Tracking the Evolutionary History of Development Genes: Implications for the Diversification of Fruits and Flowers in the Brassicaceae and Cleomaceae

by

Kelsey Brock

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science
in
Plant Biology

Department of Biological Sciences
University of Alberta

Abstract

Flowers of Brassicaceae are remarkably similar across all species, whereas their fruits vary in almost all conceivable traits, particularly in the tribe Brassiceae. In contrast, Brassicaceae's sister family, Cleomaceae, exhibits substantial variation in flowers but are more uniform in their fruits. These diversifications represent either variation in pollen transfer or seed dispersal, which are important reproductive traits that likely affect survival. The history of both families involves four shared polyploidy events as well as each independently experiencing separate additional polyploid events. Thus, these families offer an excellent opportunity to investigate whether additional genetic materials from polyploid events are correlated with the evolution of novel flower and fruit morphologies. Based on knowledge from model plants, *FRUITFULL* (*FUL*), a fruit development gene important for dehiscence (fruit opening), and *TEOSINTE BRANCHED 1/ CYCLOIDEA/ PROLIFERATING CELL FACTOR 1* (*TCP1*), a gene known to affect floral symmetry, were chosen as candidate genes to examine the evolutionary history and retention of gene duplicates alongside morphological novelty in Brassicaceae and Cleomaceae. A gene phylogeny of *FUL* was generated to determine if fruit diversity in the tribe Brassiceae (Brassicaceae) could be correlated with *FUL* copy number, structure, or evolutionary history. Similarly, *TCP1* was assessed to identify differences in gene evolutionary history between the florally diverse Cleomaceae and the florally uniform Brassicaceae. Both *FUL* and *TCP1* were found to exhibit complex evolutionary histories, with multiple copies of these genes found in both taxa with and without morphological novelty. However, evaluation of which copies were retained and the rates of selection acting on these genes suggest their involvement in generating morphological diversity of reproductive

structures. This thesis presents a strong correlative framework to direct future hypothesis testing using gene expression and functional approaches to further unravel the genetic changes underlying flower and fruit diversification.

Acknowledgements

I thank my supervisor Dr. Jocelyn Hall for her advice, enthusiasm, and for always showing optimism and unyielding support of my career goals. I would also like to thank Dr. Felix Sperling and Dr. Janice Cooke for offering a wealth of experience and ideas throughout the completion of this thesis. Additionally, I thank Dr. David Coltman for generously agreeing to be an arms-length examiner during the defense of my thesis. I thank Drs. Patrick Edger and J. Chris Pires for kindly providing access to the unpublished *Cleome violacea* genome and Dr. Suzanne Warwick for providing DNA of *Brassica napus*. I'd also like to thank the University of Alberta Molecular Biology Service Unit staff for their advice and assistance.

I thank Dorothy Fabijan and Karen Romanyk for being an endlessly impressive source of information and for offering guidance during my teaching assistantships. My lab mates, Brittney Miller, Melanie Patchell, Mónica Higuera, Peter Mankowski, and Stephanie McCrea are thanked for their wisdom and humour. Lastly, I thank my parents Cheryl Langford-Brock and Dewey Brock and my husband, Jared Bernard, for their constant help, support and perspective.

Table of Contents

List of Tables

List of Figures

Chapter 1 – Introduction

The diversity of angiosperms

The evolution of angiosperms represents one of the largest and most rapid terrestrial radiations with extant species estimates ranging from 250,000-400,000 (Soltis et al., 2009), a number which represents more than all other land plant groups combined (Crane et al., 1995; Davies et al., 2004). Angiosperms' radiation resulted in their domination of almost all terrestrial ecosystems since the end of the Cretaceous (Lidgard and Crane, 1990) as well as a variety of morphologies that were modified for agriculture (Seymour et al., 2013). Furthermore, angiosperms were likely involved in the diversification of other organisms, including co-radiations with amphibians (Roelants et al., 2007) and some insect groups (Farrell, 1998; Moreau et al., 2006; but see Mckenna, 2011).

Although many hypotheses exist to explain their diversification, variation of traits affecting reproductive success is correlated with speciation in angiosperms (Crepet and Niklas, 2009). The sessile nature of plants prompted the development of multiple ways to facilitate pollen transfer through insect, bird, mammal or abiotic vectors, producing a great range of floral forms (Crepet and Niklas, 2009). In addition, the inability of plants to move resulted in an equally impressive diversity in how plants disperse offspring via fruit variation (Eriksson et al., 2000; Bolmgren and Eriksson, 2005; Lorts et al., 2008). To investigate the origins of reproductive diversity in angiosperms, characterization of genetic changes responsible for morphological variation is necessary. Because no other group of organisms can lay claim to such a rampant history of polyploidization when compared to angiosperms, the raw genetic material created by these events is thought to have a key role

in facilitating adaptive evolution underlying morphological novelty (Flagel and Wendel, 2009; Soltis et al., 2009).

Gene duplications as drivers of morphological diversity

An estimated 70-80% of angiosperms have undergone polyploidization since the divergence of the group, prompting discussion on the fates of these duplicated genes (Otto and Whitton, 2000; Simillion et al., 2002; Blanc et al., 2003; Bowers et al., 2003; Paterson et al., 2004). Few morphological changes result from tandem duplications, where single sections of DNA are duplicated rather than whole genomes (Veit et al., 1990; Park et al., 2004). Instead, polyploid events facilitate morphological complexity and speciation more so than other duplications because multiple copies of genomes increase the chance of at least one gene incurring a unique functional fate (Moore and Purugganan, 2005; Edger and Pires, 2009; Rodgers-Melnick et al., 2012; Oliveira et al., 2013). Classical theories specify that the loss of duplicates through deleterious mutations is most common (pseudogenization), whereas some duplicated genes may be selectively maintained by being co-opted for new functions (neofunctionalization) (Ohno, 1970). Although rare, neofunctionalization has been documented to underlie unique floral morphologies (Rosin and Kramer, 2009). In *Aquilegia* (Ranunculaceae), neofunctionalization of one of three replicates of *APETALA3* (*AP3*) resulted in the novel formation of staminodes (sterile stamens) located between the corolla and the stamens (Kramer et al., 2007). Neofunctionalization is also demonstrated in *Helianthus* (Asteraceae), where expression of one of the ten replicates of *CYCLOIDEA2* (*CYC2*) was specific to the development of ray

florets (specialized flowers located around the edge of sunflower inflorescences) (Chapman et al., 2008).

More recent theoretical work integrated additional complexity by including the concept of subfunctionalization, where partial redundancy is achieved by two sister copies partitioning the ancestral functions of the original gene (Lynch and Conery, 2000). Current gene duplication models predict that redundancy and subfunctionalization are necessary during early functional divergence to retain loci in the genome long enough to accrue the mutations necessary for neofunctionalization (Moore and Purugganan, 2005). A prominent example of ancient and persistent subfunctionalization is the diversification of *AGAMOUS* (*AG*), a MADS-box transcription factor important for reproductive development. In gymnosperms, an *AG*-like homologue is expressed in the ovule as well as the female reproductive structure (megasporophyll) (Tandre et al., 1995; Rutledge et al., 1998; Jager et al., 2003; Zhang et al., 2004). In contrast, a duplication at the base of angiosperms has produced C-class and D-class lineages that exhibit both subfunctionalization and redundancy (Pinyopich et al., 2003). In angiosperms, C-class genes are mainly involved in stamen and carpel development while D-class genes are recruited for ovule development (Theissen, 2001). Remarkably, further partitioning of this lineage also serves as an example of a comparatively young subfunctionalization event. In maize, duplicates *ZEA MAYS MADS1 (ZMM1)* and *ZEA AGAMOUS1 (ZAG1)* have partitioned the ancestral role of *AG,* where *ZMM1* is expressed primarily in the stamens while *ZAG1* is expressed primarily in the carpels (Mena et al., 1996).

A history of polyploidization and morphological novelty in two sister families

Within the Brassicales lie two sister families, Brassicaceae and Cleomaceae, that are ideal to study how gene duplications affect the evolutionary origins of morphological novelty. Brassicaceae is a model for understanding polyploid events because it contains the model plant *Arabidopsis thaliana* (Blanc et al., 2003; Thomas et al., 2006; Wang et al., 2011). More recent genomic investigation into Cleomaceae, which diverged from Brassicaceae only ~38 million years ago (Schranz and Mitchell-Olds, 2006; Couvreur et al., 2010), has provided more context for associations between morphological traits and polyploid events (Cheng et al., 2013). Brassicaceae and Cleomaceae share four known ancient polyploid events: one tetraploidization during the origin of seed plants (Jiao et al., 2011), one tetraploidization preceding the origin of angiosperms (Jiao et al., 2011), one hexaploidization near the origin of most eudicots (Jaillon et al., 2007; Vekemans et al., 2012), and one tetraploidization during the radiation of the Brassicales (after the divergence of papaya (*Carica*) within the Brassicales) (Ming et al., 2008). After the divergence of the two lineages, Brassicaceae and Cleomaceae each underwent separate whole genome duplication and triplication events, respectively, at the base of each family (Schranz and Mitchell-Olds, 2006; Barker et al., 2009; Cheng et al., 2013). Additionally, another hexaploid event took place at the base of a tribe within the Brassicaceae called Brassiceae (Lysak et al., 2005; Parkin et al., 2005; Lysak et al., 2007).

Brassicaceae contains approximately 3700 species that exist on every continent except Antarctica but largely occupy northern temperate regions and are most diverse in the Mediterranean (Warwick and Sauder, 2005). Cleomaceae contains approximately 300

species that are mostly restricted to the tropics or arid desert regions. Brassicaceae and Cleomaceae have several morphological traits in common including the same basic floral blueprint (four sepals, four petals, six stamens and two fused carpels), fruits that possess a replum, and a propensity towards herbaceous growth habit (Hall et al., 2002; Iltis et al., 2011). Flowers of Brassicaceae are remarkably similar across all species, whereas their fruits vary in almost all conceivable traits (Koch and Mummenhoff, 2006; Franzke et al., 2011). In contrast, Cleomaceae exhibit substantial variation in flowers despite having fewer species in this lineage, although fruits are less diverse than observed in Brassicaceae (Endress, 1992; Kers, 2003).

Cheng et al. (2013) provides comparative analysis of the genomes of *Tarenaya hassleriana* (Cleomaceae), *Arabidopsis thaliana* (Brassicaceae), *Arabidopsis lyrata* (Brassicaceae) and *Brassica rapa* (Brassiceae polyploid hybrid; Brassicaceae) that emphasizes that genome evolution and retention of gene copies are unique to each family. Although, flower and fruit variability have not yet been investigated. Cleomaceae are dominated by monosymmetric (one plane of symmetry) flowers while almost all flowers in Brassicaceae are dissymmetric (two planes of symmetry; Iltis et al., 2011). The calyx and corolla of Brassicaceae flowers are polysymmetric, but the androecium has two planes of symmetry, resulting in the predominant symmetry of this family being called dissymmetric (Endress, 1992; Busch et al., 2012). The variable symmetry between the two families invites question as to whether the hexaploidization coinciding with the early radiation of Cleomaceae provided the raw genetic material for development of monosymmetric flowers and remarkable variation in petal shape, size, position and colour, as well as androecium and gynoecium elongation. Furthermore, the tribe Brassiceae exhibits great variability in

fruit morphology within Brassicaceae, most notably in positioning where mature fruits open to release seeds, called dehiscence (Hall et al., 2011). This questions whether triplication of fruit development genes played a role in the significant fruit variation found across the tribe.

The evolution of fruit

Variable fruit morphology in angiosperms reflects the vast range by which tissue is modified to ensure the protection and dispersal of offspring. Surprisingly, fruits are considerably less studied in an evolutionary development context than flowers, even though dispersal of offspring is equally as important as pollination and the majority of calories consumed by humans are from eating fruits (Hall and Donohue, 2012; Seymour et al., 2013). Fruits can be divided into dry and fleshy fruits. Examples of fleshy fruits include tomatoes and peaches, which possess a fleshy pericarp (mature ovary wall) and do not open to release seeds at maturity. Fleshy fruits are thought to have evolved as a method of supplementary carbohydrate storage for the seedling in environments where sunlight was restricted due to closed forest canopies (Eriksson et al., 2000; Lorts et al., 2008). Subsequently, these fleshy tissues were recruited to attract frugivorous animals to aid in seed dispersal (Eriksson et al., 2000). Dry fruits are the ancestral fruit type of angiosperms, and can be further divided in dehiscent, where the fruit opens to release seeds at maturity, and indehiscent, where the fruit remains closed (Eriksson et al., 2000; Lewis et al., 2006; Lorts et al., 2008). Examples of dehiscent fruits include capsules, follicles and legumes while examples of indehiscent fruits include nuts and grains. Dispersal syndromes for dry fruits are variable and include gravity, wind dispersal, scatter hoarding (where animals

store seeds in food caches), and water dispersal (Beck and Vander Wall, 2010; Vander Wall and Beck, 2012; Seymour et al., 2013).

In addition to being diverse, fruit morphology within angiosperms is evolutionarily labile. Lorts et al. (2008) mapped fleshy, indehiscent dry, and dehiscent dry fruits onto a phylogeny of angiosperm orders and discovered that there was no association between fruit type and lineage. This pattern suggests that fleshiness and dehiscence evolved independently multiple times and that there is little phylogenetic constraint on fruit morphology at the ordinal level. The presence of both fleshy and dry fruits which do not form monophyletic lineages in the Solanaceae, Bignoniaceae and Verbenaceae families suggests that this phenomenon likely extends beyond the ordinal level (Olmstead, 2013).

Studies of *Arabidopsis* provide in-depth knowledge of a complex pathway of gene interactions leading to fruit maturation and dehiscence. *Arabidopsis* has a dry fruit type (called a silique) in which the ovary wall consists of two valves that dehisce from a persistent placental tissue called the replum along a valve margin (Dinneny et al., 2005). This genetic pathway appears to be conserved across the family (Girin et al., 2010; Lenser and Theissen, 2013; Muehlhausen et al., 2013). Strikingly, recent studies revealed that elements of the genetic pathway for fruit development and dehiscence in *Arabidopsis* are also employed in the fleshy, indehiscent, berry-type fruit of tomato (*Solanum lycopersicum*) to control color changes, sugar metabolism and tissue softening associated with ripening. Although the entire gene pathways for each species will not be discussed here, key parallel roles are played by the following transcription factors: *APETALA2* (*AP2*), *SHATTERPROOF1/SHATTERPROOF2* (*SHP1/2*; homologue is called *TOMATO AGAMOUS-LIKE1* (*TAGL1*) in tomato), and *FRUITFULL* (*FUL*) (Bemer et al., 2012; Seymour et al., 2013).

Utilization of conserved genetic elements despite having radically different fruit types indicates that even slight modifications may be sufficient to generate substantial variation in morphology. In *Arabidopsis* fruits, *AP2* suppresses replum formation genes, preventing replum tissue overgrowth and ensuring proper fruit dehiscence (Ripoll et al., 2011). In tomato, *AP2* is a negative regulator of ripening under the control of ethylene to ensure seeds mature at the same time the fruit becomes palatable (Chung et al., 2010; Karlova et al., 2011). Paralogues *SHP1/2* in *Arabidopsis* control the activity of other genes ensuring the lignification and separation of the valves from the replum, whereas *TAGL1* in tomato is necessary for pericarp fleshiness, adding mass by cell division and promoting ripening (Itkin et al., 2009). *FUL* restricts the actions of *SHP1/2* in the *Arabidopsis* silique to the valve margin, such that normal valves may form. In tomato, the two paralogues *FUL1* and *FUL2* act redundantly to affect fruit pigment and taste independent of ethylene presence, although it is still unclear how *TAGL1* and *FUL* interact (Bemer et al., 2012). The parallels between tomato and *Arabidopsis* indicate that studies on the evolution of fruit development genes in dry fruits present in the Brassicaceae will likely provide valuable information that may be applicable to studies on distantly related angiosperms with radically different fruit morphologies.

The evolution floral symmetry

Monosymmetry is a key modification to achieve an attractive state for animal pollinators (Rosin and Kramer, 2009; Martin-Trillo and Cubas, 2010; Preston et al., 2011; Busch et al., 2012). Generally, flowers can be roughly categorized as polysymmetric, which is the ancestral state of angiosperms, where flowers have multiple planes of symmetry, or

monosymmetric, where flowers have only one plane of symmetry (Endress, 2001; Specht and Bartlett, 2009). Monosymmetric flowers are thought to be an adaptation to insect pollination (Giurfa et al., 1999; Rodriguez et al., 2004) by providing a "landing pad" to orient pollinators for effective pollen transfer while often having embellishments advertising nectary position (Moller and Eriksson, 1995; Johnson et al., 1998). The importance of monosymmetry is exemplified by its independent origins amongst distantly related lineages within angiosperms and its prevalence in the top three most diverse flowering plant families: Asteraceae (22,750 species), Orchidaceae (21,950 species) and Fabaceae (19,400 species) (Bremer et al., 2009; Stevens, 2012). Developmental shifts to monosymmetry involve repositioning of the floral organs by lengthening through cell addition or curving by cell proliferation on one side of the organ (Finlayson, 2007; Martin-Trillo and Cubas, 2010).

TCP1 is a member of a large family of transcription factors for which expression studies have implicated its homologues in the development of monosymmetry in several angiosperm lineages including: *Antirrhinum* (Plantaginaceae; Luo et al., 1996), *Byrsonima* (Malpighiaceae; Zhang et al., 2010), *Commelina* (Commelinaceae; Preston and Hileman, 2012), *Capnoides* (Papaveraceae; Damerval et al., 2013), *Gerbera* (Asteraceae; Broholm et al., 2008), *Helianthus* (Asteraceae; Chapman et al., 2012), *Iberis* (Brassicaceae; Busch and Zachgo, 2007), *Janusia* (Malpighiaceae, Zhang et al., 2010), *Lonicera* (Caprifoliaceae; Howarth et al., 2011), *Lotus* (Fabaceae; Feng et al., 2006) and *Pisum* (Fabaceae; Wang et al., 2008). *TCP1*, as it is called in *Arabidopsis* (Finlayson, 2007), was named after the simultaneous study of four proteins containing an identifying TCP domain including: *Teosinte Branched1* (*TB1*) in maize (*Zea mays*; Poaceae; Doebley et al., 1997), *CYCLOIDEA*

(*CYC*) from snapdragon (*Antirrhinum majus;* Plantaginaceae; Luo et al., 1996), and *PROLIFERATING CELL FACTORS 1* and *2 (PCF1* and *PCF2*) from rice (*Oryza sativa,* Poaceae; Kosugi and Ohashi, 1997). Genes belonging to the TCP family encode proteins with a basic helix-loop-helix motif that allows for protein-protein or protein-DNA interactions (Cubas et al., 1999). Interestingly, TCP proteins are more likely to dimerize with TCP proteins that are closely related (Kosugi and Ohashi, 2002). *TCP1* has the opportunity to play a large role in morphological change over evolutionary time because it is associated with cell proliferation and expansion in many plant organs. The elongation of petioles, rosette leaves, and inflorescence stems has been associated with *TCP1* in *Arabidopsis* (Koyama et al., 2010) as well as the shape of organs in many flowers (Rosin and Kramer, 2009; Martin-Trillo and Cubas, 2010; Hileman, 2014).

Developmental genetic studies in *Antirrhinum* (snapdragon) across 20 years of investigation provide a strong foundation for genetic interactions underlying monosymmetry (reviewed in Hileman, 2014). *Antirrhinum* flowers consist of a corolla tube where the two lateral and single bottom petals fuse to form an abaxial (bottom) lip while two upper petals fuse to form an adaxial (top) lip (Luo et al., 1996). Monosymmetry is the result of partially redundant actions of duplicates *CYCLOIDEA* (*CYC*) and *DICHOTOMA* (*DICH*), homologues of *TCP1* in *Arabidopsis*, acting on the adaxial region of the flower (Luo et al., 1996; Luo et al., 1999; Hileman and Baum, 2003; Gubitz et al., 2003; Martin-Trillo and Cubas, 2010). Although present at different stages of floral development, the expression of these paralogues from budding to anthesis result in retarded cell division and smaller cell size of adaxial organs. Flowers are polysymmetric in *cyc/dich* double mutants and all adaxial petals become morphologically similar to abaxial petals (Luo et al., 1996; Luo et al.,

1999). *CYC* and *DICH* regulate downstream MYB-like transcription factors *RADIALIS* (*RAD*) and *DIVARICATA* (*DIV*), which are responsible for the shape and identity of the abaxial lip (Almeida et al., 1997; Galego and Almeida, 2002; Corley et al., 2005). Currently, nothing is known about the upstream regulators of *TCP1* homologues (Martin-Trillo and Cubas, 2010).

TCP1 is also expressed in Brassicaceae flowers, although no studies have explored Cleomaceae. In *Arabidopsis*, a single known copy of *TCP1* is transiently expressed adaxially in floral buds, resulting in a short-lived enlargement of the abaxial sepal early in development (Cubas et al., 2001). However, the absence of *TCP1* expression during later stages of floral development results in the mature flower being dissymmetric. Furthermore, *tcp1* mutants produce flowers with no discernable differences from wild type flowers (Cubas, 2004). In contrast, *IaTCP1* expression in *Iberis amara* is associated with reduced adaxial petal growth during the later stages of floral development (Busch and Zachgo, 2007; Busch et al., 2012), resulting in it being one of the few monosymmetric species within the Brassicaceae. A complementary study comparing expression patterns across additional monosymmetric Brassicaceae taxa found similar expression domains (Busch et al., 2012). Furthermore, *TCP1* expression was greatly reduced in rare but naturally occurring dissymmetric *I. amara* flowers. Additionally, transformation of the *I. amara TCP1* locus into *Arabidopsis* plants resulted in dissymmetric flowers with four shortened petals. These studies exemplify the importance of where and when *TCP1* expression occurs during floral development.

Studying development genes in a phylogenetic context

In the study of evolution and development (evo-devo), investigating the evolutionary history of a gene involves the construction of a phylogeny where the primary interest is not to discern species relationships, but instead determine the evolutionary history of the gene itself (Maddison, 1997; Arvestad et al., 2009; Szoellosi and Daubin, 2012). Thus, these studies often focus on candidate genes putatively important for morphological or physiological traits. Furthermore, the history of these genes is particularly interesting if they have undergone duplication events or directional selection (Hileman and Baum, 2003; Kramer et al., 2004; Zhang et al., 2008; Airoldi and Davies, 2012). In contrast to gene evolution studies, phylogenies constructed to infer species relationships are often created with the notion that phylogenetic accuracy will increase with the number of genes used (Phillips et al., 2004; Collins et al., 2005). Additionally, single-copy genes are sought out and signals in the sequence data that are not indicative of ancestry are avoided (Collins et al., 2005). This contrast emphasizes the need to approach candidate gene evolution studies with a focus on appreciating the history the gene itself, rather than the more common practice of assessing congruence with hypothesized species phylogenies. Selective forces may be acting on candidate genes to conserve functions crucial for survival or, contrastingly, be under directional selection reflecting the evolution of significant and novel modifications that may have arisen independently more than once (Gubitz et al., 2003; Litt and Irish, 2003). Thus, interpreting gene phylogenies requires careful consideration of which branches indicate common ancestry versus convergent evolution of sequences utilized during phylogenetic analysis.

Generating a phylogeny of a candidate gene whose functional role has been identified in model organisms can significantly aid the development of further gene expression and function studies (Citerne et al., 2003; Arvestad et al., 2009; Dunn et al., 2013). When this phylogeny is mapped alongside an already established species phylogeny, we can investigate whether gene sequence characteristics or retention of duplicate genes is correlated with interesting morphologies that exist amongst a certain group of species (Cronk et al., 2002). For example, Bartlett and Specht (2010) created a phylogeny of *GLOBOSA* (*GLO*) homologues from species within Zingerberales and concluded that the duplication and subsequent expression divergence of these genes were likely responsible for considerable floral diversity in the order. In addition, these analyses can further be used to propose phylogenetic placement of duplication events (Howarth and Donoghue, 2006). Thus, phylogenies of candidate genes provide a powerful foundation for testable hypotheses about function of gene copies, especially when combined with species phylogenies and morphological variation of focal taxa.

A phylogenetic approach to assessing whether retention of duplicated candidate genes is associated with evolutionary novelty also offers an opportunity to study the subsequent selective pressure after duplication events. Evolution of genes is typically measured using the ratio of non-synonymous substitutions (dN), where base pair changes result in a different amino acid, to synonymous substitutions (dS), where base pair changes result in the same amino acid (Nei and Gojobori, 1986). Approximately equal synonymous and non-synonymous substitutions resulting in dN/dS close to 1 is characteristic of neutral or relaxed selective pressure. Deviation from this scenario results in either preservation of the amino acid sequence via purifying selection (dN/dS close to 0) or positive selection

including a change in amino acid sequence from a disproportionate incorporation of non-synonymous substitutions (dN/dS >1). Detection of multiple gene copies alongside the measurement of selective pressures allows us to make hypotheses about the role of each duplicate in generating evolutionary novelties (Zhang et al., 2008). For example, strong purifying selection would indicate constrained divergence between the copies, which suggests a stable and functionally redundant relationship (Moore and Purugganan, 2005).

Goals

The goal of this thesis is to investigate three characteristics of angiosperms that are hypothesized to be involved in their rapid diversification and dominance of terrestrial ecosystems: their possession of flowers, fruits and their propensity to retain duplicate genes after polyploidization. This thesis is comprised of two data chapters concerning the origins of morphological variation in Brassicales using a phylogenetic approach. The first data chapter correlates the occurrence of variably dehiscent fruits with the evolution of a fruit development gene in the tribe Brassiceae. In the same manner, the second data chapter compares the evolution of a floral symmetry gene between the Brassicaceae and Cleomaceae to understand relationships between sequence evolution and monosymmetry. Retention and loss of duplicate genes from polyploid events as well as rates of gene evolution will be explored in both chapters. This research will establish a framework for the evolution of two development genes that are currently of interest in plant research and will aid in directing future comparative expression and function studies.

References

AIROLDI C. A. AND B. DAVIES. 2012. Gene duplication and the evolution of plant MADS-box transcription factors. *Journal of Genetics and Genomics* 39: 157-165.

ALMEIDA J., M. ROCHETA, AND L. GALEGO. 1997. Genetic control of flower shape in *Antirrhinum majus*. *Development* 124: 1387-1392.

ARVESTAD L., J. LAGERGREN, AND B. SENNBLAD. 2009. The gene evolution model and computing its associated probabilities. *Journal of the ACM* 56: 7.

BARKER M. S., H. VOGEL, AND M. E. SCHRANZ. 2009. Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biology and Evolution* 1: 391-399.

BARTLETT M. E. AND C. D. SPECHT. 2010. Evidence for the involvement of *GLOBOSA*-like gene duplications and expression divergence in the evolution of floral morphology in the Zingiberales. *New Phytologist* 187: 521-541.

BECK M. J. AND S. B. VANDER WALL. 2010. Seed dispersal by scatter-hoarding rodents in arid environments. *Journal of Ecology* 98: 1300-1309.

BEMER M., R. KARLOVA, A. R. BALLESTER, Y. M. TIKUNOV, A. G. BOVY, M. WOLTERS-ARTS, P. D. B. ROSSETTO, G. C. ANGENENT, AND R. A. DE MAAGD. 2012. The tomato *FRUITFULL* homologs *TDR4/FUL1* and *MBP7/FUL2* regulate ethylene-independent aspects of fruit ripening. *Plant Cell* 24: 4437-4451.

BLANC G., K. HOKAMP, AND K. H. WOLFE. 2003. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome research* 13: 137-144.

BOLMGREN K. AND O. ERIKSSON. 2005. Fleshy fruits - origins, niche shifts, and diversification. *Oikos* 109: 255-272.

BOWERS J. E., B. A. CHAPMAN, J. K. RONG, AND A. H. PATERSON. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422: 433-438.

BREMER B., K. BREMER, M. W. CHASE, M. F. FAY, J. L. REVEAL, D. E. SOLTIS, P. S. SOLTIS, P. F. STEVENS, A. A. ANDERBERG, M. J. MOORE, R. G. OLMSTEAD, P. J. RUDALL, K. J. SYTSMA, D. C. TANK, K. WURDACK, J. Q. -. XIANG, S. ZMARZTY, AND ANGIOSPERM PHYLOGENY GRP. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105-121.

BROHOLM S. K., S. TAHTIHARJU, R. A. E. LAITINEN, V. A. ALBERT, T. H. TEERI, AND P. ELOMAA. 2008. A TCP domain transcription factor controls flower type specification along the radial axis of the *Gerbera* (Asteraceae) inflorescence. *Proceedings of the National Academy of Sciences of the United States of America* 105: 9117-9122.

BUSCH A., S. HORN, A. MUEHLHAUSEN, K. MUMMENHOFF, AND S. ZACHGO. 2012. Corolla monosymmetry: evolution of a morphological novelty in the Brassicaceae family. *Molecular Biology and Evolution* 29: 1241-1254.

BUSCH A. AND S. ZACHGO. 2007. Control of corolla monosymmetry in the Brassicaceae Iberis amara. *Proceedings of the National Academy of Sciences of the United States of America* 104: 16714-16719.

CHAPMAN M. A., J. H. LEEBENS-MACK, AND J. M. BURKE. 2008. Positive selection and expression divergence following gene duplication in the sunflower *CYCLOIDEA* gene family. *Molecular Biology and Evolution* 25: 1260-1273.

CHAPMAN M. A., S. TANG, D. DRAEGER, S. NAMBEESAN, H. SHAFFER, J. G. BARB, S. J. KNAPP, AND J. M. BURKE. 2012. Genetic analysis of floral symmetry in Van Gogh's sunflowers reveals independent recruitment of *CYCLOIDEA* genes in the Asteraceae. *PLoS Genetics* 8: e1002628.

CHENG S., E. VAN DEN BERGH, P. ZENG, X. ZHONG, J. XU, X. LIU, J. HOFBERGER, S. DE BRUIJN, A. S. BHIDE, C. KUELAHOGLU, C. BIAN, J. CHEN, G. FAN, K. KAUFMANN, J. C. HALL, A. BECKER, A. BRAEUTIGAM, A. P. M. WEBER, C. SHI, Z. ZHENG, W. LI, M. LV, Y. TAO, J. WANG, H. ZOU, Z. QUAN, J. M. HIBBERD, G. ZHANG, X. ZHU, X. XU, AND M. E. SCHRANZ. 2013. The *Tarenaya hassleriana* genome provides insight into reproductive trait and genome evolution of Crucifers. *Plant Cell* 25: 2813-2830.

CHUNG M., J. VREBALOV, R. ALBA, J. LEE, R. MCQUINN, J. CHUNG, P. KLEIN, AND J. GIOVANNONI. 2010. A tomato (*Solanum lycopersicum*) *APETALA2/ERF* gene, *SlAP2a*, is a negative regulator of fruit ripening. *Plant Journal* 64: 936-947.

CITERNE H., D. LUO, R. PENNINGTON, E. COEN, AND Q. CRONK. 2003. A phylogenomic investigation of *CYCLOIDEA*-like TCP genes in the Leguminosae. *Plant Physiology* 131: 1042-1053.

COLLINS T. M., O. FEDRIGO, AND G. J. P. NAYLOR. 2005. Choosing the best genes for the job: The case for stationary genes in genome-scale phylogenetics. *Systematic Biology* 54: 493-500.

CORLEY S. B., R. CARPENTER, L. COPSEY, AND E. COEN. 2005. Floral asymmetry involves an interplay between TO and MYB transcription factors in *Antirrhinum*. *Proceedings of the National Academy of Sciences of the United States of America* 102: 5068-5073.

COUVREUR T. L. P., A. FRANZKE, I. A. AL-SHEHBAZ, F. T. BAKKER, M. A. KOCH, AND K. MUMMENHOFF. 2010. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustardfamily (Brassicaceae). *Molecular Biology and Evolution* 27: 55-71.

CRANE P., E. FRIIS, AND K. PEDERSEN. 1995. The origin and early diversification of angiosperms. *Nature* 374: 27-33.

CREPET W. L. AND K. J. NIKLAS. 2009. Darwin's second "abominable mystery": why are there so many angiosperm species? *American Journal of Botany* 96: 366-381.

CRONK, Q.C.B., BATEMAN, R.M. AND HAWKINS, J.A. 2002. *Developmental Genetics and Plant Evolution.* Taylor and Francis, London.

CUBAS P. 2004. Floral zygomorphy, the recurring evolution of a successful trait. *Bioessays* 26: 1175-1184.

CUBAS P., E. COEN, AND J. M. M. ZAPATER. 2001. Ancient asymmetries in the evolution of flowers. *Current Biology* 11: 1050-1052.

CUBAS P., C. VINCENT, AND E. COEN. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401: 157-161.

DAMERVAL C., H. CITERNE, M. LE GUILLOUX, S. DOMENICHINI, J. DUTHEIL, L. R. DE CRAENE, AND S. NADOT. 2013. Asymmetric morphogenetic cues along the transverse plane: shift from dissymmetry to zygomorphy in the flower of Fumarioideae. *American Journal of Botany* 100: 391-402.

DAVIES T. J., T. G. BARRACLOUGH, M. W. CHASE, P. S. SOLTIS, D. E. SOLTIS, AND V. SAVOLAINEN. 2004. Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* 101: 1904-1909.

