. identified the *DRIFs* through a yeast-2 hybrid screen using RAD as bait; only the *DRIFs* were recovered with this protocol (Raimundo et al. 2013).

Raimundo et al. recovered *DRIF-like* sequences from several taxa across angiosperms and built a phylogeny, using *Physcomitrella patens* sequences as an outgroup (Raimundo et al. 2013). This work allowed the authors to distinguish two groups of *DRIFs*. Group 1 includes *A. majus DRIF1* and *DRIF2*, while group 2 includes a previously characterized MYB-related gene from *Solanum lycopersicum* (tomato), *SFSB1*. In *S. lycopersicum*, the SFSB1 protein interacts directly with RAD and DIV proteins as in *A. majus*, but plays a role in fruit development rather than floral symmetry (MacHemer et al. 2011). DRIF protein interactions have only been characterized in *A. majus* and *S. lycopersicum*, and *DRIF* group 1 gene expression (those implicated in floral symmetry) have only been studied in *A. majus* and *Orchis italica.* In *O. italica,* OITA_10599, a *DRIF* Group 1 ortholog, was universally expressed in developing flowers (Valoroso et al. 2017).

Although the tree produced by Raimundo et al. helps illuminate the diversity of *DRIF-like* genes present in plants, the tree contains only 7 taxa, lacking sampling that would make it useful to

identify *DRIF* homologs in other groups of interest. The monocot sequences included a few sequences each from *Oryza sativa* and *Brachypodium distachyon*. In order to find putative homologs to *DRIFs* in the Zingiberales, as well as to illuminate putative orthologs in other groups of interest, we recovered *DRIF-like* sequences from additional taxa and built phylogenetic trees containing taxa across angiosperms. Additionally, in order to test the hypothesis that *DRIFs* may be involved in floral zygomorphy in monocots, we assayed the expression of these genes in *Musa basjoo* (Musaceae) and analyzed the sequences found for the motifs described in *A. majus DRIFs*.

Methods:

Sequence retrieval and alignment

Undergraduate honors researcher Michelle Liu recovered the initial *DRIF* and *DRIF-like* sequences published by Raimundo et al. (2013) from the National Center for Biotechnology (NCBI) using their accession numbers. These included *Antirrhinum majus, Oryza sativa, Brachypodium distachyon, Solanum lycopersicum, Physcomitrella patens, Arabidopsis thaliana* and *Lotus japonicas*. We used the monocot *DRIF-like* sequences as queries in BLAST to recover *Musa acuminata* genomic sequences from the Banana Genome Hub (Droc et al. 2013). Student Research Mentorship Teams undergraduate researcher Annie Zell recovered additional *DRIF-like* sequences through targeted BLAST searches against the NCBI database using closest relative sequences as queries (Altschup et al. 1990). Identical duplicate sequences, and those with less than 50% of the gene were removed.

Zingiberales sequences, as in Chapter 1, were recovered from the cleaned raw reads of transcriptomes for 5 of the 8 families as well as *Dichorisandra,* from the sister order Commelinales. BLAST databases were generated from each set of cleaned reads, which were then queried using the short sequence option with *Musa acuminata DRIF-like* sequences as queries. BLAST results were assembled in Geneious version 6.6.1 (Kearse et al. 2012) using the De Novo assembler with low sensitivity/fastest default settings with exceptions of 10% maximum mismatches per read, 80% minimum overlap identity, and a maximum gap length of 1. Resulting contiguous sequences or "contigs" supported with at least 10X coverage were included in the main alignment. The two *Musa basjoo* sequences used as templates for RT-PCR (arrows, Figure 1) were verified by Sanger sequencing at the Museum of Vertebrate Zoology facility at the University of California at Berkeley.

Initial alignments were made using MUSCLE (Edgar 2004) in Geneious. Sequences were trimmed to the first conserved codon of the protein (the first tryptophan) then re-aligned by translation with MUSCLE in Geneious. Alignments were manually edited in Mesquite (Maddison and Maddison 2014). Regions at the start, end, and center of the alignment where un-alignable were trimmed before generating phylogenies. Mesquite was used to export an amino acid translation of the nucleotide alignment.

Phylogenetic trees

The final *DRIF-like* nucleotide and amino acid alignments were used to produce a phylogeny in RAxML-HPC v.8 on XSEDE (Stamatakis 2014) on the CIPRES server (Miller, Pfeiffer, and Schwartz 2010) with 100 bootstrap replicates. These phylogenies were rooted on the monophyletic group of *Physcomitrella patens DRIF-like* sequences, as in Raimundo et al. (2013).

Expression analysis

Floral tissue was collected, RNA extracted, and cDNA generated as in Chapter 2. RT-PCR was performed on cDNA samples using Phire taq and 37 cycles as in Chapter 2. An actin loading control was used as in Chapter 2. RT-PCR reactions were performed in Bio RAD My Cycler thermocyclers. Primers used are in table 1.

Results:

Phylogenetic Trees

The resulting phylogenetic trees closely match the tree published in Raimundo et al. (2013) (Figures 1 and 2). As previously found, we recovered two monophyletic groups of *DRIF-like* sequences within the angiosperms. Group 1 contains *DRIF1* and *DRIF2* (purple stars, Figure 1), from *Antirrhinum majus*, while group 2 contains *SFSB1* and *SFSB2,* the *Solanum lycopersicum* genes found to be involved in fruit development (MacHemer et al. 2011). *DRIF* genes from both group 1 and group 2 were present in all organisms queried. Each group is highly supported, with 100 and 99 bootstrap support respectively. Relationships in the amino acid tree and nucleotide trees are similar, but as expected, clades in the amino acid tree are supported by fewer bootstrap replicates overall. With the additional taxa, we were able to recover a single gene duplication event within the eudicots after the divergence of the monocots resulting in two *DRIF* Group 1 clades. One of the *DRIF* Group 1 clades contains *A. majus DRIF1* while the other contains *A. majus DRIF2.* Within each *DRIF* group and subset, gene relationships closely mirror species relationships.

Each Zingiberales taxon may have two copies of *DRIF* Group 1 (*DRIF1a* and *DRIF1b*), indicating a duplication event within the order. The presence of multiple *Musa basjoo* (musacomp sequences) for each *Musa acuminata* gene may indicate a species-specific duplication, but more likely indicates allelic variants recovered from transcriptome data. Our inability to recover *Costus spicatus DRIF-like* sequences is likely a result of poor sequence recovery in the transcriptome data for that species rather than lack of *DRIF-like* genes in the family. Unfortunately, *DRIF-like* sequences were not included in the 2017 sequence capture that yielded the MYB-related sequences reported in Chapter 1. Within the Zingiberales *DRIF-like* sequences, there is one monophyletic group (*DRIF1b*) including taxa from four families (Musaceae, Lowiaceae, Marantaceae, and Zingiberaceae) sister to a second, smaller clade (*DRIF1a*) containing sequences from only Lowiaceae and Musaceae. This may represent a banana group specific duplication, a ginger group specific gene loss, or failure to recover the *DRIF1a* gene copy from ginger families due to potential sequence divergence.

Sequence Analysis

Motifs recovered in a subset of *DRIF* Group 1 sequences using MEME (Bailey and Elkan 1994) are shown in Table 2. MEME recovered two motifs overlapping the MYB-like and DUF3755 domains recovered in Raimundo et al. (Raimundo et al. 2013). Across Group 1 *DRIF-like* sequences, the MYB-like domain contains the characteristic regularly spaced tryptophans (W), with the central tryptophan replaced with a tyrosine (Y) (arrows, Figure 3). Figure 3 illustrates the domains as well as the motifs recovered with MEME on the alignment of *DRIF* Group 1 sequences. We used FIMO (Grant, Bailey, and Noble 2011) to search *DRIF* Group 1 sequences for the MYB and DUF3755 motifs found in *A. majus DRIF1* and *DRIF2*. The motifs were found in all full-length sequences with significant *p* values (Data not shown). These results confirm that recovered sequences are likely *DRIF* orthologs, and that these domains are likely conserved across flowering plants.

Expression Analysis

Results of reverse-transcription PCR (RT-PCR) for two copies of *DRIF1b* (arrows, Figure 1) in *Musa basjoo* (Figure 4) indicate universal expression in developing *Musa basjoo* flowers, consistent with the model from *A. majus*.

Discussion:

DRIF-like proteins have so far been functionally described only in *Antirrhinum majus* and *Solanum lycopersicum*. In *A. majus* the DRIF proteins form complexes with either RAD or DIV. In the latter heterodimer, DIV+DRIF perform transcriptional functions leading to ventral petal identity (Raimundo et al. 2013). A similar relationship was found in *S. lycopersicum*, though the DIV-DRIF complex acted in fruit development rather than floral development (MacHemer et al. 2011). In both cases, the DIV+DRIF complex initiated cellular expansion (Galego and Almeida 2002; MacHemer et al. 2011; Raimundo et al. 2013).

The low *p* values found for the MYB domain of the recovered *DRIF-like* sequences as well as the sequence similarity (65% mean pairwise identity across the MYB domain for all included sequences) among *DRIF-like* sequences for the taxa studied indicate conservation of this domain. Interestingly, multiple copies of *DRIF-like* genes are present throughout angiosperms and even in *Physcomitrella patens*, despite the lack of zygomorphy or even flowers in many of these lineages. The only binding partners found for RAD and RAD-like proteins have been DRIF-like proteins, though in the case of SFB1 in *Solanum lycopersicum*, the DRIF-like protein is a Group 2 rather than a Group 1 protein (Raimundo et al. 2013) indicating the potential for functional diversification following duplication. Additional copies of *RAD-like* and *DIV-like* genes are present in most taxa (see Chapter 1) including *A. majus,* though few have been functionally characterized. If the RAD/DRIF protein partnership is conserved across flowering plants, sets of duplicated copies may play different roles throughout the plant.

Despite apparent monogamy and widespread sequence conservation, the RAD/DIV/DRIF relationship is likely not restricted to floral development and instead plays roles in other differentially developing structures. Ectopically expressing a homolog to *RAD* in *Arabidopsis thaliana* hypocotyls suppressed expansion in central hypocotyl cells, retarding the formation of

the hook typical in developing seedlings (MacHemer et al. 2011). In addition to floral tissue, *RAD-like* expression was detected through RT-PCR in *Costus spicatus* leaf tissue (data not shown), consistent with a role in vegetative structures. *CYC-like* expression has also been recovered in vegetative tissue (Damerval et al. 2007; Chapman, Leebens-Mack, and Burke 2008; Howarth et al. 2011) as have *RAD-like* and *DIV-like* expression (Howarth and Donoghue 2009; Boyden, Donoghue, and Howarth 2014). The likelihood of additional functions for these MYB-related genes is consistent with a growing literature showing multiple functions of protein complexes (Specht and Howarth 2014), and helps to explain the surprising conservation of these genes across diverse lineages of land plants.

Future Directions:

In order to recover more Zingiberales *DRIF-like* sequences, *DRIF-like* genes reported here should be used to generate baits for a sequence capture as described in Chapter 1 (Sass et al. 2016). Retrieving additional sequences would allow us to determine copy number, discover whether there has truly been a duplication or gene loss between the banana and ginger groups, and to develop primers to assess expression of *DRIF-like* copies in additional taxa.

In *A. majus* and in *S. lycopersicum*, endogenous levels of RAD and RAD-like protein are sufficient to suppress the function of DIV and DIV-like proteins early in flower or fruit development. This is likely due to an abundance of protein that surpasses the protein levels of DRIF and DRIF-like protein binding partners, allowing most or all of these proteins to be sequestered and unable to bind to DIV or DIV-like (Cui et al. 2010; MacHemer et al. 2011). This being true, we expect levels of *RAD-like* expression in Zingiberales flowers to surpass that of *DRIF-like* early in development. Furthermore, expression patterns can change throughout development, with proteins serving different functions at different developmental stages, as with *DIV* in *A. majus* (Galego and Almeida 2002). With the importance of timing and levels of expression, qPCR series at different points in floral development could help elucidate the role of *DRIF-like* genes in Zingiberales floral development. The asymmetrically expressed *RAD-like* genes identified in Chapters 1 and 2 paired with the *Musa basjoo DRIF1b* sequences here are candidates for qPCR analysis. Ddditional copies of *DRIF1a* and *1b* in *Musa basjoo* and *Costus spicatus* would likewise be assayed if recovered.