DINNENY J., D. WEIGEL, AND M. YANOFSKY. 2005. A genetic framework for fruit patterning in *Arabidopsis thaliana*. *Development* 132: 4687-4696.

DOEBLEY J., A. STEC, AND L. HUBBARD. 1997. The evolution of apical dominance in maize. *Nature* 386: 485-488.

DUNN C. W., X. LUO, AND Z. WU. 2013. Phylogenetic analysis of gene expression. *Integrative and Comparative Biology* 53: 847-856.

EDGER P. P. AND J. C. PIRES. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research* 17: 699-717.

ENDRESS P. K. 1992. Evolution and floral diversity - the phylogenetic surroundings of *Arabidopsis* and *Antirrhinum*. *International Journal of Plant Sciences* 153: S106-S122.

ENDRESS P. K. 2001. Evolution of floral symmetry. *Current Opinion in Plant Biology* 4: 86-91.

ERIKSSON O., E. M. FRIIS, AND P. LOFGREN. 2000. Seed size, fruit size, and dispersal systems in angiosperms from the early cretaceous to the late tertiary. *American Naturalist* 156: 47-58.

FARRELL B. D. 1998. "Inordinate fondness" explained: why are there so many beetles? *Science* 281: 555-559.

FENG X. Z., Z. ZHAO, Z. X. TIAN, S. L. XU, Y. H. LUO, Z. G. CAI, Y. M. WANG, J. YANG, Z. WANG, L. WENG, J. H. CHEN, L. Y. ZHENG, X. Z. GUO, J. H. LUO, S. S. SATO, S. TABATA, W. MA, X. L. CAO, X. H. HU, C. R. SUN, AND D. LUO. 2006. Control of petal shape and floral zygomorphy in *Lotus japonicus*. *Proceedings of the National Academy of Sciences of the United States of America* 103: 4970-4975.

FINLAYSON S. A. 2007. *Arabidopsis TEOSINTE BRANCHED1-LIKE 1* regulates axillary bud outgrowth and is homologous to monocot *TEOSINTE BRANCHED1*. *Plant and Cell Physiology* 48: 667-677.

FLAGEL L. E. AND J. F. WENDEL. 2009. Gene duplication and evolutionary novelty in plants. *New Phytologist* 183: 557-564.

FRANZKE A., M. A. LYSAK, I. A. AL-SHEHBAZ, M. A. KOCH, AND K. MUMMENHOFF. 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends in Plant Science* 16: 108-116.

GALEGO L. AND J. ALMEIDA. 2002. Role of *DIVARICATA* in the control of dorsoventral asymmetry in *Antirrhinum* flowers. *Genes & Development* 16: 880-891.

GIRIN T., P. STEPHENSON, C. M. P. GOLDSACK, S. A. KEMPIN, A. PEREZ, N. PIRES, P. A. SPARROW, T. A. WOOD, M. F. YANOFSKY, AND L. OSTERGAARD. 2010. Brassicaceae *INDEHISCENT* genes specify valve margin cell fate and repress replum formation. *Plant Journal* 63: 329-338.

GIURFA M., A. DAFNI, AND P. R. NEAL. 1999. Floral symmetry and its role in plant-pollinator systems. *International Journal of Plant Sciences* 160: S41-S50.

GUBITZ T., A. CALDWELL, AND A. HUDSON. 2003. Rapid molecular evolution of *CYCLOIDEA*-like genes in *Antirrhinum* and its relatives. *Molecular Biology and Evolution* 20: 1537-1544.

HALL J. C., K. J. SYTSMA, AND H. H. ILTIS. 2002. Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. *American Journal of Botany* 89: 1826-1842.

HALL J. C. AND K. DONOHUE. 2012. The genetics of plant dispersal. *In* Clobert J., M. Baguette, T. Benton and J. Bullock [eds.], *Dispersal Ecology and Evolution*, Oxford University Press, .

HALL J. C., T. E. TISDALE, K. DONOHUE, A. WHEELER, M. A. AL-YAHYA, AND E. M. KRAMER. 2011. Convergent evolution of a complex fruit structure in the tribe Brassiceae (Brassicaceae). *American Journal of Botany* 98: 1989-2003.

HILEMAN L. C. AND D. A. BAUM. 2003. Why do paralogs persist? Molecular evolution of *CYCLOIDEA* and related floral symmetry genes in antirrhineae (Veronicaceae). *Molecular Biology and Evolution* 20: 591-600.

HILEMAN L. C. 2014. Bilateral flower symmetry - how, when and why? *Current Opinion in Plant Biology* 17: 146-152.

HOWARTH D. G. AND M. J. DONOGHUE. 2006. Phylogenetic analysis of the "ECE" (CYC/TB1) clade reveals duplications predating the core eudicots. *Proceedings of the National Academy of Sciences of the United States of America* 103: 9101-9106.

HOWARTH D. G., T. MARTINS, E. CHIMNEY, AND M. J. DONOGHUE. 2011. Diversification of *CYCLOIDEA* expression in the evolution of bilateral flower symmetry in Caprifoliaceae and *Lonicera* (Dipsacales). *Annals of Botany* 107: 1521-1532.

ILTIS H. H., J. C. HALL, T. S. COCHRANE, AND K. J. SYTSMA. 2011. Studies in the Cleomaceae I. on the separate recognition of Capparaceae, Cleomaceae, and Brassicaceae. *Annals of the Missouri Botanical Garden* 98: 28-36.

ITKIN M., H. SEYBOLD, D. BREITEL, I. ROGACHEV, S. MEIR, AND A. AHARONI. 2009. *TOMATO AGAMOUS-LIKE 1* is a component of the fruit ripening regulatory network. *Plant Journal* 60: 1081-1095.

JAGER M., A. HASSANIN, M. MANUEL, H. LE GUYADER, AND J. DEUTSCH. 2003. MADS-box genes in *Ginkgo biloba* and the evolution of the *AGAMOUS* family. *Molecular biology and evolution* 20: 842-854.

JAILLON O., J. AURY, B. NOEL, A. POLICRITI, C. CLEPET, A. CASAGRANDE, N. CHOISNE, S. AUBOURG, N. VITULO, C. JUBIN, A. VEZZI, F. LEGEAI, P. HUGUENEY, C. DASILVA, D. HORNER, E. MICA, D. JUBLOT, J. POULAIN, C. BRUYERE, A. BILLAULT, B. SEGURENS, M. GOUYVENOUX, E. UGARTE, F. CATTONARO, V. ANTHOUARD, V. VICO, C. DEL FABBRO, M. ALAUX, G. DI GASPERO, V. DUMAS, N. FELICE, S. PAILLARD, I. JUMAN, M. MOROLDO, S. SCALABRIN, A. CANAGUIER, I. LE CLAINCHE, G. MALACRIDA, E. DURAND, G. PESOLE, V. LAUCOU, P. CHATELET, D. MERDINOGLU, M. DELLEDONNE, M. PEZZOTTI, A. LECHARNY, C. SCARPELLI, F. ARTIGUENAVE, M. E. PE, G. VALLE, M. MORGANTE, M.

CABOCHE, A. ADAM-BLONDON, J. WEISSENBACH, F. QUETIER, P. WINCKER, AND FRENCH-ITALIAN PUBLIC. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463-U5.

JIAO Y., N. J. WICKETT, S. AYYAMPALAYAM, A. S. CHANDERBALI, L. LANDHERR, P. E. RALPH, L. P. TOMSHO, Y. HU, H. LIANG, P. S. SOLTIS, D. E. SOLTIS, S. W. CLIFTON, S. E. SCHLARBAUM, S. C. SCHUSTER, H. MA, J. LEEBENS-MACK, AND C. W. DEPAMPHILIS. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97-113.

JOHNSON S. D., H. P. LINDER, AND K. E. STEINER. 1998. Phylogeny and radiation of pollination systems in *Disa* (Orchidaceae). *American Journal of Botany* 85: 402-411.

KARLOVA R., F. M. ROSIN, J. BUSSCHER-LANGE, V. PARAPUNOVA, P. T. DO, A. R. FERNIE, P. D. FRASER, C. BAXTER, G. C. ANGENENT, AND R. A. DE MAAGD. 2011. Transcriptome and metabolite profiling show that *APETALA2a* is a major regulator of tomato fruit ripening. *Plant Cell* 23: 923-941.

KERS L. E. 2003. Capparaceae. *In* Kubitzki K. and C. Bayer [eds.], *The Families and Genera of Vascular Plants*, 36-56. Springer-Verlag, Berlin.

KOCH M. A. AND K. MUMMENHOFF. 2006. Evolution and phylogeny of the Brassicaceae. *Plant Systematics and Evolution* 259: 81-83.

KOSUGI S. AND Y. OHASHI. 1997. *PCF1* and *PCF2* specifically bind to *cis* elements in the rice proliferating cell nuclear antigen gene. *Plant Cell* 9: 1607-1619.

KOSUGI S. AND Y. OHASHI. 2002. DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant Journal* 30: 337-348.

KOYAMA T., F. SATO, AND M. OHME-TAKAGI. 2010. A role of *TCP1* in the longitudinal elongation of leaves in *Arabidopsis*. *Bioscience Biotechnology and Biochemistry* 74: 2145-2147.

KRAMER E. M., L. HOLAPPA, B. GOULD, M. A. JARAMILLO, D. SETNIKOV, AND P. M. SANTIAGO. 2007. Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia*. *Plant Cell* 19: 750-766.

KRAMER E., M. JARAMILLO, AND V. DI STILIO. 2004. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* 166: 1011-1023.

LENSER T. AND G. THEISSEN. 2013. Conservation of fruit dehiscence pathways between *Lepidium campestre* and *Arabidopsis thaliana* sheds light on the regulation of *INDEHISCENT*. *Plant Journal* 76: 545-556.

LEWIS M. W., M. E. LESLIE, AND S. J. LILJEGREN. 2006. Plant separation: 50 ways to leave your mother. *Current Opinion in Plant Biology* 9: 59-65.

LIDGARD S. AND P. R. CRANE. 1990. Angiosperm diversification and Cretaceous floristic trends - a comparison of palynofloras and leaf macrofloras. *Paleobiology* 16: 77-93.

LITT A. AND V. IRISH. 2003. Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: Implications for the evolution of floral development. *Genetics* 165: 821-833.

LORTS C. M., T. BRIGGEMAN, AND TAO SANG. 2008. Evolution of fruit types and seed dispersal: A phylogenetic and ecological snapshot. *Journal of Systematics and Evolution* 46: 396-404.

LUO D., R. CARPENTER, L. COPSEY, C. VINCENT, J. CLARK, AND E. COEN. 1999. Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* 99: 367-376.

LUO D., R. CARPENTER, C. VINCENT, L. COPSEY, AND E. COEN. 1996. Origin of floral asymmetry in *Antirrhinum*. *Nature* 383: 794-799.

LYNCH M. AND J. CONERY. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151-1155.

LYSAK M., M. KOCH, A. PECINKA, AND I. SCHUBERT. 2005. Chromosome triplication found across the tribe Brassiceae. *Genome Research* 15: 516-525.

LYSAK M. A., K. CHEUNG, M. KITSCHKE, AND P. BURES. 2007. Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiology* 145: 402-410.

MADDISON W. 1997. Gene trees in species trees. *Systematic Biology* 46: 523-536.

MARTIN-TRILLO M. AND P. CUBAS. 2010. TCP genes: a family snapshot ten years later. *Trends in Plant Science* 15: 31-39.

MCKENNA D. D. 2011. Towards a temporal framework for "inordinate fondness": reconstructing the macroevolutionary history of beetles (Coleoptera). *Entomologica Americana* 117: 28-36.

MENA M., B. A. AMBROSE, R. B. MEELEY, S. P. BRIGGS, M. F. YANOFSKY, AND R. J. SCHMIDT. 1996. Diversification of C-function activity in maize flower development. *Science* 274: 1537-1540.

MING R., S. HOU, Y. FENG, Q. YU, A. DIONNE-LAPORTE, J. H. SAW, P. SENIN, W. WANG, B. V. LY, K. L. T. LEWIS, S. L. SALZBERG, L. FENG, M. R. JONES, R. L. SKELTON, J. E. MURRAY, C. CHEN, W. QIAN, J. SHEN, P. DU, M. EUSTICE, E. TONG, H. TANG, E. LYONS, R. E.

PAULL, T. P. MICHAEL, K. WALL, D. W. RICE, H. ALBERT, M. WANG, Y. J. ZHU, M. SCHATZ, N. NAGARAJAN, R. A. ACOB, P. GUAN, A. BLAS, C. M. WAI, C. M. ACKERMAN, Y. REN, C. LIU, J. WANG, J. WANG, J. NA, E. V. SHAKIROV, B. HAAS, J. THIMMAPURAM, D. NELSON, X. WANG, J. E. BOWERS, A. R. GSCHWEND, A. L. DELCHER, R. SINGH, J. Y. SUZUKI, S. TRIPATHI, K. NEUPANE, H. WEI, B. IRIKURA, M. PAIDI, N. JIANG, W. ZHANG, G. PRESTING, A. WINDSOR, R. NAVAJAS-PEREZ, M. J. TORRES, F. A. FELTUS, B. PORTER, Y. LI, A. M. BURROUGHS, M. LUO, L. LIU, D. A. CHRISTOPHER, S. M. MOUNT, P. H. MOORE, T. SUGIMURA, J. JIANG, M. A. SCHULER, V. FRIEDMAN, T. MITCHELL-OLDS, D. E. SHIPPEN, C. W. DEPAMPHILIS, J. D. PALMER, M. FREELING, A. H. PATERSON, D. GONSALVES, L. WANG, AND M. ALAM. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452: 991-1007.

MOLLER A. P. AND M. ERIKSSON. 1995. Pollinator preference for symmetrical flowers and sexual selection in plants. *Oikos* 73: 15-22.

MOORE R. AND M. PURUGGANAN. 2005. The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology* 8: 122-128.

MOREAU C. S., C. D. BELL, R. VILA, S. B. ARCHIBALD, AND N. E. PIERCE. 2006. Phylogeny of the ants: diversification in the age of angiosperms. *Science* 312: 101-104.

MUEHLHAUSEN A., T. LENSER, K. MUMMENHOFF, AND G. THEISSEN. 2013. Evidence that an evolutionary transition from dehiscent to indehiscent fruits in *Lepidium* (Brassicaceae) was caused by a change in the control of valve margin identity genes. *Plant Journal* 73: 824-835.

NEI M. AND T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 3: 418-426.

OHNO S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, New York.

OLIVEIRA A. E. A., D. BELLIENY-RABELO, AND T. M. VENANCIO. 2013. Impact of whole-genome and tandem duplications in the expansion and functional diversification of the F-Box family in legumes (Fabaceae). *PLoS One* 8 (2): e55127.

OLMSTEAD R. G. 2013. Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: a comparison of continental and intercontinental diversification patterns. *Botanical Journal of the Linnean Society* 171: 80-102.

OTTO S. P. AND J. WHITTON. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401-437.

PARK K. I., J. D. CHOI, A. HOSHINO, Y. MORITA, AND S. IIDA. 2004. An intragenic tandem duplication in a transcriptional regulatory gene for anthocyanin biosynthesis confers

pale-colored flowers and seeds with fine spots in *Ipomoea tricolor*. *Plant Journal* 38: 840-849.

PARKIN I., S. GULDEN, A. SHARPE, L. LUKENS, M. TRICK, T. OSBORN, AND D. LYDIATE. 2005. Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171: 765-781.

PATERSON A. H., J. E. BOWERS, AND B. A. CHAPMAN. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences of the United States of America* 101: 9903-9908.

PHILLIPS M. J., F. DELSUC, AND D. PENNY. 2004. Genome-scale phylogeny and the detection of systematic biases. *Molecular Biology and Evolution* 21: 1455-1458.

PINYOPICH A., G. S. DITTA, B. SAVIDGE, S. J. LILJEGREN, E. BAUMANN, E. WISMAN, AND M. F. YANOFSKY. 2003. Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424: 85-88.

PRESTON J. C. AND L. C. HILEMAN. 2012. Parallel evolution of TCP and B-class genes in Commelinaceae flower bilateral symmetry. *Evodevo* 3: 6.

PRESTON J. C., L. C. HILEMAN, AND P. CUBAS. 2011. Reduce, reuse, and recycle: developmental evolution of trait diversification. *American Journal of Botany* 98: 397-403.

RIPOLL J. J., A. H. K. ROEDER, G. S. DITTA, AND M. F. YANOFSKY. 2011. A novel role for the floral homeotic gene *APETALA2* during *Arabidopsis* fruit development. *Development* 138: 5167-5176.

RODGERS-MELNICK E., S. P. MANE, P. DHARMAWARDHANA, G. T. SLAVOV, O. R. CRASTA, S. H. STRAUSS, A. M. BRUNNER, AND S. P. DIFAZIO. 2012. Contrasting patterns of evolution following whole genome versus tandem duplication events in Populus. *Genome Research* 22: 95-105.

RODRIGUEZ I., A. GUMBERT, N. H. DE IBARRA, J. KUNZE, AND M. GIURFA. 2004. Symmetry is in the eye of the 'beeholder': innate preference for bilateral symmetry in flower-naive bumblebees. *Naturwissenschaften* 91: 374-377.

ROELANTS K., D. J. GOWER, M. WILKINSON, S. P. LOADER, S. D. BIJU, K. GUILLAUME, L. MORIAU, AND F. BOSSUYT. 2007. Global patterns of diversification in the history of modern amphibians. *Proceedings of the National Academy of Sciences of the United States of America* 104: 887-892.

ROSIN F. M. AND E. M. KRAMER. 2009. Old dogs, new tricks: Regulatory evolution in conserved genetic modules leads to novel morphologies in plants. *Developmental Biology* 332: 25-35.

RUTLEDGE R., S. REGAN, O. NICOLAS, P. FOBERT, C. COTE, W. BOSNICH, C. KAUFFELDT, G. SUNOHARA, A. SEGUIN, AND D. STEWART. 1998. Characterization of an *AGAMOUS* homologue from the conifer black spruce (*Picea mariana*) that produces floral homeotic conversions when expressed in *Arabidopsis*. *Plant Journal* 15: 625-634.

SCHRANZ M. E. AND T. MITCHELL-OLDS. 2006. Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* 18: 1152-1165.

SEYMOUR G. B., L. OSTERGAARD, N. H. CHAPMAN, S. KNAPP, AND C. MARTIN. 2013. Fruit development and ripening. *Annual Review of Plant Biology* 64: 219-241.

SIMILLION C., K. VANDEPOELE, M. C. E. VAN MONTAGU, M. ZABEAU, AND Y. VAN DE PEER. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* 99: 13627-13632.

SOLTIS D. E., V. A. ALBERT, J. LEEBENS-MACK, C. D. BELL, A. H. PATERSON, C. ZHENG, D. SANKOFF, C. W. DEPAMPHILIS, P. K. WALL, AND P. S. SOLTIS. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336-348.

SPECHT C. D. AND M. E. BARTLETT. 2009. Flower evolution: the origin and subsequent diversification of the angiosperm flower. *Annual Review of Ecology Evolution and Systematics* 40: 217-243.

STEVENS P. F. 2012. Missouri Botanical Garden. http://www.mobot.org/MOBOT/research/APweb/ [Accessed April/01 2014].

SZOELLOSI G. J. AND V. DAUBIN. 2012. Modeling gene family evolution and reconciling phylogenetic discord. *Evolutionary Genomics: Statistical and Computational Methods, Vol 2* 856: 29-51.

TANDRE K., V. A. ALBERT, A. SUNDAS, AND P. ENGSTROM. 1995. Conifer homologs to genes that control floral development in angiosperms. *Plant Molecular Biology* 27: 69-78.

THEISSEN G. 2001. Development of floral organ identity: stories from the MADS house. *Current Opinion in Plant Biology* 4: 75-85.

THOMAS B. C., B. PEDERSEN, AND M. FREELING. 2006. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homolog leaving clusters enriched in dose-sensitive genes. *Genome Research* 16: 934-946.

VANDER WALL S. B. AND M. J. BECK. 2012. A comparison of frugivory and scatter-hoarding seed-dispersal syndromes. *Botanical Review* 78: 10-31.

VEIT B., E. VOLLBRECHT, J. MATHERN, AND S. HAKE. 1990. A tandem duplication causes the Kn1-0 allele of *KNOTTED*, a dominant morphological mutant of maize. *Genetics* 125: 623-631.

VEKEMANS D., S. PROOST, K. VANNESTE, H. COENEN, T. VIAENE, P. RUELENS, S. MAERE, Y. VAN DE PEER, AND K. GEUTEN. 2012. Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-box gene and species diversification. *Molecular Biology and Evolution* 29: 3793-3806.

WANG X., H. WANG, J. WANG, R. SUN, J. WU, S. LIU, Y. BAI, J. MUN, I. BANCROFT, F. CHENG, S. HUANG, X. LI, W. HUA, J. WANG, X. WANG, M. FREELING, J. C. PIRES, A. H. PATERSON, B. CHALHOUB, B. WANG, A. HAYWARD, A. G. SHARPE, B. PARK, B. WEISSHAAR, B. LIU, B. LI, B. LIU, C. TONG, C. SONG, C. DURAN, C. PENG, C. GENG, C. KOH, C. LIN, D. EDWARDS, D. MU, D. SHEN, E. SOUMPOUROU, F. LI, F. FRASER, G. CONANT, G. LASSALLE, G. J. KING, G. BONNEMA, H. TANG, H. WANG, H. BELCRAM, H. ZHOU, H. HIRAKAWA, H. ABE, H. GUO, H. WANG, H. JIN, I. A. P. PARKIN, J. BATLEY, J. KIM, J. JUST, J. LI, J. XU, J. DENG, J. A. KIM, J. LI, J. YU, J. MENG, J. WANG, J. MIN, J. POULAIN, K. HATAKEYAMA, K. WU, L. WANG, L. FANG, M. TRICK, M. G. LINKS, M. ZHAO, M. JIN, N. RAMCHIARY, N. DROU, P. J. BERKMAN, Q. CAI, Q. HUANG, R. LI, S. TABATA, S. CHENG, S. ZHANG, S. ZHANG, S. HUANG, S. SATO, S. SUN, S. KWON, S. CHOI, T. LEE, W. FAN, X. ZHAO, X. TAN, X. XU, Y. WANG, Y. QIU, Y. YIN, Y. LI, Y. DU, Y. LIAO, Y. LIM, Y. NARUSAKA, Y. WANG, Z. WANG, Z. LI, Z. WANG, Z. XIONG, AND Z. ZHANG. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature genetics* 43: 1035-1040.

WANG Z., Y. LUO, X. LI, L. WANG, S. XU, J. YANG, L. WENG, S. SATO, S. TABATA, M. AMBROSE, C. RAMEAU, X. FENG, X. HU, AND D. LUO. 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: 10414-10419.

WARWICK S. AND C. SAUDER. 2005. Phylogeny of tribe Brassiceae (Brassicaceae) based on chloroplast restriction site polymorphisms and nuclear ribosomal internal transcribed spacer and chloroplast *trnL* intron sequences. *Canadian Journal of Botany-Revue Canadienne De Botanique* 83: 467-483.

ZHANG P. Y., H. T. W. TAN, K. H. PWEE, AND P. P. KUMAR. 2004. Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. *Plant Journal* 37: 566-577.

ZHANG W., E. M. KRAMER, AND C. 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: 6388-6393.

ZHANG W., Q. XIANG, D. T. THOMAS, B. M. WIEGMANN, M. W. FROHLICH, AND D. E. SOLTIS. 2008. Molecular evolution of *PISTILLATA*-like genes in the dogwood genus *Cornus* (Cornaceae). *Molecular Phylogenetics and Evolution* 47: 175-195.

Chapter 2 – *FRUITFULL* evolution in Brassiceae (Brassicaceae): insight into the diversification of fruit morphology

Introduction

The sessile nature of plants has resulted in incredible variation of fruit and floral morphology by which plants disperse their genes and offspring. Although floral variation is better studied, fruit structure is also a remarkably labile trait within angiosperms that provides a natural opportunity to identify the genetic alterations responsible for morphological change (Bolmgren and Eriksson, 2005; Lorts et al., 2008; Fourquin et al., 2013). Applying phylogenetic analyses to candidate genes identified in model organisms is a starting point for understanding whether gene duplications or subsequent gene loss (Lee and Irish, 2011) may be correlated with variable morphologies (Cronk et al., 2002). When combined with organismal phylogenies, this approach reconciles the evolutionary history of a candidate gene with a map of morphological characters exhibited amongst closely related species. This method has been useful in discerning the phylogenetic placement of duplication events (Howarth and Donoghue, 2006), identifying whether gene families are prone to retention (Litt and Irish, 2003; Kramer et al., 2004; Nei and Rooney, 2005) or loss (Baum et al., 2005), and planning future investigations into the functional differences between retained copies (Citerne et al., 2003). Furthermore, software allowing us to measure selective pressures is a valuable tool for generating hypotheses for the role of duplicated genes (Zhang et al., 2008) because different rates and patterns of evolution are often associated with variation in function amongst paralogues (Pabon-Mora et al., 2013a). For example, studies associating relaxed selection on B-class floral organ identity genes with petal loss in the Piperales (Jaramillo and Kramer, 2007) and positive selection on

CYCLOIDEA duplicates with diversification of floral structure in the Asteraceae (Chapman et al., 2008) demonstrate the value of identifying selective pressures on genes contributing to morphological change.

The tribe Brassiceae (Brassicaceae), containing 50 genera and 240 species (Warwick and Sauder, 2005) is an excellent model to examine gene duplications because a hexaploid event (genome triplication) occurred at the base of this tribe 7.9-14.6 million years ago (Lysak et al., 2005; Parkin et al., 2005; Lysak et al., 2007). Brassiceae has long been considered monophyletic (Hedge, 1976; Koch et al., 2001; Koch et al., 2003; Appel and Al-Shehbaz, 2003; Al-Shehbaz et al., 2006), and uncontestably holds the most economic importance of all tribes in the family, containing canola, cole-crops (broccoli, cauliflower, kale, Brussels sprouts, cabbage), mustard, radish and salad green species. Furthermore, the close phylogenetic position of this tribe to the model species *Arabidopsis thaliana* (Camelineae; Brassicaceae) (O'Kane and Al-Shehbaz, 2003; Al-Shehbaz et al., 2006; Beilstein et al., 2006; Al-Shehbaz, 2012), allows for more informed supposition of which genes may be involved in the evolution of morphologically diverse traits in wild relatives. Relationships within the Brassiceae are becoming better understood, although outstanding issues remain. Chloroplast and nuclear phylogenies are incongruent, although eight lineages including the Cakile, Crambe, Henophyton, Nigra, Rapa/Oleracea, Savignya, Vella, and Zilla lineages have been identified in both topologies (Warwick and Sauder, 2005; Hall et al., 2011; Arias and Pires, 2012).

Within the Brassiceae, over half of the members possess a unique fruit type called heteroarthrocarpy, in which a joint laterally divides the fruit into a proximal segment and a distal segment (Appel, 1999; Hall et al., 2006; Hall et al., 2011) (Figure 2-1). This trait has

evolved independently within the tribe at least twice with subsequent losses (Hall et al., 2011) and is hypothesized to be the result of a downward shift of the distal portion of the valve margin such that the valves do not extend over the distal portion of the fruit (Hall et al., 2006). Most non-heteroarthrocarpic members of the tribe exhibit the same fully dehiscent silique fruit as in *Arabidopsis* where the valves extend along the entire length of the fruit (Dinneny et al., 2005), which is the ancestral state for the tribe (Hall et al., 2011). Heteroarthrocarpy results in ovules being present in both the proximal and distal segments, allowing the plant two potential methods of seed dispersal due to the presence of the joint (Figure 2-1). The joint abscises (dehisces laterally) in some species, while the joint remains intact in other species. Furthermore, variable longitudinal dehiscence exists in heteroarthrocarpic fruits where the proximal segment is either dehiscent or has become fully indehiscent, in contrast to the distal segment which is invariably indehiscent. Thus, heteroarthrocarpic fruits fit into two categories regarding longitudinal dehiscence: partially indehiscent, with a dehiscent proximal segment and an indehiscent distal segment, and fully indehiscent, with both proximal and distal segments indehiscent. However, indehiscence is not limited to heteroarthrocarpic taxa; species such as *Raffenaldia primuloides* and *Zilla spinosa* have non-heteroarthrocarpic fruits that are fully indehiscent (Figure 2-1). Indehiscence and joint abscission reveal a complex evolutionary history with apparent multiple gains and losses of both across the tribe (Hall et al., 2011). This emphasizes that evolution of the joint and longitudinal indehiscence is independent, although these traits are correlated.

The thorough characterization of the fruit dehiscence gene pathway in *Arabidopsis* provides a framework to study evolution of these genes in related species (Hall and

Donohue, 2012). Fruits in *Arabidopsis* have an ovary wall consisting of two valves that dehisce from a persistent placental tissue called the replum (Dinneny et al., 2005). This boundary where the valves meet the replum is called the valve margin and is characterized by two narrow strips of cells called the lignification and separation layers, which are necessary for proper dehiscence when the valve margin develops into the dehiscence zone at fruit maturity (Figure 2-2). Previous studies identified four genes expressed in the valve margin: *SHATTERPROOF1* (*SHP1*) and *SHATTERPROOF2* (*SHP2*), which are recently-duplicated genes that act redundantly to regulate *INDEHISCENT* (*IND*) and *ALCATRAZ* (*ALC*), which are responsible for the formation of the lignification and separation layers (Liljegren et al., 2000; Liljegren et al., 2004; Dinneny et al., 2005). *FRUITFULL* (*FUL*), which is expressed in the valves, and *REPLUMLESS* (*RPL*), which is expressed in the replum, act as valve margin positioning genes by negatively regulating and restricting *SHP1/2*, *IND* and *ALC* expression to the valve margin between the valves and the replum (Ferrandiz et al., 2000c; Roeder et al., 2003; Liljegren et al., 2004). Together, these six genes collectively establish the dehiscence zone and mutant forms of any gene member (double mutant in the case of *SHP1/*2) result in an indehiscent silique. Studies show that expression of *IND, RPL* and *FUL* is conserved in Brassiceae member *Brassica napus,* although copy number of *FUL* was not investigated and the entire pathway is yet to be uncovered (Tan et al., 2009; Hua et al., 2009; Girin et al., 2010). Upstream genes promoting *FUL* and valve margin identity gene expression have been identified as *JAGGED* (*JAG*), *YABBY* (*YAB3*) and *FILAMENTOUS FLOWER* (*FIL*), which are negatively regulated by *RPL* (Dinneny et al., 2005).

With rare exceptions like *R. primuloides* and *Z. spinosa* that have non-heteroarthrocarpic but fully indehiscent fruits, some form of dehiscence is always present

across Brassiceae as even fully indehiscent heteroarthrocarpic fruits maintain joint dehiscence. Thus, genes of the fruit dehiscence pathway are likely candidates controlling the remarkable variability of dehiscence patterning found within the tribe. Of the six *Arabidopsis* dehiscence pathway genes investigated in heteroarthrocarpic species *Erucaria erucarioides* and *Cakile lanceolata*, only *FUL* was present in two copies (aside from the recently duplicated *SHP1/*2) (Avino et al., 2012). Two copies of *SHP2* and *ALC* have been found in the heteroarthrocarpic *Brassica napus*, although this is likely due to a second duplication event that occurred by hybridizing *B. rapa* and *B. oleracea* during cultivation to produce *B. napus* (Hua et al., 2009; Tan et al., 2009; Girin et al., 2010). The retention of multiple copies of *FUL* warrants further examination because it has been proposed that gene families that retain duplicates may underlie ecologically significant traits (Wang et al., 2011). Importantly, the expression domain of two *FUL* genes found in the partially indehiscent *E. erucarioides* differed from the expression domain of the single *FUL* copy in *Arabidopsis*, whereas *SHP1/SHP2, IND* and *ALC* revealed conservation in expression patterns (Avino et al., 2012). Interestingly, the expression domain of one copy covered only the dehiscent proximal segment, mirroring the downward shift of the distal portion of the valve margin in heteroarthrocarpic fruits. In contrast, no expression of *FUL* was observed in the valve of fully indehiscent *Cakile*. These data indicate that the function of *FUL* is partially conserved in heteroarthrocarpic Brassiceae and is likely necessary for dehiscence in the proximal segment, although the origin of the joint and variable longitudinal dehiscence in heteroarthrocarpic fruits has not been fully explained.

FUL structure has been described in *Arabidopsis*, which contains eight exons and seven introns (Swarbreck et al., 2008). As introns account for more than two-thirds the

length of the genomic sequence, Brassicaceae *FUL* genes make good candidates for the study of intron evolution. Furthermore, cDNA or genes with few introns are commonly employed in gene phylogenies (Citerne et al., 2003; Skipper et al., 2005; Howarth and Donoghue, 2006; Jaramillo and Kramer, 2007) such that mapping presence/absence and length of introns in a phylogenetic context amongst closely related species is conspicuously absent in the literature, with the exception of He et al. (2013). No attempts have been made to correlate differences in gene structure with the evolution of morphological characters, although several studies have implicated introns as having important regulatory functions (Deyholos and Sieburth, 2000; Osnato et al., 2010; Parra et al., 2011), and a few studies have documented a complex history of intron gain and loss in model organisms over broad evolutionary scales (Nielsen et al., 2004; Knowles and McLysaght, 2006; Roy and Penny, 2006; Coulombe-Huntington and Majewski, 2007).

The duplication event resulting in the presence of *FUL* and *APETALA1* (*AP1*) in core eudicots has been comprehensively investigated (Litt and Irish, 2003). Studies comparing the expression and function of these genes before and after this important duplication have shown that *FUL/AP1* genes maintained functions similar to *FUL-like* genes found in the monocots and basal eudicots but have subfunctionalized their roles in flowering time, floral meristem and perianth identity, leaf morphogenesis and fruit development (Gu et al., 1998; Ferrandiz et al., 2000a; Alvarez-Buylla et al., 2006; Litt, 2007; Pabon-Mora et al., 2012; Pabon-Mora et al., 2013b). In *Arabidopsis,* expression and function data confirm the role of *FUL* in the development of organs other than carpel/fruit development, including flowering time, leaf development and meristem identity (Gu et al., 1998; Ferrandiz et al., 2000a; Chu et al., 2010).

Examination of *FUL* in a phylogenetic context will address four questions: 1) Are patterns of *FUL* evolution in the Brassiceae consistent with the proposed hexaploid event at the base of the tribe (Lysak et al., 2005), or have there been additional duplication events? 2) Is there a correlation between copy number and heteroarthrocarpy? 3) Is there a correlation between gene structure and fruit morphology? 4) Do patterns of molecular evolution indicate variable selection amongst *FUL* lineages, suggesting alternative fates for different copies, and furthermore are patterns of molecular evolution correlated with heteroarthrocarpy or type of longitudinal dehiscence? Mapping fruit diversity present in the Brassiceae alongside the evolutionary history of *FUL* is an important first step to determining whether the duplication or sequence divergence of *FUL* genes is an important factor driving fruit evolution in the Brassiceae.

Methods

Taxon sampling

Twenty-nine species were sampled (Table 2-1) from the Cakile, Crambe, Henophyton, Nigra, Rapa/Oleracea, Savignya, Vella, and Zilla lineages (Warwick and Sauder, 2005; Hall et al., 2011; Arias and Pires, 2012). Sampled taxa consisted of 14 species with heteroarthrocarpic fruits and 15 species with non-heteroarthrocarpic fruits, reflecting the overall diversity in longitudinal dehiscence in the tribe. *FUL* cDNA sequences were obtained from GenBank for *Sinapis alba*, *Brassica oleracea* var. *botrytis*, *Cakile lanceolata* and *Erucaria erucarioides* to supplement Brassiceae sampling and to determine exon/intron boundaries. Four Brassicaceae GenBank sequences from outside of Brassiceae were included in the phylogeny: 1) *Lepidium appelianum* for its indehiscent fruits, 2) *Lepidium campestre* for its dehiscent fruits (Mummenhoff et al., 2009), 3) *Aethionema*

carneum, because this genus is sister to all remaining Brassicaceae (Al-Shehbaz et al., 2006; Al-Shehbaz, 2012) and 4) *Arabidopsis thaliana* due to knowledge about its dehiscence pathway. One further cDNA sequence from *Cleome violacea,* belonging to the Cleomaceae*,* was also included as an outgroup because this family is sister to Brassicaceae (Iltis et al., 2011). GenBank accession numbers are listed in Appendix 2-1.

Amplification, cloning and sequencing

FUL was amplified from genomic DNA from plants grown from seed in a greenhouse and extracted according to Hall et al. (2011). A fragment ranging from 1255 – 2671 bp (Figure 2-3) was generated using PCR with degenerate primers FULexon2F and FULexon8R (Table 2-2, Figure 2-3). These primers were designed to anneal to regions of the 2nd and 8th exons that are conserved amongst the available Brassicaceae GenBank sequences: *Brassica napus*, *Brassica oleracea* var*. botrytis* and Brassiceae members *E. erucarioides* and *Cakile lanceolata* (Appendix 2-1).

The PCR contained 1X Phire II reaction buffer, 200μM dNTPs, 0.5μM forward and reverse primer (Table 2-2), 1 unit of Phire Hot Start II polymerase (Thermo Fisher Scientific, Waltham, Massachusetts, USA), 30ng of total genomic DNA, adjusted to a final volume of 20μL using sterile distilled water. Six samples per template were placed in a thermocycler under hot start conditions with the cycling settings as follows: initial denaturation for 30 seconds at 98°C, followed by 30 cycles of 98°C for 5 seconds, 55–65°C temperature gradient for 20 seconds, 72°C for 3 minutes, followed by a final extension of 72°C for 10 minutes. Due to a high degree of variation in length of *FUL* copies, the number of bands visible after gel electrophoresis was used as an indicator of successful copy

number retrieval. If zero or 1 band was visible, another PCR using primers FULexon2F-2 and FULexon8R-2 (Table 2-2) was completed using the same protocol as above. These degenerate primers were designed to anneal to different regions along the 2nd and 8th exons to potentially capture additional copies containing mutations at FULexon2F and FULexon8R binding sites. Products from both the 1st and 2nd PCRs were then pooled. PCR products were cleaned using QIAquick PCR purification kit (Qiagen, Hilden, Germany) and a-tailed using 2 units of Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, California, USA), 1X PCR buffer, 1.5 mM magnesium chloride, 5.5μL of cleaned PCR product of varying concentration and 10μL of distilled water. This reaction was placed in the thermocycler at 72°C for 45 minutes.

A-tailed products were then ligated into a pCR4-TOPO vector (Invitrogen, Carlsbad, California, USA) and transformed into OneShot Top10 chemically competent *E. coli* cells (Invitrogen, Carlsbad, California, USA) as per manufacturer protocol. Twenty-five to 50μL of transformed cells were plated onto LB agar plates containing 0.5μL carbomycin /1 mL LB agar and 0.8mg X-Gal (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and incubated overnight at 37°C. Twenty-four colonies were screened using M13F and M13R primers and *Taq* DNA Polymerase (Qiagen, Hilden, Germany) as per the manufacturer's instructions while using a small amount of the colony as template. Products of this PCR were then digested using *Eco*RI and *Bgl*II restriction enzymes because use of these enzymes resulted in different patterns of digested DNA after gel electrophoresis and, thus, provided a useful screening tool when selecting which clones to sequence. Eight – 16 clones were grown for 16 hours in 3 mL LB before purifying the plasmids using GeneJet Plasmid miniprep kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Plasmids

were then sequenced using the M13F primer on an ABI 3730 capillary sequencer (Applied Biosystems, Inc., Foster City, California, USA) and assessed for base-calling errors using MacVector v. 12.1 (Rastogi, 2000). *FUL* sequences that differed by a single base pair were sequenced fully using M13R, FULexon4-F, and FULexon6-F primer (Table 2-2). If FULexon4-F and FULexon6-F failed, then combinations of FULexon3-F, FULexon4-R, FULexon5-F, FULexon5-R were used to read the entire sequence (Table 2-2). At least four sequences were entirely sequenced per taxon.

To assess this approach for recovering multiple copies of *FUL*, an additional 72 colonies were PCR-screened from each of eight randomly selected species (Table 2-1) and digested using restriction enzymes. Clones with a different pattern of digested DNA after gel electrophoresis when compared to already sequenced clones were sequenced in attempt to identify copies that may have been missed during initial sequencing. This approach has been shown to be effective in finding gene copies when compared to southern blot hybridizations in other families (Howe et al., 1998; Li and Chinnappa, 2003; Zhang et al., 2010). Restriction enzymes were chosen based on available sequence information. In all instances, clones that had non-recognizable patterns were either identical to copies that were already sequenced or were non-target amplicons. A pictorial representation of amplification, cloning, and sequencing methods can be viewed in Appendix 2-2. Southern blot hybridizations to confirm copy number (Southern, 1975) were attempted without success; the protocol for this technique as well as the sampled taxa and probe primer sequence is presented in Appendices 2-3 to 2-5.

Sequence alignment

Four alignments were generated for subsequent analyses: 1) nucleotide coding plus intron sequence including all fully sequenced clones, hereon called "all clone coding + intron;" 2) nucleotide coding sequence including only one representative sequence per monophyletic group of clones per accession, hereon called "coding only;" 3) nucleotide translated to amino acid sequence including only one representative sequence per monophyletic group of clones, hereon called "amino acid;" and 4) nucleotide coding sequence including only one representative sequence per monophyletic group of clones plus introns, hereon called "coding + intron". Intron/exon boundaries were inferred using coding sequences obtained from GenBank and introns were removed to prepare the nucleotide coding and amino acid alignments. Nucleotide coding sequences were positioned in the proper reading frame and were transformed into amino acids using the "align DNA to protein" command in Mesquite v2.75 (Maddison and Maddison, 2009). Amino acid sequences were aligned by hand using a published *FUL* alignment (Litt and Irish, 2003) as a reference. Nucleotide coding sequences were then aligned according to the amino acid alignment in Mesquite v2.75, allowing easier identification of positional homology. Intron sequences were aligned individually using Clustal X (Thompson et al., 1997) and re-inserted into the coding alignment because large gaps were problematic and interrupted the reading frame of the coding sequence. Introns 3 and 6 were discarded from all phylogenetic analyses because they were too divergent to generate a reliable alignment.

Phylogenetic analysis

An initial analysis using the all clone coding + intron alignment was performed using the alignment of 136 ingroup and five outgroup sequences in MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001). A model of sequence evolution was determined separately for codon positions of the coding sequence and introns using the Akaike Information Criterion in ModelTest v3.06 (Posada and Crandall, 1998). The models selected for all codon positions and for introns were GTR + I and GTR + I + Γ, respectively. The Bayesian Markov chain Monte-Carlo (MCMC) analysis was run using eight chains and temperature of 0.1. The analysis ran for 10,000,000 generations at which the standard deviation of the split frequency was < 0.01 and the potential scale reduction factors were approximately 1. Stationarity of the run was confirmed using Tracer v1.4.1 (Rambaut and Drummond, 2007). Posterior probabilities (PP) were obtained from a 50% majority rule consensus tree wherein the first 25% of the trees were discarded as burnin. This tree was used to choose one sequence from each monophyletic group to be included in the other three alignments (see sequence alignments above). Thus, all subsequent analyses were completed using only one representative sequence.

A second Bayesian analysis was run using the coding + intron alignment using GTR + I model of evolution for the coding sequence and GTR + I + Γ for introns, with 4 chains and default temperature. The coding + intron alignment was subjected to PartitionFinder (data not shown) to ensure that potential over- or under-parameterization during *a priori* partitioning of the dataset did not affect the topology (Lanfear et al., 2012). Both the coding only and amino acid alignments were subjected to Bayesian analyses to determine how the inclusion of introns or amino acid translation may influence resolution topology.

Because the coding + intron alignment provided best resolution, maximum parsimony (MP) and maximum likelihood (ML) bootstrapping analyses were also conducted on this alignment to provide additional measures of branch support. MP heuristic searches were conducted in PAUP* v4.0β10 (Swofford, 2003) starting from random trees and using 100 random addition replicates, tree-bisection-reconnection (TBR) branch swapping and holding 10 trees per replicate. Bootstrapping (BS) was calculated using 1000 replicates, TBR swapping, random addition and saving no more than 1000 trees per replicate. ML analyses using the same partitioning scheme as in Bayesian analyses were run through GARLi v1.0 (Zwickl, 2006) allowing for estimation of model parameters and using 10 000 trees / search and starting from random trees. Bootstrap values were determined from 100 replicates of the ML search. A Templeton test comparing the coding + intron topology to an alternate topology in which two *FUL* lineages were united was conducted to test the possibility of further duplications in the tribe since triplication (Wang et al., 2000).

Descriptive statistics

A Mann-Whitney U test was conducted to assess statistical support for intron length differences in species with heteroarthrocarpic fruits compared to species with non-heteroarthrocarpic fruits. A Kruskal–Wallis one-way analysis of variance was performed in SPSS v19.0 (IBM Corp., 2012) to assess statistical support for whether intron lengths were different among clades and whether intron lengths varied among plants with different types of longitudinal dehiscence. Post hoc Mann-Whitney U tests were then conducted to verify which factors, if any, were responsible for significance. Bonferroni adjustments were applied to account for multiple comparisons (Dunn, 1961). Analysis of the data using

phylogenetic independent contrasts (Felsenstien, 1985) was not conducted due to lack of resolution amongst species relationships within the Brassiceae.

Molecular evolution tests

PAML v4.4 (Yang et al., 2000) was used to estimate the ratio of nonsynonymous to synonymous substitutions (dN/dS) in a tree-based likelihood framework to address ten questions: 1a) Do major identified lineages have different rates of evolution after duplication? 1b) Are there different rates of evolution among lineages as in 1a, but do putative sister clades remain under similar selective pressures? 1c) Does one particular lineage stand out as having a significantly different rate of evolution when compared to other lineages? 2a) Do branches with heteroarthrocarpic taxa have a different rate of evolution than branches with non-heteroarthrocarpic taxa, indicating that the origin of the joint is an important trait influencing selective pressure on *FUL*? 2b) Do branches with heteroarthrocarpic taxa evolve differently and have the multiple copies of *FUL* possessed by these taxa been subjected to different selective pressures? 3) Do taxa that have any variation in dehiscence, including joint dehiscence or longitudinal dehiscence, have a different rate of evolution than entirely indehiscent taxa *Zilla spinosa* and *Raffenaldia primuloides*? 4a) Do taxa with full longitudinal dehiscence, partial longitudinal dehiscence and full longitudinal indehiscence have different rates of evolution, indicating that longitudinal dehiscence of the valve is an important factor influencing selection? 4b) Do the three types of longitudinal dehiscence have different rates of evolution, and have the multiple copies of *FUL* possessed by these taxa been subjected to different selective pressures? A "fixed" model of evolution wherein all branches were constrained to the same dN/dS ratio as well as a "free" model wherein each branch was allowed to have its own

ratio was also conducted. Comparisons were done using likelihood ratio tests (LRTs), which allow evaluation of whether one model is favoured or whether two models are equally supported by the data while penalizing models with additional parameters (Baum et al., 2005; Jaramillo and Kramer, 2007; Zhang et al., 2008).

Results

Statistics on the data sets

The all clone coding + intron aligned dataset comprising 136 ingroup sequences from 29 species was 1317 bp long. The coding + intron, coding only and amino acid datasets comprised a total of 53 terminal sequences. The aligned length of the coding + intron dataset was 1310 bp with 67% variable characters and 545 (42%) parsimony informative characters. The aligned length of the coding only dataset was 450 bp long with 58% variable characters and 172 (38%) parsimony informative characters.

Phylogenetic analyses

Analyses of all four data sets revealed monophyly of *FUL* Brassiceae sequences (Figure 2-4, Appendix 2-6). In addition, we identified and informally named four clades that were recovered and supported by all analyses: *FULa*, *FULb*, *FULc*, and *FULd.* Bayesian analysis of the coding + intron alignment yielded the most resolution and support of all phylogenetic analyses (Figure 2-4). *FULa*, *FULb*, *FULc*, and *FULd* are well supported by Bayesian (PP=100%) and MP/ML (BS > 90%) analyses. The topology generated from the coding + intron alignment includes a weak branch uniting *FULb* and *FULc* as sister in MP (BS=62%), ML (BS=55%), Bayesian (PP=69%) analyses. The Templeton test comparing the coding + intron topology to an alternative topology constraining *FULa* and *FULb* as

sister, indicating subsequent duplication events after the ancestral hexaploidization event, failed to reject the alternative topology (p-value = 0.29). Bayesian analyses of the coding only and amino acid alignment (not shown) yielded topologies similar to the coding plus intron alignment, but the branch representing *FULb* and *FULc* as sister is collapsed and all within-clade species relationships are unresolved or are weakly supported (PP<78%). Bayesian analysis of the all clone coding + intron alignment resulted in all clones from within each species being monophyletic with 85-100% sequence similarity (Appendix 2-6), with one exception. One sequence from amphiploid *Brassica napus* did not form a monophyletic group with other *B*. *napus* clones but instead was sister to *B*. *napus* and *B. oleracea* within the *FULa* (Figure 2-4) and thus, was not excluded from subsequent analyses. This sequence was 78% similar to the other *B. napus FULa* with 28 differing base pairs in the coding sequence.

Copy number and intron length

Intron lengths amongst *FULa*, *FULb*, *FULc*, and *FULd*, were most variable in the longest introns, with introns 3 and 6 having the largest standard deviations (Figure 2-3, Appendix 2-7). Lineage appeared to influence lengths of introns 3, 4 and 6, as supported by Kruskal-Wallis tests(intron 3: χ^2=10.3, p < 0.044; intron 4: χ^2=13.2, p=0.019;intron 6: χ^2=8.76, p < 0.047). Intron 6 is significantly shorter in *FULb* than other *FUL* lineages and intron 3 is significantly shorter in *FULb* and *FULc* when compared to *FULa* and *FULd*, as indicated by Post hoc Mann-Whitney U tests. Intron 4 is significantly shorter in *FULd* when compared to other *FUL* lineages. Average intron lengths tended to be longer in heteroarthrocarpic taxa, with the exception of intron 4, although Mann-Whitney U tests

revealed that this trend was only significant in intron 5 (p=0.026, r=0.62; Figure 2-5). There was no significant association between intron length or copy number and type of longitudinal dehiscence.

All 15 non-heteroarthrocarpic taxa had either one copy from any *FUL* lineage or two copies from *FULb* and *FULc* (Table 2-3). Having two or three copies from any *FUL* lineage is significantly associated with heteroarthrocarpy . There is no unique combination of genes possessed by heteroarthrocarpic taxa that is consistent, although of the heteroarthrocarpic taxa with two copies, both copies were never recovered from *FULb* and *FULc*. However, of the 14 heteroarthrocarpic taxa sampled in this study, only one *FUL* gene was retrieved from *Coincya monensis*, *Diplotaxis assurgens*, *Guiraoa arvensis*, and *Hemicrambe fruticulosa* (Table 2-3). Of the other heteroarthrocarpic taxa sampled, five taxa had two copies, and five taxa had three copies. These numbers exclude more recent duplications noted in *B. napus* which is a known hybrid and thus, retention of features from the A genome of the *B. rapa* parent as well as the C genome of the *B. oleracea* parent is a unique circumstance in the tribe (Parkin et al., 1995).

Selection tests

Maximum likelihood tests of selection reveal that the free model was not a significantly better fit than the fixed model, indicating that if dN/dS heterogeneity exists in the data it cannot be explained by random divergence of every lineage (Table 2-4). Further model comparisons show that dN/dS is significantly influenced by *FUL* lineage, although this is only indicated when *FULb* and *FULc* were constrained to the same rate of evolution because dN/dS ratios between those lineages are similar. That is, model 1b is a

significantly better fit than the fixed ratio model, while the similar model allowing *FULb* and *FULc* to have separate rates (model 1a) was not a significantly better fit than the fixed model. Model 1c, which allowed *FULa* to be free while *FULb*, *FULc*, and *FULd* were constrained to the same rate of evolution, was not a significantly better fit than the fixed model. This result indicates that too much evolutionary rate variation exists between *FULb/FULc* and *FULd* for this to be a suitable model.

Model comparisons indicate that heteroarthrocarpy is not a significant factor influencing dN/dS. Both model 2a, which allows heteroarthrocarpic branches to differ from non-heteroarthrocarpic branches, and model 2b, which is the same as model 2a but allows for different rates amongst copies, failed to provide a significantly greater fit than the fixed model (Table 2-4).

Model 3, in which taxa with any type of joint or longitudinal dehiscence were allowed to differ from entirely indehiscent taxa, failed to fit the dataset better than the fixed model. Model 4a, allowing for different dN/dS among different types of longitudinal dehiscence, was not significantly different from the fixed model. Model 4b, which was the same as 4a but allowed for different rates amongst multiple copies, was strongly favoured above the fixed model and was the most favoured of all maximum likelihood tests of selection. A second comparison revealed that model 4b is a significantly better fit than model 1b ($p=0.031$), indicating that allowing rates to vary among the types of longitudinal dehiscence in addition to sequence divergence amongst lineages is a better explanation of the data than accounting for sequence divergence alone. Significance of model fit when compared to the fixed model is shown in Table 2-4.

Discussion

Phylogenies focusing on gene and gene family evolution are typically generated from cDNA, thus eliminating opportunities to explore post-duplication evolutionary patterns in introns. Here we present a detailed look at the *FUL* gene in the Brassiceae based on genomic DNA, examining patterns of gene evolution following a known hexaploidization event. Moreover, these data enable us to test hypotheses relating copy number and intron lengths to the presence of a highly specialized fruit type. Similar to other studies, these data reveal a dynamic history of gene loss and gene retention (Palmer et al., 2000; Baum et al., 2005; Coate and Doyle, 2011; Airoldi and Davies, 2012). Specifically, we find (1) four major lineages of *FUL* are identified in the Brassiceae, (2) heteroarthrocarpic taxa tend to retain more copies of the gene than non-heteroarthrocarpic species, and (3) rates of molecular evolution can be partially explained by natural selection acting on patterns of variable dehiscence in the tribe. However, our results reveal trends rather than absolutes, which reflect the complex patterns of fruit morphology and *FUL* evolution in Brassiceae.

Evolution and copy number of *FUL* in Brassiceae

The recovery of four clades, not three, of *FUL* appears contradictory to a single triplication at the base of the tribe that was identified (Lysak et al., 2005; Lysak et al., 2007). We propose two alternatives to the topology derived from the coding + intron alignment in light of this finding. First, only three major lineages of *FUL* exist and the weakly supported unification of *FULb* and *FULc* is an artefact of within-tribe sequence variation. We argue that the data presented here suggest that all copies of *FUL* retrieved

from Brassiceae taxa likely originated from a single hexaploid ancestor (Lysak et al., 2005). This inference is aided by taxon sampling overlap between this study and genomic investigations by Lysak et al. (Lysak et al., 2005) where *Cordylocarpus muricatus*, *M. arvensis*, and *Psychine stylosa*, which are present in either *FULb* or *FULc* in this study, were found to contain just three copies of a ~8.7Mb DNA segment from the *Arabidopsis* genome. No further polyploid events are inferred if *FULa* and *FULb* are sister because both lineages contain only one representative species (e.g., there is no species found in both lineages). Three copies recovered in *Brassica napus, Brassica oleracea, Crambella teretifolia, Enarthrocarpus lyratus* and *Muricaria prostrata* support this hypothesis, as do our data suggesting no single taxon has four copies of *FUL*. In addition, taxa with three copies possessed either *FULa* or *FULb* along with copies from the remaining *FUL* lineages. Furthermore, selection tests revealed that rates of evolution were similar for *FULb* and *FULc*, indicating that natural selection acting on these two clades may be influencing phylogenetic analyses. Alternatively, post-hexaploidization event duplications may have occurred as indicated by the weak branch supporting *FULb* and *FULc* as sister and the presence of *Henophyton deserti*, *Moricandia arvensis*, *Rytidocarpus moricandioides, Pseuderucaria teretifolia* and *Zilla spinosa* in both clades. This duplication can be inferred to be near the origin of the tribe given the distribution of these taxa across many clades in species phylogenies derived from both nuclear and cpDNA data (Hall et al., 2011; Arias and Pires, 2012). Additional studies on *FUL* in other Brassicaceae including *Lepidium* (Muehlhausen et al., 2013) and *Arabidopsis* (Gu et al., 1998; Ferrandiz et al., 2000a; Chu et al., 2010) have found only a single *FUL* copy, implying that retention of multiple *FUL* copies may be exclusive to the Brassiceae tribe despite opportunity for multiple *FUL* copy

retention from previous polyploid events (Vision et al., 2000; Koch et al., 2001; Simillion et al., 2002; Blanc et al., 2003). However, a more thorough examination of *FUL* copy number within the Brassicaceae is needed to ensure whether the hexaploid event in the Brassiceae is the only source of multiple *FUL* copies within the family. Regardless, the combination of multiple copies of *FUL* and the presence of heteroarthrocarpy, which is unique to the Brassiceae, supports the hypothesis that the origin of heteroarthrocarpy and its complex evolution across the tribe may be related to this early hexaploid event (Hall et al., 2011).

In this study, the inclusion of introns in phylogenetic analysis provided improved resolution for species relationships, although this can be challenging to interpret with conflicting topologies of cpDNA and nuclear-based phylogenies. Similar to the *Phytochrome A* nuclear phylogeny (Hall et al., 2011), the *FUL* topology showed that Nigra and Rapa/Oleracea lineages are not monophyletic. The two representatives included from the Zilla lineage, however, are monophyletic, as is the Cakile lineage with the exception of one *Cakile lanceolata* sequence in *FULc* that is united with the Rapa/Oleracea and Nigra lineages. With only one member from the Crambe, Vella and Savignya lineages included, their monophyly cannot be assessed. The Henophyton lineage identified by Arias and Pires (2012) is not monophyletic based on these data because *Raffenaldia primuloides* (Nigra) is sister to *Pseuderucaria teretifolia* in *FULb*. As expected for a MADS-box gene in general (Purugganan et al., 1995; Kramer et al., 1998; Cubas et al., 1999), portions of the coding sequences are quite conserved and therefore fail to provide resolution in phylogenetic analyses of closely related species. However, the use of gDNA rather than cDNA in phylogenetic analyses of developmental genes has the benefit of providing additional

characters with the caveat that some introns may not be alignable (Charlesworth et al., 2003; Zhang et al., 2008).

Copy number and heteroarthrocarpy

One *FUL* copy was retrieved in about half of the 13 non-heteroarthrocarpic taxa sampled here. This indicates that these species lost two *FUL* copies since hexaploidization, although there was no consistency in which *FUL* copy was retained. The discovery that any *FUL* lineage may be recruited in order to exhibit non-heteroarthrocarpic fruits insinuates that all copies are sufficiently conserved to perform similar functions during fruit development. The other half of the non-heteroarthrocarpic taxa possess two *FUL* copies belonging to the *FULb* and *FULc* clades. This implies that *FUL* genes may act redundantly in fruit development or may have undergone subfunctionalization such that the role of duplicate *FUL* copies is recruited for non-fruit functions. These data are consistent with research done on multiple *FUL* copies from recent polyploid events exhibiting both subfunctionalization and redundancy in *Avena sativa* (Poaceae; Preston and Kellogg, 2006; Kinjo et al., 2012), *Aquilegia coerulea* (Ranunculaceae; Pabon-Mora et al., 2013b) and *Papaver somniferum* (Papaveraceae; Pabon-Mora et al., 2012).

Heteroarthrocarpic taxa maintain multiple paralogues of *FUL* as evident in the majority of heteroarthrocarpic taxa having two or more copies. Only one copy was found in heteroarthrocarpic taxa belonging to the Nigra lineage: *Coincya, Diplotaxis*, *Guiroa*, and *Hemicrambe*. Similar to non-heteroarthrocarpic taxa, there was no consistency in which *FUL* copies were retained although two copies were never from *FULb* and *FULc,* in stark contrast to which copies were retained in non-heteroarthrocarpic species. Taxa that have

three copies rather than two showed no perceivable difference in fruit morphology. The fact that the majority of heteroarthrocarpic taxa have more than two copies while *in situ* gene experiments by Avino et al. (2012) were unable to determine expression of the second copy in the ovary wall in *Erucaria erucarioides* or *Cakile lanceolata*, implies opposing scenarios for the fate of these genes. The question of whether more than one *FUL* copy is necessary to generate a heteroarthrocarpic fruit or the second *FUL* copy is being recruited for functions unrelated to ovary wall development could be answered by gathering further gene expression and function data.

Ultimately, these data reveal intriguing trends but not definite relationships regarding *FUL* copy number and heteroarthrocarpy. Two alternatives exist to explain this pattern: (1) the PCR based method of *FUL* copy retrieval missed some copies or (2) the evolution of heteroarthrocarpy is produced by changes in how *FUL* copies interact with upstream regulators. Dinneny et al. (2005) determined that *FILAMENTOUS FLOWER* (*FIL*) and *YABBY3* (*YAB3*) work in concert to control *FUL* expression in *Arabidopsis* fruits. These genes are partially redundant with *JAGGED* (*JAG*) and *fil/yab3/jag* triple mutants result in complete loss of dehiscence. Interestingly, *fil*/*yab3* mutants were indehiscent in the distal region of the fruit, bearing a striking similarity to partially indehiscent heteroarthrocarpic fruits in the Brassiceae, although no tissue reminiscent of the joint was noted. However, *jag* fruits exhibited an inward sloping of the upper valve margin associated with lack of *FUL* expression, inviting investigation into whether this phenomenon is developmentally similar to the constriction of tissue resulting in joint formation (Hall et al., 2006). These findings support the hypothesis that alterations upstream of *FUL* in conjunction with

complexity arising from their control of multiple *FUL* copies may explain the origins of heteroarthrocarpy.

Gene structure and fruit morphology

Patterns of greater intron lengths in heteroarthrocarpic taxa are difficult to interpret because little work has been done attempting to discern the effects of intron length in evolutionary development. Studies on three species, including *Arabidopsis*, have shown that highly expressed genes have significantly shorter introns than less expressed genes (Castillo-Davis et al., 2002; Seoighe et al., 2005; Li et al., 2007). This pattern is presumably due to the negative effect of additional time needed to transcribe a greater number of nucleotides. Thus, shorter introns in non-heteroarthrocarpic taxa could be the result of natural selection acting against intron lengthening such that longitudinal dehiscence is maintained. Considering that some portion of the fruit is indehiscent in heteroarthrocarpic fruits, it is plausible that relaxed natural selection on intron length may exist in one of the *FUL* copies such that average intron length would be longer for heteroarthrocarpic taxa. Although short introns are thought to be the result of natural selection acting against intron lengthening (Castillo-Davis et al., 2002; Li et al., 2007), it is conceivable that positive selection could also be producing longer introns as a mechanism to regulate dosage after duplication events (Edger and Pires, 2009). As this trend was significant for only intron 5, transformation experiments determining the effect of presence or absence of this intron on fruit morphology would be an interesting start to investigating their regulatory potential.

Molecular evolution models and longitudinal dehiscence

Sequence evolution appears to be associated with both *FUL* lineage and patterns of longitudinal dehiscence. The best fitting model (model 4b) partitions dN/dS amongst partially indehiscent, fully indehiscent, and fully dehiscent fruits while allowing each copy within heteroarthrocarpic taxa to evolve independently. Because partially indehiscent and fully indehiscent fruits can be thought of as subcategories of heteroarthrocarpy, the fact that both models 2a (heteroarthrocarpic vs non-heteroarthrocarpic) and model 2b (heteroarthrocarpic vs non-heteroarthrocarpic + independent copy evolution) were rejected as a significantly better fitting model indicates that evolution of *FUL* lineages is governed more by longitudinal indehiscence than by simple presence/absence of the joint. This is confirmed by the rejection of model 3 where taxa showing any kind of dehiscence, be it joint dehiscence or longitudinal dehiscence, was paired against taxa exhibiting absolutely no dehiscence. In this instance, the dN/dS of lineages with any kind of dehiscence was too heterogeneous to allow good fit of the model to the dataset. The fact that model 4b is a significantly better fit than the model simply allowing different rates among *FUL* lineages (models 1a and 1b) indicates that while dN/dS varies amongst *FUL* lineages, a better fitting model is produced by partitioning ratios amongst longitudinal dehiscence patterns. This indicates that dissimilarity of *FUL* sequences amongst Brassiceae may be governed by both divergence of *FUL* lineages as well as natural selection on fruit dehiscence.

These findings suggest that the genetic pathway interacting with *FUL* is involved in the evolution of variable fruit dehiscence in the Brassiceae, but because no positive selection (dN/dS >1) was detected during these analyses, natural selection is likely not

acting on *FUL* directly or may be acting on the MADS-domain. This premise coincides with our previous hypothesis that the evolutionary transition to heteroarthrocarpy may be caused by modifications in how upstream regulators such as *FIL, JAG* or *YAB3* control *FUL* and other valve margin identity genes (Dinneny et al., 2005). In this case, positive selection on the upstream regulator, resulting in reduced ability to form binding complexes or interact with regulatory regions, would present itself as relaxed selection in downstream genes because strong purifying selection no longer acts to conserve the sequence. The data presented by this study offer preliminary support for this hypothesis because higher dN/dS ratios, indicative of relaxed selection, are correlated with indehiscence given that dN/dS of fully indehiscent fruits > dN/dS for partially indehiscent fruits > dN/dS of fully dehiscent fruits (Table 2-4).

Conclusions

Data presented here show that *FUL* exhibits complex molecular evolution within the different lineages, which is perhaps unsurprising given the evolutionary lability in fruit morphology exhibited across the tribe (Hall et al., 2011). The fact that joint dehiscence remains in heteroarthrocarpic fruits even while longitudinal dehiscence is lost signifies that *FUL* genes may be retained and expressed in differing capacities (Avino et al., 2012), which may further contribute to the convoluted history of this gene. *FUL* genes likely contribute to the evolution of variable dehiscence in the Brassiceae and further expression and functional data may discern their precise involvement in heteroarthrocarpy. In addition to highlighting trends between heteroarthrocarpic and non-heteroarthrocarpic species, the *FUL* phylogeny provided by this study offers a basis for future hypothesis

testing. For example, expression data would determine whether both *FUL* copies retained in some non-heteroarthrocarpic taxa are functionally analogous to the single *FUL* copy reported for *Arabidopsis* (Gu et al., 1998; Ferrandiz et al., 2000a). This study has revealed the phylogenetic relationships of *FUL* copies in previously studied taxa *Erucaria* and *Cakile* (Avino et al., 2012) as belonging to *FULc* and *FULd*, leaving the role of *FULa*/*FULb* unexamined in heteroarthrocarpic taxa. Analysis of expression patterns for taxa that have retained all three copies of *FUL* from the hexaploid ancestor would provide a full picture of the fate of each *FUL* lineage.