Yeast 2-hybrids are necessary to test the hypothesis that the RAD/DIV/DRIF protein interaction is conserved in monocots. We expect that RAD/DRIF and DIV/DRIF complexes will form, judging from the sequence conservation observed as well as apparent maintenance of dimerization in studied species of Brassicales and Lamiales (Feller et al. 2011; Zhou et al. 2008). The *DRIF1b* sequences assayed here as well as *Musa basjoo DRIF1a* sequences should be tested alongside the *DIV-like* and asymmetrically expressed *RAD-like* genes identified in Chapters 1 and 2.

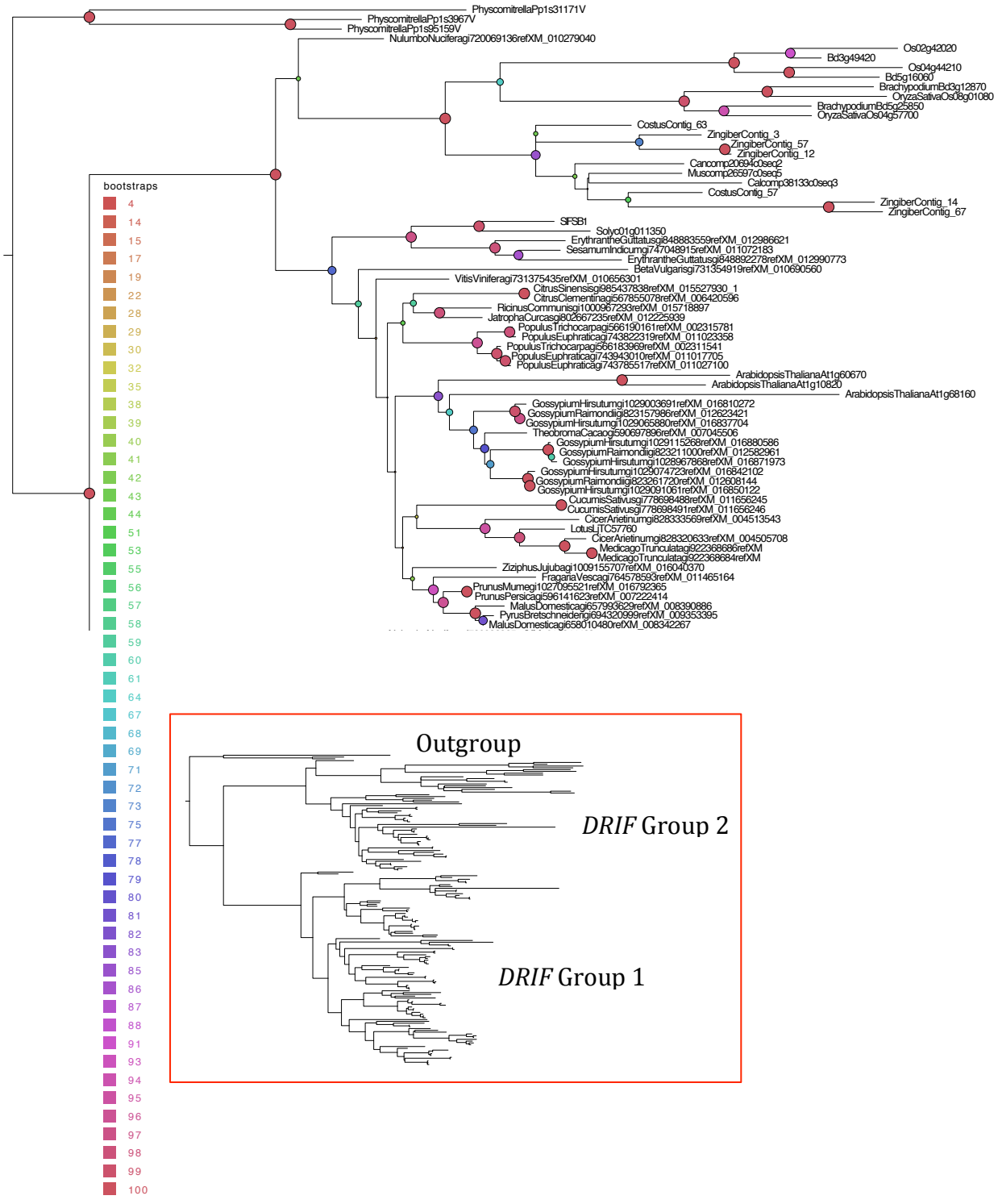


Figure 1a: RAxML tree of Group 2 *DRIF-like* nucleotide alignment with *Physcomitrella* outgroup sequences. Node symbols represent bootstrap values out of 100 with size proportional to % out of 100 and color as in legend to the left. Boxed inset represents entire *DRIF-like* phylogeny (1a and 1b) labeled with Groups. Most tips are labeled with genus, species, and accession number. Scale indicates substitutions per site. Continued in 1b.

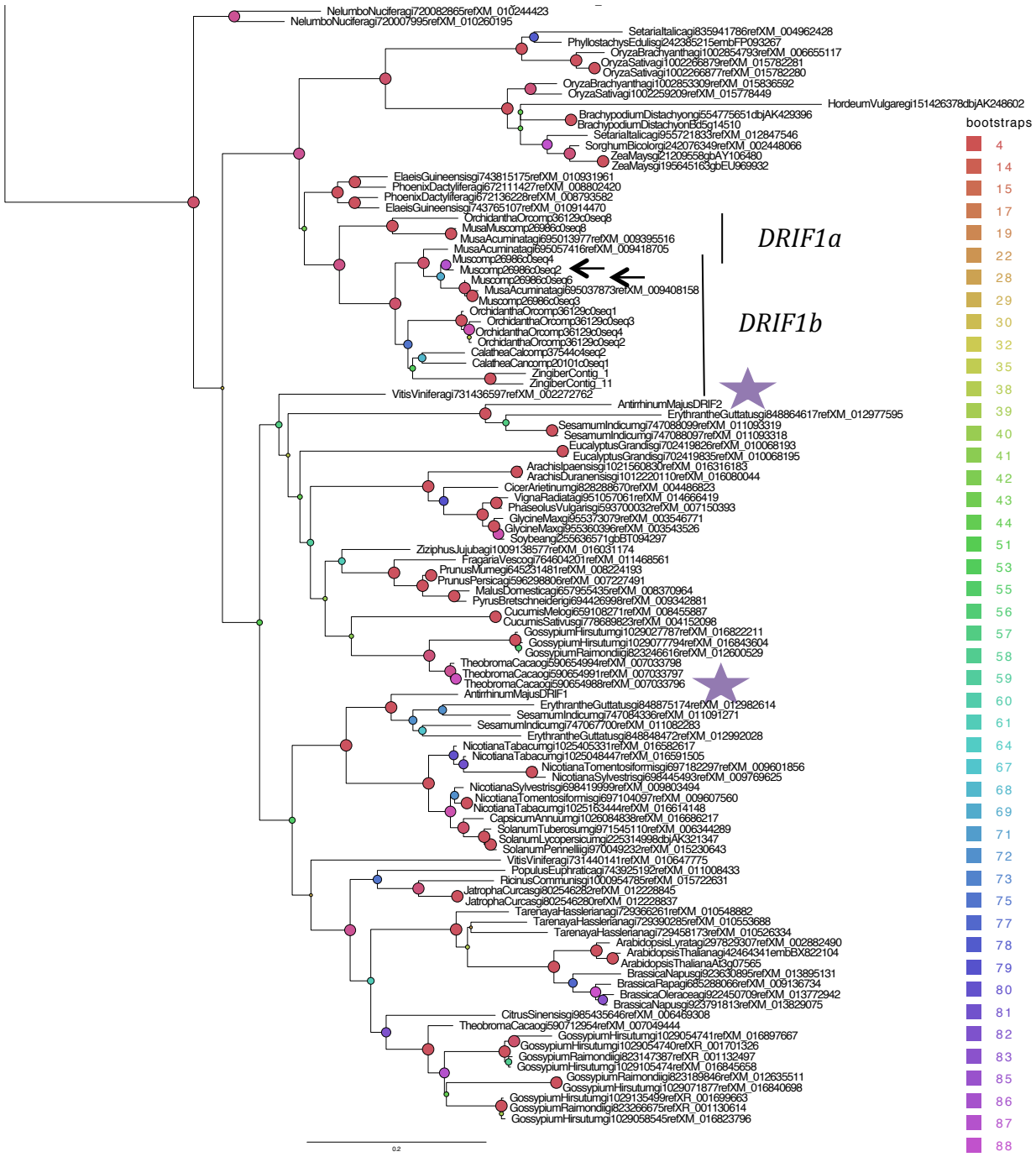
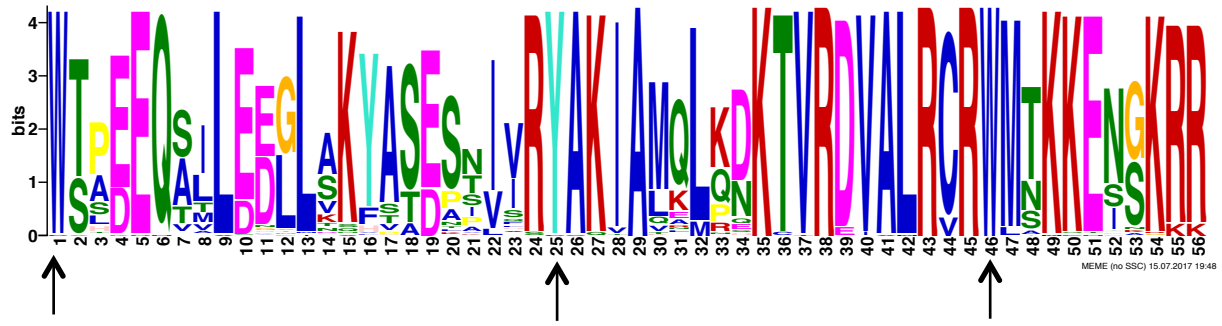


Figure 1b: RAxML tree of Group 1 *DRIF-like* nucleotide alignment. Node symbols represent bootstrap values out of 100 with size proportional to % out of 100 and color as in legend to the right. *Antirrhinum DRIF1* and *DRIF2* are labeled with purple stars. Arrows point to *Musa basjoo* sequences amplified by RT-PCR. Most tips are labeled with genus, species, and accession number. Scale indicates substitutions per site.

Gene Amplified	Primer Name	Primer Sequence (5'-3')
DRIF group 1a	DRIF_Musa1a_F	TTC CTG GGA ACC CTA GCA AT
	DRIF_Musa1a_R	ACT GCA CAG GAT GCA ACA AA
DRIF group 1b	DRIF_Musa1b_F	CCC GGA AAC CCT AGC AAT
	DRIF_Musa1b_R	TAA TCA TCG GAT GCC AGA GG

Table 1: Primers used to amplify *DRIF-like* genes in *Musa basjoo*.

A: e-value 2.6e-



B: e-value 3.9e-4884

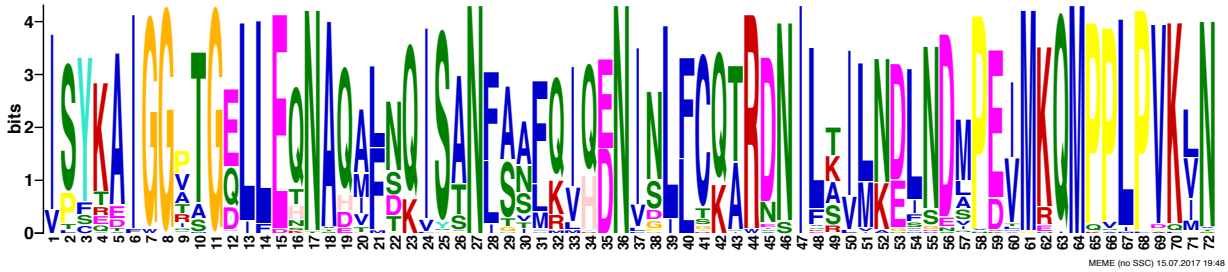


Table 2: MEME output from http://meme-suite.org/ Shown are 2 motifs discovered in DRIF Group 1 sequences, with the probability that a random sequence would contain as close a match to the motif as the sequences queried. A: MEME motif 1. Positions 1-46 contain the MYB domain. Arrows indicate the regularly spaced tryptophans (W) and tyrosines (Y) diagnostic of the MYB-like domain. B: MEME motif 2. Positions 29-72 overlap with previously recovered DUF3755.