References

AIROLDI C. A. AND B. DAVIES. 2012. Gene duplication and the evolution of plant MADS-box transcription factors. *Journal of Genetics and Genomics* 39: 157-165.

AL-SHEHBAZ I. A., M. A. BEILSTEIN, AND E. A. KELLOGG. 2006. Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Systematics and Evolution* 259: 89-120.

AL-SHEHBAZ I. A. 2012. A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon* 61: 931-954.

ALVAREZ-BUYLLA E. R., B. GARCIA-PONCE, AND A. GARAY-ARROYO. 2006. Unique and redundant functional domains of *APETALA1* and *CAULIFLOWER*, two recently duplicated *Arabidopsis thaliana* floral MADS-box genes. *Journal of Experimental Botany* 57: 3099-3107.

APPEL O. 1999. The so-called 'beak', a character in the systematics of Brassicaceae? *Botanische Jahrbücher für Systematik Pflanzengeschichte und Pflanzengeographie* 121: 85-98.

APPEL O. AND I. A. AL-SHEHBAZ. 2003. Cruciferae. *In* Kubitzki K. [ed.], *Families and Genera of Vascular Plants.* Springer-Verlag, Heidelberg.

ARIAS T. AND J. C. PIRES. 2012. A fully resolved chloroplast phylogeny of the *Brassica* crops and wild relatives (Brassicaceae: Brassiceae): novel clades and potential taxonomic implications. *Taxon* 61: 980-988.

AVINO M., E. M. KRAMER, K. DONOHUE, A. J. HAMMEL, AND J. C. HALL. 2012. Understanding the basis of a novel fruit type in Brassicaceae: conservation and deviation in expression patterns of six genes. *Evodevo* 3: 20.

BAUM D., H. YOON, AND R. OLDHAM. 2005. Molecular evolution of the transcription factor *LEAFY* in Brassicaceae. *Molecular phylogenetics and evolution* 37: 1-14.

BEILSTEIN M., I. AL-SHEHBAZ, AND E. KELLOGG. 2006. Brassicaceae phylogeny and trichome evolution. *American Journal of Botany* 93: 607-619.

BLANC G., K. HOKAMP, AND K. H. WOLFE. 2003. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Research* 13: 137-144.

BOLMGREN K. AND O. ERIKSSON. 2005. Fleshy fruits - origins, niche shifts, and diversification. *Oikos* 109: 255-272.

CASTILLO-DAVIS C., S. MEKHEDOV, D. HARTL, E. KOONIN, AND F. KONDRASHOV. 2002. Selection for short introns in highly expressed genes. *Nature Genetics* 31: 415-418.

CHAPMAN M. A., J. H. LEEBENS-MACK, AND J. M. BURKE. 2008. Positive selection and expression divergence following gene duplication in the sunflower CYCLOIDEA gene family. *Molecular biology and evolution* 25: 1260-1273.

CHARLESWORTH D., C. BARTOLOME, M. SCHIERUP, AND B. MABLE. 2003. Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. *Molecular Biology and Evolution* 20: 1741-1753.

CHU T., H. XIE, Y. XU, AND R. MA. 2010. Regulation pattern of the *FRUITFULL* (*FUL*) gene of *Arabidopsis thaliana*. *Sheng wu gong cheng xue bao = Chinese Journal of Biotechnology* 26: 1546-54.

CITERNE H., D. LUO, R. PENNINGTON, E. COEN, AND Q. CRONK. 2003. A phylogenomic investigation of *CYCLOIDEA*-like TCP genes in the Leguminosae. *Plant Physiology* 131: 1042-1053.

COATE J. E. AND J. J. DOYLE. 2011. Divergent evolutionary fates of major photosynthetic gene networks following gene and whole genome duplications. *Plant Signaling and Behavior* 6: 594-597.

COULOMBE-HUNTINGTON J. AND J. MAJEWSKI. 2007. Intron loss and gain in *Drosophila*. *Molecular Biology and Evolution* 24: 2842-2850.

CRONK, Q.C.B., BATEMAN, R.M. AND HAWKINS, J.A. 2002. *Developmental Genetics and Plant Evolution*. Taylor and Francis, London.

CUBAS P., N. LAUTER, J. DOEBLEY, AND E. COEN. 1999. The TCP domain: a motif found in proteins regulating plant growth and development. *Plant Journal* 18: 215-222.

DEYHOLOS M. K. AND L. E. SIEBURTH. 2000. Separable whorl-specific expression and negative regulation by enhancer elements within the *AGAMOUS* second intron. *Plant Cell* 12: 1799-1810.

DINNENY J., D. WEIGEL, AND M. YANOFSKY. 2005. A genetic framework for fruit patterning in *Arabidopsis thaliana*. *Development* 132: 4687-4696.

DUNN, O. J. 1961. Multiple Comparisons Among Means. *Journal of the American Statistical Association* 56: 52–64.

EDGER P. P. AND J. C. PIRES. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research* 17: 699-717.

FELSENSTEIN, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1-15.

FERRANDIZ C., Q. GU, R. MARTIENSSEN, AND M. F. YANOFSKY. 2000a. Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. *Development* 127: 725-734.

FERRANDIZ C., S. LILJEGREN, AND M. YANOFSKY. 2000b. Negative regulation of the *SHATTERPROOF* genes by *FRUITFULL* during *Arabidopsis* fruit development. *Science* 289: 436-438.

FOURQUIN C., C. DEL CERRO, F. C. VICTORIA, A. VIALETTE-GUIRAUD, A. C. DE OLIVEIRA, AND C. FERRANDIZ. 2013. A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in *Medicago* genus. *Plant physiology* 162: 907-917.

GIRIN T., P. STEPHENSON, C. M. P. GOLDSACK, S. A. KEMPIN, A. PEREZ, N. PIRES, P. A. SPARROW, T. A. WOOD, M. F. YANOFSKY, AND L. OSTERGAARD. 2010. Brassicaceae *INDEHISCENT* genes specify valve margin cell fate and repress replum formation. *Plant Journal* 63: 329-338.

GU Q., C. FERRANDIZ, M. F. YANOFSKY, AND R. MARTIENSSEN. 1998. The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125: 1509-1517.

HALL J. C. AND K. DONOHUE. 2012. The genetics of plant dispersal. *In* Clobert J., M. Baguette, T. Benton and J. Bullock [eds.], *Dispersal Ecology and Evolution*. Oxford University Press, Oxford.

HALL J. C., T. E. TISDALE, K. DONOHUE, AND E. M. KRAMER. 2006. Developmental basis of an anatomical novelty: heteroarthrocarpy in *Cakile lanceolata* and *Erucaria erucarioides* (Brassicaceae). *International Journal of Plant Sciences* 167: 771-789.

HALL J. C., T. E. TISDALE, K. DONOHUE, A. WHEELER, M. A. AL-YAHYA, AND E. M. KRAMER. 2011. Convergent Evolution of a complex fruit structure in the tribe Brassiceae (Brassicaceae). *American Journal of Botany* 98: 1989-2003.

HE L., M. ZHAO, Y. WANG, J. GAI, AND C. HE. 2013. Phylogeny, structural evolution and functional diversification of the plant PHOSPHATE1 gene family: a focus on *Glycine max*. *BMC Evolutionary Biology* 13: 103.

HEDGE I. C. 1976. A systematic and geographical survey of the old world Cruciferae. *In* Vaughn J. G., A. J. MacLeod and B. M. G. Jones [eds.], *The Biology and Chemistry of the Cruciferae*. Academic Press, London.

HOWARTH D. G. AND M. J. DONOGHUE. 2006. Phylogenetic analysis of the "'ECE" (CYC/TB1) clade reveals duplications predating the core eudicots. *Proceedings of the National Academy of Sciences of the United States of America* 103: 9101-9106.

HOWE G., P. BUCCIAGLIA, W. HACKETT, G. FURNIER, M. CORDONNIER-PRATT, AND G. GARDNER. 1998. Evidence that the phytochrome gene family in black cottonwood has one *PHYA* locus and two *PHYB* loci but lacks members of the PHYC/F and PHYE subfamilies. *Molecular Biology and Evolution* 15: 160-175.

HUA S., I. H. SHAMSI, Y. GUO, H. PAK, M. CHEN, C. SHI, H. MENG, AND L. JIANG. 2009. Sequence, expression divergence, and complementation of homologous *ALCATRAZ* loci in *Brassica napus*. *Planta* 230: 493-503.

HUELSENBECK J. AND F. RONQUIST. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.

IBM Corp.: IBM SPSS statistics for Windows. Released 2012, Version 19.0.

ILTIS H. H., J. C. HALL, T. S. COCHRANE, AND K. J. SYTSMA. 2011. Studies in the Cleomaceae I. on the separate recognition of Capparaceae, Cleomaceae, and Brassicaceae. *Annals of the Missouri Botanical Garden* 98: 28-36.

JARAMILLO M. A. AND E. M. KRAMER. 2007. Molecular evolution of the petal and stamen identity genes, *APETALA3* and *PISTILLATA*, after petal loss in the Piperales. *Molecular Phylogenetics and Evolution* 44: 598-609.

KINJO H., N. SHITSUKAWA, S. TAKUMI, AND K. MURAI. 2012. Diversification of three *APETALA1/FRUITFULL*-like genes in wheat. *Molecular Genetics and Genomics* 287: 283-294.

KNOWLES D. G. AND A. MCLYSAGHT. 2006. High rate of recent intron gain and loss in simultaneously duplicated *Arabidopsis* genes. *Molecular Biology and Evolution* 23: 1548-1557.

KOCH M., I. AL-SHEHBAZ, AND K. MUMMENHOFF. 2003. Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae). *Annals of the Missouri Botanical Garden* 90: 151-171.

KOCH M., B. HAUBOLD, AND T. MITCHELL-OLDS. 2001. Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *Chs* sequences. *American Journal of Botany* 88: 534-544.

KRAMER E., R. DORIT, AND V. IRISH. 1998. Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. *Genetics* 149: 765-783.

KRAMER E., M. JARAMILLO, AND V. DI STILIO. 2004. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* 166: 1011-1023.

LANFEAR R., B. CALCOTT, S. Y. W. HO, AND S. GUINDON. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695-1701.

LEE H. AND V. F. IRISH. 2011. Gene duplication and loss in a MADS-box gene transcription factor circuit. *Molecular Biology and Evolution* 28: 3367-3380.

LI S., L. FENG, AND D. NIU. 2007. Selection for the miniaturization of highly expressed genes. *Biochemical and Biophysical Research Communications* 360: 586-592.

LI W. AND C. CHINNAPPA. 2003. The phytochrome gene family in the *Stellaria longipes* complex. *International Journal of Plant Sciences* 164: 657-673.

LILJEGREN S., G. DITTA, H. ESHED, B. SAVIDGE, J. BOWMAN, AND M. YANOFSKY. 2000. *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404: 766-770.

LILJEGREN S., A. ROEDER, S. KEMPIN, K. GREMSKI, L. OSTERGAARD, S. GUIMIL, D. REYES, AND M. YANOFSKY. 2004. Control of fruit patterning in *Arabidopsis* by *INDEHISCENT*. *Cell* 116: 843-853.

LITT A. AND V. IRISH. 2003. Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* 165: 821-833.

LITT A. 2007. An evaluation of A-function: evidence from the *APETALA1* and *APETALA2* gene lineages. *International Journal of Plant Sciences* 168: 73-91.

LORTS C. M., T. BRIGGEMAN, AND TAO SANG. 2008. Evolution of fruit types and seed dispersal: a phylogenetic and ecological snapshot. *Journal of Systematics and Evolution* 46: 396-404.

LYSAK M., M. KOCH, A. PECINKA, AND I. SCHUBERT. 2005. Chromosome triplication found across the tribe Brassiceae. *Genome Research* 15: 516-525.

LYSAK M. A., K. CHEUNG, M. KITSCHKE, AND P. BURES. 2007. Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiology* 145: 402-410.

MADDISON, W. P. AND D.R. MADDISON. 2009. Mesquite: a modular system for evolutionary analysis. Version 2.71.

MUEHLHAUSEN A., T. LENSER, K. MUMMENHOFF, AND G. THEISSEN. 2013. Evidence that an evolutionary transition from dehiscent to indehiscent fruits in *Lepidium* (Brassicaceae) was caused by a change in the control of valve margin identity genes. *Plant Journal* 73: 824-835.

MUMMENHOFF K., A. POLSTER, A. MUEHLHAUSEN, AND G. THEISSEN. 2009. *Lepidium* as a model system for studying the evolution of fruit development in Brassicaceae. *Journal of Experimental Botany* 60: 1503-1513.

NEI M. AND A. ROONEY. 2005. Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics* 39: 121-152.

NIELSEN C. B., B. FRIEDMAN, B. BIRREN, C. B. BURGE, AND J. E. GALAGAN. 2004. Patterns of intron gain and loss in fungi. *PLoS Biology* 2(12): e422.

O'KANE S. AND I. AL-SHEHBAZ. 2003. Phylogenetic position and generic limits of *Arabidopsis* (Brassicaceae) based on sequences of nuclear ribosomal DNA. *Annals of the Missouri Botanical Garden* 90: 603-612.

OSNATO M., M. R. STILE, Y. WANG, D. MEYNARD, S. CURIALE, E. GUIDERDONI, Y. LIU, D. S. HORNER, P. B. F. OUWERKERK, C. POZZI, K. J. MULLER, F. SALAMINI, AND L. ROSSINI. 2010. Cross talk between the *KNOX* and ethylene pathways is mediated by intron-binding transcription factors in barley. *Plant Physiology* 154: 1616-1632.

PABON-MORA N., B. A. AMBROSE, AND A. LITT. 2012. Poppy *APETALA1/FRUITFULL* orthologs control flowering time, branching, perianth identity, and fruit development. *Plant Physiology* 158: 1685-1704.

PABON-MORA N., O. HIDALGO, S. GLEISSBERG, AND A. LITT. 2013a. Assessing duplication and loss of *APETALA1/FRUITFULL* homologs in Ranunculales. *Frontiers in Plant Science* 4: 358-358.

PABON-MORA N., B. SHARMA, L. D. HOLAPPA, E. M. KRAMER, AND A. LITT. 2013b. The *Aquilegia FRUITFULL*-like genes play key roles in leaf morphogenesis and inflorescence development. *Plant Journal* 74: 197-212.

PALMER J. D., K. L. ADAMS, Y. CHO, C. L. PARKINSON, Y. QIU, AND K. SONG. 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences of the United States of America* 97: 6960-6966.

PARKIN I. A., A. G. SHARPE, D. J. KEITH, AND D. J. LYDIATE. 1995. Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). *Genome* 38: 1122-1131.

PARKIN I., S. GULDEN, A. SHARPE, L. LUKENS, M. TRICK, T. OSBORN, AND D. LYDIATE. 2005. Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171: 765-781.

PARRA G., K. BRADNAM, A. B. ROSE, AND I. KORF. 2011. Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. *Nucleic Acids Research* 39: 5328-5337.

POSADA D. AND K. CRANDALL. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

PRESTON J. C. AND E. A. KELLOGG. 2006. Reconstructing the evolutionary history of paralogous *APETALA1/FRUITFULL*-like genes in grasses (Poaceae). *Genetics* 174: 421-437.

PURUGGANAN M., S. ROUNSLEY, R. SCHMIDT, AND M. YANOFSKY. 1995. Molecular evolution of flower development - diversification of the plant MADS-box regulatory gene family. *Genetics* 140: 345-356.

RAMBAUT A. AND A.J. DRUMMOND. 2007. Tracer. Version 1.4.

RASTOGI P. A. 2000. MacVector. Integrated sequence analysis for the Macintosh. *Methods in Molecular Biology* 132: 47-69.

ROEDER A., C. FERRANDIZ, AND M. YANOFSKY. 2003. The role of the *REPLUMLESS* homeodomain protein in patterning the *Arabidopsis* fruit. *Current Biology* 13: 1630-1635.

ROY S. W. AND D. PENNY. 2006. Smoke without fire: most reported cases of intron gain in nematodes instead reflect intron losses. *Molecular Biology and Evolution* 23: 2259-2262.

SEOIGHE C., C. GEHRING, AND L. HURST. 2005. Gametophytic selection in *Arabidopsis thaliana* supports the selective model of intron length reduction. *PLoS Genetics* 1: 154-158.

SIMILLION C., K. VANDEPOELE, M. C. E. VAN MONTAGU, M. ZABEAU, AND Y. VAN DE PEER. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* 99: 13627-13632.

SKIPPER M., K. B. PEDERSEN, L. B. JOHANSEN, S. FREDERIKSEN, V. F. IRISH, AND B. B. JOHANSEN. 2005. Identification and quantification of expression levels of three *FRUITFULL*-like MADS-box genes from the orchid *Dendrobium thyrsiflorum* (Reichb. f.). *Plant Science* 169: 579-586.

SOUTHERN, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98: 503–517

SWARBRECK D., C. WILKS, P. LAMESCH, T. Z. BERARDINI, M. GARCIA-HERNANDEZ, H. FOERSTER, D. LI, T. MEYER, R. MULLER, L. PLOETZ, A. RADENBAUGH, S. SINGH, V. SWING, C. TISSIER, P. ZHANG, AND E. HUALA. 2008. The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research* 36: 1009-1014.

SWOFFORD D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0β10.

TAN X., Z. XIA, L. ZHANG, Z. ZHANG, Z. GUO, AND C. QI. 2009. Cloning and sequence analysis of oilseed rape (*Brassica napus*) *SHP2* gene. *Botanical Studies* 50: 403-412.

THOMPSON J., T. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. HIGGINS. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876-4882.

VISION T. J., D. G. BROWN, AND S. D. TANKSLEY. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290: 2114-2117.

WANG X., H. WANG, J. WANG, R. SUN, J. WU, S. LIU, Y. BAI, J. -. MUN, I. BANCROFT, F. CHENG, S. HUANG, X. LI, W. HUA, J. WANG, X. WANG, M. FREELING, J. C. PIRES, A. H. PATERSON, B. CHALHOUB, B. WANG, A. HAYWARD, A. G. SHARPE, B. -. PARK, B. WEISSHAAR, B. LIU, B. LI, B. LIU, C. TONG, C. SONG, C. DURAN, C. PENG, C. GENG, C. KOH, C. LIN, D. EDWARDS, D. MU, D. SHEN, E. SOUMPOUROU, F. LI, F. FRASER, G. CONANT, G. LASSALLE, G. J. KING, G. BONNEMA, H. TANG, H. WANG, H. BELCRAM, H. ZHOU, H. HIRAKAWA, H. ABE, H. GUO, H. WANG, H. JIN, I. A. P. PARKIN, J. BATLEY, J. -. KIM, J. JUST, J. LI, J. XU, J. DENG, J. A. KIM, J. LI, J. YU, J. MENG, J. WANG, J. MIN, J. POULAIN, K. HATAKEYAMA, K. WU, L. WANG, L. FANG, M. TRICK, M. G. LINKS, M. ZHAO, M. JIN, N. RAMCHIARY, N. DROU, P. J. BERKMAN, Q. CAI, Q. HUANG, R. LI, S. TABATA, S. CHENG, S. ZHANG, S. ZHANG, S. HUANG, S. SATO, S. SUN, S. -. KWON, S. -. CHOI, T. -. LEE, W. FAN, X. ZHAO, X. TAN, X. XU, Y. WANG, Y. QIU, Y. YIN, Y. LI, Y. DU, Y. LIAO, Y. LIM, Y. NARUSAKA, Y. WANG, Z. WANG, Z. LI, Z. WANG, Z. XIONG, AND Z. ZHANG. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature genetics* 43: 1035-1040.

WANG X., D. C. TANK, AND T. SANG. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular biology and evolution* 17: 773-781.

WARWICK S. AND C. SAUDER. 2005. Phylogeny of tribe Brassiceae (Brassicaceae) based on chloroplast restriction site polymorphisms and nuclear ribosomal internal transcribed spacer and chloroplast *trnL* intron sequences. *Canadian Journal of Botany-Revue Canadienne De Botanique* 83: 467-483.

YANG Z., R. NIELSEN, N. GOLDMAN, AND A. PEDERSEN. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431-449.

ZHANG W., E. M. KRAMER, AND C. 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: 6388-6393.

ZHANG W., Q. XIANG, D. T. THOMAS, B. M. WIEGMANN, M. W. FROHLICH, AND D. E. SOLTIS. 2008. Molecular evolution of *PISTILLATA*-like genes in the dogwood genus *Cornus* (Cornaceae). *Molecular Phylogenetics and Evolution* 47: 175-195.

ZWICKL D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Dissertation. University of Texas at Austin.

Tables

Table 2-1. PCR screening effort, copy number and sequencing of *FUL* in this study.

Species	Primers	clones PCR screened	clones M13F sequenced	No. of copies	Gene Name	clones entirely sequenced
Brassica napus	FULexon2F, FULexon8R	120	13	3	*FULa1* *FULa2* *FULc* *FULd*	*1* *1* *1* *3*
Brassica nigra	FULexon2F, FULexon8R	120	18	1	*FULd*	*4*
Brassica oleracea var. *acephala*	FULexon2F, FULexon8R	24	12	2	*FULa* *FULc* *FULd*	*1* *3* *0**
Brassica spinescens	FULexon2F, FULexon8R	24	12	2	*FULd*	*4*
Cakile lanceolata subsp. *fusiformis*	FULexon2F, FULexon8R	24	12	2	*FULc* *FULd*	*4* *3*
Coincya monensis	FULexon2F, FULexon8R, FULexon2F-2, FULexon8R-2	24	16	1	*FULc*	*4*
Cordylocarpus muricatus	FULexon2F, FULexon8R	24	12	2	*FULa* *FULc*	*2* *2*
Crambe orientalis	FULexon2F, FULexon8R	120	15	2	*FULa* *FULc*	*1* *3*
Crambella teretifolia	FULexon2F, FULexon8R	24	12	3	*FULa* *FULc* *FULd*	*1* *2* *2*
Didesmus bipinnatus	FULexon2F, FULexon8R	24	12	2	*FULa* *FULd*	*2* *3*
Diplotaxis assurgens	FULexon2F, FULexon8R, FULexon2F-2, FULexon8R-2	24	16	1	*FULd*	*4*
Enarthrocarpus lyratus	FULexon2F, FULexon8R	24	12	2	*FULa* *FULc* *FULd*	*1* *2* *2*
Eruca vesicaria	FULexon2F, FULexon8R	24	12	1	*FULa*	*4*
Erucaria erucarioides	FULexon2F, FULexon8R, FULexon2F-2, FULexon8R-2	24	12	2	*FULc* *FULd*	*3* *1*
Erucastrum gallicum	FULexon2F, FULexon8R, FULexon2F-2, FULexon8R-2	120	12	1	*FULd*	*4*
Guiraoa arvensis	FULexon2F, FULexon8R, FULexon2F-2, FULexon8R-2	24	16	1	*FULc*	*4*

Hemicrambe fruticulosa	FULexon2F, FULexon8R, FULexon2F-2, FULexon8R-2	24	16	1	*FULc*	*4*
Henophyton deserti	FULexon2F, FULexon8R, FULexon2F-2, FULexon8R-2	120	17	2	*FULb* *FULc*	*2* *2*
Hirschfeldia incana	FULexon2F, FULexon8R	24	12	2	*FULa* *FULc*	*2* *2*
Moricandia arvensis	FULexon2F, FULexon8R	24	12	2	*FULb* *FULc*	*4* *3*
Muricaria prostrata	FULexon2F, FULexon8R	24	12	3	*FULb* *FULc* *FULd*	*1* *2* *2*
Pseuderucaria teretifolia	FULexon2F, FULexon8R, FULexon2F-2, FULexon8R-2	120	17	2	*FULb* *FULc*	*4* *2*
Psychine stylosa	FULexon2F, FULexon8R	24	12	1	*FULc*	*4*
Raffenaldia primuloides	FULexon2F, FULexon8R	120	14	2	*FULb* *FULc*	*2* *2*
Rytidocarpus moricandiodes	FULexon2F, FULexon8R	24	12	2	*FULb* *FULc*	*2* *2*
Schouwia thebaica	FULexon2F, FULexon8R, FULexon2F-2, FULexon8R-2	24	12	1	*FULc*	*4*
Sinapidendron angustifolium	FULexon2F, FULexon8R	24	12	1	*FULd*	*4*
Vella spinosa	FULexon2F, FULexon8R, FULexon2F-2, FULexon8R-2	24	12	1	*FULc*	*4*
Zilla spinosa	FULexon2F, FULexon8R	120	13	2	*FULb* *FULc*	*2* *3*

Column two: primers that were used to obtain *FUL* genes during PCR. Column three: number of clones screened using PCR methods. Column four: number of clones that were sequenced using the M13F primer. Column five: number of *FUL* copies retrieved. Column seven: number of clones entirely sequenced for each *FUL* copy.

* This sequence was retrieved from GenBank and is from *Brassica oleracea* var. *botrytis*

Table 2-2. Primers used for *FUL* amplification and sequencing.

Purpose	Primer Name	Sequence (5'-3')
Amplification	FULexon2F	CRGAYAAACAACTTGTNGGHMGMGABRYTTC
	FULexon8R	RGCTTGGATGYTRCGYCCDRCDAC
	FULexon2F-2	GAYCGNTARTAYTAYTCRGACAAACAAC
	FULexon8R-2	CCYCAACYCTMYCCAMRARGYCATCT
Sequencing	M13F	GTAAAACGACGGCCAG
	M13R	CAGGAAACAGCTATGAC
	FULexon3-F	GCTSAAGGCWAGTDGAG
	FULexon4-F	GAHCTYGRWWCNTTGAGYYTRAAGG
	FULexon4-R	CTTGAYCTRATGYTYTTRATDGCWG
	FULexon6-F	CAAGAKCAHAAYAATDYDCTTCTMAAAAAG
	FULexon5-F	TATGTTCGAAWCCATATCRGCGC
	FULexon5-R	GCGCYGATATGGWTTCGAACATA

Table 2-3. Fruit morphology and number of *FUL* copies present in each taxon.

Species	No. of copies	*FUL a*	*FUL b*	*FUL c*	*FULd*	Fruit type	Longitudinal dehiscence
Brassica napus	3	✓		✓	✓	heteroarthrocarpic	partially indehiscent
Brassica nigra	2				✓	non-heteroarthrocarpic	fully dehiscent
Brassica oleracea	2	✓		✓	✓	heteroarthrocarpic	partially indehiscent
Brassica spinescens	2				✓	heteroarthrocarpic	partially indehiscent
Cakile lanceolata subsp. *fusiformis*	2			✓	✓	heteroarthrocarpic	fully indehiscent
Coincya monensis	1			✓		heteroarthrocarpic	partially indehiscent
Cordylocarpus muricatus	2	✓		✓		heteroarthrocarpic	fully indehiscent
Crambe orientalis	2	✓		✓		heteroarthrocarpic	fully indehiscent
Crambella teretifolia	3	✓		✓	✓	heteroarthrocarpic	fully indehiscent
Didesmus bipinnatus	2	✓			✓	heteroarthrocarpic	fully indehiscent
Diplotaxis assurgens	1				✓	heteroarthrocarpic	partially indehiscent
Enarthrocarpus lyratus	3	✓		✓	✓	heteroarthrocarpic	fully indehiscent
Eruca vesicaria	1	✓				non-heteroarthrocarpic	fully dehiscent
Erucaria erucarioides	2			✓	✓	heteroarthrocarpic	partially indehiscent
Erucastrum gallicum	1				✓	non-heteroarthrocarpic	fully dehiscent
Guiraoa arvensis	1			✓		heteroarthrocarpic	partially indehiscent
Hemicrambe fruticulosa	1			✓		heteroarthrocarpic	partially indehiscent

Henophyton deserti	2		✓	✓		non-heteroarthrocarpic	fully dehiscent
Hirschfeldia incana	2	✓		✓		heteroarthrocarpic	partially indehiscent
Moricandia arvensis	2		✓	✓		non-heteroarthrocarpic	fully dehiscent
Muricaria prostrata	3	✓		✓	✓	heteroarthrocarpic	fully indehiscent
Pseuderucaria teretifolia	1		✓	✓		non-heteroarthrocarpic	fully dehiscent
Psychine stylosa	2			✓		non-heteroarthrocarpic	fully dehiscent
Raffenaldia primuloides	2		✓	✓		non-heteroarthrocarpic	fully indehiscent
Rytidocarpus moricandiodes	2		✓	✓		non-heteroarthrocarpic	fully dehiscent
Schouwia thebaica	1			✓		non-heteroarthrocarpic	fully dehiscent
Sinapidendron angustifolium	1				✓	non-heteroarthrocarpic	fully dehiscent
Vella spinosa	1			✓		non-heteroarthrocarpic	fully dehiscent
Zilla spinosa	2		✓	✓		non-heteroarthrocarpic	fully indehiscent

Table 2-4. Maximum likelihood tests of selection from PAML4.4 and estimated dN/dS ratios.

Model	-lnL	Partition	dN/dS
fixed	4151.47	All branches	0.2600
free	4100.50	-	-
1a: all copies separate	4148.00	*FULa*	0.3973
		FULb	0.2205
		FULc	0.2153
		FULd	0.2564
1b: *FULb/FULc* pooled	4146.04*	*FULa*	0.3973
		FULb/FULc	0.2198
		FULd	0.2564
1c: *FULb/FULc/FULa* pooled	4149.41	*FULa*	0.3973
		FULb/FULc/FULd	0.2320
2a: heteroarthrocarpic vs non-heteroarthrocarpic	4150.80	Heteroarthrocarpic	0.2855
		non-heteroarthrocarpic	0.2370
2b: heteroarthrocarpic vs non-heteroarthrocarpic + copies independent	4147.50	non-heteroarthrocarpic	0.2372
		heteroarthrocarpic *FULa*	0.4116
		heteroarthrocarpic *FULb/FULc*	0.2372
		heteroarthrocarpic *FULd*	0.2664
3: no dehiscence vs dehiscence any kind	4149.48	no dehiscence	0.3903
		any kind of dehiscence	0.2491
4a: fully dehiscent vs partially indehiscent vs fully indehiscent ignoring joint dehiscence	4147.89	fully dehiscent	0.2095
		partially indehiscent	0.2328
		fully indehiscent	0.3472
4b: fully dehiscent vs partially indehiscent vs fully indehiscent ignoring joint dehiscence + copies independent	4142.79**	fully dehiscent all copies	0.2096
		partially indehiscent *FULa*	0.3408
		fully indehiscent *FULa*	0.7675
		partially indehiscent *FULb/FULc*	0.2165
		fully indehiscent *FULb/FULc*	0.2745
		partially indehiscent *FULd*	0.2376
		fully indehiscent *FULd*	0.2623

* indicates significant at α=0.05 and ** indicates significant at α=0.01 when compared to the fixed model.

Figures

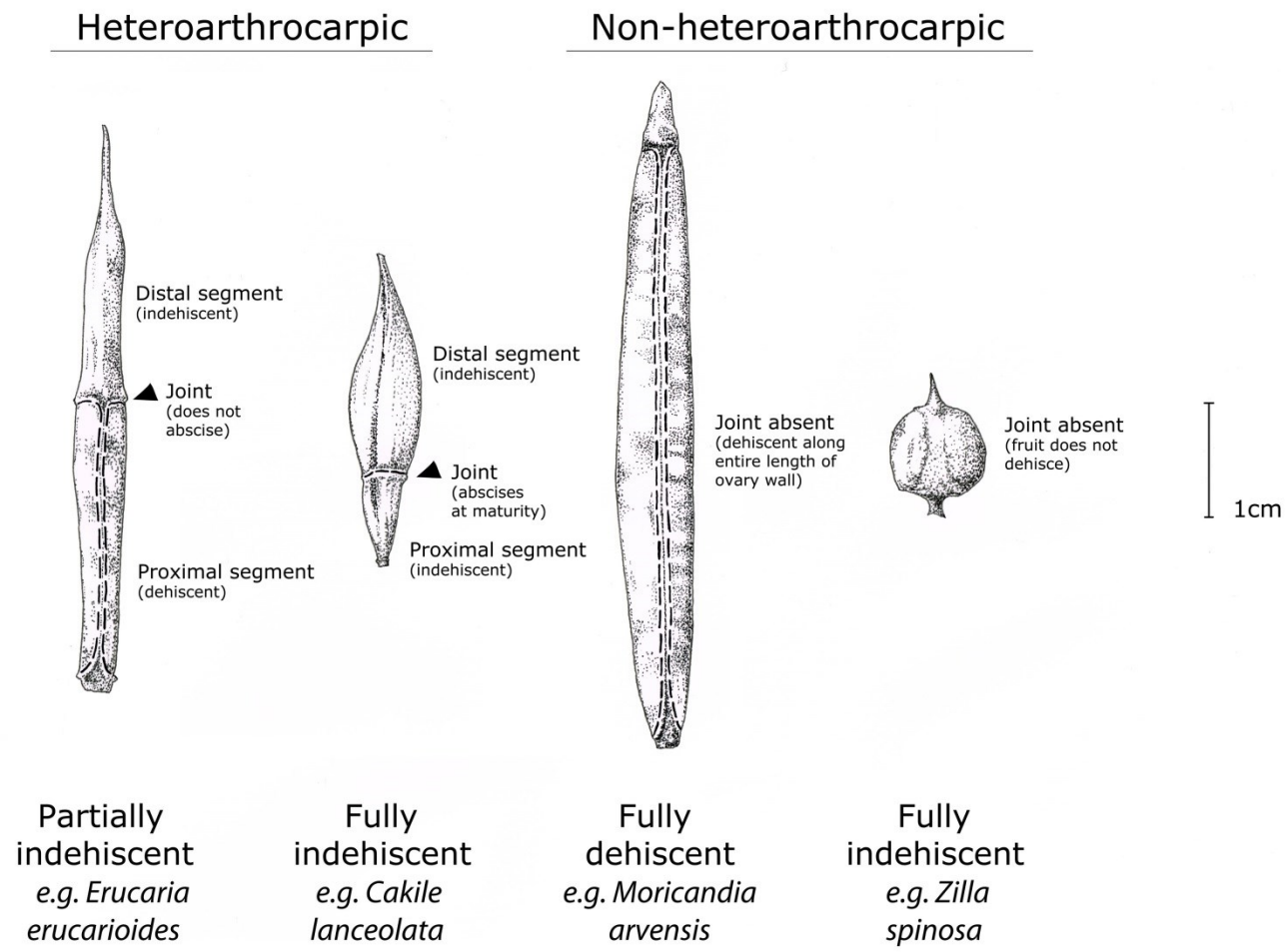


Figure 2-1. Overview of fruit diversity within the Brassiceae. Patterns of longitudinal dehiscence amongst heteroarthrocarpic and non-heteroarthrocarpic fruits are shown as black dashed lines. The joint occurring in heteroarthrocarpic fruits is indicated by a solid black arrow.