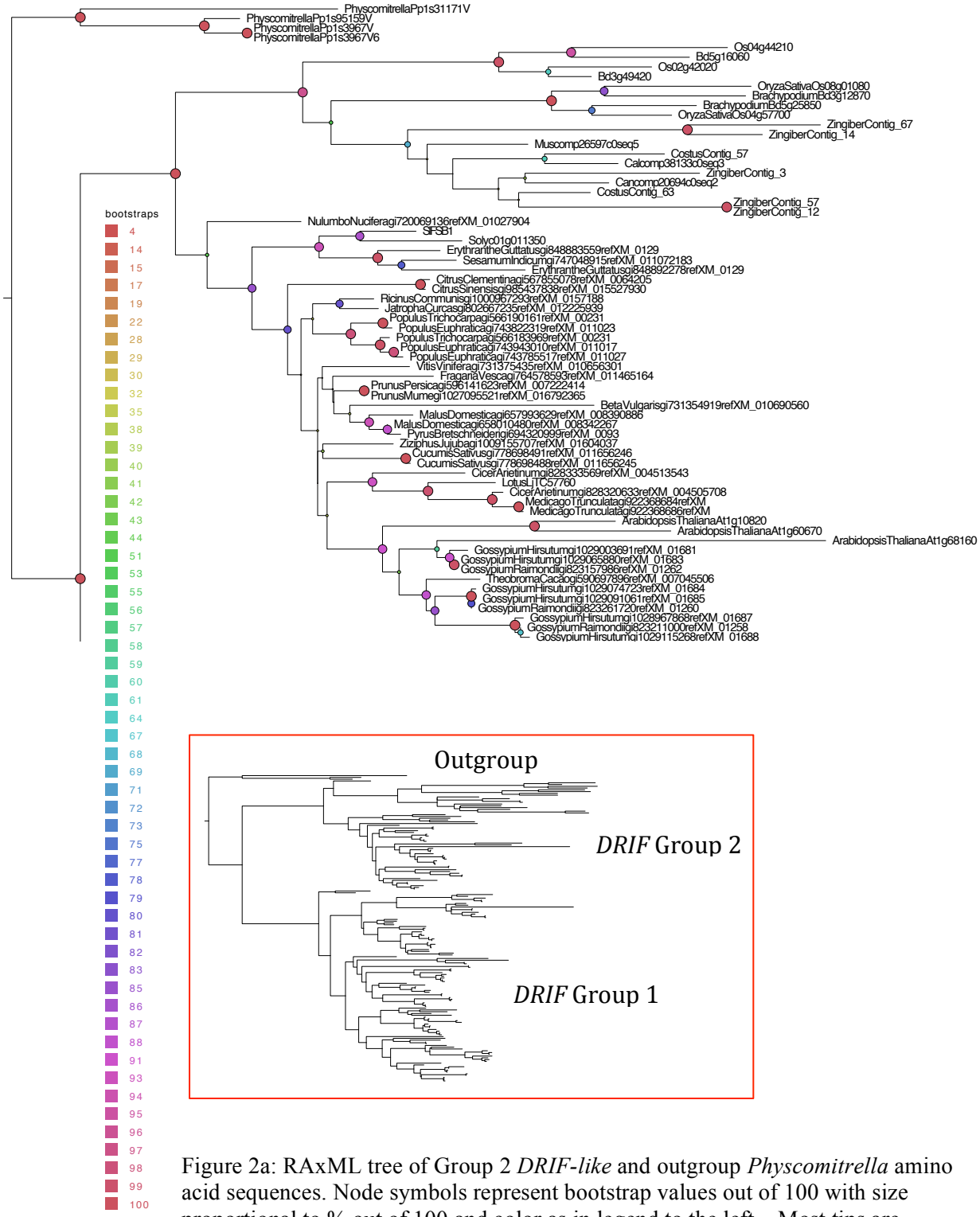


Figure 2a: RAxML tree of Group 2 *DRIF-like* and outgroup *Physcomitrella* amino acid sequences. Node symbols represent bootstrap values out of 100 with size proportional to % out of 100 and color as in legend to the left. Most tips are labeled with genus, species, and accession number. Boxed inset represents entire *DRIF-like* phylogeny with groups labeled. Continued in 2b.

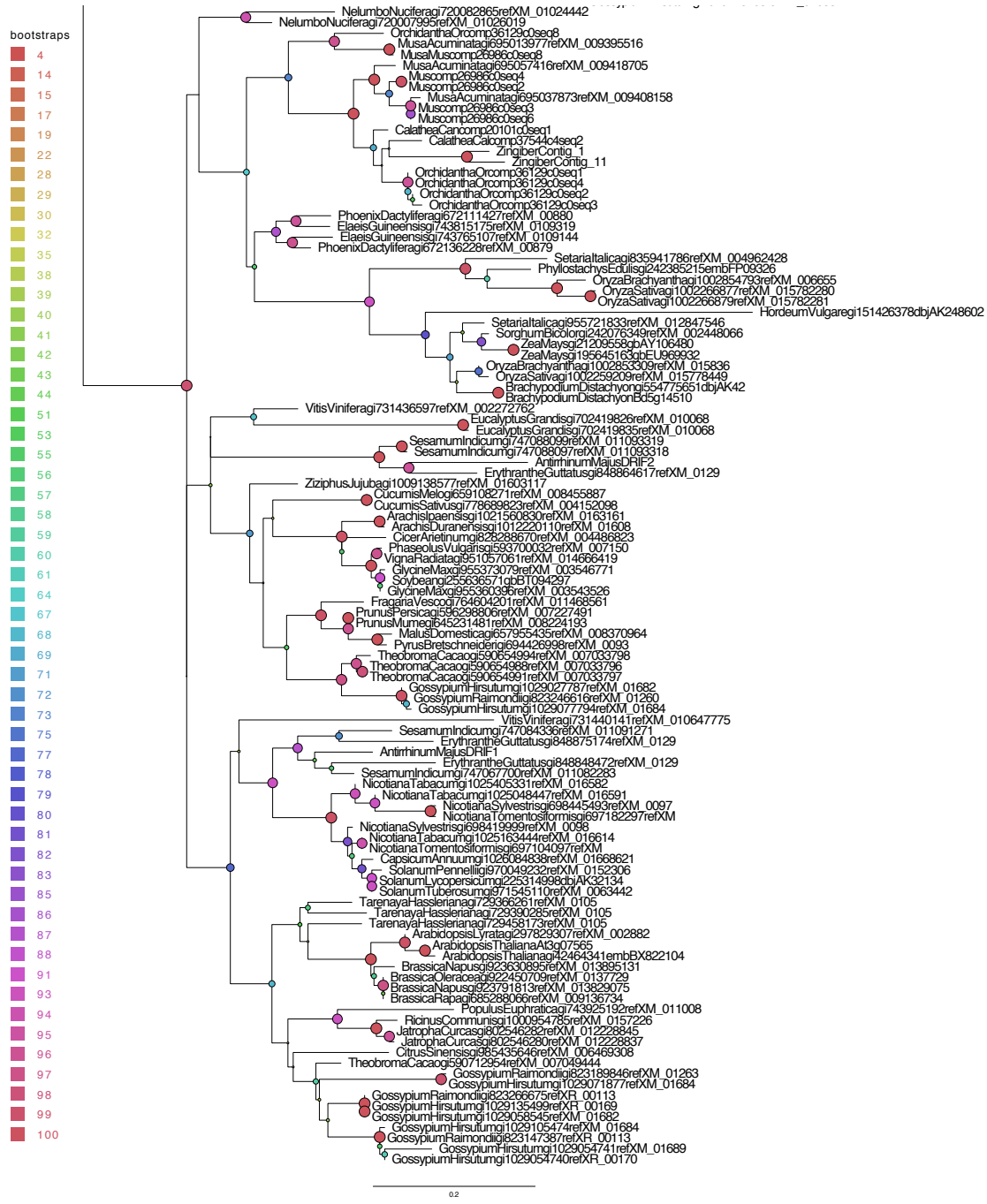


Figure 2b: RAxML tree of Group 1 *DRIF-like* amino acid alignment. Node symbols represent bootstrap values out of 100 with size proportional to % out of 100 and color as in legend to the left. Most tips are labeled with genus, species, and accession number. Scale indicates substitutions per site.

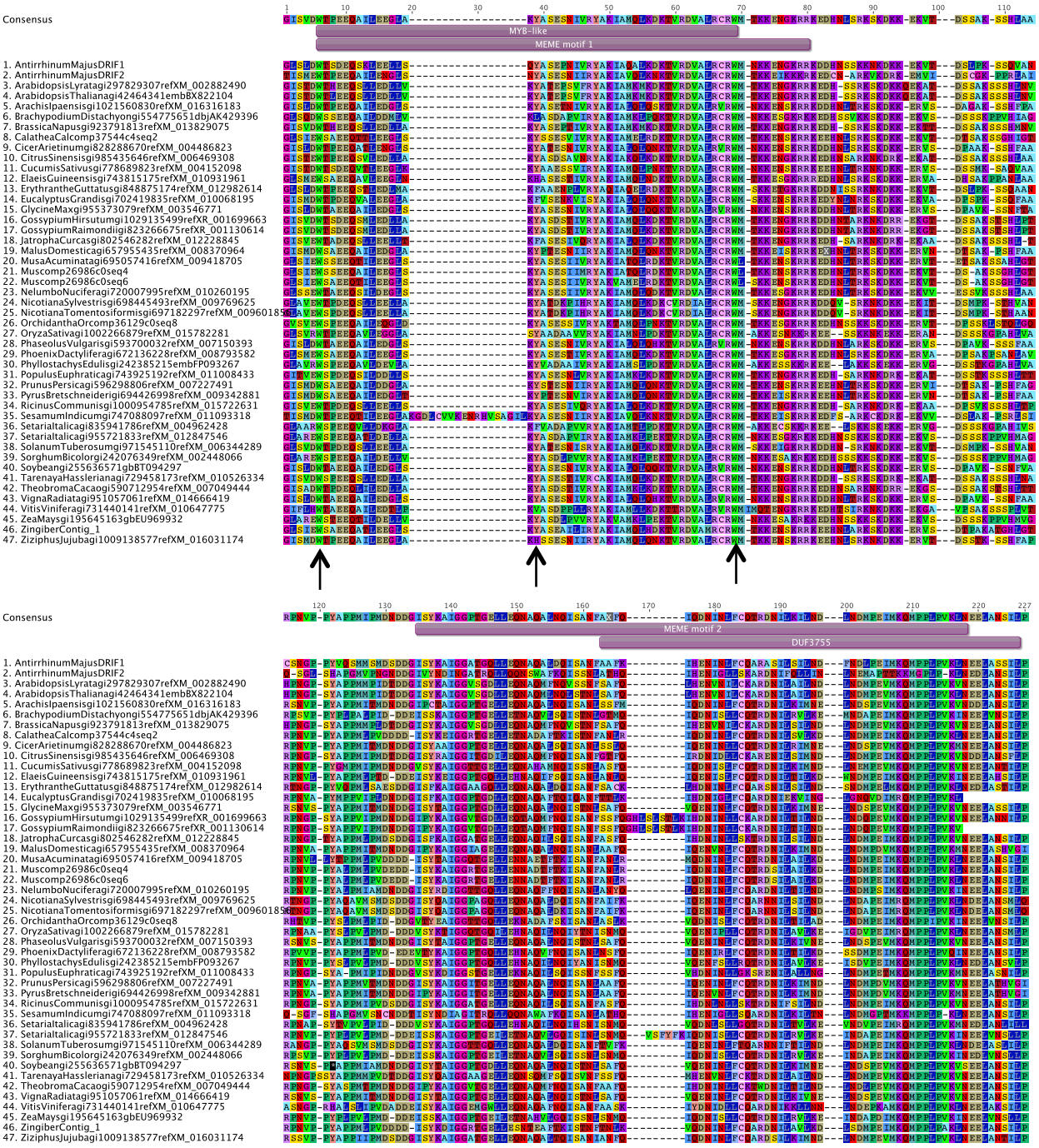


Figure 3: Amino acid alignment annotated with motifs found in MEME (Bailey and Elkan 1994) and domains found previously (Raimundo et al. 2013). Taxa included are a representative subset of group 1 *DRIF-like* genes. Arrows show the regularly spaced tryptophans and tyrosine of the MYB-like domain.