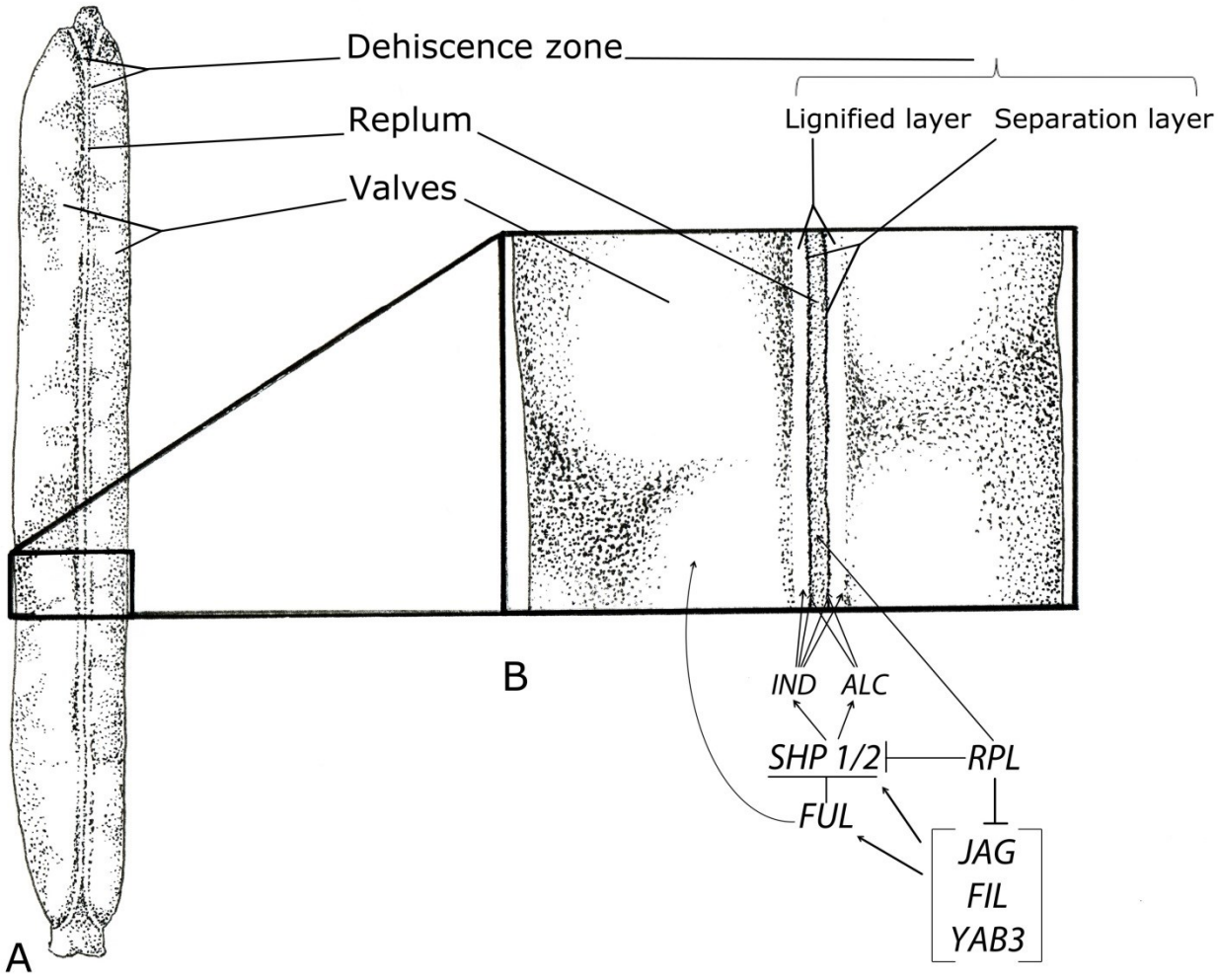


Figure 2-2. A) Mature silique of Arabidopsis B) Valve margin anatomy alongside genetic pathway establishing the dehiscence zone. Recreated from Dinneny et al. (2005).

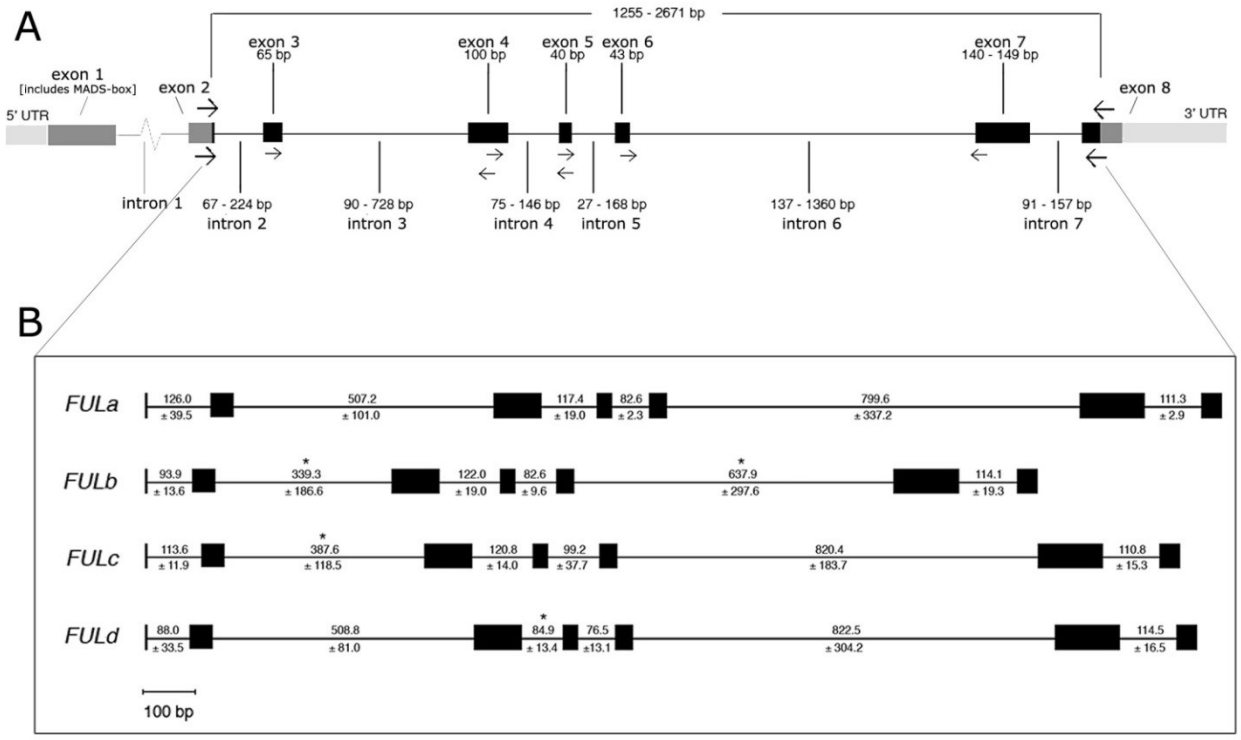


Figure 2-3. Gene structure and variation of *FUL* in the Brassiceae and B) mean intron lengths amongst copies. Values shown in diagram A represent variation in exon and intron lengths found amongst all copies and taxon sampled in this study. Black regions are those amplified in this study. Large arrows designate the annealing sites for amplification primers and small arrows signify the locations of sequencing primers. Values listed in diagram B represent mean intron length ± one standard deviation from the mean. The asterisks identify significantly different intron lengths (p-value <0.05).

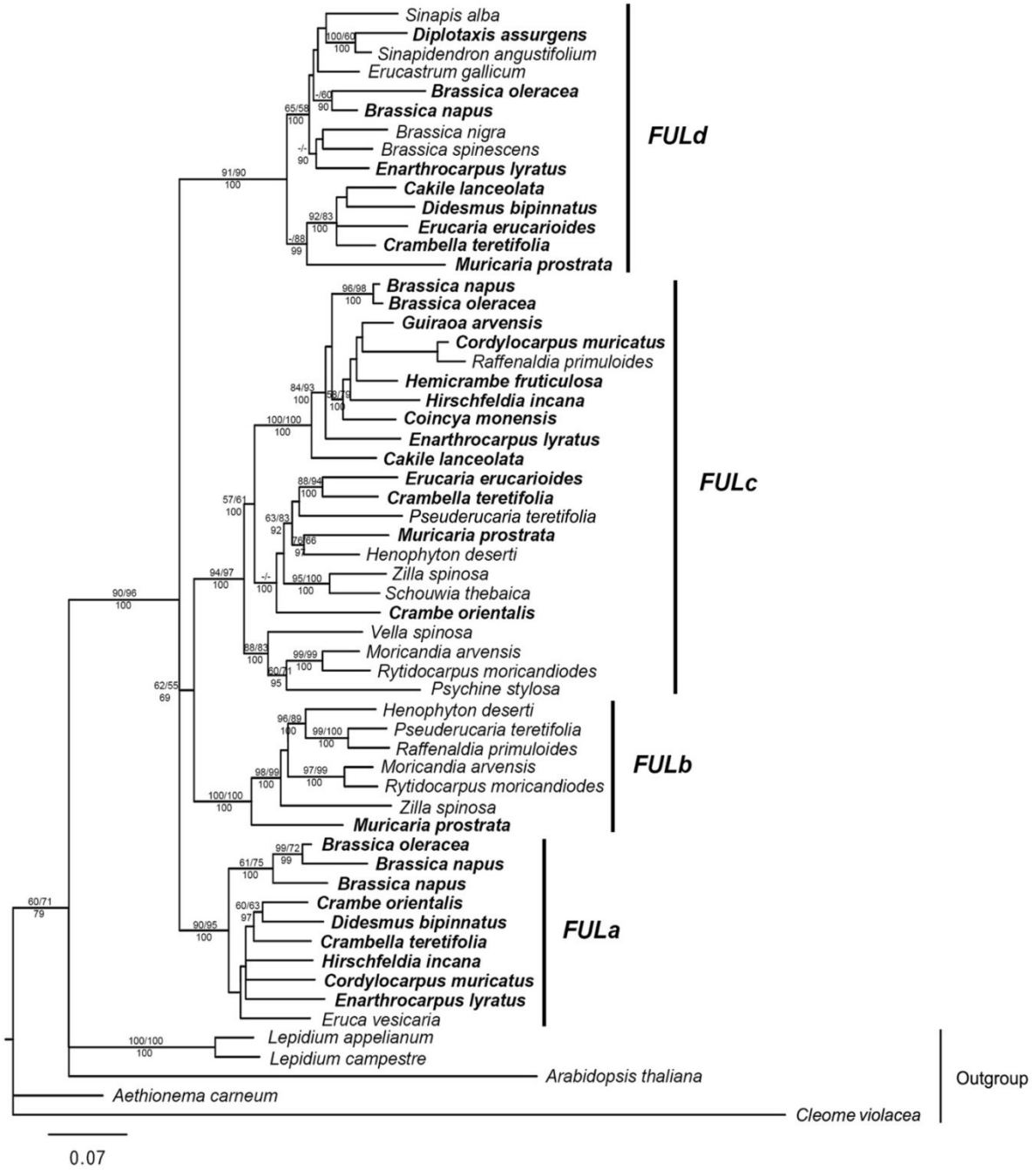


Figure 2-4. Bayesian consensus tree including branch lengths of representative *FUL* sequences using the coding + intron nucleotide dataset. Bayesian posterior probabilities are listed below branches and maximum parsimony / maximum likelihood bootstrap values are listed above branches. Heteroarthrocarpic taxa are bolded.

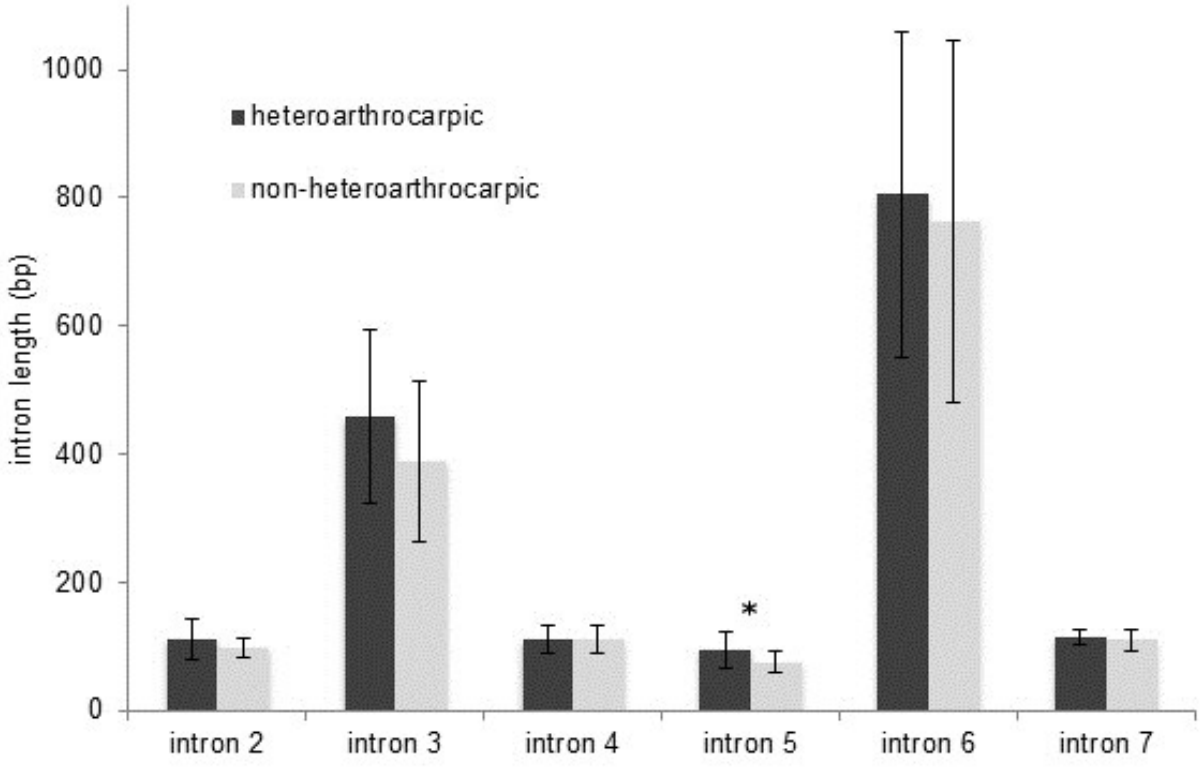


Figure 2-5. Mean intron length in heteroarthrocarpic and non-heteroarthrocarpic taxa. Error bars are equal to one standard deviation from the mean. * denotes statistical significance at p-value < 0.05

Chapter 3 – The Evolution of *TCP1* genes in Brassicaceae and Cleomaceae: setting the stage for floral monosymmetry studies

Introduction

The diversification of angiosperms is marked by many independent shifts from polysymmetric flowers to monosymmetric flowers, as evidenced by the occurrence of this trait in distantly related families (Endress, 2001; Bremer et al., 2009). Some of the largest plant families have monosymmetric flowers and transitions from monosymmetry back to polysymmetry are comparatively rare, indicating that understanding this trait is crucial to understanding the evolution and success of angiosperms (Citerne et al., 2006; Wang et al., 2008; Zhang et al., 2010; Howarth et al., 2011; Preston et al., 2011). Furthermore, floral symmetry is associated with pollinator appeal, signifying that transitions to monosymmetry may be associated with specialized pollinator interactions (Kampny, 1995; Fenster et al., 2009; Ushimaru et al., 2009).

TCP1 is a gene identified in *Arabidopsis thaliana* belonging to the large TCP family of transcription factors in which homologues have been repeatedly associated with changes in floral symmetry (Cubas et al., 1999b; Endress, 2001; Howarth and Donoghue, 2006; Rosin and Kramer, 2009; Busch et al., 2012). Much of our understanding of the genetic regulation of monosymmetry is based on *Antirrhinum majus*, where closely related copies *CYCLOIDEA* (*CYC*) and *DICHOTOMA* (*DICH*) act in a partially redundant fashion to regulate the pathway necessary for the development of a monosymmetric corolla (Luo et al., 1996). *TCP1*, *CYC* and *DICH* are all members of the TCP gene family that belong to the "CYC2" branch within a larger "ECE" clade (Howarth and Donoghue, 2006). In addition to the conserved TCP domain that is common across all TCP genes, the "ECE" clade also contains a

semi-conserved R domain (Cubas et al., 1999a). The presence of the R domain is of interest because multiple events of independent acquisition and loss are the most parsimonious explanation of its reoccurring presence in the TCP gene family, possibly reflecting repeated evolution of similar protein structure and function (Citerne et al., 2003; Damerval and Manuel, 2003).

Almost all studies of taxa with monosymmetric flowers reported the presence of multiple *TCP1* copies, and investigations of Asteraceae, Caprifoliaceae and Malpighiaceae have correlated shifts to monosymmetry with duplication of the *TCP1* lineage (Howarth and Donoghue, 2006; Chapman et al., 2008; Zhang et al., 2010). In monosymmetric species with multiple *TCP1* copies, a common pattern is of partial subfunctionalization amongst the adaxial and abaxial portion of the flower, although this is mostly demonstrated by correlative expression studies (Luo et al., 1999; Feng et al., 2006; Wang et al., 2008; Song et al., 2009; Zhang et al., 2010). Despite a prevalence of *TCP1* phylogenies in the literature, few studies, with the notable exception of Gesneriaceae (Moller et al., 1999; Citerne et al., 2000; Smith et al., 2004; Smith et al., 2006), have described evolutionary patterns of *TCP1* sequences corresponding to shifts in floral symmetry. This omission is partially due to lack of opportunity, as many previously studied taxa belong to purely monosymmetric clades for which monosymmetry is most likely the ancestral trait (Luo et al., 1996; Wang et al., 2008; Preston et al., 2009; Preston et al., 2011). Additional sequence evolution analyses are required to indicate if *TCP1* sequences are subjected to the same evolutionary trends multiple times throughout the history of angiosperms.

The sister families Brassicaceae and Cleomaceae (order Brassicales) offer an opportunity to investigate *TCP1* evolution because this lineage contains both dissymmetric

and monosymmetric representatives (Hall et al., 2002; Hall, 2008). These families represent the only lineage that exhibits three distinct developmental trajectories of achieving monosymmetric flowers, to the best of our knowledge (Busch and Zachgo, 2007; Patchell et al., 2011; Busch et al., 2012). In addition to sharing four known polyploid events, both Brassicaceae and Cleomaceae each underwent an additional tetraloidization and hexaploidization, respectively (Schranz and Mitchell-Olds, 2006; Barker et al., 2009a; Cheng et al., 2013). Thus, these sister families are a good model system to investigate the retention of floral development genes and the relative contribution of duplicate genes in producing monosymmetric flowers (Schranz and Mitchell-Olds, 2006; Jaillon et al., 2007; Ming et al., 2008; Jiao et al., 2011; Vekemans et al., 2012; Cheng et al., 2013).

Most Brassicaceae have dissymmetric flowers, where the corolla is polysymmetric but unequal stamen positioning allow flowers to be divided by only two planes of symmetry (Endress, 2001). The Brassicaceae ancestor is presumed to be dissymmetric, with six out of 330 genera having developed a monosymmetric corolla (Busch and Zachgo, 2007; Busch et al., 2012). *TCP1* is transiently expressed adaxially during early developing *Arabidopsis* floral buds, which seemingly results in a short-lived enlargement of the abaxial sepal (Cubas et al., 2001). However, the mature flower is ultimately dissymmetric. This study shaped the hypothesis that early monosymmetry in Brassicaceae is the ancestral condition that was modified in monosymmetric taxa such that *TCP1* expression persisted into later stages of development (Cubas et al., 2001; Busch and Zachgo, 2007; Busch et al., 2012). *TCP1* expression patterns have been described across one Brassicaceae lineage including *Calepina, Iberis*, and *Teesdalia* (Busch et al., 2012). These are the only monosymmetric taxa from which only a single copy of *TCP1* was identified (Busch et al.,

2012). These taxa attain monosymmetry by having adaxial petals that are shorter than the abaxial petals, which is caused by increased *TCP1* expression in the adaxial petals during later stages of development that restricts cell proliferation (Busch and Zachgo, 2007). In contrast to *Arabidopsis*, *TCP1* expression was not detected in the early sepal primordia in these species. Comparisons of *TCP1* expression domains in both monosymmetric and dissymmetric Brassicaceae revealed a broad expression domain across all floral whorls in both symmetry types (Busch et al., 2012). In monosymmetric taxa, expression was localized to the adaxial petals during later stages of development.

Unlike Brassicaceae, Cleomaceae is comprised almost entirely of monosymmetric taxa whose flowers vary remarkably in petal shape, size, position and colour, as well as stamen number and degree of gynoecium elongation (Endress, 2001; Hall et al., 2002; Patchell et al., 2011). Upward curvature of the petal bases, androecium and gynoecium further enhance the monosymmetry of the flower. Variation in petal size is a trait that is similar to monosymmetric Brassicaceae, although organ curvature is unique to the Cleomaceae. Further, some species of Cleomaceae have additional colour and shape differences between adaxial and abaxial petals (Endress, 2001; Patchell et al., 2011). Additionally, Patchell et al. (2011) identified two separate developmental trajectories resulting in monosymmetric flowers across 11 taxa. Species belonging to the early dissymmetry clade exhibit equally sized sepal and petal primordia early in development that become monosymmetric after emergence of the stamens via upward curvature of the sepals, petals, stamens and pistil before maturation. In contrast, the development of early monosymmetry was described in a grade of taxa. Here, the abaxial sepal grows rapidly to overarch the floral bud early in development and petal primordia are shifted upward to

emerge in the corners of a trapezoid-shaped floral apex. Furthermore, adaxial stamen primordia development is delayed relative to the early emergence of the stamen primordia adjacent to the enlarged abaxial sepal. Monosymmetry of the floral bud continues into the late stages of development where monosymmetry is exaggerated by the persistence of unequal stamen size and upward curvature of sepals, petals, stamens and pistil before anthesis. This monosymmetric pattern early in development is reminiscent of *Arabidopsis* development, suggestive of an ancient monosymmetry shared between the families (Patchell et al., 2011).

The Brassicaceae and Cleomaceae present the opportunity to not only study the evolution of *TCP1* in relation to the acquisition of monosymmetry in this group, but also allow for investigation into the genetic root of three developmental paths leading to monosymmetry. This study will examine the evolution of *TCP1* (CYC2 clade) in these families to address three questions: 1) Are sequences amplified during this study *TCP1* homologues (i.e. not representatives of other members of the ECE clade)? 2) Are multiple copies detected in Brassicaceae and Cleomaceae, and, if so, are multiple copies associated with the two developmental paths to monosymmetry seen in Cleomaceae? 3) Are patterns of molecular evolution correlated with monosymmetry? Comparing the evolutionary history of *TCP1* genes alongside the diversity of monosymmetric flowers in Brassicaceae and Cleomaceae is an important step in assessing the potential role of *TCP1* in floral symmetry.

Methods

Taxon sampling

TCP1 sequences were obtained from twenty species comprising 10 Brassicaceae and 10 Cleomaceae taxa (Table 3-1). Although sequences were obtained for 31 taxa, 11 species were removed due to presence of a contamination whose origin was traced to a previous PCR contamination event in the fall of 2010. For future researchers continuing to work on this project, a summary of taxa removed from this study, attempted methods to resolve the contamination issue, and sequence of the suspect amplicon are presented in Appendices 3-1 to 3-3. Excluding these taxa nonetheless retains broad taxonomic sampling across both families. Cleomaceae taxa comprised representatives from seven out of 15 clades within the family, including: African, Clade 6, Cleome s.str, Droserifolia, Melidiscus, Polanisia and Tarenaya lineages (Feodorova et al., 2010; Patchell et al., 2014). Additionally, *TCP1* sequences were included from the genomic libraries of *Cleome violacea* and *Tarenaya hassleriana.* This sampling of Cleomaceae includes five taxa exhibiting early monosymmetry, three taxa exhibiting early dissymmetry and two taxa where the development pathway to monosymmetry has not been assessed (Table 3-1). Brassicaceae taxa comprised representatives from seven out of 49 tribes, including: Aethionemeae, Arabideae, Brassiceae, Iberideae, Isatideae, Streptantheae, and Thlaspideae tribes (Al-Shehbaz, 2012). Inclusion of Aethionemeae is particularly important as this tribe is sister to all remaining Brassicaceae. GenBank sequences were retrieved for an additional 12 Brassicaceae species to include the Anastaticeae, Calepineae, Camelineae, Cardamineae, Dontostemoneae, and Eutremeae lineages (Al-Shehbaz, 2012). In total, the sampled taxa

for the *TCP1* lineage consisted of 18 dissymmetric Brassicaceae, four monosymmetric Brassicaceae and 10 monosymmetric Cleomaceae.

This sampling was supplemented by mining available genomic libraries, which permits testing of amplified sequences for monophyly. In addition to *TCP1* from *Carica papaya* (papaya; Caricaceae), a basal taxon in the Brassicales (Bremer et al., 2009), members of the "ECE" clade including five *TCP12* and six *TCP18* sequences from both Brassicaceae and Cleomaceae representatives were included (Table 3-1). Furthermore, two *TCP17* sequences from outside of the "ECE" clade were included as outgroups.

Amplification, cloning and sequencing

TCP1 genes were amplified from genomic DNA extracted according to Hall et al. (2011). Primers TCP1-fwd 5' and TCP1-rev 5' designed by Busch et al. (2007) or TCP1 F degen.2 and TCP1 R degen.2 designed for this study were used to generate a fragment ranging from 798 – 1140bp (Table 3-2). TCP1 F degen.2 and TCP1 R degen.2 were designed ~350 bp upstream of the TCP domain and ~350bp downstream of the R domain, respectively. Sequence information from two Brassicaceae (*Arabidopsis thaliana* and *I. amara*) and three Cleomaceae (*Cleome viridiflora, C. violacea*, *Tarenaya spinosa*,) was used to determine degenerate sites for primer design. Takara Ex Taq™ polymerase and associated reagents were used in the PCR mix according to manufacturer's instructions (Takara Bio Inc., Otsu, Shiga, Japan). Six samples per taxon were placed along a temperature gradient in an Eppendorf Mastercycler Pro thermocycler (Eppendorf Canada, Mississauga, Ontario) under hot start conditions with the cycling settings as follows: initial denaturation for 10 minutes at 94°C, followed by 36 cycles of 94°C for 30 seconds, 50–60°C

for 1 minute, 72°C for 2 minutes, followed by a final extension of 72°C for 10 minutes. PCR products were then ligated into a PGEM easy T vector and transformed into PGEM JM109 chemically competent E. coli cells (Promega Corporation, Fitchburg, Wisconsin, USA) as per manufacturer protocol using half reactions. Twenty-five μL of transformed cells were plated onto LB agar plates containing 0.5μL carbomycin /1 mL LB agar and 0.8mg X-Gal (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and incubated overnight at 37°C. Twenty-four or more colonies were screened using M13F and M13R primers and Taq DNA Polymerase (Qiagen, Hilden, Germany) as per the manufacturer's instructions while using a small amount of the colony as template. Eight or more clones were grown for 16 hours in 3 mL LB before purifying the plasmids using GeneJet Plasmid miniprep kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Plasmids were sequenced using M13F and M13R primers on an ABI 3730 capillary sequencer (Applied Biosystems Inc., Foster City, California, USA). Base-calling errors were assessed using MacVector v12.1 (Rastogi, 2000). Two – 11 sequences were obtained for each taxon, with the exception of *C. viridiflora*, which only yielded one usable sequence. A summary of amplification efforts are presented in Appendix 3-4.

Sequence alignment and phylogenetic analysis

Two alignments were analyzed: 1) a nucleotide alignment containing all sequenced *TCP1* clones, hereafter called "all *TCP1* clones"; and 2) a nucleotide alignment including only the TCP domain from one representative *TCP1* sequence per monophyletic group of clones, hereafter referred to as "representative TCP domain". The later alignment included additional TCP family members: *TCP12*, *TCP17* and *TCP18*. Both matrices were aligned

using MAFFT v7.0 (Katoh and Standley, 2013) using default settings then manually adjusted such that they were positioned in the proper reading frame.

A Bayesian analysis of the "all *TCP1* clones" alignment was conducted in MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001). A GTR + I + Γ model of sequence evolution determined by ModelTest v3.06 using Akaike information criterion (Posada and Crandall, 1998). The Bayesian Markov chain Monte-Carlo (MCMC) analysis used four chains and default temperature and priors. The analysis ran for 10,000,000 generations at which the potential scale reduction factors were approximately 1 and standard deviation of the split frequency was < 0.01. Stationarity of the run was confirmed using Tracer v1.4.1 (Rambaut and Drummond, 2007). The first 25% of the trees were discarded as burnin and posterior probabilities (PP) were obtained from a 50% majority rule consensus tree. This tree was used to choose one representative sequence from each monophyletic group of fully sequenced clones to construct the "representative TCP domain" alignment. All subsequent analyses were conducted on the "representative TCP domain" alignment which included all 177 bp of the TCP domain.

A Bayesian analysis on the "representative TCP domain" alignment was conducted using the same conditions as in the "all TCP1 clones" alignment. Maximum parsimony (MP) bootstrapping (BS) analyses were also conducted on this alignment to provide additional measures of branch support. MP heuristic searches were conducted in PAUP* v4.0β10 (Swofford, 2003) starting from random trees and using 100 random addition replicates, tree-bisection-reconnection (TBR) branch swapping and holding 10 trees per replicate. Bootstrapping (BS) was calculated using 1000 replicates, TBR swapping, random addition and saving no more than 1000 trees per replicate.

Molecular evolution tests

Evolution models were designed in PAML v4.4 (Yang et al., 2000) to partition the data in the following ways: 1) Brassicaceae vs Cleomaceae to assess differences in rates of *TCP1* evolution associated with divergence of the families; 2) Dissymmetric vs monosymmetric taxa to assess differences in rates of *TCP1* evolution associated with floral symmetry; 3) Early monosymmetric vs early dissymmetric Cleomaceae taxa to assess differences in rates of *TCP1* evolution associated with the two different Cleomaceae monosymmetry development pathways. The nonsynonymous to synonymous substitution ratios (dN/dS) of these models were compared to a "fixed" model of evolution where all branches were constrained to the same ratio and a "free" model where each branch was allowed to have its own ratio. Significance of model favourability was measured by conducting likelihood ratio tests (LRTs), which penalizes models with additional parameters (Baum et al., 2005; Jaramillo and Kramer, 2007; Zhang et al., 2008).

Results

Statistics on the data sets

The "all *TCP1* clones" aligned dataset comprised 119 ingroup sequences and two out group sequences from 21 species and was 1260 bp long. It contained 86% variable characters and 68% parsimony informative characters. The "representative TCP domain" alignment comprised 68 sequences from the *TCP1*, *TCP12*, and *TCP18* clades and two sequences from the *TCP17* out group. This data set was 177 bp long with 76% variable characters and 64% parsimony informative characters.

Phylogenetic analyses

Both alignments revealed *TCP1* sequences were monophyletic (100% MP BS, 100% PP) for both alignments; Figures 3-1 and 3-2). Within this clade, I designated two groups. The first is called Group I, which represents a monophyletic clade (90% MP BS; 100% PP) of sequences from only Cleomaceae. In contrast, Group II is a paraphyletic grade of sequences from both Brassicaceae and Cleomaceae (Figure 3-2). A subsequent Bayesian analysis on an alignment excluding the Group II Cleomaceae using the same model of evolution and parameters as applied to the other two alignments was also conducted (data not shown), but failed to represent the Brassicaceae as monophyletic. The topology from the "all *TCP1* clones" Bayesian analyses resulted in monophyly of clones from each taxon, aside from when multiple copies were suspected (Figure 3-1). However, relationships among clones were not resolved between the closely related *C. arabica* and *C. africana*. The single representative sequence chosen for the TCP domain alignment was >99% similar to the other sequences amplified for that species.

Whereas a single *TCP1* copy was retrieved from most taxa, there are three Brassicaceae and three Cleomaceae exceptions. Two possible copies of *TCP1* from within Group II were present in Brassicaceae taxa *Brassica napus*, *Caulanthus amplexicaulis* and *Erucaria erucarioides*. This trend was confirmed by the retrieval of two sequences from the *B. rapa* genome (Lyons and Freeling, 2008) which are similar to the two putative *TCP1* copies found in *B. napus.* One copy from each of these species was resolved as a single clade ("all *TCP1* clones" alignment: 100% MP BS, 99% PP; "representative TCP domain" alignment: 100% MP BS, 92% PP). However, relationships amongst other Brassicaceae clones are unresolved. Three copies were found in Cleomaceae taxa *T. spinosa* and *Cleome*

violacea, including two copies from Group I and one copy from Group II. Additionally, two copies from Group I were found in *T. hassleriana.* The Group II copies of these taxa are located amongst Brassicaceae taxa in the "all *TCP1* clones" alignment (Figures 3-1 and 3-2).

Sequence divergence between Group I and II resulted in several regions along the entire amplified sequence that were difficult to align (Appendix 3-5). The R domain was identified in all *TCP1*-like sequences, although only the first ten out of 18 amino acids in Group I were alignable to Group II. A summary of sequence similarity of the TCP domain between Group I and Group II taxa as well as length of the coding sequence is presented in Table 3-3. A divergent sequence, *C. violacea* 1b from Group I, was removed from the sequence similarity assessment.

Selection tests

Maximum likelihood tests of selection revealed that purifying selection is the primary evolutionary force acting on these genes because the dN/dS was low (<0.4) in all models (Table 3-4). The free model was the least favoured model. The model partitioning dN/dS heterogeneity between dissymmetric and monosymmetric taxa was the most favoured model, although it was not significantly different than the fixed model (α=0.05; p-value = 0.07). Models partitioning dN/dS between Cleomaceae and Brassicaceae, and early monosymmetric vs late monosymmetric Cleomaceae taxa were not significantly different than the fixed model (p-value 0.84 and 0.88, respectively).

Discussion

The evolution of *TCP1* genes has not yet been studied in Brassicaceae and Cleomaceae despite the group's variation in monosymmetry (Patchell et al., 2011),

inclusion of the model plant *Arabidopsis* (O'Kane and Al-Shehbaz, 2003), and recent genomic work on *T. hassleriana* (Barker et al., 2009a; Braeutigam et al., 2011; Cheng et al., 2013; Bhide et al., 2014). The investigation of *TCP1* genes in Brassicaceae and Cleomaceae presented here has yielded a valuable framework and some interesting trends regarding the involvement of these genes in floral development. Moreover, these data will enable future hypothesis testing relating copy number and sequence divergence to presence of differential expression domains and gene function.

TCP1 clade

TCP1 sequences from Cleomaceae and Brassicaceae were strongly supported as monophyletic, although intriguing differences amongst sequences within this clade were observed. While Group II contains both Brassicaceae and Cleomaceae sequences, Group I represents only species from Cleomaceae. The differences between these two groups are exemplified by partial loss of conservation in the R domain and only 67-82% sequence similarity of the TCP domain between Groups I and II. Furthermore, Group I coding sequences are an average of 258 bp longer (Table 3-3; Appendix 3-5). This level of divergence in the TCP domain is comparable to the 71-86% sequence similarity observed between *CYC* and *DICH* in the Antirrhineae (Veronicaceae), which are known to perform important partially subfunctional roles in monosymmetry development (Luo et al., 1996; Hileman and Baum, 2003). The exclusivity of Group I sequences to the Cleomaceae is supported by the fact that all Brassicaceae sequences retrieved from outside genomic resources (Lyons and Freeling, 2008), in addition to sequences amplified in this study, belonged to Group II. Studies showing the involvement of *TCP1* in the monosymmetry in

Brassicaceae lead to the logical hypothesis that the same gene is involved in monosymmetry in Cleomaceae (Busch and Zachgo, 2007; Busch et al., 2012; Cheng et al., 2013). The presence of Group I depicted here could reflect that the same gene was recruited in Cleomaceae, but that sequence divergence may underlie the evolution of more elaborate floral symmetry in this family.

Contrary to the pattern where Group I sequences are conserved, the largely dissimilar *C. violacea* 1b gene retrieved from an unpublished draft of the full genome was included in the *TCP1* Group I lineage (Figures 3-1 and 3-2). The sequence of the TCP domain is 69% similar to *C. violacea* 1a, which also belongs in Group I. In comparison, all remaining sequences in Group I display 88.6% - 100% sequence similarity of the TCP domain (Table 3-3). This divergence may indicate that this is a duplicate gene that has undergone pseudogenization, the process by which a gene accrues deleterious mutations until it is no longer a functional member of the genome (Moore and Purugganan, 2005). However, the TCP domain from this unpublished sequence is still conserved enough to be retrieved via blast search using a representative sequence of the *TCP1* Group I lineage and no frame shifts or stop codons were detected in this sequence. The divergence of the coding sequence outside of the TCP domain likely explains why this gene was not amplified as primers were not located in the TCP domain. As selection tests are based on sequence homology, positive selection on this gene could not be assessed due to the fact that it was only alignable along 55% of the length of other retrieved *TCP1* sequences, resulting in alignment gaps. Future expression and functional experiments would be useful to assess whether this gene plays a role in floral development.

Relationships amongst *TCP1* are not well-resolved in the presented phylogenies (Figures 3-1 and 3-2), making comparisons with species phylogenies derived from nuclear genes difficult (Al-Shehbaz, 2012; Patchell et al., 2014). This finding is not surprising as similar *TCP1* studies in other angiosperms also possess a considerable number of nodes with low support (Citerne et al., 2003; Howarth and Donoghue, 2005; Smith et al., 2006; Carlson et al., 2011). Comparing the taxon quantity used here to other studies suggests that additional taxon sampling may not improve resolution, although a more rigorous exploration for *TCP1* copies may help resolve relationships within Brassicaceae and Cleomaceae.

These data do not support the Brassicaceae *TCP1* genes as monophyletic, although monophyly of this family is well accepted (Hall et al., 2002; Iltis et al., 2011; Al-Shehbaz, 2012). These results also contradict a strongly supported *CYC2*/*TCP1* clade of Brassicaceae sequences in a previous investigation (Busch et al., 2012), although it is important to note that those authors did not sample Cleomaceae and only included 11 taxa of Brassicaceae (compared to 22 in this study). The paraphyly of Brassicaceae can be explained by two possible scenarios: 1) the *TCP1* alignment is causing systematic error, giving the appearance of untrue relationships, or 2) one of the copies retained by Cleomaceae taxa originated from a duplication before the Brassicaceae and Cleomaceae diverged. However, scenario two would likely require longer branch lengths than observed in this study in order to reflect sequence dissimilarity acquired through divergence of the families. When compared to the "representative TCP domain" alignment (Figure 3-2), inclusion of the entire *TCP1* sequence in the "all *TCP1* clones" alignment (Figure 3-1) increased anomalies that are likely the result of problematic alignment (Appendix 3-5). For example, when

questionable regions of alignment were included in the "all *TCP1* clones" alignment, the topology revealed Brassicaceae taxa *Dontostemon integrifolius* within the Group I Cleomaceae clade (100% MP BS, 99% PP) and Brassicaceae members *Brassica rapa*, *Caulanthus amplexicaulis*, *Erucaria erucarioides* and *Myagrum perfoliatum* in the same Group II clade as Cleomaceae taxa (100% MP BS, 70% PP). Other published *TCP1* gene phylogenies have cited a high degree of variability as well as unreliable alignment outside of the TCP domain (Reeves and Olmstead, 2003; Citerne et al., 2003; Damerval et al., 2007; Chapman et al., 2008; Carlson et al., 2011). However, lack of monophyly of the Brassicaceae is found in both the "all *TCP1* clones" entire sequence alignment as well as the "representative TCP domain" alignment topologies, indicating that the evolutionary pattern of *TCP1* genes is different than that of previously utilized phylogenetic markers for species tree construction. Again, this is consistent with some studies where *TCP1* phylogenies in the Dipsacales and Ranunculales do not support families as monophyletic despite evidence of monophyly from analyses of other loci (Howarth and Donoghue, 2005; Damerval et al., 2007; Carlson et al., 2011).

Copy number in Brassicaceae

In conjunction with the single known *TCP1* copy in *Arabidopsis* (Cubas et al., 1999b; Finlayson, 2007) and previously studied Brassicaceae (Busch and Zachgo, 2007; Busch et al., 2012), the presence of two *TCP1* copies in four Brassicaceae represents a novel finding of this study as it implies variable *TCP1* copy retention throughout the family. Although a known hexaploid event at the base of the Brassiceae tribe may explain additional copies found in *E. erucarioides*, *B. napus* and *B. rapa* (Lysak et al., 2005; Lysak et al., 2007), *C.*

amplexicaulis belongs to Schizopetaleae for which there is no known polyploid event (Al-Shehbaz, 2012). Furthermore, one copy from each of these taxa form a monophyletic group with additional Brassiceae and Schizopetaleae taxa, including *Cakile lanceolata* and *Stanleya bipinnata*, respectively (Beilstein et al., 2006; Bailey et al., 2006). The presence of this relationship in representatives from two Brassicaceae tribes indicates that this copy may be a duplicate maintained from the ancestral genome doubling at the base of the family, rather than subsequent tribe specific duplications (Schranz and Mitchell-Olds, 2006; Barker et al., 2009a).

Findings of multiple copies in Brassicaceae may also have implications for previously published floral symmetry development studies. Despite monosymmetric *I. amara* having similar sepal primordia development when compared to the ultimately dissymmetric *Arabidopsis* (Cubas et al., 2001), Busch and Zachgo (2007) did not detect monosymmetric *TCP1* expression in *I. amara* during early development. Additional expression data from more monosymmetric and dissymmetric Brassicaceae species also failed to detect early onset monosymmetric *TCP1* expression, but rather revealed broad expression domains across all whorls of the developing bud in early stages (Busch et al. 2012). This led to the conclusion that *TCP1* expression associated with monosymmetry early in development is not a prerequisite for the evolution of monosymmetry in Brassicaceae. However, it is possible that subfunctionalization of two copies, particularly via partitioning expression during early and late stages of floral development, may account for its absence in their study. This scenario is somewhat consistent with *Antirrhinum*, where *CYC* and *DICH* perform partially redundant functions but *CYC* is expressed from early to late stages of development whereas *DICH* expression is restricted to the inner adaxial

petal lobe during late-stage development (Luo et al., 1999). These data exemplify the need for gene phylogenies as a first step in evo-devo studies, and indicate that early expression of *TCP1* in the flowers of Brassicaceae needs to be revisited alongside an exhaustive search for copy number in these taxa.

Copy number in Cleomaceae

Only two Cleomaceae taxa, *T. spinosa* and *C. violacea,* were found to have three copies despite the known hexaploid event in the family (Schranz and Mitchell-Olds, 2006; Barker et al., 2009; Braeutigam et al., 2011; Cheng et al., 2013). These included two copies from Group I and one copy from Group II. Additionally, two copies from Group I were found in *T. hassleriana.* However, copies retrieved with our primers did not match the number of putative copies retrieved from genomic resources. For instance, copy 1b (Group I) from *C. violacea* was only retrieved from genomic resources and only one clone was obtained for copy 1a (Group I) from *T. spinosa*. In both of these instances, apparent mutations at primer binding sites may have prevented amplification. Furthermore, only one Group II sequence was retrieved from *C. viridiflora*, despite all other Cleomaceae having Group I sequences (Appendix 3-4). This indicates that the primers used in this study may not have retrieved all copies, resulting in an underestimate of copy number. However, genomic analysis of *T. hassleriana* found that less than half of genes were retained in three copies (Cheng et al., 2013), demonstrating a high possibility of copy loss following polyploid events. Thus, retention of less than three *TCP1* copies is possible.

Genomic comparisons amongst *Arabidopsis*, *B. rapa* and *T. hassleriana*, revealed that *T. hassleriana* retained more copies of the B-class floral development genes *PISTILLATA*

(*PI*) and *APETALA3* (*AP3*) when compared to Brassicaceae representatives. As *AP3* has been associated with monosymmetry of monocot flowers (Tsai et al., 2004; Tsai et al., 2008; Mondragon-Palomino and Theissen, 2009; Bartlett and Specht, 2010; Preston and Hileman, 2012), and *PI* has been linked to floral variation in Asterids (Viaene et al., 2009), the authors speculate that B-class genes may be responsible for floral diversification in Cleomaceae. Although these data represent a promising basis for future hypothesis testing, the specific focus on *TCP1* in this study has shown that although copy number may be comparable between the two families, patterns of *TCP1* sequence differentiation indicate that this gene still stands as a likely candidate. Because Cleomaceae sequences within Group II align easily with those of Brassicaceae, the comparatively divergent Cleomaceae-specific Group I sequences are likely players underlying the unique floral morphology exhibited in this family. Monosymmetry of some Brassicaceae, despite their lack of a Group I-like sequence, may be explained by anatomical differences by which Brassicaceae and Cleomaceae obtain monosymmetry (Busch and Zachgo, 2007; Patchell et al., 2011). The additional complexity of organ curvature found in Cleomaceae is likely a reflection of further complexity at the genetic level.

In situ expression data of copy 1b (Group I) from *C. violacea* reveal uniform expression in early to late stages of floral development (Patchell, 2012). Although these data are consistent with *TCP1* expression found in both early developing dissymmetric and monosymmetric Brassicaceae (Busch et al., 2012), it does not explain what is driving the persistent floral monosymmetry found in all Cleomaceae. Aside from Brassicaceae examples, nearly all studies of monosymmetric taxa have documented differential expression of multiple *TCP1* copies where one copy is restricted to the adaxial portion of

the developing flower relative to the other (Howarth et al., 2011). Remarkably, this pattern of expression in multiple copies is repeatedly found in distantly related angiosperms including *Pisum* and *Lotus* (Feng et al., 2006; Wang et al., 2008), *Lonicera* (Zhang et al., 2010), *Opithandra* (Song et al., 2009) and *Antirrhinum* (Luo et al., 1999). This represents an increased likelihood that a similar parallelism may exist in Cleomaceae, where expression studies on additional copies detected in this study may reveal an adaxially restricted expression domain. In particular, the sequence divergence in copy 1b from *C. violacea* is intriguing because a similar sequence was not identified in the genomic library of *T. hassleriana.* While *C. violacea* exhibits early monosymmetry floral development, *T. hassleriana* likely displays early dissymmetry based on studies of closely related *T. spinosa* (Patchell et al., 2011; Patchell et al., 2014). This indicates that this copy may be a good candidate to investigate the cause of early dissymmetry in *C. violacea*.

Models of Evolution

Purifying selection found across the branches of the Brassicaceae and Cleomaceae *TCP1* phylogeny indicate a history of gene preservation associated with a necessary function in both the Group I and Group II sequences. Positive selection, as found in the evolution of multiple copies of *TCP1* homologues controlling the development of ray and discoid florets in sunflower (*Helianthus annuus*) (Chapman et al., 2008; Chapman et al., 2012), was absent in this study. Although positive selection cannot be discounted on a few specific sites (Hileman and Baum, 2003), these evolutionary trends imply partial redundancy or subfunctionalization. This is consistent with purifying selection found in the paralogues *CYC* and *DICH* in the Antirrhineae (Veronicaceae) that is associated with

their partially redundant roles in establishing floral monosymmetry (Hileman and Baum, 2003; Preston et al., 2009).

Although the model testing for evolutionary differences between dissymmetric and monosymmetric taxa was not significant at $\alpha = 0.05$, the favourability of this model when compared to the others, in addition to the small p-value (0.07), indicates that the sequence evolution of these genes may be correlated with floral monosymmetry. However, because the Brassicaceae vs Cleomaceae model was not similarly favoured, this indicates that the sequence divergence between monosymmetric and dissymmetric Brassicaceae members is driving the favourability of this model. The lack of significance for this model, along with the findings of purifying selection, may be due to the large insertion and deletion gaps required to align the Cleomaceae-specific Group I *TCP1* sequences to the Group II sequences. Due to the nature by which dN/dS is calculated, codons that were not present in all taxa because of alignment restrictions were deleted. Thus, comparing evolution of sequences in this context can only measure sequence change caused by point mutations, but not insertion or deletion events. This may have resulted in the deletion of some of the most important novel functional changes to the *TCP1* sequence (Yang et al., 2000; Hileman and Baum, 2003). Contrarily, both Group I and Group II may play important roles outside of determining floral symmetry, resulting in low variation in dN/dS in order to perform other functions. For instance, Cubas et al. (2001) found that *TCP1* expression was adaxially expressed during lateral shoot elongation. Other studies in Gesneraceae that compare *TCP1* evolution between polysymmetric and monosymmetric taxa have revealed a contradictory interpretation, with LRT and p-values supporting no significant difference in sequence evolution associated with floral symmetry. Smith et al. (2006) attributed this

lack of significance to activity at the transcriptional level which cannot be detected by measuring dN/dS, and because they suspect *TCP1* has other functions that constrain sequence evolution.

Conclusions

This study represents the first step of future experiments intending to uncover the genetic complexity behind interesting developmental pathways generating floral monosymmetry in Brassicaceae and Cleomaceae. Likely *TCP1* duplicates in both families have been identified and will serve as candidates for further expression and functional studies and will help decipher the intricacies of temporal and spatial expression of these genes. Despite difficulties with *TCP1* alignment in this study as well as others (Reeves and Olmstead, 2003; Howarth and Donoghue, 2006), this study exemplifies the importance of establishing developmental gene evolution and copy number as an initial step in evolutionary development research in angiosperms. As angiosperms are particularly well-known for their propensity for polyploidization, establishing copy number is a significant factor affecting accurate conclusions.

References

AL-SHEHBAZ I. A. 2012. A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon* 61: 931-954.

BAILEY C. D., M. A. KOCH, M. MAYER, K. MUMMENHOFF, S. L. O'KANE Jr., S. I. WARWICK, M. D. WINDHAM, AND I. A. AL-SHEHBAZ. 2006. Toward a global phylogeny of the Brassicaceae. *Molecular Biology and Evolution* 23: 2142-2160.

BARKER M. S., H. VOGEL, AND M. E. SCHRANZ. 2009. Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biology and Evolution* 1: 391-399.

BARTLETT M. E. AND C. D. SPECHT. 2010. Evidence for the involvement of *GLOBOSA*-like gene duplications and expression divergence in the evolution of floral morphology in the Zingiberales. *New Phytologist* 187: 521-541.

BAUM D., H. YOON, AND R. OLDHAM. 2005. Molecular evolution of the transcription factor *LEAFY* in Brassicaceae. *Molecular Phylogenetics and Evolution* 37: 1-14.

BEILSTEIN M., I. AL-SHEHBAZ, AND E. KELLOGG. 2006. Brassicaceae phylogeny and trichome evolution. *American Journal of Botany* 93: 607-619.

BHIDE A., S. SCHLIESKY, M. REICH, A. P. M. WEBER, AND A. BECKER. 2014. Analysis of the floral transcriptome of *Tarenaya hassleriana* (Cleomaceae), a member of the sister group to the Brassicaceae: towards understanding the base of morphological diversity in Brassicales. *BMC Genomics* 15: 140.

BRAEUTIGAM A., K. KAJALA, J. WULLENWEBER, M. SOMMER, D. GAGNEUL, K. L. WEBER, K. M. CARR, U. GOWIK, J. MASS, M. J. LERCHER, P. WESTHOFF, J. M. HIBBERD, AND A. P. M. WEBER. 2011. An mRNA blueprint for C-4 photosynthesis derived from comparative transcriptomics of closely related C-3 and C-4 species. *Plant Physiology* 155: 142-156.

BREMER B., K. BREMER, M. W. CHASE, M. F. FAY, J. L. REVEAL, D. E. SOLTIS, P. S. SOLTIS, P. F. STEVENS, A. A. ANDERBERG, M. J. MOORE, R. G. OLMSTEAD, P. J. RUDALL, K. J. SYTSMA, D. C. TANK, K. WURDACK, J. Q. XIANG, S. ZMARZTY, AND ANGIOSPERM PHYLOGENY GRP. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105-121.

BUSCH A., S. HORN, A. MUEHLHAUSEN, K. MUMMENHOFF, AND S. ZACHGO. 2012. Corolla monosymmetry: evolution of a morphological novelty in the Brassicaceae family. *Molecular Biology and Evolution* 29: 1241-1254.

BUSCH A. AND S. ZACHGO. 2007. Control of corolla monosymmetry in the Brassicaceae *Iberis amara*. *Proceedings of the National Academy of Sciences of the United States of America* 104: 16714-16719.

CARLSON S. E., D. G. HOWARTH, AND M. J. DONOGHUE. 2011. Diversification of *CYCLOIDEA*-like genes in Dipsacaceae (Dipsacales): implications for the evolution of capitulum inflorescences. *BMC Evolutionary Biology* 11: 325.

CHAPMAN M. A., J. H. LEEBENS-MACK, AND J. M. BURKE. 2008. Positive selection and expression divergence following gene duplication in the sunflower CYCLOIDEA gene family. *Molecular Biology and Evolution* 25: 1260-1273.

CHAPMAN M. A., S. TANG, D. DRAEGER, S. NAMBEESAN, H. SHAFFER, J. G. BARB, S. J. KNAPP, AND J. M. BURKE. 2012. Genetic analysis of floral symmetry in Van Gogh's sunflowers reveals independent recruitment of *CYCLOIDEA* genes in the Asteraceae. *PLoS Genetics* 8: e1002628.

CHENG S., E. VAN DEN BERGH, P. ZENG, X. ZHONG, J. XU, X. LIU, J. HOFBERGER, S. DE BRUIJN, A. S. BHIDE, C. KUELAHOGLU, C. BIAN, J. CHEN, G. FAN, K. KAUFMANN, J. C. HALL, A. BECKER, A. BRAEUTIGAM, A. P. M. WEBER, C. SHI, Z. ZHENG, W. LI, M. LV, Y. TAO, J. WANG, H. ZOU, Z. QUAN, J. M. HIBBERD, G. ZHANG, X. ZHU, X. XU, AND M. E. SCHRANZ. 2013. The *Tarenaya hassleriana* genome provides insight into reproductive trait and genome evolution of crucifers. *Plant Cell* 25: 2813-2830.

CITERNE H. L., M. MOLLER, AND Q. C. B. CRONK. 2000. Diversity of *CYCLOIDEA*-like genes in Gesneriaceae in relation to floral symmetry. *Annals of Botany* 86: 167-176.

CITERNE H. L., R. T. PENNINGTON, AND Q. C. B. CRONK. 2006. An apparent reversal in floral symmetry in the legume *Cadia* is a homeotic transformation. *Proceedings of the National Academy of Sciences of the United States of America* 103: 12017-12020.

CITERNE H., D. LUO, R. PENNINGTON, E. COEN, AND Q. CRONK. 2003. A phylogenomic investigation of *CYCLOIDEA*-like TCP genes in the Leguminosae. *Plant Physiology* 131: 1042-1053.

CUBAS P., E. COEN, AND J. M. M. ZAPATER. 2001. Ancient asymmetries in the evolution of flowers. *Current Biology* 11: 1050-1052.

CUBAS P., N. LAUTER, J. DOEBLEY, AND E. COEN. 1999a. The TCP domain: a motif found in proteins regulating plant growth and development. *Plant Journal* 18: 215-222.

CUBAS P., C. VINCENT, AND E. COEN. 1999b. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401: 157-161.

DAMERVAL C. AND M. MANUEL. 2003. Independent evolution of *CYCLOIDEA*-like sequences in several angiosperm taxa. *Comptes Rendus Palevol* 2: 241-250.

DAMERVAL C., M. LE GUILLOUX, M. JAGER, AND C. CHARON. 2007. Diversity and evolution of *CYCLOIDEA*-like TCP genes in relation to flower development in Papaveraceae. *Plant Physiology* 143: 759-772.

ENDRESS P. K. 2001. Evolution of floral symmetry. *Current Opinion in Plant Biology* 4: 86-91.

FENG X. Z., Z. ZHAO, Z. X. TIAN, S. L. XU, Y. H. LUO, Z. G. CAI, Y. M. WANG, J. YANG, Z. WANG, L. WENG, J. H. CHEN, L. Y. ZHENG, X. Z. GUO, J. H. LUO, S. S. SATO, S. TABATA, W. MA, X. L. CAO, X. H. HU, C. R. SUN, AND D. LUO. 2006. Control of petal shape and floral zygomorphy in *Lotus japonicus*. *Proceedings of the National Academy of Sciences of the United States of America* 103: 4970-4975.

FENSTER C. B., W. S. ARMBRUSTER, AND M. R. DUDASH. 2009. Specialization of flowers: is floral orientation an overlooked first step? *New Phytologist* 183: 502-506.

FEODOROVA T. A., E. V. VOZNESENSKAYA, G. E. EDWARDS, AND E. H. ROALSON. 2010. Biogeographic patterns of diversification and the origins of C-4 in *Cleome* (Cleomaceae). *Systematic Botany* 35: 811-826.

FINLAYSON S. A. 2007. *Arabidopsis TEOSINTE BRANCHED1-LIKE 1* regulates axillary bud outgrowth and is homologous to monocot *TEOSINTE BRANCHED1*. *Plant and Cell Physiology* 48: 667-677.

HALL J. C., K. J. SYTSMA, AND H. H. ILTIS. 2002. Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. *American Journal of Botany* 89: 1826-1842.

HALL J. C. 2008. Systematics of Capparaceae and Cleomaceae: An evaluation of the generic delimitations of *Capparis* and *Cleome* using plastid DNA sequence data. *Botany-Botanique* 86: 682-696.

HALL J. C., T. E. TISDALE, K. DONOHUE, A. WHEELER, M. A. AL-YAHYA, AND E. M. KRAMER. 2011. Convergent evolution of a complex fruit structure in the tribe Brassiceae (Brassicaceae). *American Journal of Botany* 98: 1989-2003.

HILEMAN L. C. AND D. A. BAUM. 2003. Why do paralogs persist? Molecular evolution of *CYCLOIDEA* and related floral symmetry genes in Antirrhineae (Veronicaceae). *Molecular Biology and Evolution* 20: 591-600.

HOWARTH D. G. AND M. J. DONOGHUE. 2005. Duplications in *CYC*-like genes from Dipsacales correlate with floral form. *International Journal of Plant Sciences* 166: 357-370.

HOWARTH D. G. AND M. J. DONOGHUE. 2006. Phylogenetic analysis of the "'ECE" (CYC/TB1) clade reveals duplications predating the core eudicots. *Proceedings of the National Academy of Sciences of the United States of America* 103: 9101-9106.

HOWARTH D. G., T. MARTINS, E. CHIMNEY, AND M. J. DONOGHUE. 2011. Diversification of *CYCLOIDEA* expression in the evolution of bilateral flower symmetry in Caprifoliaceae and *Lonicera* (Dipsacales). *Annals of Botany* 107: 1521-1532.

HUELSENBECK J. AND F. RONQUIST. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.

ILTIS H. H., J. C. HALL, T. S. COCHRANE, AND K. J. SYTSMA. 2011. Studies in the Cleomaceae I. on the separate recognition of Capparaceae, Cleomaceae, and Brassicaceae. *Annals of the Missouri Botanical Garden* 98: 28-36.

JAILLON O., J. AURY, B. NOEL, A. POLICRITI, C. CLEPET, A. CASAGRANDE, N. CHOISNE, S. AUBOURG, N. VITULO, C. JUBIN, A. VEZZI, F. LEGEAI, P. HUGUENEY, C. DASILVA, D. HORNER, E. MICA, D. JUBLOT, J. POULAIN, C. BRUYERE, A. BILLAULT, B. SEGURENS, M. GOUYVENOUX, E. UGARTE, F. CATTONARO, V. ANTHOUARD, V. VICO, C. DEL FABBRO, M. ALAUX, G. DI GASPERO, V. DUMAS, N. FELICE, S. PAILLARD, I. JUMAN, M. MOROLDO, S. SCALABRIN, A. CANAGUIER, I. LE CLAINCHE, G. MALACRIDA, E. DURAND, G. PESOLE, V. LAUCOU, P. CHATELET, D. MERDINOGLU, M. DELLEDONNE, M. PEZZOTTI, A. LECHARNY, C. SCARPELLI, F. ARTIGUENAVE, M. E. PE, G. VALLE, M. MORGANTE, M. CABOCHE, A. ADAM-BLONDON, J. WEISSENBACH, F. QUETIER, P. WINCKER, AND FRENCH-ITALIAN PUBLIC. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463-475.

JARAMILLO M. A. AND E. M. KRAMER. 2007. Molecular evolution of the petal and stamen identity genes, *APETALA3* and *PISTILLATA*, after petal loss in the Piperales. *Molecular Phylogenetics and Evolution* 44: 598-609.

JIAO Y., N. J. WICKETT, S. AYYAMPALAYAM, A. S. CHANDERBALI, L. LANDHERR, P. E. RALPH, L. P. TOMSHO, Y. HU, H. LIANG, P. S. SOLTIS, D. E. SOLTIS, S. W. CLIFTON, S. E. SCHLARBAUM, S. C. SCHUSTER, H. MA, J. LEEBENS-MACK, AND C. W. DEPAMPHILIS. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97-113.

KAMPNY C. M. 1995. Pollination and flower diversity in Scrophulariaceae. *Botanical Review* 61: 352-366.

KATOH K. AND D. M. STANDLEY. 2013. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772-780.

LUO D., R. CARPENTER, L. COPSEY, C. VINCENT, J. CLARK, AND E. COEN. 1999. Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* 99: 367-376.

LUO D., R. CARPENTER, C. VINCENT, L. COPSEY, AND E. COEN. 1996. Origin of floral asymmetry in *Antirrhinum*. *Nature* 383: 794-799.

LYONS E. AND M. FREELING. 2008. How to usefully compare homologous plant genes and chromosomes as DNA sequences. *The Plant Journal* 53: 661-673.

LYSAK M., M. KOCH, A. PECINKA, AND I. SCHUBERT. 2005. Chromosome triplication found across the tribe Brassiceae. *Genome Research* 15: 516-525.

LYSAK M. A., K. CHEUNG, M. KITSCHKE, AND P. BURES. 2007. Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiology* 145: 402-410.

MING R., S. HOU, Y. FENG, Q. YU, A. DIONNE-LAPORTE, J. H. SAW, P. SENIN, W. WANG, B. V. LY, K. L. T. LEWIS, S. L. SALZBERG, L. FENG, M. R. JONES, R. L. SKELTON, J. E. MURRAY, C. CHEN, W. QIAN, J. SHEN, P. DU, M. EUSTICE, E. TONG, H. TANG, E. LYONS, R. E. PAULL, T. P. MICHAEL, K. WALL, D. W. RICE, H. ALBERT, M. WANG, Y. J. ZHU, M. SCHATZ, N. NAGARAJAN, R. A. ACOB, P. GUAN, A. BLAS, C. M. WAI, C. M. ACKERMAN, Y. REN, C. LIU, J. WANG, J. WANG, J. NA, E. V. SHAKIROV, B. HAAS, J. THIMMAPURAM, D. NELSON, X. WANG, J. E. BOWERS, A. R. GSCHWEND, A. L. DELCHER, R. SINGH, J. Y. SUZUKI, S. TRIPATHI, K. NEUPANE, H. WEI, B. IRIKURA, M. PAIDI, N. JIANG, W. ZHANG, G. PRESTING, A. WINDSOR, R. NAVAJAS-PEREZ, M. J. TORRES, F. A. FELTUS, B. PORTER, Y. LI, A. M. BURROUGHS, M. LUO, L. LIU, D. A. CHRISTOPHER, S. M. MOUNT, P. H. MOORE, T. SUGIMURA, J. JIANG, M. A. SCHULER, V. FRIEDMAN, T. MITCHELL-OLDS, D. E. SHIPPEN, C. W. DEPAMPHILIS, J. D. PALMER, M. FREELING, A. H. PATERSON, D. GONSALVES, L. WANG, AND M. ALAM. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452: 991-1007.

MOLLER M., M. CLOKIE, P. CUBAS, AND Q. C. B. CRONK. 1999. Integrating molecular phylogenies and developmental genetics: a Gesneriaceae case study. *Molecular Systematics and Plant Evolution* 57: 375-402.

MONDRAGON-PALOMINO M. AND G. THEISSEN. 2009. Why are orchid flowers so diverse? reduction of evolutionary constraints by paralogues of class B floral homeotic genes. *Annals of Botany* 104: 583-594.

MOORE R. AND M. PURUGGANAN. 2005. The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology* 8: 122-128.

O'KANE S. AND I. AL-SHEHBAZ. 2003. Phylogenetic position and generic limits of *Arabidopsis* (Brassicaceae) based on sequences of nuclear ribosomal DNA. *Annals of the Missouri Botanical Garden* 90: 603-612.