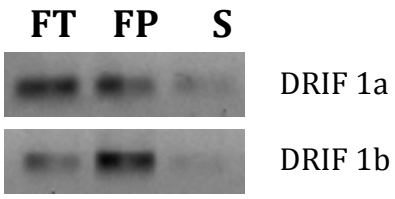


Figure 4: RT-PCR amplification of *DRIF-like* genes in *Musa basjoo* floral tissue. FT = floral tube, FP = free petal, S = stamens.

Chapter 4: Interactive Tools for Teaching Evolution (ArborEd)

Arbor (available at http://www.arborworkflows.com/) is an online program researchers can use to perform evolutionary analyses using previously generated phylogenies and character data. Arbor and services like it are designed to empower researchers to use evolutionary methods easily without independently finding and learning to use suites of programs to analyze their data. Arbor includes tutorials, and ready-to-use "Arbor Apps" with preset common analyses. Users can also develop their own save-able and share-able workflows by combining existing analysis tools, and/or coding their own.

Members of the Arbor team are currently developing ArborEd – an instance of Arbor to be used by college and high school students to get hands on experience with evolutionary analysis. Hands-on experience with science not only contributes to achievement in science (Stohr-hunt 1996), but interactive successful experiences with science can help students identify with the field and pursue higher education (Aschbacher, Li, and Roth 2010). With advice from Anastasia Thanukos, the principal editor of the University of California at Berkeley Understanding Evolution site http://evolution.berkeley.edu/, and Yael Wyner, an Assistant Professor at the City College of New York, I developed a tutorial to be used by high school and community college students to understand how researchers study the coevolution of organisms. I hope to further develop this module in the future to include student-led analysis of *HIV/SIV* coevolution with human and simian hosts.

What follows is a tutorial on coevolution using data from sharpshooters and their gut microbes, *Baumannia* and *Sulcia.*

Welcome to the coevolution module of ArborEd!

Coevolution is what happens when two or more species are living in the same place at the same time, and interactions between them influence the way they change over time. Species might be interacting in a mutually beneficial (**symbiotic**) way like plants and their pollinators, or a competitive way like predator and prey or host and parasite. In each of these types of relationships, changes in one species impact the other. For example, hawk moths (like in the image on the left) are moths with extremely long tongues they use to drink nectar from the deep nectar spurs of some flowers. The hawk moths with the longest tongues can drink nectar from even the deepest nectar spurs, so they tend to get the most food, survive, and reproduce more than shorter tongued hawk moths. Meanwhile, the flowers need the bodies of the moths to touch their pollen to transfer that pollen to other flowers and complete pollination. The deepest nectar spurs cause moths to get the closest in order to drink their nectar, so those flowers reproduce the most. Through natural selection, moths evolve longer and longer tongues to reach nectar, and plants evolve deeper and deeper nectar spurs so moths will get closer to their pollen.

By IronChris - Wikipedia. CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=1008123

In this way, the hawk moths and the plants they pollinate mutually impact the evolution of the other species (Haber and Frankie 1989).

How can you tell when species are coevolving?

Remember that coevolution is when two or more groups of organisms change over time (evolve) in response to one another. Researchers test for signatures of coevolution by comparing changes in two groups (like tongue length and nectar spur length) in different geographic locations and/or over time. Correlations between these changes supports coevolution, while lack of correlation refutes coevolution. Researchers might also test whether the two groups have **co-speciated**, meaning that when new species arose in one group, the same thing happened in the other group of organisms. Co-speciation is common in coevolving groups. Can you think of other ways to test a coevolutionary hypothesis?

By the end of this tutorial you will have some answers to the questions we have asked. You may even have lots of new questions to answer on your own in the future.

In this tutorial, you will learn about a mutualistic relationship between a group of insects and their symbiotic gut microbes.

By the end of this ArborEd module, you should be able to:

1. Define coevolution
2. Form hypotheses to explain inter-species relationships
3. Formulate and carry out a plan to test these hypotheses
4. Interpret your results / evaluate whether they support or refute your hypotheses

We will start with a published case study of sharpshooters and their gut microbes.

The glassy-winged sharpshooter, *Homalodisca coagulata,* is a dark brown winged insect about ½ an inch long that feeds on xylem, the water and nutrient transfer system in plants (Varela et. al. 2007). Sharpshooters are dangerous agricultural pests. Similar to how mosquitoes can transmit malaria to their human victims, sharpshooters can transmit diseases fatal to the plants they feed on (Varela et. al. 2007). In an effort to learn how to fight against sharpshooters, a group of researchers discovered a fascinating 3-way symbiosis. The group, led by Nancy Moran at the University of Arizona, wanted to find out how sharpshooters were using xylem as a food source. Xylem is very similar to water, and is low on the materials that insects usually need to survive. Moran's group hypothesized that the microbes living inside the guts of *Homalodisca coagulata* provided the insects with vitamins, cofactors and essential amino acids to make xylem a viable food source. When organisms live inside others and both benefit from the relationship, it is called **endosymbiosis**. Jonathan Eisen at the University of California at Davis helped Moran analyze the DNA of all the bacteria living inside sharpshooters. They discovered a common insect microbe, *Baumannia cicadellinicola* (Jensen 2006). Relatives of *Baumannia cicadellinicola* are endosymbionts of many other insects. Sharpshooters need amino acids, vitamins, and cofactors to survive. The researchers found that *Baumannia* was producing vitamins and cofactors for the sharpshooter, but was not producing essential amino acids. Puzzled, the researchers returned to the DNA data and found evidence for another organism, *Sulcia mulleri.* Like *Baumannia,* relatives of *Sulcia mulleri* are found in other insects as well. The researchers found that in sharpshooters, *Sulcia* was making the amino acids *Baumannia* was missing (Jensen 2006). The researchers even found some evidence suggesting that *Sulcia* and *Baumannia* were trading materials with each other! The sharpshooters, *Sulcia,* and *Baumannia* are all necessary for the others' survival, and form a 3-way mutualistic relationship (Wu et al. 2006).

Is this an example of coevolution?

Because this is a tutorial, we will provide a testable hypothesis and walk through how we test it. One hypothesis might be:

1. *Baumannia* and *Sulcia* are coevolving with sharpshooters and with each other. This means that:
    a. The bacteria have changed over time in response to sharpshooters and each other
    b. Sharpshooters have changed over time in response to the bacteria

How would you test this hypothesis? What data would you need?

Let's start with point a. Bacteria have changed in response to sharpshooters. To test this point, we might seek data to answer the following questions:

1. Do *Baumannia* and *Sulcia* species living in sharpshooters have different metabolic processes than *Baumannia* and *Sulcia* species living outside of sharpshooters?
    a. If endosymbiotic *Baumannia* and *Sulcia* have different metabolic processes, this supports the coevolutionary hypothesis, as these processes may have evolved in response to life inside Sharpshooters.
    b. If endosymbiotic *Baumannia* and *Sulcia* do not have different metabolic processes than relatives outside of sharpshooters, then these processes were likely not an adaptation to life inside of Sharpshooters and instead already existed prior to colonizing the insects. This would not support the coevolutionary hypothesis.
2. Are *Baumannia* and *Sulcia* species living in sharpshooters more closely related to other *Baumannia* and *Sulcia* species in sharpshooters than to those living outside of sharpshooters?
    a. If the closest relatives also live in sharpshooters, these bacteria may have diverged long ago from ancestors that do not colonize sharpshooters. This would support our coevolutionary hypothesis, as these endosymbionts would have co-existed with sharpshooters for long enough to form new species.
    b. If the closest relatives are species that live outside of sharpshooters, *Baumannia* and *Sulcia* endosymbionts may have colonized sharpshooters relatively recently or may have jumped into and out of sharpshooters many times, which would not support our coevolutionary hypothesis.

Next, let's look at point b. Sharpshooters have changed in response to bacteria. To test this point, we might look for data to answer these questions:

1. Do sharpshooters containing *Baumannia* and *Sulcia* use these bacteria to do something the insects cannot do without them?
    a. If so, these new abilities may have evolved due to the presence of the bacteria, supporting the coevolutionary hypothesis.
    b. If not, there would not be evidence that the sharpshooters have gained new abilities in response to the bacteria, refuting the coevolutionary hypothesis.
2. Are sharpshooters containing *Baumannia* and *Sulcia* are more closely related to other sharpshooters with *Baumannia* and *Sulcia* bacteria than to relatives without these bacteria?
    a. If so, the sharpshooters have diverged from insects without the bacteria and have co-existed with the bacteria long enough to form new species, supporting the coevolutionary hypothesis.
    b. If not, the sharpshooters may have recently gained the bacteria, which would not support the coevolutionary hypothesis.

A third question we might pursue is whether *Baumannia, Sulcia,* and sharpshooters have co-speciated with one another. If so, the organisms have lived together for a long time (long enough to influence one another's evolution), and the bacteria are not jumping into and out of sharpshooters or between groups of sharpshooters. Cospeciation, though it cannot alone prove coevolution, shows that two groups have been closely associated for a long time.

Let's start by looking at the **phylogenies** of the three organisms in Figure 1. A phylogeny, or phylogenetic tree, is a diagram that represents the evolutionary relationships between organisms (The Understanding Evolution Team). If you follow the branches (lines) towards the left from a species label until you find a cross-roads (node), the other branches to species labels coming off of that node to the right represent the closest relatives of the species you started with.

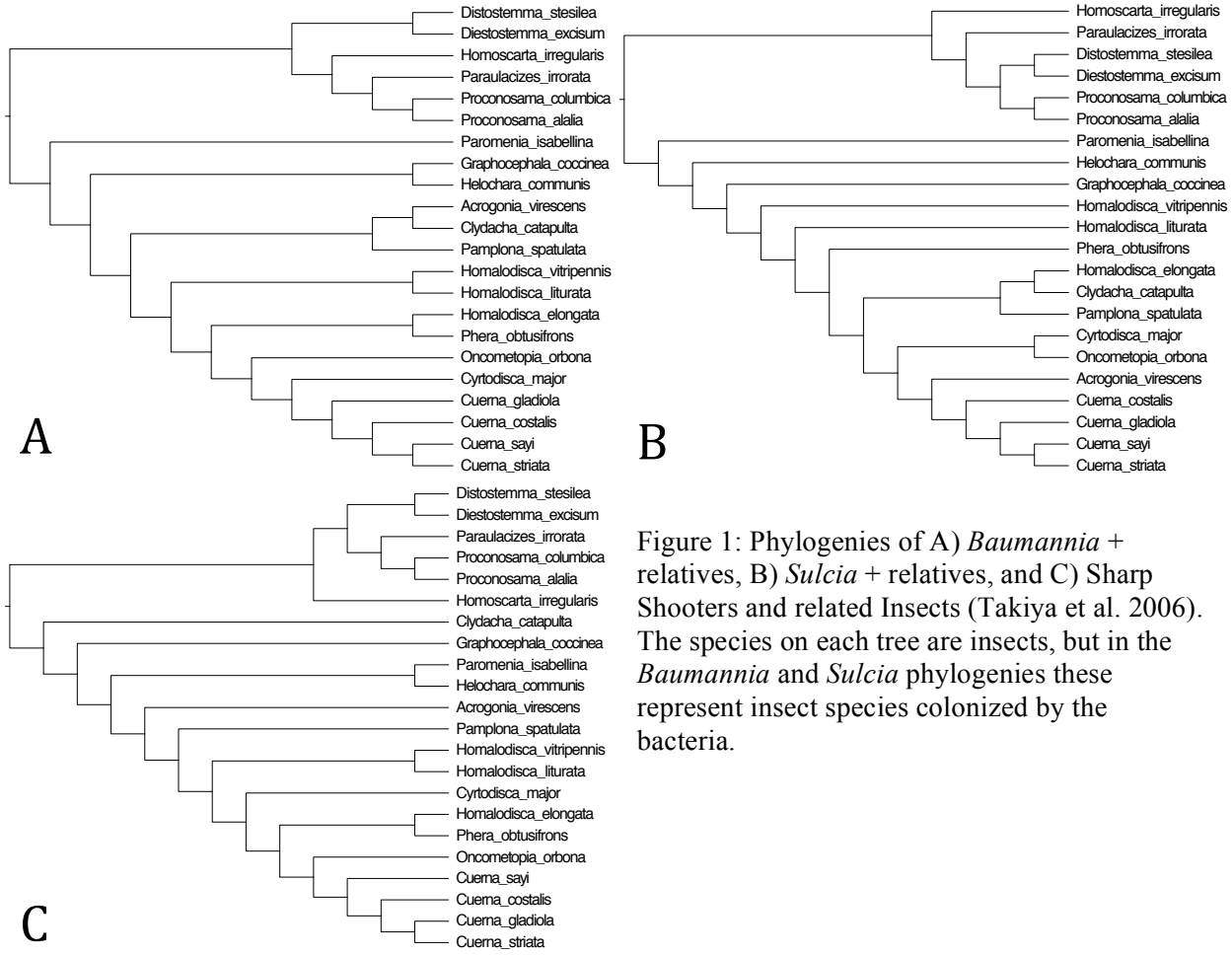


Figure 1: Phylogenies of A) *Baumannia* + relatives, B) *Sulcia* + relatives, and C) Sharp Shooters and related Insects (Takiya et al. 2006). The species on each tree are insects, but in the *Baumannia* and *Sulcia* phylogenies these represent insect species colonized by the bacteria.

For example, in the Insect tree (Figure 1C) *Diestemma stesilea* (on top) is most closely related to *Diestemma excisum*. The group (clade) of *D. stesilea* + *D. excisum* is most closely related to the clade *Paraulacizes irrorata* + *Proconosama columbica* + *Proconosama alalia*.

If you want more information on how to read phylogenies, visit the Understanding Evolution site at evolution.berkeley.edu.

The *Baumannia* and *Sulcia* trees are labeled with the insect each bacterial sample was collected from. This means that we can directly compare the three trees to find similarities and differences. If every time the insects speciated the bacteria did too, the three trees would look identical. If there are differences, it could mean that the bacteria jumped from one species of insect to

another, that the insects or bacteria speciated independently, that the bacteria failed to continue living in the descendants of a group of insects, or that the bacteria living in a group of insects died out. Compare the relationships presented in the three trees. Can you find a relationship that is the same? What about one that is different?

It can be hard to find all the similarities and differences between trees by eye. Researchers often use **tanglegrams** to see these differences better. Tanglegrams are a way to visualize the congruence, or similarities, of phylogenies. To make a tanglegram, we need two phylogenies with the same taxa. We can use ArborEd to line up the same taxa in the two trees. This helps us compare them. An example is in Figure 2.

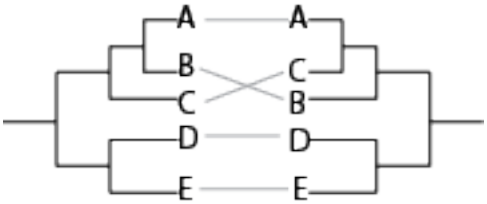


Figure 2: Example tanglegram. Grey lines in the center of the diagram indicate the same taxa.

This tanglegram helps us see that taxa A, B, and C have different relationships in the two trees. We can tell because the lines connecting B and C cross, illuminating a difference between the phylogenies. In the phylogeny on the left, A and B are most closely related. In the phylogeny on the right, it is A and C that are most closely related.

To generate tanglegrams on ArborEd using our data:

1. Click the “Analysis” tab.
2. In the dropdown menu for “Select Analysis,” choose “tanglegram.”
3. Click “Set up and run.”
4. For the data, load 2 trees. Start with “SapSuckersAndRelatedInsects.tre” and “BaumanniaAndRelatives.tre”
5. For the output, name the file you will produce. Something like “SapSuckerBaumanniaTanglegram” would be informative.
6. Click “Run.”

To see the results:

1. Click the “Visualization” tab.
2. In the dropdown menu, select the file you generated.
3. Click “Update.

You should now be able to see your tanglegram! If you wish to save it, click “Download.” Repeat these steps until you have tanglegrams of the 3 combinations of trees.

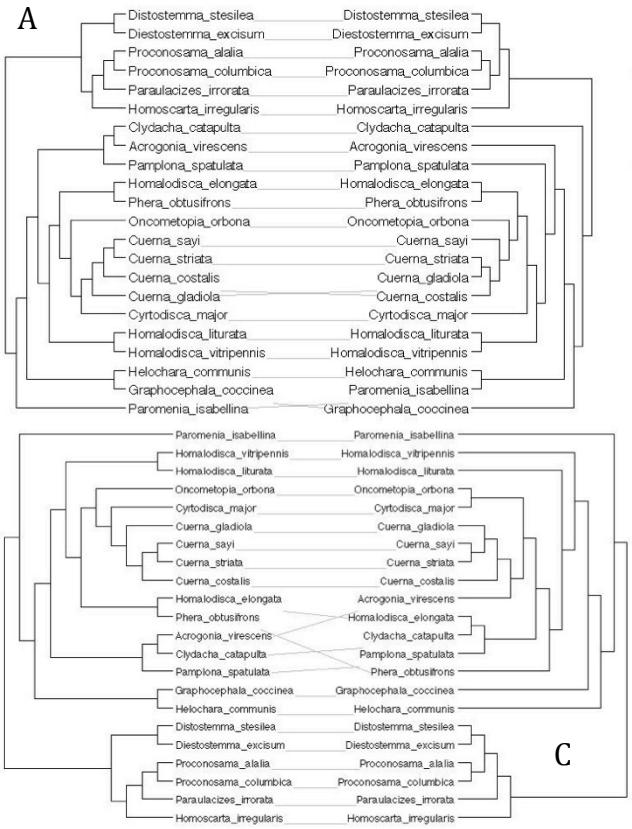


Figure 3: Tanglegrams of relevant species. Taxon names shown are those of insects. For the bacterial phylogenies, insect names represent the names of the hosts of bacteria. A: *Baumannia* + relatives vs. Insects B: Insects vs. *Sulcia* + relatives C: *Baumannia* + relatives vs. *Sulcia* + relatives. Phylogenies included (Takiya et al. 2006).

Figure 3 contains tanglegrams of the three pairs of phylogenies. The lines through the middle mark the corresponding species in the two trees. Lots of horizontal lines means that trees are very congruent, or share many of the same relationships. Diagonal lines indicate a difference between the two trees. Find the relationships you identified when looking at Figure 1. The tanglegrams should help make similarities and differences easier to spot.

By observing these tanglegrams, we can see that the phylogenies of *Sulcia, Baumannia,* and sharpshooters match very closely. Notice all of the horizontal lines and the small number of diagonal ones. Figure 3A, showing *Baumannia* + relatives compared with sharpshooters + relatives has the fewest diagonal lines, meaning that those groups have likely cospeciated the closest with the fewest jumps of bacteria between different sharpshooters.

As a comparison, Figure 4 shows a tanglegram made between two randomly generated trees of our taxa. You can see that horizontal lines are very rare, and there are many diagonal ones.

Let's interpret our results. Although the phylogenies do not match perfectly, they are highly congruent. This is an indication of cospeciation with some deviations, likely from host switching. These data indicate that the bacteria and insects have been living together closely and consistently over time. This pattern of cospeciation is consistent with coevolution.

What does it mean for endosymbionts to cospeciate with their hosts? This means that when host insects diverged into two species, their endosymbiotic bacteria did as well. Speciation is often caused by separation, which eventually leads to loss of gene flow (or loss of reproduction

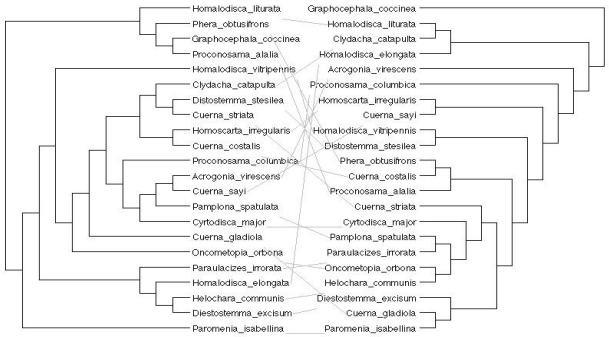


Figure 4: A comparison. These are two randomly generated trees using the taxa from Figure 3.

between separated groups). Loss of gene flow means that the separated groups may eventually accumulate enough differences to be considered two species. When the ancestors of the sapsucker species separated, their bacterial endosymbionts were separated as well, also losing gene flow. In this way, cospeciation may have been a byproduct of the close association of these organisms, rather than a result of coevolution. In order to really test our hypothesis of coevolution, we need to find out whether the bacteria and the insects mutually influenced each other's evolution. Let's pursue the other portions of our hypothesis.

How can we test whether the endosymbiotic bacteria and insects have influenced each other's evolution? We can investigate whether either group has evolved differences in function.

In order to investigate point 1, that the bacteria and sharpshooters have evolved functions that do not exist in ancestors without the symbiosis, let's look at the evolution of xylem-feeding in sharpshooters. Do sharpshooters need both bacteria to feed on xylem, or do other sharpshooters without these endosymbionts also feed on xylem?

We have a table of insects, whether they have bacterial endosymbionts, and whether they feed on xylem. Let's map each of these characters onto our tree and perform an ancestral state reconstruction. Ancestral state reconstruction estimates the ancient character states of ancestors in a phylogenetic tree. To perform the analysis, follow these instructions:

To perform an ancestral state reconstruction on ArborEd:

7. Click the "Analysis" tab.
8. In the dropdown menu for "Select Analysis," choose "ancestral state reconstruction."
9. Click "Set up and run."
10. For the data, load the sharpshooter tree and the character matrix.
11. In the drop-down menu for "column" choose one of the characters. Start with "*Baumannia* endosymbiont."
12. For the output, name the file you will produce. Something like "Ancestral state reconstruction Baumannia" would be informative.
13. Click "Run."

To see the results:

4. Click the "Visualization" tab.
5. In the dropdown menu, select the file you generated.
6. Click "Update."

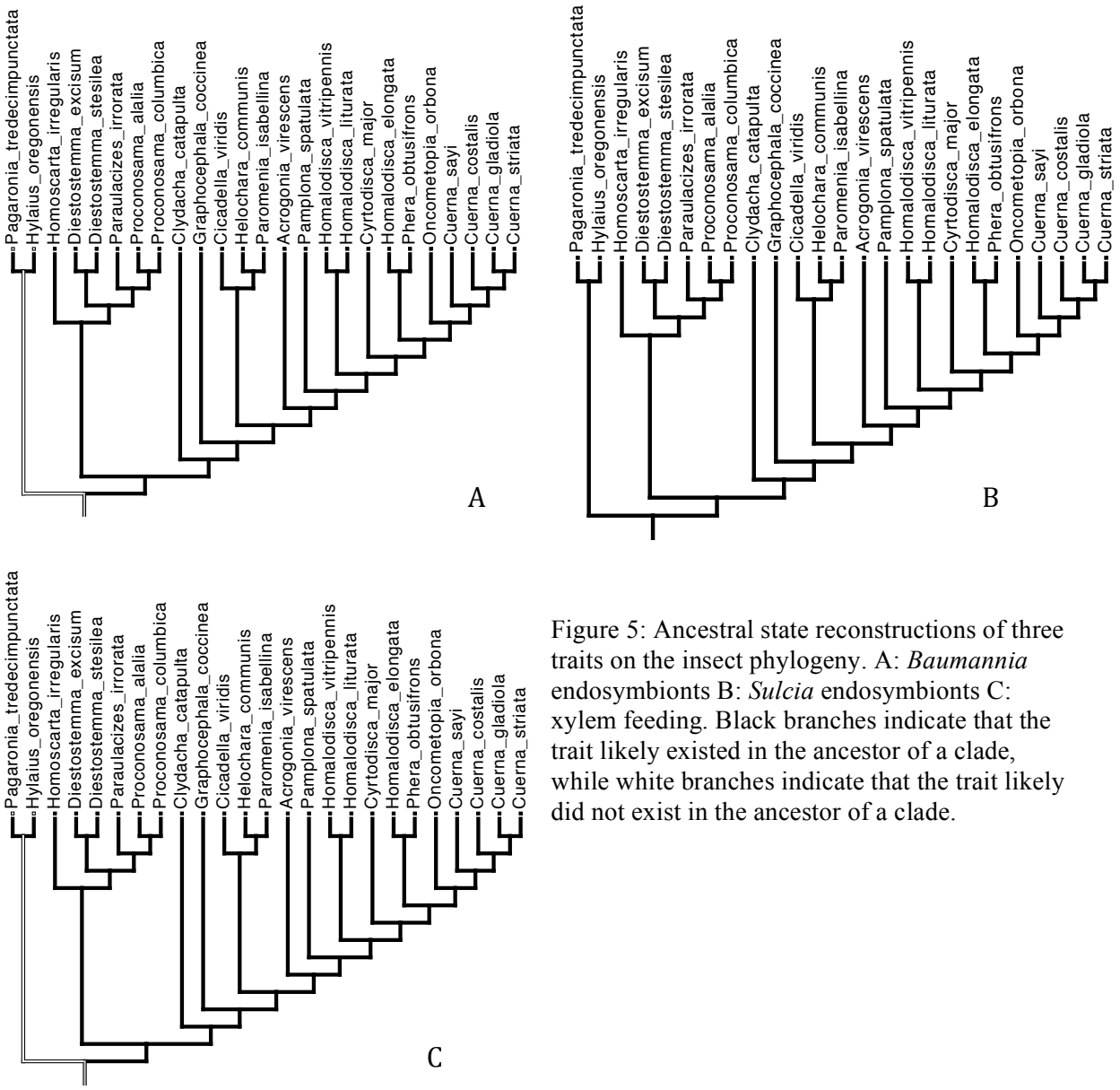


Figure 5: Ancestral state reconstructions of three traits on the insect phylogeny. A: *Baumannia* endosymbionts B: *Sulcia* endosymbionts C: xylem feeding. Black branches indicate that the trait likely existed in the ancestor of a clade, while white branches indicate that the trait likely did not exist in the ancestor of a clade.