PATCHELL M. J. 2012. Connecting molecules to morphology: evolutionary developmental genetics of floral monosymmetry in *Cleome violacea* (Cleomaceae). MSc Thesis, University of Alberta, Edmonton.

PATCHELL M. J., M. C. BOLTON, P. MANKOWSKI, AND J. C. HALL. 2011. Comparative floral development in Cleomaceae reveals two distinct pathways leading to monosymmetry. *International Journal of Plant Sciences* 172: 352-365.

PATCHELL M. J., E. H. ROALSON, AND J. C. HALL. 2014. Resolved phylogeny of the Cleomaceae based on all three genomes. *Taxon* 63: 315-328.

POSADA D. AND K. CRANDALL. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

PRESTON J. C. AND L. C. HILEMAN. 2012. Parallel evolution of TCP and B-class genes in Commelinaceae flower bilateral symmetry. *Evodevo* 3: 6.

PRESTON J. C., M. A. KOST, AND L. C. HILEMAN. 2009. Conservation and diversification of the symmetry developmental program among close relatives of snapdragon with divergent floral morphologies. *New Phytologist* 182: 751-762.

PRESTON J. C., C. C. MARTINEZ, AND L. C. HILEMAN. 2011. Gradual disintegration of the floral symmetry gene network is implicated in the evolution of a wind-pollination syndrome. *Proceedings of the National Academy of Sciences of the United States of America* 108: 2343-2348.

RAMBAUT A. AND A.J. DRUMMOND. 2007. Tracer. Version 1.4.

RASTOGI P. A. 2000. MacVector. Integrated sequence analysis for the Macintosh. *Methods in Molecular Biology* 132: 47-69.

REEVES P. A. AND R. G. OLMSTEAD. 2003. Evolution of the TCP gene family in Asteridae: cladistic and network approaches to understanding regulatory gene family diversification and its impact on morphological evolution. *Molecular Biology and Evolution* 20: 1997-2009.

ROSIN F. M. AND E. M. KRAMER. 2009. Old dogs, new tricks: regulatory evolution in conserved genetic modules leads to novel morphologies in plants. *Developmental Biology* 332: 25-35.

SCHRANZ M. AND T. MITCHELL-OLDS. 2006. Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* 18: 1152-1165.

SMITH J. F., M. M. FUNKE, AND V. L. WOO. 2006. A duplication of *GCYC* predates divergence within tribe Coronanthereae (Gesneriaceae): phylogenetic analysis and evolution. *Plant Systematics and Evolution* 261: 245-256.

SMITH J. F., L. C. HILEMAN, M. P. POWELL, AND D. A. BAUM. 2004. Evolution of *GCYC*, a Gesneriaceae homolog of *CYCLOIDEA*, within Gesnerioideae (Gesneriaceae). *Molecular Phylogenetics and Evolution* 31: 765-779.

SONG C., Q. LIN, R. LIANG, AND Y. WANG. 2009. Expressions of ECE-CYC2 clade genes relating to abortion of both dorsal and ventral stamens in *Opithandra* (Gesneriaceae). *BMC Evolutionary Biology* 9: 244.

SWOFFORD D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0β10.

TSAI W. C., C. S. KUOH, M. H. CHUANG, W. H. CHEN, AND H. H. CHEN. 2004. Four *DEF*-Like MADS-box genes displayed distinct floral morphogenetic roles in *Phaldenopsis* orchid. *Plant and Cell Physiology* 45: 831-844.

TSAI W., Z. PAN, Y. HSIAO, M. JENG, T. WU, W. CHEN, AND H. CHEN. 2008. Interactions of B-class complex proteins involved in tepal development in *Phalaenopsis* orchid. *Plant and Cell Physiology* 49: 814-824.

USHIMARU A., I. DOHZONO, Y. TAKAMI, AND F. HYODO. 2009. Flower orientation enhances pollen transfer in bilaterally symmetrical flowers. *Oecologia* 160: 667-674.

VEKEMANS D., S. PROOST, K. VANNESTE, H. COENEN, T. VIAENE, P. RUELENS, S. MAERE, Y. VAN DE PEER, AND K. GEUTEN. 2012. Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-box gene and species diversification. *Molecular Biology and Evolution* 29: 3793-3806.

VIAENE T., D. VEKEMANS, V. F. IRISH, A. GEERAERTS, S. HUYSMANS, S. JANSSENS, E. SMETS, AND K. GEUTEN. 2009. *PISTILLATA*-duplications as a mode for floral diversification in (basal) Asterids. *Molecular Biology and Evolution* 26: 2627-2645.

WANG Z., Y. LUO, X. LI, L. WANG, S. XU, J. YANG, L. WENG, S. SATO, S. TABATA, M. AMBROSE, C. RAMEAU, X. FENG, X. HU, AND D. LUO. 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: 10414-10419.

YANG Z., R. NIELSEN, N. GOLDMAN, AND A. PEDERSEN. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431-449.

ZHANG W., E. M. KRAMER, AND C. 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: 6388-6393.

ZHANG W., Q. XIANG, D. T. THOMAS, B. M. WIEGMANN, M. W. FROHLICH, AND D. E. SOLTIS. 2008. Molecular evolution of *PISTILLATA*-like genes in the dogwood genus *Cornus* (Cornaceae). *Molecular Phylogenetics and Evolution* 47: 175-195.

Tables

Table 3-1. List of taxa used in the *TCP1* phylogeny.

Species	Symmetry of species*	Molecule type	Source of DNA sequence	Accession No.	Voucher and herbarium information
TCP1					
Out group					
Carica papaya L.	-	Genomic scaffold	GenBank	ABIM01022054	-
Brassicaceae					
Aethionema grandiflorum Boiss & Hohen	Dissymmetry	gDNA	This study	-	3775-79E, (HUH)
Arabidopsis lyrata (L.) O'Kane & Al-Shehbaz	Dissymmetry	cDNA	GenBank	XM002888537.1	-
Arabidopsis thaliana (L.) Heynh	Dissymmetry	cDNA	GenBank	NM001084312.1	-
Arabis lyrata L.	Dissymmetry	gDNA	This study	-	Not available
Brassica napus L.	Dissymmetry	gDNA	This study	-	Excel canola; Ex-19
Brassica rapa subsp. pekinensis (Lour.) Hanelt ex Mansf.	Dissymmetry	Genomic scaffold	GenBank	AC189200.2	-
Cakile lanceolata subsp. *fusiformis* (Greene) Rodman	Dissymmetry	gDNA	This study	-	GK1, (HUH)
Calepina irregularis (Asso) Thellung	Monosymmetry	cDNA	GenBank	JQ264770.1	-
Capsella rubella Reut.	Dissymmetry	cDNA	GenBank	XM006302401.1	-
Cardamine hirsuta L.	Dissymmetry	gDNA	GenBank	JQ290067.1	-
Caulanthus amplexicaulis var. amplexicaulis S. Wats.	Dissymmetry	gDNA	This study	-	Not available
Dontostemon integrifolius (L.) C.A.Mey.	Dissymmetry	gDNA	GenBank	JQ290068.1	-
Erucaria erucarioides (Coss. & Durieu) Müll. & Berol.	Dissymmetry	gDNA	This study	-	1944-71, (HUH)
Eutrema salsugineum (Pall.) Al-Shehbaz & Warwick	Dissymmetry	cDNA	GenBank	XM006391443.1	-
Iberis grandiflora L.	Monosymmetry	gDNA	This study	-	WIS
Iberis umbellata L.	Monosymmetry	cDNA	GenBank	JQ290074.1	-
*Lobularia maritima (*L.)	Dissymmetry	gDNA	GenBank	JQ290069.1	-

Desv.					
Myagrum perfoliatum L.	Dissymmetry	gDNA	This study	-	103-R29 (HUH)
Stanleya pinnata (Pursh) Britton	Dissymmetry	gDNA	This study	-	1735-69F (HUH)
Teesdalia coronopifolia (J.P.Bergeret) Thell.	Dissymmetry	gDNA	GenBank	JQ290070.1	-
Teesdalia nudicaulis (L.) W.T.Aiton	Monosymmetry	cDNA	GenBank	JQ290073.1	-
Thlaspi arvense L.	Dissymmetry	gDNA	This study	-	C9 (WIS)
Cleomaceae					
Cleome africana Botsch.	Early Monosymmetry	gDNA	This study		142768 (ALTA)
Cleome ambylocarpa Baratte & Murb.	Early Monosymmetry	gDNA	This study		151485 (ALTA)
Cleome arabica L.	Early Monosymmetry	gDNA	This study		810 (ALTA)
Cleome brachycarpa Vahl. ex DC.	Early Dissymmetry	gDNA	This study		85014 (ALTA)
Cleome droserifolia (Forssk.) Delile	Monosymmetry	gDNA	This study		6397 (WIS)
Cleome hirta (Klotzch) Oliv.	Monosymmetry	gDNA	This study		74520 (ALTA)
Cleome viridiflora Schreb.	Early Dissymmetry	gDNA	This study		8803 (WIS)
Cleome violacea L.	Early Monosymmetry	gDNA	This study		813 (ALTA)
Cleome violacea L.	Early Monosymmetry	Genomic scaffold	CoGe	Scaffold 169, bps 289897-291101 (copy 1a) and Scaffold 59, bps 352604-353930 (copy 1b)	-
Polanisia dodecandra DC.	Early Monosymmetry	gDNA	This study		s.n. (ALTA)
Tarenaya hassleriana (Chodat) Iltis	Monosymmetry	Genomic scaffold	CoGe	Chromosome Cleome11, bps 2113112-2114315 (copy 1a)[1] and Chromosome Cleome4, bps 4049225-4050431 (copy 1b)[2]	-
Tarenaya spinosa(L.) Iltis	Early Dissymmetry	gDNA	This study		91-11 (WIS)
TCP12					
Brassicaceae					
Arabidopsis thaliana	Dissymmetry	cDNA	GenBank	NM105554.1	-
Brassica rapa subsp.	Dissymmetry	Genomic	CoGe	Scaffold 144, bps	-

pekinensis (Lour.) Hanelt ex Mansf.		scaffold		304340-304684	
Capsella rubella Reut.	Dissymmetry	cDNA	GenBank	XM006302353	-
Eutrema salsugineum (Pall.) Al-Shehbaz & Warwick	Dissymmetry	cDNA	GenBank	XM006391033	-
Cleomaceae					
Cleome violacea L.	Early Monosymmetry	Genomic scaffold	CoGe	Scaffold 6, bps 1226109-1226571	-
***TCP17* (Out group)**					
Brassicaceae					
Arabidopsis thaliana	Dissymmetry	cDNA	GenBank	NM120889	
Cleomaceae					
Tarenaya hassleriana (Chodat) Iltis	Monosymmetry	Genomic scaffold	CoGe	Chromosome 21, bps 979496-979862	
TCP18					
Brassicaceae					
Arabidopsis thaliana	Dissymmetry	cDNA	GenBank	NM1125184	
Brassica rapa subsp. pekinensis (Lour.) Hanelt ex Mansf.	Dissymmetry	Genomic scaffold	CoGe	Scaffold 144, bps 304340-304684	-
Capsella rubella Reut.	Dissymmetry	cDNA	GenBank	XM006302353	-
Eutrema salsugineum (Pall.) Al-Shehbaz & Warwick	Dissymmetry	cDNA	GenBank	XM006391033	-
Cleomaceae					
Cleome violacea L.	Early Monosymmetry	Genomic scaffold	CoGe	Scaffold197, bps 120840-121271	-
Tarenaya hassleriana (Chodat) Iltis	Early Dissymmetry	Genomic scaffold	CoGe	Chromosome 16, Bps 1630402_1630867	

* Species are categorised as having either dissymmetry or monosymmetry except in taxa where the developmental pathway to monosymmetry was assessed by Patchell et al. (2011). These species are more specifically characterized as having early monosymmetry or early dissymmetry. CoGe refers to the Comparative Genomic database (Lyons and Freeling, 2008). [1] Gene called Th24587 in Cheng et al. (2013). [2] Gene called Th21666 in Cheng et al. (2013).

Table 3-2. Primers used for *TCP1* amplification and sequencing.

Purpose	Primer Name	Sequence (5'-3')
Amplification	TCP1-fwd 5'	ACAATCGAGTGTACCCTCTCTCTCTTTACC
	TCP1-rev 5'	TTATAGTTGCTGCTAGAACTCTGATCTACC
	TCP1 F degen.2	ACAATGGAGYGTACCCTYTCTCTYTTT
	TCP1 R degen.2	TTGTARTTGCTGCTASAACHHTGATC
Sequencing	M13F	GTAAAACGACGGCCAG
	M13R	CAGGAAACAGCTATGAC

Table 3-3. Summary of sequence similarity (%) of the TCP domain and length of the coding sequence among and between Group I and Group II TCP1 sequences. Coding sequence length refers to the total length of the amplified fragment and as such, excludes sequences that were not originally retrieved in this study.

	Within Group I variation	Within Group II variation	Between Group I and Group II variation
TCP domain sequence similarity	88.6% - 100%	79.8% - 97.7%	68.9% - 81.6%
Coding sequence length	1071bp - 1134bp	789bp – 888bp	258bp*

* value represents mean length of Group I sequences minus mean length of Group II sequences.

Table 3-4. Maximum likelihood tests of selection from PAML4.4 and estimated dN/dS ratios.

Model	-lnL	Partition	dN/dS
Fixed	5006.068	All branches	0.2722
Free	5111.036	-	-
Brassicaceae vs Cleomaceae	5006.048	Brassicaceae	0.2747
		Cleomaceae	0.2667
Monosymmetry vs dissymmetry	5004.823	Monosymmetric taxa	0.3090
		Disymmetric taxa	0.2376
Early monosymmetry vs Early Dissymmetry	5007.012	Brassicaceae	0.2747
		Early monosymmetry Cleomaceae	0.2633
		Early dissymmetry Cleomaceae	0.2698

Figures

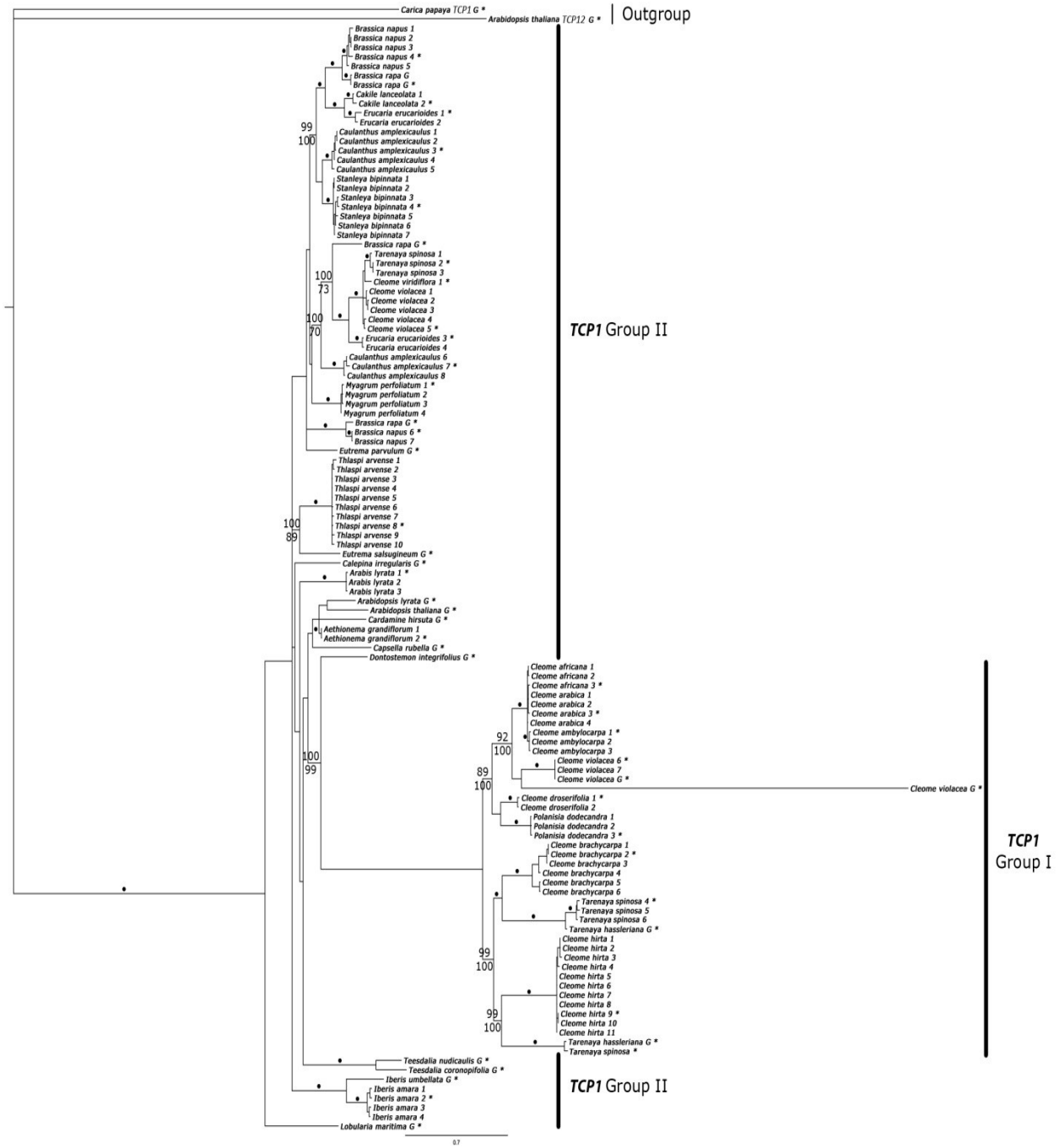


Figure 3-1. Bayesian consensus including branch lengths of *TCP1* sequences from the "all *TCP1* clones" dataset. Bayesian posterior probabilities >70% are listed below branches and maximum parsimony likelihood bootstrap values >70% are listed above branches. Branches with 100% support for both Bayesian and Maximum Parsimony analyses are indicated by •. "G" indicates sequences that were retrieved from genomic resources. Sequences that were selected for subsequent analyses are indicated by an asterisk.

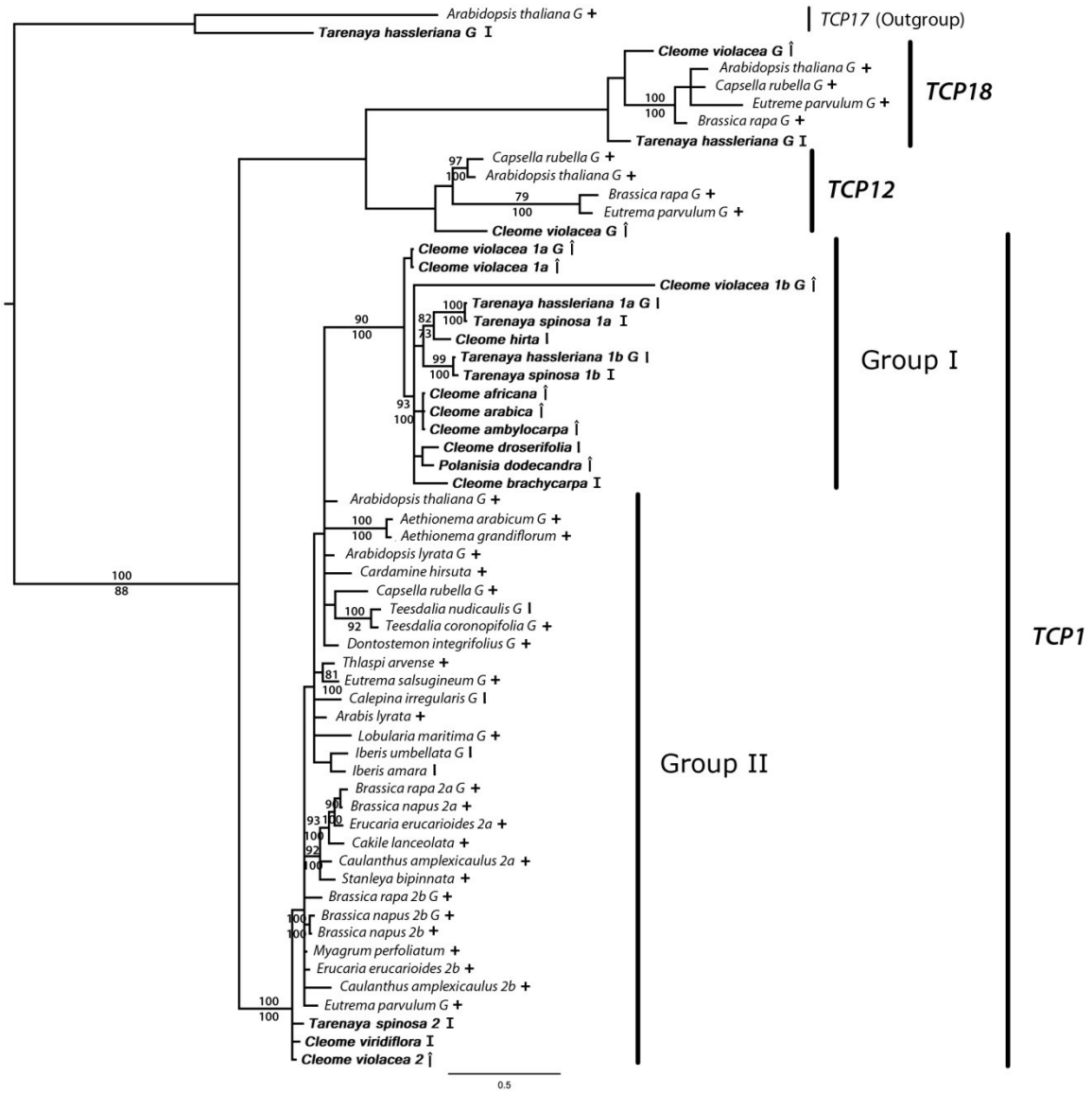


Figure 3-2. Bayesian consensus including branch lengths of *TCP* sequences using the "representative TCP domain" nucleotide dataset. Bayesian posterior probabilities >70% are listed below branches and maximum parsimony likelihood bootstrap values >70% are listed above branches. Cleomaceae taxa are bolded. Symbols beside taxon names indicate if the sequence was mined from a genomic resource "G" and the following symbols pertain to floral symmetry: "+" = dissymmetric, "|" = monosymmetric, "î" = early monosymmetry, and "I" = early dissymmetry.

Chapter 4 – Conclusion

The findings of this thesis implicate involvement of candidate genes in the evolution of novel morphologies. These data support the commonly used tactic in evolutionary development studies whereby researchers characterize genes in model plants and then hypothesize a similar role in related species while investigating new modifications that result in morphological change (Frohlich, 2006; Soltis et al., 2009b; de Bruijn et al., 2012). Many studies have demonstrated that gene families are repeatedly recruited in distantly related groups despite wildly different morphologies (Flagel and Wendel, 2009; Bemer et al., 2012). My thesis sought to answer whether the evolution of multiple *FUL* copies is associated with the diversity of fruit types found in the Brassiceae and whether the evolution of multiple *TCP1* copies appears correlated with the evolution of floral symmetry in the Brassicaceae and Cleomaceae. This chapter will assess how effective my chosen research approaches were in answering these questions while discussing alternative methods. Additionally, I will suggest future means of uncovering the genetics of fruit and flower diversification in these groups.

Effectiveness of research methods and alternative approaches

Rather than presenting simple resolutions to the questions, my research unveiled complex patterns regarding copy number retention and evolution with multiple exceptions to general trends. Data presented in this thesis demonstrate that some species retain multiple copies although they are not associated with a detected morphological novelty. My investigation of *FUL* in Brassiceae found that shifts between non-heteroarthrocarpy and heteroarthrocarpy are likely associated with number of copies; however, types of copies

retained are equally important. The difference between the monosymmetric Cleomaceae and the predominantly dissymmetric Brassicaceae was not found to be associated with copy number, but instead with the presence of a divergent *TCP1* copy. Analyses using molecular evolution models revealed that the history of these genes likely reflects their involvement in fruit and flower diversification, although interpretation was complicated by the independent evolution of copy lineages in the case of *FUL*, and by substantial sequence divergence affecting alignment in the case of *TCP1*. Retention of multiple copies in taxa lacking morphological novelty may be explained by the need for polyploids to balance the ratios of all duplicated genes in a given pathway, producing functionally redundant genes (Edger and Pires, 2009). Additionally, duplicate genes may be recruited for functions outside of fruit and flower development. This emphasizes the need for comparative gene expression and functional studies to elucidate whether *TCP1* or *FUL* copies have redundant, subfunctional or neofunctional roles (Tonsor et al., 2005; Stratmann and Hind, 2011).

Many evo-devo studies focus on functional and expressional data in a few taxa and forgo generating extensive gene phylogenies or detailing their copy number retrieval efforts (Tani et al., 2009; Busch et al., 2012; Preston and Hileman, 2012; Sun et al., 2014). An alternative, however, is to examine gene evolution across a wide range of taxa to provide broader picture on gene evolutionary patterns (Howarth and Donoghue, 2005; Zhang et al., 2010). Although expression and functional data provide evidence of causation as opposed to merely correlation, conclusions may be misled without possessing vital information regarding copy number and evolutionary history of the gene (Doyle, 1994). Given that angiosperms possess the highest incidence of polyploidy, the seemingly insufficient effort given to characterize copy number and gene history is counterintuitive,

particularly because polyploidization is thought to be responsible for new morphologies and speciation (Flagel and Wendel, 2009; Soltis et al., 2009a). One must be especially cognizant of this when dealing with known hexaploid events, as documented in focal taxa Brassiceae and Cleomaceae.

A major drawback in the degenerate-primer sequencing of multiple clones approach taken here is that copies may be missed, thus obscuring trends and possible associations. Another alternative to constructing a gene phylogeny while still accounting for copy number is quantitative PCR (qPCR), which can indicate the number of target loci in gDNA by measuring relative rates of amplification by PCR (De Preter et al., 2002; D'haene et al., 2010). Although qPCR is common in evo-devo studies for assessing patterns of gene expression, it has not been used to detect copy number with the purpose of finding a correlation with morphological change. However, utilization of qPCR to analyze plant physiological traits is starting to appear in the literature. For instance, the increased ability of corn plants to withstand high amounts of soil aluminum was shown to be positively correlated with copy number of *MULTIDRUG AND TOXIC COMPOUND EXTRUSION 1* (*MATE1*) gene using this method (Maron et al., 2013). Although this approach also relies on PCR technology, the smaller 80-150bp amplicon avoids some of the challenges of conventional PCR methods, including highly repetitive regions, variable length of different copies or secondary DNA structure impeding copy number retrieval (D'haene et al., 2010). Pending troubleshooting, the qPCR method has the potential to retrieve results faster than creating a gene phylogeny. Additionally, this would eliminate the need for extensive characterization of amplified fragments by enzyme digest or re-sequencing of clones, which is particularly attractive if the gene in question is too long to be amplified with a single set

of forward and reverse primers. However, construction of sufficiently degenerate primers to retrieve multiple copies is a concern in both methods, and identification of pseudogenes and trends in introns are not possible with qPCR.

I argue that although the qPCR method successfully retrieves copy number information, lack of information regarding the evolutionary history of the gene may affect future hypothesis testing. For example, chapter two of this thesis revealed an apparent retention of two copies in both heteroarthrocarpic and non-heteroarthrocarpic taxa. Acquiring these data via qPCR may have halted further hypothesis testing as it implies that copy number has no effect on fruit morphology. Conversely, assessment of the evolutionary history of these copies in a phylogenetic context revealed that non-heteroarthrocarpic taxa always retain the same two copies of *FUL,* whereas heteroarthrocarpic taxa have retained different copies. This allows us to make hypotheses about possible functional redundancy of the copies retained in non-heteroarthrocarpic taxa and focus investigations on *FUL* lineages that are retained in heteroarthrocarpic taxa. Importantly, these findings will guide taxon sampling for comparative gene expression and functional studies. Likewise, the largely dissymmetric Brassicaceae were sometimes found to have two copies of *TCP1*, similar to the entirely monosymmetric Cleomaceae. However, the ability to assess sequence similarity via our approach allowed for the identification of a clade of sequences unique to the Cleomaceae that were intriguingly divergent in terms of both base pair differences and sequence length. Hence, constructing a gene phylogeny and investigating molecular evolution through model selection tests was more beneficial than simply assessing copy number by qPCR.

Future experiments

Although *FUL* appears to be involved in the development of heteroarthrocarpy, investigation of other genes included in the development pathway may yield additional players. Aside from the upstream regulators of *FUL* in *Arabidopsis* discussed in chapter two of this thesis, recent work on the abscission zone (AZ) in tomato has identified candidates worthy of investigation that could be involved in the joint of heteroarthrocarpic Brassiceae (Nakano et al., 2012; Liu et al., 2014). Abscission refers to the process by which mature plants shed their leaves, flower organs and fruits (Dinneny et al., 2005; Hall et al., 2006; Nakano et al., 2012). Nakano et al. (2012) analyzed the AZ of the pedicel to study which genes control the release of fruits. They discovered that *MACROCALYX* (*MC*), which interestingly belongs to the *FUL*/*AP1* group, interacts with another MADS-box gene called *JOINTLESS* to form the AZ. *MC* is most similar to *AGAMOUS 79* (*AGL79*) in *Arabidopsis*, whose function is largely unknown although it appears slightly more expressed in the flowers and siliques than roots, shoots or leaves (Shan et al., 2007). *JOINTLESS* is most similar to *SHORT VEGETATIVE PHASE* (*SVP*) in *Arabidopsis* (Cohen et al., 2012), which Balanz et al. (2014) propose forms a heterodimer with *FUL* that is important for regulating meristem identity. However, similar anatomy between the joint and the apical portion of the valve margin in heteroarthrocarpic fruits supports prioritization of investigating the role of valve margin identity genes in development of the joint (Hall et al., 2006). Expression of valve margin identity genes *SHP1*/*2*, *IND* and *ALC* were absent in the joint of two heteroarthrocarpic species, but technological shortcomings of *in situ* expression experiments make these data difficult to confirm (Avino et al., 2012). It is notable that some similarities at the cellular level exist between the pedicel AZ in Brassiceae member

Raphanus raphanistrum and the joint, including lignified cells and a separation layer (Hall et al., 2006; Taghizadeh et al., 2009). This indicates that pedicel AZ genes discovered in tomato may be worth investigating if future expression and functional studies of *FUL* and its upstream regulators yield inconclusive results.

Although *TCP1* homologues have been investigated in many plant families, other members of the "ECE" clade have not been focused on despite their close relationship to *TCP1* inviting intrigue into whether these genes also influence floral development. An investigation into expression patterns of *TCP12* and *TCP18* homologues, the other two members of the "ECE" clade as they are identified in *Arabidopsis,* yielded interesting preliminary results in monosymmetric *Lonicera* (Caprifoliaceae) flowers (Howarth and Donoghue, 2006). The authors found that although *TCP18* was not detected in floral tissue, *TCP12* expression persisted into later stages of floral development, where it was more strongly expressed in the abaxial petals. Interestingly, *tcp12* mutants of Arabidopsis, which has dissymmetric flowers, showed no notable difference in phenotype when compared to the wild type (Finlayson, 2007). These findings suggest that interactions amongst "ECE" members of the TCP may be possible and investigation of expression and functional data may reveal interesting results in the Cleomaceae.

The Cleomaceae family and Brassiceae tribe offer a good system to study the genetic mechanisms behind reoccurring evolution of similar floral and fruit morphologies, respectively. The independent evolution of a similar phenotype due to the same reoccurring mutation in a development pathway is known as parallelism whereas a different mutation necessitating entirely new genetic interactions is called convergence (Yoon and Baum, 2004; Scotland, 2011). Distinguishing the difference is important

because convergent evolution implies that the environmental pressures that influence the evolution of a trait are strong enough to induce any mutation as long as it produces a necessary phenotype that is preferentially retained in the population. An excellent example of these dynamics is revealed by transformation experiments by Yoon and Baum (2004), where two-thirds of their study Brassicaceae species had changes at the *LEAFY* (*LFY*) locus causing rosette flowering while the third taxon exhibited changes upstream of *LFY* to attain the same inflorescence type.

Multiple independent heteroarthrocarpic origins, which are underlain by differential *FUL* copy number retention from an ancestral hexaploid event, grant us an excellent opportunity to study parallelism/convergence (Lysak et al., 2005; Lysak et al., 2007; Hall et al., 2011). The gene phylogenies presented here are valuable for directing studies of parallelism because they provide information on the putative copy number and evolutionary history of copies retained in each taxon. For instance, comparisons of expression and functional data between the three *FUL* copies in *Enarthrocarpus lyratus* and two copies in *Cakile lanceolata* could reveal whether all copies are necessary to create a longitudinally indehiscent fruit that abscises at the joint. Importantly, these taxa have evolved this fruit type independently. Another potential comparison for parallelism would be between *Hirschfeldia incana* and *Erucaria erucarioides*, two partially longitudinally indehiscent Brassiceae whose joints do not abscise. This thesis reveals that *E. erucarioides* possesses copies from *FULc* and *FULd* lineages while *H. incana* possesses copies from *FULa* and *FULc*. It would be interesting to determine whether these genes have similar expression domains despite their different evolutionary histories. For this parallelism experiment, fruit anatomy of *E*. *erucariodes* and *C*. *lanceolata* are well examined in Hall et

al. (2006) and expression data of *ALC*, *FUL*, *IND*, *SHP* and *RPL* homologues are analysed by Avino et al. (2012). These data provide an anatomical and genetic basis for comparisons to the *Cakile* lineage of the Brassiceae. Furthermore, neither *Moricandia arvensis* nor *Arabidopsis thaliana* possess heteroarthrocarpy, yet the *FUL* gene phylogeny shows that *M. arvensis* has two copies of *FUL* compared to the single *Arabidopsis* copy, which could help inform hypotheses of subfunctionalization or redundancy.