Figure 5 contains our ancestral state reconstructions for *Baumannia* endosymbionts, *Sulcia* endosymbionts, and xylem feeding.

Compare Figure 5A and 5B. We can see that sharpshooters and *Sulcia* have had a relationship earlier in time (more branches are black than in the *Baumannia* reconstruction). Now compare Figure 5B and 5C. We can see that sharpshooter relatives that do not feed on xylem also have a relationship with *Sulcia* (*P. tredecimpunctata* and *H. oregonensis*). Sharpshooters have more recently developed a mutualism with *Baumannia*. The start of mutualism with *Baumannia* coincides with the sharpshooters switching to xylem as a food source. These data support our coevolutionary hypothesis in the sense that *Baumannia* and sharpshooters have diverged from ancestors without the symbiosis, and that *Baumannia* may have influenced the evolution of sharpshooters by enabling them to feed on xylem. *Sulcia,* however does not follow the same pattern and therefore the *Sulcia* data do not support a coevolutionary hypothesis.

We now have evidence pointing to co-speciation of all three groups and to *Baumannia* influencing the evolution of sharpshooters. We have evidence that refutes the hypothesis that *Sulcia* influenced the evolution of sharpshooters. How can we test whether sharpshooters have influenced the evolution of *Baumannia* and *Sulcia*?

In order to directly test our hypothesis, we would need data showing metabolic processes of *Baumannia, Sulcia,* and sharpshooters compared with relatives without the symbiosis. We have data on *Baumannia* and relatives ability to produce amino acids vital to the survival of sharpshooters, but we do not have complementary studies for insects and *Sulcia*..

We have a phylogeny of *Baumannia* + relatives and a data matrix of association with xylem-feeding sharpshooters and ability to produce amino acids vital to the survival of insects. Perform an ancestral state reconstruction of these two characters, following the instructions used to produce figure 5.

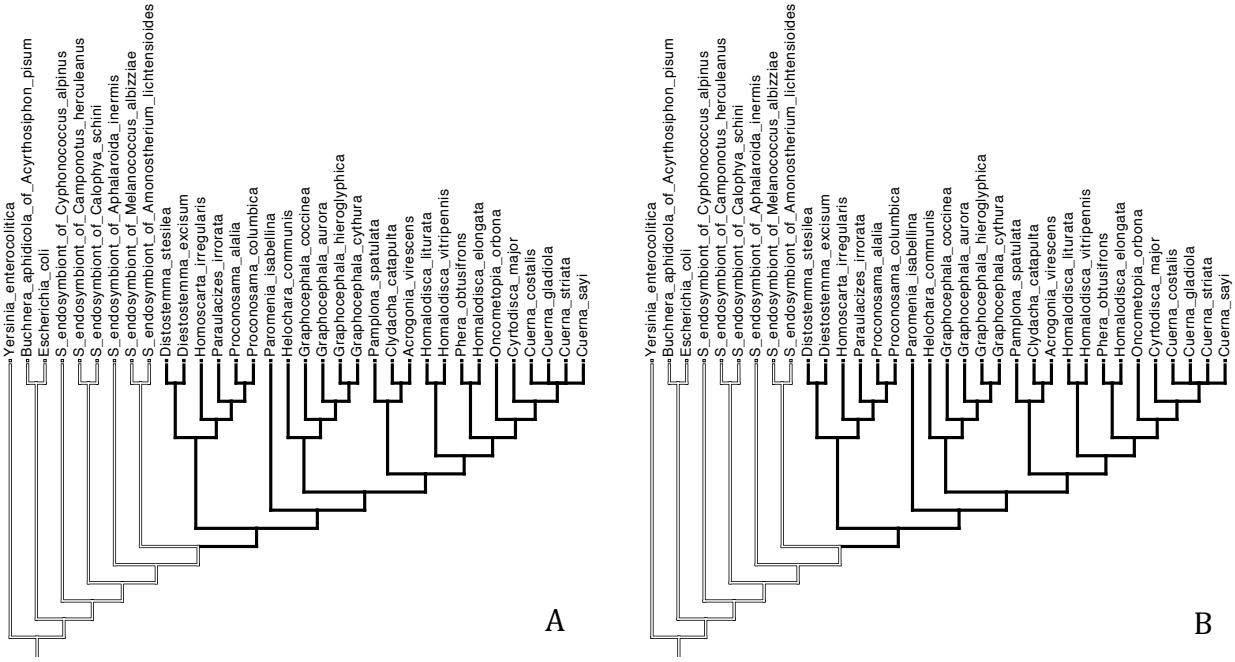


Figure 6: Ancestral state reconstructions on the *Baumannia* phylogeny (Takiya et al. 2006). Unless labeled 'endosymbiont' or *Yerenia, Buchnera,* or *Escherichia,* taxon labels are those of insect hosts. A: association with xylem-feeding sharpshooters. B: loss of amino acid biosynthetic pathways.

Figure 6 contains ancestral state reconstructions of association with xylem-feeding sharpshooters (A) and loss of amino acid biosynthesis capability (B). Notice that association with xylem-feeding sapsuckers coincides with the loss of ability to produce vital amino acids.

Although we have a comprehensive phylogeny of *Sulcia* + relatives, we do not have data on the metabolic abilities of *Sulcia* relatives not associated with sharpshooters and *Baumannia*. Instead we can compare the capabilities of the host and the two bacteria (Table 2).

Produced by *Sulcia muelleri*	Produced by *Baumannia cicadellinicola*
fabF	fatty acids (except for fabF)
amino acids (except for methionine and histidine)	methionine and histidine
menaquinone	cofactors (except for menaquinone)

Table 2: Materials required by sharpshooters to metabolize xylem sap (McCutcheon and Moran 2007).

Sulcia and *Baumannia* not only produce materials that sharpshooters require to feed off of xylem, they also supply each other with necessary products. Do these data support or refute our hypothesis that the three organisms have coevolved?

Let's bring our data together and reach a conclusion. Remember, our hypothesis was

1. *Baumannia* and *Sulcia* are coevolving with sharpshooters. This means that:
    a. The bacteria have changed over time in response to sharpshooters
    b. Sharpshooters have changed over time in response to the bacteria
    c. cospeciation of the three organisms

Has *Baumannia* changed in response to sharpshooters?

- Figure 6 shows that *Baumannia* associated with sharpshooters and *Sulcia* have lost the ability to make most amino acids compared with relatives without the associations.

Has *Sulcia* changed in response to sharpshooters?

- We do not have any data to support or refute this point.

Has *Baumannia* changed in response to *Sulcia*?

- Figure 6 and Table 2 show that *Baumannia* have lost the ability to produce vital amino acids, and that *Sulcia* complements this loss. Table 2 further shows that *Baumannia* rely on *Sulcia* for a cofactor and a fatty acid.

Has *Sulcia* changed in response to *Baumannia*?

- Although we lack ancestral state reconstructions, Table 2 shows that *Sulcia* is dependent on *Baumannia* for an amino acid, fatty acids, and cofactors.

Have sharpshooters changed in response to *Baumannia*?

- Figure 5 shows that only sharpshooters associated with *Baumannia* feed on xylem. This supports the hypothesis that the association with *Baumannia* has allowed sharpshooters to perform new functions. Table 2 shows that sharpshooters are reliant on *Baumannia* for vital materials.

Have sharpshooters changed in response to *Sulcia*?

- Figure 5 shows that sharpshooter relatives that do not feed on xylem also have *Sulcia* endosymbionts. This refutes the coevolutionary hypothesis. Table 2 shows that sharpshooters are reliant on *Sulcia* for vital materials, which supports the coevolutionary hypothesis. These data are inconclusive and do not satisfy our questions.

Have sharpshooters, *Sulcia*, and *Baumannia* cospeciated?

- Figures 1 and 3 support cospeciation.

In conclusion, our analyses support the hypothesis that *Baumannia* and sharpshooters have coevolved, and we have some weaker evidence consistent with *Sulcia* coevolving with both organisms, though not enough to answer our research questions.

What other evidence could you look for to test the coevolutionary hypothesis? Are there alternative hypotheses that might explain these relationships?

Bibliography

Almeida, J, M Rocheta, and L Galego. 1997. “Genetic Control of Flower Shape in Antirrhinum Majus.” *Development (Cambridge, England)* 124 (7): 1387–92. http://www.ncbi.nlm.nih.gov/pubmed/9118809.

Altschup, Stephen F, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology* 215: 403–10. doi:10.1016/S0022-2836(05)80360-2.

Aschbacher, Pamela R, Erika Li, and Ellen J Roth. 2010. “Is Science Me? High School Students’ Identities, Participation and Aspirations in Science, Engineering, and Medicine.” *Journal of Research in Science Teaching* 47 (5). Wiley Subscription Services, Inc., A Wiley Company: 564–82. doi:10.1002/tea.20353.

Bailey, Timothy L, and Charles Elkan. 1994. “Fitting A Mixture Model By Expectation Maximization To Discover Motifs In Biopolymers.” *UCSD Technical Report*.

Bartlett, Madelaine E. 2010. “The Evolution of Floral Morphology in the Zingiberales : An Investigation into Possible Roles for the GLOBOSA -like and TEOSINTE BRANCHED 1 -like Genes.” University of California at Berkeley.

Bartlett, Madelaine E, Bruce K Kirchoff, and Chelsea D Specht. 2008. “Epi-Illumination Microscopy Coupled to in Situ Hybridization and Its Utility in the Study of Evolution and Development in Non-Model Species.” *Development Genes and Evolution* 218 (5): 273–79. doi:10.1007/s00427-008-0211-6.

Bartlett, Madelaine E, and Chelsea D Specht. 2010. “Evidence for the Involvement of Globosa-like Gene Duplications and Expression Divergence in the Evolution of Floral Morphology in the Zingiberales.” *The New Phytologist* 187 (2): 521–41. doi:10.1111/j.1469-8137.2010.03279.x.

———. 2011. “Changes in Expression Pattern of the Teosinte branched1-like Genes in the Zingiberales Provide a Mechanism for Evolutionary Shifts in Symmetry across the Order.” *American Journal of Botany* 98 (2): 227–43. doi:10.3732/ajb.1000246.

Boyden, Geraldine S, Michael J Donoghue, and Dianella G Howarth. 2014. “Duplications and Expression of RADIALIS -Like Genes in Dipsacales.” *International Journal of Plant Sciences* 173 (9): 971–83. doi:10.1086/667626.

Busch, Andrea, Stefanie Horn, Andreas Mühlhausen, Klaus Mummenhoff, and Sabine Zachgo. 2012. “Corolla Monosymmetry: Evolution of a Morphological Novelty in the Brassicaceae Family.” *Molecular Biology and Evolution* 29 (4): 1241–54. doi:10.1093/molbev/msr297.

Chapman, Mark a, James H Leebens-Mack, and John M Burke. 2008. “Positive Selection and Expression Divergence Following Gene Duplication in the Sunflower CYCLOIDEA Gene Family.” *Molecular Biology and Evolution* 25 (7): 1260–73. doi:10.1093/molbev/msn001.

Charlton, W A, A D Macdonald, U Posluszny, and C P Wilkins. 1989. “Additions to the Technique of Epi-Illumination Light Microscopy for the Study of Floral and Vegetative Apices.” *Canadian Journal of Botany* 67: 1739–43.

Citerne, Helene, Florian Jabbour, Sophie Nadot, and Catherine Damerval. 2010. “The Evolution of Floral Symmetry.” *Advances in Botanical Research* 54 (10): 85–137. doi:10.1016/S0065-2296(10)54003-5.

Clark, Jennifer I, and Enrico S Coen. 2002. “The Cycloidea Gene Can Respond to a Common Dorsoventral Prepattern in Antirrhinum.” *The Plant Journal : For Cell and Molecular Biology* 30 (6): 639–48. http://www.ncbi.nlm.nih.gov/pubmed/12061896.

Corley, Susie B, Rosemary Carpenter, Lucy Copsey, and Enrico Coen. 2005. "Floral Asymmetry Involves an Interplay between TCP and MYB Transcription Factors in Antirrhinum." *Proceedings of the National Academy of Sciences of the United States of America* 102 (14): 5068–73. doi:10.1073/pnas.0501340102.

Costa, Maria Manuela R, Samantha Fox, Andy I Hanna, Catherine Baxter, and Enrico Coen. 2005. "Evolution of Regulatory Interactions Controlling Floral Asymmetry." *Development (Cambridge, England)* 132 (22): 5093–5101. doi:10.1242/dev.02085.

Cubas, Pilar. 2004. "Floral Zygomorphy, the Recurring Evolution of a Successful Trait." *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* 26 (11): 1175–84. doi:10.1002/bies.20119.

Cui, Min-long, Lucy Copsey, Amelia A Green, J Andrew Bangham, and Enrico Coen. 2010. "Quantitative Control of Organ Shape by Combinatorial Gene Activity." *PloS Biology* 8 (11). doi:10.1371/journal.pbio.1000538.

Damerval, Catherine, Martine Le Guilloux, Muriel Jager, and Céline Charon. 2007. "Diversity and Evolution of CYCLOIDEA-like TCP Genes in Relation to Flower Development in Papaveraceae." *Plant Physiology* 143 (2): 759–72. doi:10.1104/pp.106.090324.

Droc, Gaetan, Delphine Lariviere, Valentin Guignon, Nabila Yahiaoui, Dominique This, Olivier Garsmeur, Alexis Dereeper, et al. 2013. "The Banana Genome Hub." *Database*, 1–14. doi:10.1093/data-Citation.

Du, Hai, Yongin Wang, Yi Xie, Zhe Liang, Sanie Jiang, S Huang Huang Zhang, and Yui Huang. 2013. "Genome-Wide Identification and Evolutionary and Expression Analyses of MYB-Related Genes in Land Plants." *DNA Research*, 1–12.

Du, Hai, Li Zhang, Lei Liu, Xiao-Feng Tang, Wen-Jie Yang, Yan-Min Wu, Yu-Bi Huang, and Yi-Xiong Tang. 2009. "Biochemical and Molecular Characterization of Plant MYB Transcription Factor Family." *Biochemistry. Biokhimiia* 74 (1): 1–11. doi:10.1134/S0006297909010015.

Dubos, Christian. 2013. "MYB Transcription Factors in Plants." *AccessScience*. doi:10.1016/S0168-9525(96)10049-4.

Edgar, Robert C. 2004. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research* 32 (5): 1792–97. doi:10.1093/nar/gkh340.

Endress, P K. 2001. "Evolution of Floral Symmetry." *Current Opinion in Plant Biology* 4 (1): 86–91. http://www.ncbi.nlm.nih.gov/pubmed/11163173.

Endress, Peter K. 1999a. "Symmetry in Flowers : Diversity and Evolution." *International Journal of Plant Sciences* 160 (S6): S3–23.

———. 1999b. "Symmetry in Flowers : Diversity and Evolution." *International Journal of Plant Sciences* 160 (S6): S3–23.

Endress, Peter K. 2012. "The Immense Diversity of Floral Monosymmetry and Asymmetry Across Angiosperms." *The Botanical Review* 78 (4): 345–97. doi:10.1007/s12229-012-9106-3.

Endress, Peter K, and James a Doyle. 2009. "Reconstructing the Ancestral Angiosperm Flower and Its Initial Specializations." *American Journal of Botany* 96 (1): 22–66. doi:10.3732/ajb.0800047.

Feller, Antje, Katja Machemer, Edward L Braun, and Erich Grotewold. 2011. "Evolutionary and Comparative Analysis of MYB and bHLH Plant Transcription Factors." *The Plant Journal : For Cell and Molecular Biology* 66 (1): 94–116. doi:10.1111/j.1365-313X.2010.04459.x.

Feuermann, Marc, Pascale Gaudet, Huaiyu Mi, Suzanna E Lewis, and Paul D Thomas. 2016.

"Large-Scale Inference of Gene Function through Phylogenetic Annotation of Gene Ontology Terms : Case Study of the Apoptosis and Autophagy Cellular Processes." *Database: The Journal of Biological Databases and Curation* 2016: 1–11. doi:10.1093/database/baw155.

Finn, Robert D, Jody Clements, and Sean R Eddy. 2011. "HMMER Web Server: Interactive Sequence Similarity Searching." *Nucleic Acids Research* 39 (Web Server issue): W29-37. doi:10.1093/nar/gkr367.

Fu, Ning, Ines Drinnenberg, Janet Kelso, Jia-Rui Wu, Svante Pääbo, Rong Zeng, and Philipp Khaitovich. 2007. "Comparison of Protein and mRNA Expression Evolution in Humans and Chimpanzees." *PLoS ONE* 2 (2): e216. doi:10.1371/journal.pone.0000216.

Galego, Lisete, and Jorge Almeida. 2002. "Role of DIVARICATA in the Control of Dorsoventral Asymmetry in Antirrhinum Flowers." *Genes & Development* 16 (7): 880–91. doi:10.1101/gad.221002.

Gaudin, V, P a Lunness, P R Fobert, M Towers, C Riou-Khamlichi, J a Murray, E Coen, and J H Doonan. 2000. "The Expression of D-Cyclin Genes Defines Distinct Developmental Zones in Snapdragon Apical Meristems and Is Locally Regulated by the Cycloidea Gene." *Plant Physiology* 122 (4): 1137–48. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=58948&tool=pmcentrez&rendertype=abstract.

Giurfa, Martin, Amots Dafni, and Paul R Neal. 1999. "Floral Symmetry and Its Role in Plant – Pollinator Systems." *International Journal of Plant Sciences* 160 (S6): S41–50.

Grant, Charles E, Timothy L Bailey, and William Stafford Noble. 2011. "FIMO : Scanning for Occurrences of a given Motif." *Bioinformatics* 27 (7): 1017–18. doi:10.1093/bioinformatics/btr064.

Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology* 59 (3): 307–21. doi:10.1093/sysbio/syq010.

Haas, Brian J, Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D Blood, Joshua Bowden, Matthew Brian Couger, et al. 2013. "De Novo Transcript Sequence Reconstruction from RNA-Seq: Reference Generation and Analysis with Trinity." *National Proceedings of the National Institute of Health* 8 (8): 1–43. doi:10.1038/nprot.2013.084.De.

Haber, W. A., and G. W. Frankie. 1989. "A Tropical Hawkmoth Community: Costa Rican Dry Forest Sphingidae." *Biotropica* 21 (2): 155–72.

Hileman, Lena C. 2014a. "ScienceDirect Bilateral Flower Symmetry — How , When and Why ?" *Current Opinion in Plant Biology* 17. Elsevier Ltd: 146–52. doi:10.1016/j.pbi.2013.12.002.

———. 2014b. "Trends in Flower Symmetry Evolution Revealed through Phylogenetic and Developmental Genetic Advances." *Philosophical Transactions of the Royal Society of Biological Sciences* 369 (June).

Hoshino, Yoichiro, Toshiya Igarashi, Masumi Ohshima, Koichi Shinoda, Naho Murata, Akira Kanno, and Masaru Nakano. 2014. "Characterization of CYCLOIDEA-like Genes in Controlling Floral Zygomorphy in the Monocotyledon Alstroemeria." *Scientia Horticulturae* 169 (April). Elsevier B.V.: 6–13. doi:10.1016/j.scienta.2014.01.046.

Howarth, Dianella G., and Michael J. Donoghue. 2009. "Duplications and Expression of DIVARICATA-like Genes in Dipsacales." *Molecular Biology and Evolution* 26 (6): 1245–

58. doi:10.1093/molbev/msp051.
Howarth, Dianella G, Tiago Martins, Edward Chimney, and Michael J Donoghue. 2011. "Diversification of CYCLOIDEA Expression in the Evolution of Bilateral Flower Symmetry in Caprifoliaceae and Lonicera (Dipsacales)." *Annals of Botany* 107: 1521–32. doi:10.1093/aob/mcr049.
Hu, Shusheng, David L Dilcher, David M Jarzen, and David Winship Taylor. 2008. "Early Steps of Angiosperm Pollinator Coevolution." *Proceedings of the National Academy of Sciences of the United States of America* 105 (1): 240–45. doi:10.1073/pnas.0707989105.
Inta, Wandee, Paweena Traiperm, and Sasivimon Chomchalow Swangpol. 2015. "Floral Micromorphology of the Genus Ensete Bruce Ex Horan . (Musaceae) in Thailand." *Taiwania* 60 (3): 99–106. doi:10.6165/tai.2015.60.99.
Javelle, Marie, Cristina F Marco, and Marja Timmermans. 2011. "In Situ Hybridization for the Precise Localization of Transcripts in Plants." *Journal of Visualized Experiments : JoVE*, no. 57: e3328. doi:10.3791/3328.
Jin, H, and C Martin. 1999. "Multifunctionality and Diversity within the Plant MYB-Gene Family." *Plant Molecular Biology* 41 (5): 577–85. http://www.ncbi.nlm.nih.gov/pubmed/23588495.
Katoh, Kazutaka, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. 2002. "MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." *Nucleic Acids Research* 30 (14): 3059–66. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=135756&tool=pmcentrez&rendertype=abstract.
Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, et al. 2012. "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." *Bioinformatics* 28 (12): 1647–49. doi:10.1093/bioinformatics/bts199.
Kirchoff, BK. 1991. "Homeosis in the Flowers of the Zingiberales." *American Journal of Botany* 78 (6): 833–37. http://www.jstor.org/stable/10.2307/2445074.
Kirchoff, Bruce K. 1987. "Inflorescence and Flower Development in Costus Scaber (Costaceae)."
———. 1988. "Floral Ontogeny and Evolution in the Ginger Group of the Zingiberales." *Aspects of Floral Development*.
———. 1991. "Homeosis in the Flowers of the Zingiberales." *American Journal of Botany* 78 (6): 833–37.
———. 2013. "Homeosis In The Flowers Of The Zingiberales '" 78 (6): 833–37.
Klempnauer, Karl Heinz, Thomas J. Gonda, and J. Michael Bishop. 1982. "Nucleotide Sequence of the Retroviral Leukemia Gene v-Myb and Its Cellular Progenitor c-Myb: The Architecture of a Transduced Oncogene." *Cell* 31 (2 PART 1): 453–63. doi:10.1016/0092-8674(82)90138-6.
Konieczny, Andrzej, and Frederick M. Ausubel. 1993. "A Procedure for Mapping Arabidopsis Mutations Using Co-Dominant Ecotype-Specific PCR-Based Markers." *The Plant Journal* 4 (2): 403–10.
Koski, L B, and G B Golding. 2001. "The Closest BLAST Hit Is Often Not the Nearest Neighbor." *Journal of Molecular Evolution* 52 (6): 540–42. doi:10.1007/s002390010184.
Langmead, B, C Trapnell, M Pop, and S L Salzberg. 2009. "Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome." *Genome Biology* 10 (3): R25.

doi:gb-2009-10-3-r25 [pii]\n10.1186/gb-2009-10-3-r25.
Li, Weizhong, and Adam Godzik. 2006. “Cd-Hit : A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences.” *Bioinformatics Applications Note* 22 (13): 1658–59. doi:10.1093/bioinformatics/btl158.
Luo, D, R Carpenter, L Copsey, C Vincent, J Clark, and E Coen. 1999. “Control of Organ Asymmetry in Flowers of Antirrhinum.” *Cell* 99 (4): 367–76. http://www.ncbi.nlm.nih.gov/pubmed/10571179.
Luo, D, Rosemary Carpenter, Coral Vincent, Lucy Copsey, and Enrico Coen. 1996. “Origin of Floral Asymmetry in Antirrhinum.” *Nature* 383 (6603): 794–99. http://dx.doi.org/10.1038/383794a0.
Lyons-weiler, James, G U Y A Hoelzer, and Robin J Tausch. 1998. “Optimal Outgroup Analysis.” *Biological Journal of the Linnean Society*, no. 1 998.
MacHemer, Katja, Oxana Shaiman, Yehiam Salts, Sara Shabtai, Irina Sobolev, Eduard Belausov, Erich Grotewold, and Rivka Barg. 2011. “Interplay of MYB Factors in Differential Cell Expansion, and Consequences for Tomato Fruit Development.” *Plant Journal* 68 (2): 337–50. doi:10.1111/j.1365-313X.2011.04690.x.
Maddison, W.P., and D.R. Maddison. 2014. *Mesquite: A Modular System for Evolutionary Analysis*. http://mesquiteproject.org.
McCutcheon, John P, and Nancy A Moran. 2007. “Parallel Genomic Evolution and Metabolic Interdependence in an Ancient Symbiosis.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (49): 19392–97. doi:10.1073/pnas.0708855104.
Meara, Brian C O, Stacey D Smith, W Scott Armbruster, D Harder, Christopher R Hardy, Lena C Hileman, Larry Hufford, et al. 2016. “Non-Equilibrium Dynamics and Floral Trait Interactions Shape Extant Angiosperm Diversity.” *Proceedings of the Royal Society Biological Sciences* 283 (May). doi:10.1098/rspb.2015.2304.
Miller, Mark A, Wayne Pfeiffer, and Terri Schwartz. 2010. “Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees.” In *Proceedings of the Gateway Computing Environments Workshop (GCE)*.
Mirarab, Siavash, R Reaz, S Bayzid, T Zimmermann, M S Swenson, and T Warnow. 2014. “ASTRAL : Genome-Scale Coalescent-Based Species Tree Estimation.” *Bioinformatics* 30: 541–48. doi:10.1093/bioinformatics/btu462.
O’Connell, Mitchell R, Benjamin L Oakes, Samuel H Sternberg, Alexandra East-Seletsky, Matias Kaplan, and Jennifer A Doudna. 2014. “Programmable RNA Recognition and Cleavage by CRISPR/Cas9.” *Nature* 516 (7530). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 263–66. http://dx.doi.org/10.1038/nature13769.
Perez-Rodriguez, Maria, Felix W Jaffe, Eugenio Butelli, Beverley J Glover, and Cathie Martin. 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 (Cambridge, England)* 132 (2): 359–70. doi:10.1242/dev.01584.
Preston, Jill C. 2010. “Evolutionary Genetics of Core Eudicot Inflorescence and Flower Development.” *International Journal of Plant Developmental Biology* 4 (Special Issue 1): 17–29.
Preston, Jill C, and Lena C Hileman. 2012. “Parallel Evolution of TCP and B-Class Genes in Commelinaceae Flower Bilateral Symmetry.” *EvoDevo* 3 (1). BioMed Central Ltd: 6.

doi:10.1186/2041-9139-3-6.
Qiu, Yin-long, Jungho Lee, Barbara A Whitlock, Fabiana Bernasconi-quadroni, and Olena Dombrovska. 2001. “Was the ANITA Rooting of the Angiosperm Phylogeny Affected by Long-Branch Attraction ?” *Molecular Biology and Evolution* 18 (9): 1745–53.
Raimundo, João, Rómulo Sobral, Paul Bailey, Herlânder Azevedo, Lisete Galego, Jorge Almeida, Enrico Coen, and Maria Manuela R Costa. 2013. “A Subcellular Tug of War Involving Three MYB-like Proteins Underlies a Molecular Antagonism in Antirrhinum Flower Asymmetry.” *The Plant Journal : For Cell and Molecular Biology* 75 (4): 527–38. doi:10.1111/tpj.12225.
Reyes, Elisabeth, Hervé Sauquet, and Sophie Nadot. 2016. “Perianth Symmetry Changed at Least 199 Times in Angiosperm Evolution.” *Taxon* 65 (5): 945–64.
Ronquist, F., and J. P. Huelsenbeck. 2003. “MrBayes 3: Bayesian Phylogenetic Inference under Mixed Models.” *Bioinformatics* 19 (12): 1572–74. doi:10.1093/bioinformatics/btg180.
Rose, Annkatrin, Iris Meier, and Udo Wienand. 1999. “The Tomato I-Box Binding Factor LeMYBI Is a Member of a Novel Class of Myb-like Proteins.” *Plant Journal* 20 (6): 641–52. doi:10.1046/j.1365-313X.1999.00638.x.
Rosinski, James A., and William R. Atchley. 1998. “Molecular Evolution of the Myb Family of Transcription Factors: Evidence for Polyphyletic Origin.” *Journal of Molecular Evolution* 46 (1): 74–83. doi:10.1007/PL00006285.
Rudall, Paula J., and Richard M. Bateman. 2004. “Evolution of Zygomorphy in Monocot Flowers: Iterative Patterns and Developmental Constraints.” *New Phytologist* 162 (April): 25–44. doi:10.1111/j.1469-8137.2004.01032.x.
Sambrook, Joseph, and David Russell. 2001. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press. citeulike-article-id:144016.
Sargent, Risa D. 2004. “Floral Symmetry Affects Speciation Rates in Angiosperms.” *Proceedings. Biological Sciences / The Royal Society* 271 (1539): 603–8. doi:10.1098/rspb.2003.2644.
Sass, Chodon, William J.D. Iles, Craig F. Barrett, Selena Y. Smith, and Chelsea D. Specht. 2016. “Revisiting the Zingiberales: Using Multiplexed Exon Capture to Resolve Ancient and Recent Phylogenetic Splits in a Charismatic Plant Lineage.” *PeerJ* 4: e1584. doi:10.7717/peerj.1584.
Schaffer, Robert, Nicola Ramsay, Alon Samach, Sally Corden, Joanna Putterill, Isabelle A Carre, and George Coupland. 1998. “The Late Elongated Hypocotyl Mutation of Arabidopsis Disrupts Circadian Rhythms and the Photoperiodic Control of Flowering.” *Cell* 93: 1219–29.
Smith, Stacey D. 2016. “Pleiotropy and the Evolution of Floral Integration.” *New Phytologist* 209: 80–85.
Smith, Stacey D, and Mark D Rausher. 2011. “Gene Loss and Parallel Evolution Contribute to Species Difference in Flower Color.” *Molecular Biology and Evolution* 28 (10): 2799–2810. doi:10.1093/molbev/msr109.
Sobral, Romulo Sacramento. 2010. “Antirrhinum Majus.” Universidade do Minho.
Specht, Chelsea D, and Dianella G Howarth. 2014. “Tansley Review Adaptation in Flower Form : A Comparative Evodevo Approach.”
Stamatakis, Alexandros. 2014. “RAxML Version 8 : A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies.” *Bioinformatics Applications Note* 30 (9): 1312–13. doi:10.1093/bioinformatics/btu033.

Stern, David L. 2013. “The Genetic Causes of Convergent Evolution.” *Nature Publishing Group* 14 (11). Nature Publishing Group: 751–64. doi:10.1038/nrg3483.

Stevenson, Clare E. M., Nicolas Burton, Maria M. R. Costa, Utpal Nath, Ray A. Dixon, Enrico S. Coen, and David M. Lawson. 2006. “Crystal Structure of the MYB Domain of the RAD Transcription Factor from Antirrhinum Majus.” *Proteins: Structure, Function, and Bioinformatics* 65: 1041–45. doi:10.1002/prot.

Stohr-hunt, Patricia M. 1996. “An Analysis of Frequency of Hands-on Experience and Science Achievement.” *Journal of Research in Science Teaching* 33 (1): 101–9.

Takiya, Daniela M., Phat L. Tran, Christopher H. Dietrich, and Nancy A. Moran. 2006. “Co-Cladogenesis Spanning Three Phyla: Leafhoppers (Insecta: Hemiptera: Cicadellidae) and Their Dual Bacterial Symbionts.” *Molecular Ecology* 15 (13): 4175–91. doi:10.1111/j.1365-294X.2006.03071.x.

Talavera, Gerard, and Jose Castresana. 2007. “Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments.” *Systematic Biology* 56 (4): 564–77. doi:10.1080/10635150701472164.

Tucker, Shirley C. 1999. “Evolutionary Lability Of Symmetry In Early Floral Development.” *International Journal of Plant Sciences* 160 (6): S25–39.

Ushimaru, Atushi, Ikumi Dohzono, Yasuoki Takami, and Fujio Hyodo. 2009. “Flower Orientation Enhances Pollen Transfer in Bilaterally Symmetrical Flowers.” *Oecologia* 160 (4): 667–74. doi:10.1007/s00442-009-1334-9.

Ushimaru, Atushi, and Fujio Hyodo. 2005. “Why Do Bilaterally Symmetrical Flowers Orient Vertically ? Flower Orientation Influences Pollinator Landing Behaviour.” *Evolutionary Ecology Research* 7 (1): 151–60.

Valoroso, Maria Carmen, Sofia De Paolo, Giovanni Iazzetti, and Serena Aceto. 2017. “Transcriptome-Wide Identification and Expression Analysis of DIVARICATA- and RADIALIS- Like Genes of the Mediterranean.” *Genome Biology and Evolution* 9 (6): 1418–31. doi:10.1093/gbe/evx101.

Wang, Zheng, Yonghai Luo, Xin Li, Liping Wang, Shilei Xu, Jun Yang, Lin Weng, et al. 2008. “Genetic Control of Floral Zygomorphy in Pea (Pisum Sativum L.).” *Proceedings of the National Academy of Sciences of the United States of America* 105 (30): 10414–19. doi:10.1073/pnas.0803291105.

Wang, Zhi-yong, and Elaine M Tobin. 1998. “Constitutive Expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) Gene Disrupts Circadian Rhythms and Suppresses Its Own Expression” 93: 1207–17.

Waser, Nikolas M, Lars Chittka, Mary V Price, Neal M Williams, and Jeff Ollerton. 2010. “New Ideas In Ecofogy.” *America* 77 (4): 1043–60. http://www.esajournals.org/doi/abs/10.2307/2265575.

Wester, Katja, Simona Digiuni, Florian Geier, Jens Timmer, Christian Fleck, and Martin Hülskamp. 2009. “Functional Diversity of R3 Single-Repeat Genes in Trichome Development.” *Development* 136: 1487–96. doi:10.1242/dev.021733.

Wu, Dongying, Sean C. Daugherty, Susan E. Van Aken, Grace H. Pai, Kisha L. Watkins, Hoda Khouri, Luke J. Tallon, et al. 2006. “Metabolic Complementarity and Genomics of the Dual Bacterial Symbiosis of Sharpshooters.” *PLoS Biology* 4 (6): 1079–92. doi:10.1002/adsc.200505252.

Yang, Ziheng. 1994. “Statistical Properties of the Maximum Likelihood Method of Phylogenetic Estimation and Comparison With Distance Matrix Methods.” *Systematic Biology* 43 (3):

329–42. http://dx.doi.org/10.1093/sysbio/43.3.329.
Yanhui, Chen, Yang Xiaoyuan, He Kun, Liu Meihua, Li Jigang, Gao Zhaofeng, Lin Zhiqiang, et al. 2006. “The MYB Transcription Factor Superfamily of Arabidopsis: Expression Analysis and Phylogenetic Comparison with the Rice MYB Family.” *Plant Molecular Biology* 60: 107–24. doi:10.1007/s11103-005-2910-y.
Zhang, Wenheng, Elena M Kramer, and Charles C Davis. 2010. “Floral Symmetry Genes and the Origin and Maintenance of Zygomorphy in a Plant-Pollinator Mutualism.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (14): 6388–93. doi:10.1073/pnas.0910155107.
Zhou, Xiu-Ren, Yin-Zheng Wang, James F Smith, and Rujin Chen. 2008. “Altered Expression Patterns of TCP and MYB Genes Relating to the Floral Developmental Transition from Initial Zygomorphy to Actinomorphy in Bournea (Gesneriaceae).” *The New Phytologist* 178 (3): 532–43. doi:10.1111/j.1469-8137.2008.02384.x.