Although all Cleomaceae taxa are monosymmetric, a parallelism/convergence study using data from dissymmetric and monosymmetric Brassicaceae provides a foundation for comparing this derived trait in the Cleomaceae (Cubas et al., 1999; Cubas et al., 2001; Busch and Zachgo, 2007; Busch et al., 2012). Within the Cleomaceae, the presence of two developmental trajectories in different lineages demonstrates two different ways to acquire monosymmetric flowers (Patchell et al., 2011). These trajectories are likely activated by differences in the genetic floral symmetry development pathway, indicating convergent evolution. The *TCP1* phylogeny presented in this thesis reveals that *Cleome violacea* and *Tarenaya spinosa* have at least three copies although these taxa exhibit early monosymmetry and early dissymmetry development pathways, respectively. This makes them ideal focal taxa to compare to the dissymmetric *Arabidopsis* (Cubas et al., 2001), and monosymmetric *Iberis* (Busch and Zachgo, 2007). Characterizing the role of *TCP1* copies in these Cleomaceae taxa, particularly copies within the comparatively divergent Group I lineage, may reveal whether similar genetic mechanisms are used to create monosymmetry, and which modifications are necessary to create different monosymmetry pathways and floral complexity in the Cleomaceae.

The families investigated in this thesis display remarkable variation in reproductive structures that could affect both pollen and seed dispersal, making them ideal study groups to uncover the origins of morphological variation. Lack of phylogenetic resolution in the Brassicaceae and the short backbone of the Cleomaceae tree are evidence of rapid radiation of these groups (Couvreur et al., 2010; Patchell et al., 2011). Rapid diversification implies rapid evolution of adventitious morphology, which is a key area of study for elucidating the success of angiosperms. This thesis represents the buildings blocks for others to further explore how angiosperms tweak their genetic blueprints to manifest diversity.

References

AVINO M., E. M. KRAMER, K. DONOHUE, A. J. HAMMEL, AND J. C. HALL. 2012. Understanding the basis of a novel fruit type in Brassicaceae: conservation and deviation in expression patterns of six genes. *Evodevo* 3: 20.

BALANZ V., I. MARTNEZ-FERNNDEZ, AND C. FERRNDIZ. 2014. Sequential action of *FRUITFULL* as a modulator of the activity of the floral regulators *SVP* and *SOC1*. *Journal of Experimental Botany* 65: 1193-1203.

BEMER M., R. KARLOVA, A. R. BALLESTER, Y. M. TIKUNOV, A. G. BOVY, M. WOLTERS-ARTS, P. D. B. ROSSETTO, G. C. ANGENENT, AND R. A. DE MAAGD. 2012. The tomato *FRUITFULL* Homologs *TDR4/FUL1* and *MBP7/FUL2* regulate ethylene-independent aspects of fruit ripening. *Plant Cell* 24: 4437-4451.

BUSCH A., S. HORN, A. MUEHLHAUSEN, K. MUMMENHOFF, AND S. ZACHGO. 2012. Corolla monosymmetry: evolution of a morphological novelty in the Brassicaceae family. *Molecular Biology and Evolution* 29: 1241-1254.

BUSCH A. AND S. ZACHGO. 2007. Control of corolla monosymmetry in the Brassicaceae *Iberis amara*. *Proceedings of the National Academy of Sciences of the United States of America* 104: 16714-16719.

COHEN O., Y. BOROVSKY, R. DAVID-SCHWARTZ, AND I. PARAN. 2012. *CaJOINTLESS* is a MADS-box gene involved in suppression of vegetative growth in all shoot meristems in pepper. *Journal of Experimental Botany* 63: 4947-4957.

COUVREUR T. L. P., A. FRANZKE, I. A. AL-SHEHBAZ, F. T. BAKKER, M. A. KOCH, AND K. MUMMENHOFF. 2010. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Molecular Biology and Evolution* 27: 55-71.

CUBAS P., E. COEN, AND J. M. M. ZAPATER. 2001. Ancient asymmetries in the evolution of flowers. *Current Biology* 11: 1050-1052.

CUBAS P., C. VINCENT, AND E. COEN. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401: 157-161.

DE BRUIJN S., G. C. ANGENENT, AND K. KAUFMANN. 2012. Plant 'evo-devo' goes genomic: from candidate genes to regulatory networks. *Trends in Plant Science* 17: 441-447.

DE PRETER K., F. SPELEMAN, V. COMBARET, J. LUNEC, G. LAUREYS, B. H. J. EUSSEN, N. FRANCOTTE, J. BOARD, A. D. J. PEARSON, A. DE PAEPE, N. VAN ROY, AND J. VANDESOMPELE. 2002. Quantification of *MYCN*, *DDX1*, and *NAG* gene copy number in

neuroblastoma using a real-time quantitative PCR assay. *Modern Pathology* 15: 159-166.

D'HAENE B., J. VANDESOMPELE, AND J. HELLEMANS. 2010. Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* 50: 262-270.

DINNENY J., D. WEIGEL, AND M. YANOFSKY. 2005. A genetic framework for fruit patterning in *Arabidopsis thaliana*. *Development* 132: 4687-4696.

DOYLE J. J. 1994. Evolution of a plant homeotic multigene family: toward connecting molecular systematics and molecular developmental genetics. *Systematic Biology* 43: 307-328.

EDGER P. P. AND J. C. PIRES. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research* 17: 699-717.

FINLAYSON S. A. 2007. *Arabidopsis TEOSINTE BRANCHED1-LIKE 1* regulates axillary bud outgrowth and is homologous to monocot *TEOSINTE BRANCHED1*. *Plant and Cell Physiology* 48: 667-677.

FLAGEL L. E. AND J. F. WENDEL. 2009. Gene duplication and evolutionary novelty in plants. *New Phytologist* 183: 557-564.

FROHLICH M. W. 2006. Recommendations and goals for evo-devo research: scenarios, genetic constraint, and developmental homeostasis. *Aliso* 22: 172-187.

HALL J. C., T. E. TISDALE, K. DONOHUE, AND E. M. KRAMER. 2006. Developmental basis of an anatomical novelty: heteroarthrocarpy in *Cakile lanceolata* and *Erucaria erucarioides* (Brassicaceae). *International Journal of Plant Sciences* 167: 771-789.

HALL J. C., T. E. TISDALE, K. DONOHUE, A. WHEELER, M. A. AL-YAHYA, AND E. M. KRAMER. 2011. Convergent evolution of a complex fruit structure in the tribe Brassiceae (Brassicaceae). *American Journal of Botany* 98: 1989-2003.

HOWARTH D. G. AND M. J. DONOGHUE. 2005. Duplications in *CYC*-like genes from Dipsacales correlate with floral form. *International Journal of Plant Sciences* 166: 357-370.

HOWARTH D. G. AND M. J. DONOGHUE. 2006. Phylogenetic analysis of the "'ECE" (CYC/TB1) clade reveals duplications predating the core eudicots. *Proceedings of the National Academy of Sciences of the United States of America* 103: 9101-9106.

LIU D., D. WANG, Z. QIN, D. ZHANG, L. YIN, L. WU, J. COLASANTI, A. LI, AND L. MAO. 2014. The *SEPALLATA* MADS-box protein SLMBP21 forms protein complexes with *JOINTLESS* and *MACROCALYX* as a transcription activator for development of the tomato flower abscission zone. *The Plant Journal for Cell and Molecular Biology* 77: 284-96.

LYSAK M., M. KOCH, A. PECINKA, AND I. SCHUBERT. 2005. Chromosome triplication found across the tribe Brassiceae. *Genome Research* 15: 516-525.

LYSAK M. A., K. CHEUNG, M. KITSCHKE, AND P. BURES. 2007. Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiology* 145: 402-410.

MARON L. G., C. T. GUIMARAES, M. KIRST, P. S. ALBERT, J. A. BIRCHLER, P. J. BRADBURY, E. S. BUCKLER, A. E. COLUCCIO, T. V. DANILOVA, D. KUDRNA, J. V. MAGALHAES, M. A. PINEROS, M. C. SCHATZ, R. A. WING, AND L. V. KOCHIAN. 2013. Aluminum tolerance in maize is associated with higher *MATE1* gene copy number. *Proceedings of the National Academy of Sciences of the United States of America* 110: 5241-5246.

NAKANO T., J. KIMBARA, M. FUJISAWA, M. KITAGAWA, N. IHASHI, H. MAEDA, T. KASUMI, AND Y. ITO. 2012. *MACROCALYX* and *JOINTLESS* interact in the transcriptional regulation of tomato fruit abscission zone development. *Plant Physiology* 158: 439-450.

PATCHELL M. J., M. C. BOLTON, P. MANKOWSKI, AND J. C. HALL. 2011. Comparative floral development in Cleomaceae reveals two distinct pathways leading to monosymmetry. *International Journal of Plant Sciences* 172: 352-365.

PRESTON J. C. AND L. C. HILEMAN. 2012. Parallel evolution of TCP and B-class genes in Commelinaceae flower bilateral symmetry. *Evodevo* 3: 6.

SCOTLAND R. W. 2011. What is parallelism? *Evolution & development* 13: 214-227.

SHAN H., N. ZHAN, C. LIU, G. XU, J. ZHANG, Z. CHEN, AND H. KONG. 2007. Patterns of gene duplication and functional diversification during the evolution of the AP1/SQUA subfamily of plant MADS-box genes. *Molecular Phylogenetics and Evolution* 44: 26-41.

SOLTIS D. E., V. A. ALBERT, J. LEEBENS-MACK, C. D. BELL, A. H. PATERSON, C. ZHENG, D. SANKOFF, C. W. DEPAMPHILIS, P. K. WALL, AND P. S. SOLTIS. 2009a. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336-348.

SOLTIS P. S., S. F. BROCKINGTON, M. YOO, A. PIEDRAHITA, M. LATVIS, M. J. MOORE, A. S. CHANDERBALI, AND D. E. SOLTIS. 2009b. Floral variation and floral genetics in basal angiosperms. *American Journal of Botany* 96: 110-128.

STRATMANN J. W. AND S. R. HIND. 2011. Gene silencing goes viral and uncovers the private life of plants. *Entomologia Experimentalis et Applicata* 140: 91-102.

SUN W., W. HUANG, Z. LI, C. SONG, D. LIU, Y. LIU, A. HAYWARD, Y. LIU, H. HUANG, AND Y. WANG. 2014. Functional and evolutionary analysis of the *AP1/SEP/AGL6* superclade of MADS-box genes in the basal eudicot *Epimedium sagittatum*. *Annals of Botany* 113: 653-668.

TAGHIZADEH M. S., S. CRAWFORD, M. E. NICOLAS, AND R. D. COUSENS. 2009. Water deficit changes the anatomy of the fruit abscission zone in *Raphanus raphanistrum* (Brassicaceae). *Australian Journal of Botany* 57: 708-714.

TANI E., A. N. POLIDOROS, E. FLEMETAKIS, C. STEDEL, C. KALLONIATI, K. DEMETRIOU, P. KATINAKIS, AND A. S. TSAFTARIS. 2009. Characterization and expression analysis of *AGAMOUS*-like, *SEEDSTICK*-like, and *SEPALLATA*-like MADS-box genes in peach (*Prunus persica*) fruit. *Plant Physiology and Biochemistry* 47: 690-700.

TONSOR S. J., C. ALONSO-BLANCO, AND M. KOORNNEEF. 2005. Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. *Plant Cell and Environment* 28: 2-20.

YOON H. S. AND D. A. BAUM. 2004. Transgenic study of parallelism in plant morphological evolution. *Proceedings of the National Academy of Sciences of the United States of America* 101: 6524-6529.

ZHANG W., E. M. KRAMER, AND C. 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: 6388-6393.

References for all chapters

AIROLDI C. A. AND B. DAVIES. 2012. Gene duplication and the evolution of plant MADS-box transcription factors. *Journal of Genetics and Genomics* 39: 157-165.

ALMEIDA J., M. ROCHETA, AND L. GALEGO. 1997. Genetic control of flower shape in *Antirrhinum majus*. *Development* 124: 1387-1392.

AL-SHEHBAZ I. A., M. A. BEILSTEIN, AND E. A. KELLOGG. 2006. Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Systematics and Evolution* 259: 89-120.

AL-SHEHBAZ I. A. 2012. A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon* 61: 931-954.

ALVAREZ-BUYLLA E. R., B. GARCIA-PONCE, AND A. GARAY-ARROYO. 2006. Unique and redundant functional domains of *APETALA1* and *CAULIFLOWER*, two recently duplicated *Arabidopsis thaliana* floral MADS-box genes. *Journal of Experimental Botany* 57: 3099-3107.

APPEL O. 1999. The so-called 'beak', a character in the systematics of Brassicaceae? *Botanische Jahrbücher für Systematik Pflanzengeschichte und Pflanzengeographie* 121: 85-98.

APPEL O. AND I. A. AL-SHEHBAZ. 2003. Cruciferae. *In* Kubitzki K. [ed.], *Families and Genera of Vascular Plants.* Springer-Verlag, Heidelberg.

ARIAS T. AND J. C. PIRES. 2012. A fully resolved chloroplast phylogeny of the *Brassica* crops and wild relatives (Brassicaceae: Brassiceae): Novel clades and potential taxonomic implications. *Taxon* 61: 980-988.

ARVESTAD L., J. LAGERGREN, AND B. SENNBLAD. 2009. The gene evolution model and computing its associated probabilities. *Journal of the ACM* 56: 7.

AVINO M., E. M. KRAMER, K. DONOHUE, A. J. HAMMEL, AND J. C. HALL. 2012. Understanding the basis of a novel fruit type in Brassicaceae: conservation and deviation in expression patterns of six genes. *Evodevo* 3: 20.

BAILEY C. D., M. A. KOCH, M. MAYER, K. MUMMENHOFF, S. L. O'KANE Jr., S. I. WARWICK, M. D. WINDHAM, AND I. A. AL-SHEHBAZ. 2006. Toward a global phylogeny of the Brassicaceae. *Molecular Biology and Evolution* 23: 2142-2160.

BALANZ V., I. MARTNEZ-FERNNDEZ, AND C. FERRNDIZ. 2014. Sequential action of *FRUITFULL* as a modulator of the activity of the floral regulators *SVP* and *SOC1*. *Journal of Experimental Botany* 65: 1193-1203.

BARKER M. S., H. VOGEL, AND M. E. SCHRANZ. 2009. Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biology and Evolution* 1: 391-399.

BARTLETT M. E. AND C. D. SPECHT. 2010. Evidence for the involvement of *GLOBOSA*-like gene duplications and expression divergence in the evolution of floral morphology in the Zingiberales. *New Phytologist* 187: 521-541.

BAUM D., H. YOON, AND R. OLDHAM. 2005. Molecular evolution of the transcription factor *LEAFY* in Brassicaceae. *Molecular Phylogenetics and Evolution* 37: 1-14.

BECK M. J. AND S. B. VANDER WALL. 2010. Seed dispersal by scatter-hoarding rodents in arid environments. *Journal of Ecology* 98: 1300-1309.

BEILSTEIN M., I. AL-SHEHBAZ, AND E. KELLOGG. 2006. Brassicaceae phylogeny and trichome evolution. *American Journal of Botany* 93: 607-619.

BEMER M., R. KARLOVA, A. R. BALLESTER, Y. M. TIKUNOV, A. G. BOVY, M. WOLTERS-ARTS, P. D. B. ROSSETTO, G. C. ANGENENT, AND R. A. DE MAAGD. 2012. The tomato *FRUITFULL* homologs *TDR4/FUL1* and *MBP7/FUL2* regulate ethylene-independent aspects of fruit ripening. *Plant Cell* 24: 4437-4451.

BHIDE A., S. SCHLIESKY, M. REICH, A. P. M. WEBER, AND A. BECKER. 2014. Analysis of the floral transcriptome of *Tarenaya hassleriana* (Cleomaceae), a member of the sister group to the Brassicaceae: towards understanding the base of morphological diversity in Brassicales. *BMC Genomics* 15: 140.

BLANC G., K. HOKAMP, AND K. H. WOLFE. 2003. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Research* 13: 137-144.

BOLMGREN K. AND O. ERIKSSON. 2005. Fleshy fruits - origins, niche shifts, and diversification. *Oikos* 109: 255-272.

BOWERS J. E., B. A. CHAPMAN, J. K. RONG, AND A. H. PATERSON. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422: 433-438.

BRAEUTIGAM A., K. KAJALA, J. WULLENWEBER, M. SOMMER, D. GAGNEUL, K. L. WEBER, K. M. CARR, U. GOWIK, J. MASS, M. J. LERCHER, P. WESTHOFF, J. M. HIBBERD, AND A. P. M. WEBER. 2011. An mRNA blueprint for C-4 photosynthesis derived from comparative transcriptomics of closely related C-3 and C-4 species. *Plant Physiology* 155: 142-156.

BREMER B., K. BREMER, M. W. CHASE, M. F. FAY, J. L. REVEAL, D. E. SOLTIS, P. S. SOLTIS, P. F. STEVENS, A. A. ANDERBERG, M. J. MOORE, R. G. OLMSTEAD, P. J. RUDALL, K. J. SYTSMA, D. C. TANK, K. WURDACK, J. Q. -. XIANG, S. ZMARZTY, AND ANGIOSPERM PHYLOGENY GRP. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105-121.

BROHOLM S. K., S. TAHTIHARJU, R. A. E. LAITINEN, V. A. ALBERT, T. H. TEERI, AND P. ELOMAA. 2008. A TCP domain transcription factor controls flower type specification along the radial axis of the *Gerbera* (Asteraceae) inflorescence. *Proceedings of the National Academy of Sciences of the United States of America* 105: 9117-9122.

BUSCH A., S. HORN, A. MUEHLHAUSEN, K. MUMMENHOFF, AND S. ZACHGO. 2012. Corolla monosymmetry: evolution of a morphological novelty in the Brassicaceae family. *Molecular Biology and Evolution* 29: 1241-1254.

BUSCH A. AND S. ZACHGO. 2007. Control of corolla monosymmetry in the Brassicaceae *Iberis amara*. *Proceedings of the National Academy of Sciences of the United States of America* 104: 16714-16719.

CARLSON S. E., D. G. HOWARTH, AND M. J. DONOGHUE. 2011. Diversification of *CYCLOIDEA*-like genes in Dipsacaceae (Dipsacales): implications for the evolution of capitulum inflorescences. *BMC Evolutionary Biology* 11: 325.

CASTILLO-DAVIS C., S. MEKHEDOV, D. HARTL, E. KOONIN, AND F. KONDRASHOV. 2002. Selection for short introns in highly expressed genes. *Nature Genetics* 31: 415-418.

CHAPMAN M. A., J. H. LEEBENS-MACK, AND J. M. BURKE. 2008. Positive selection and expression divergence following gene duplication in the sunflower CYCLOIDEA gene family. *Molecular Biology and Evolution* 25: 1260-1273.

CHAPMAN M. A., S. TANG, D. DRAEGER, S. NAMBEESAN, H. SHAFFER, J. G. BARB, S. J. KNAPP, AND J. M. BURKE. 2012. Genetic analysis of floral symmetry in Van Gogh's sunflowers reveals independent recruitment of *CYCLOIDEA* genes in the Asteraceae. *PLoS Genetics* 8: e1002628.

CHARLESWORTH D., C. BARTOLOME, M. SCHIERUP, AND B. MABLE. 2003. Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. *Molecular Biology and Evolution* 20: 1741-1753.

CHENG S., E. VAN DEN BERGH, P. ZENG, X. ZHONG, J. XU, X. LIU, J. HOFBERGER, S. DE BRUIJN, A. S. BHIDE, C. KUELAHOGLU, C. BIAN, J. CHEN, G. FAN, K. KAUFMANN, J. C. HALL, A. BECKER, A. BRAEUTIGAM, A. P. M. WEBER, C. SHI, Z. ZHENG, W. LI, M. LV, Y. TAO, J. WANG, H. ZOU, Z. QUAN, J. M. HIBBERD, G. ZHANG, X. ZHU, X. XU, AND M. E. SCHRANZ. 2013. The Tarenaya hassleriana Genome provides insight into reproductive trait and genome evolution of crucifers. *Plant Cell* 25: 2813-2830.

CHU T., H. XIE, Y. XU, AND R. MA. 2010. Regulation pattern of the *FRUITFULL* (*FUL*) gene of *Arabidopsis thaliana*. *Sheng wu gong cheng xue bao = Chinese Journal of Biotechnology* 26: 1546-54.

CHUNG M., J. VREBALOV, R. ALBA, J. LEE, R. MCQUINN, J. CHUNG, P. KLEIN, AND J. GIOVANNONI. 2010. A tomato (*Solanum lycopersicum*) *APETALA2/ERF* gene, *SlAP2a*, is a negative regulator of fruit ripening. *Plant Journal* 64: 936-947.

CITERNE H. L., M. MOLLER, AND Q. C. B. CRONK. 2000. Diversity of *CYCLOIDEA*-like genes in Gesneriaceae in relation to floral symmetry. *Annals of Botany* 86: 167-176.

CITERNE H. L., R. T. PENNINGTON, AND Q. C. B. CRONK. 2006. An apparent reversal in floral symmetry in the legume *Cadia* is a homeotic transformation. *Proceedings of the National Academy of Sciences of the United States of America* 103: 12017-12020.

CITERNE H., D. LUO, R. PENNINGTON, E. COEN, AND Q. CRONK. 2003. A phylogenomic investigation of *CYCLOIDEA*-like TCP genes in the Leguminosae. *Plant Physiology* 131: 1042-1053.

COATE J. E. AND J. J. DOYLE. 2011. Divergent evolutionary fates of major photosynthetic gene networks following gene and whole genome duplications. *Plant Signaling and Behavior* 6: 594-597.

COHEN O., Y. BOROVSKY, R. DAVID-SCHWARTZ, AND I. PARAN. 2012. *CaJOINTLESS* is a MADS-box gene involved in suppression of vegetative growth in all shoot meristems in pepper. *Journal of Experimental Botany* 63: 4947-4957.

COLLINS T. M., O. FEDRIGO, AND G. J. P. NAYLOR. 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. *Systematic Biology* 54: 493-500.

CORLEY S. B., R. CARPENTER, L. COPSEY, AND E. COEN. 2005. Floral asymmetry involves an interplay between TO and MYB transcription factors in *Antirrhinum*. *Proceedings of the National Academy of Sciences of the United States of America* 102: 5068-5073.

COULOMBE-HUNTINGTON J. AND J. MAJEWSKI. 2007. Intron loss and gain in *Drosophila*. *Molecular Biology and Evolution* 24: 2842-2850.

COUVREUR T. L. P., A. FRANZKE, I. A. AL-SHEHBAZ, F. T. BAKKER, M. A. KOCH, AND K. MUMMENHOFF. 2010. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Molecular Biology and Evolution* 27: 55-71.

CRANE P., E. FRIIS, AND K. PEDERSEN. 1995. The origin and early diversification of angiosperms. *Nature* 374: 27-33.

CREPET W. L. AND K. J. NIKLAS. 2009. Darwin's second "abominable mystery": why are there so many angiosperm species? *American Journal of Botany* 96: 366-381.

CRONK, Q.C.B., BATEMAN, R.M. AND HAWKINS, J.A. 2002. *Developmental Genetics and Plant Evolution.* Taylor and Francis, London.

CUBAS P. 2004. Floral zygomorphy, the recurring evolution of a successful trait. *Bioessays* 26: 1175-1184.

CUBAS P., E. COEN, AND J. M. M. ZAPATER. 2001. Ancient asymmetries in the evolution of flowers. *Current Biology* 11: 1050-1052.

CUBAS P., N. LAUTER, J. DOEBLEY, AND E. COEN. 1999a. The TCP domain: a motif found in proteins regulating plant growth and development. *Plant Journal* 18: 215-222.

CUBAS P., C. VINCENT, AND E. COEN. 1999b. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401: 157-161.

DAMERVAL C. AND M. MANUEL. 2003. Independent evolution of *CYCLOIDEA*-like sequences in several angiosperm taxa. *Comptes Rendus Palevol* 2: 241-250.

DAMERVAL C., H. CITERNE, M. LE GUILLOUX, S. DOMENICHINI, J. DUTHEIL, L. R. DE CRAENE, AND S. NADOT. 2013. Asymmetric morphogenetic cues along the transverse plane: shift from disymmetry to zygomorphy in the flower of Fumarioideae. *American Journal of Botany* 100: 391-402.

DAMERVAL C., M. LE GUILLOUX, M. JAGER, AND C. CHARON. 2007. Diversity and evolution of *CYCLOIDEA*-like TCP genes in relation to flower development in Papaveraceae. *Plant Physiology* 143: 759-772.

DAVIES T. J., T. G. BARRACLOUGH, M. W. CHASE, P. S. SOLTIS, D. E. SOLTIS, AND V. SAVOLAINEN. 2004. Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* 101: 1904-1909.

DE BRUIJN S., G. C. ANGENENT, AND K. KAUFMANN. 2012. Plant 'evo-devo' goes genomic: from candidate genes to regulatory networks. *Trends in Plant Science* 17: 441-447.

DE PRETER K., F. SPELEMAN, V. COMBARET, J. LUNEC, G. LAUREYS, B. H. J. EUSSEN, N. FRANCOTTE, J. BOARD, A. D. J. PEARSON, A. DE PAEPE, N. VAN ROY, AND J. VANDESOMPELE. 2002. Quantification of *MYCN*, *DDX1*, and *NAG* gene copy number in neuroblastoma using a real-time quantitative PCR assay. *Modern Pathology* 15: 159-166.

DEYHOLOS M. K. AND L. E. SIEBURTH. 2000. Separable whorl-specific expression and negative regulation by enhancer elements within the *AGAMOUS* second intron. *Plant Cell* 12: 1799-1810.

D'HAENE B., J. VANDESOMPELE, AND J. HELLEMANS. 2010. Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* 50: 262-270.

DINNENY J., D. WEIGEL, AND M. YANOFSKY. 2005. A genetic framework for fruit patterning in *Arabidopsis thaliana*. *Development* 132: 4687-4696.

DOEBLEY J., A. STEC, AND L. HUBBARD. 1997. The evolution of apical dominance in maize. *Nature* 386: 485-488.

DOYLE J. J. 1994. Evolution of a plant homeotic multigene family: toward connecting molecular systematics and molecular developmental genetics. *Systematic Biology* 43: 307-328.

DUNN C. W., X. LUO, AND Z. WU. 2013. Phylogenetic analysis of gene expression. *Integrative and Comparative Biology* 53: 847-856.

DUNN, O. J. 1961. Multiple Comparisons Among Means. *Journal of the American Statistical Association* 56: 52–64.

EDGER P. P. AND J. C. PIRES. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research* 17: 699-717.

ENDRESS P. K. 1992. Evolution and floral diversity - the phylogenetic surroundings of *Arabidopsis* and *Antirrhinum*. *International Journal of Plant Sciences* 153: S106-S122.

ENDRESS P. K. 2001. Evolution of floral symmetry. *Current Opinion in Plant Biology* 4: 86-91.

ERIKSSON O., E. M. FRIIS, AND P. LOFGREN. 2000. Seed size, fruit size, and dispersal systems in angiosperms from the early cretaceous to the late tertiary. *American Naturalist* 156: 47-58.

FARRELL B. D. 1998. "Inordinate fondness" explained: why are there so many beetles? *Science* 281: 555-559.

FELSENSTEIN, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1-15.

FENG X. Z., Z. ZHAO, Z. X. TIAN, S. L. XU, Y. H. LUO, Z. G. CAI, Y. M. WANG, J. YANG, Z. WANG, L. WENG, J. H. CHEN, L. Y. ZHENG, X. Z. GUO, J. H. LUO, S. S. SATO, S. TABATA, W. MA, X. L. CAO, X. H. HU, C. R. SUN, AND D. LUO. 2006. Control of petal shape and floral

zygomorphy in *Lotus japonicus*. *Proceedings of the National Academy of Sciences of the United States of America* 103: 4970-4975.

FENSTER C. B., W. S. ARMBRUSTER, AND M. R. DUDASH. 2009. Specialization of flowers: is floral orientation an overlooked first step? *New Phytologist* 183: 502-506.

FEODOROVA T. A., E. V. VOZNESENSKAYA, G. E. EDWARDS, AND E. H. ROALSON. 2010. Biogeographic patterns of diversification and the origins of C-4 in *Cleome* (Cleomaceae). *Systematic Botany* 35: 811-826.

FERRANDIZ C., Q. GU, R. MARTIENSSEN, AND M. F. YANOFSKY. 2000a. Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. *Development* 127: 725-734.

FERRANDIZ C., S. LILJEGREN, AND M. YANOFSKY. 2000b. Negative regulation of the *SHATTERPROOF* genes by *FRUITFULL* during *Arabidopsis* fruit development. *Science* 289: 436-438.

FINLAYSON S. A. 2007. *Arabidopsis TEOSINTE BRANCHED1-LIKE 1* regulates axillary bud outgrowth and is homologous to monocot *TEOSINTE BRANCHED1*. *Plant and Cell Physiology* 48: 667-677.

FLAGEL L. E. AND J. F. WENDEL. 2009. Gene duplication and evolutionary novelty in plants. *New Phytologist* 183: 557-564.

FOURQUIN C., C. DEL CERRO, F. C. VICTORIA, A. VIALETTE-GUIRAUD, A. C. DE OLIVEIRA, AND C. FERRANDIZ. 2013. A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in *Medicago* genus. *Plant Physiology* 162: 907-917.

FRANZKE A., M. A. LYSAK, I. A. AL-SHEHBAZ, M. A. KOCH, AND K. MUMMENHOFF. 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends in Plant Science* 16: 108-116.

FROHLICH M. W. 2006. Recommendations and goals for evo-devo research: scenarios, genetic constraint, and developmental homeostasis. *Aliso* 22: 172-187.

GALEGO L. AND J. ALMEIDA. 2002. Role of *DIVARICATA* in the control of dorsoventral asymmetry in *Antirrhinum* flowers. *Genes & Development* 16: 880-891.

GIRIN T., P. STEPHENSON, C. M. P. GOLDSACK, S. A. KEMPIN, A. PEREZ, N. PIRES, P. A. SPARROW, T. A. WOOD, M. F. YANOFSKY, AND L. OSTERGAARD. 2010. Brassicaceae *INDEHISCENT* genes specify valve margin cell fate and repress replum formation. *Plant Journal* 63: 329-338.

GIURFA M., A. DAFNI, AND P. R. NEAL. 1999. Floral symmetry and its role in plant-pollinator systems. *International Journal of Plant Sciences* 160: S41-S50.

GU Q., C. FERRANDIZ, M. F. YANOFSKY, AND R. MARTIENSSEN. 1998. The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125: 1509-1517.

GUBITZ T., A. CALDWELL, AND A. HUDSON. 2003. Rapid molecular evolution of *CYCLOIDEA*-like genes in *Antirrhinum* and its relatives. *Molecular Biology and Evolution* 20: 1537-1544.

HALL J. C., K. J. SYTSMA, AND H. H. ILTIS. 2002. Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. *American Journal of Botany* 89: 1826-1842.

HALL J. C. 2008. Systematics of Capparaceae and Cleomaceae: An evaluation of the generic delimitations of *Capparis* and *Cleome* using plastid DNA sequence data. *Botany-Botanique* 86: 682-696.

HALL J. C. AND K. DONOHUE. 2012. The genetics of plant dispersal. *In* Clobert J., M. Baguette, T. Benton and J. Bullock [eds.], *Dispersal Ecology and Evolution*. Oxford University Press, Oxford.

HALL J. C., T. E. TISDALE, K. DONOHUE, AND E. M. KRAMER. 2006. Developmental basis of an anatomical novelty: heteroarthrocarpy in *Cakile lanceolata* and *Erucaria erucarioides* (Brassicaceae). *International Journal of Plant Sciences* 167: 771-789.

HALL J. C., T. E. TISDALE, K. DONOHUE, A. WHEELER, M. A. AL-YAHYA, AND E. M. KRAMER. 2011. Convergent evolution of a complex fruit structure in the tribe Brassiceae (Brassicaceae). *American Journal of Botany* 98: 1989-2003.

HE L., M. ZHAO, Y. WANG, J. GAI, AND C. HE. 2013. Phylogeny, structural evolution and functional diversification of the plant PHOSPHATE1 gene family: a focus on *Glycine max*. *BMC Evolutionary Biology* 13:103.

HEDGE I. C. 1976. A systematic and geographical survey of the old world Cruciferae. *In* Vaughn J. G., A. J. MacLeod and B. M. G. Jones [eds.], *The Biology and Chemistry of the Cruciferae*, Academic Press, London.

HILEMAN L. C. AND D. A. BAUM. 2003. Why do paralogs persist? Molecular evolution of *CYCLOIDEA* and related floral symmetry genes in Antirrhineae (Veronicaceae). *Molecular Biology and Evolution* 20: 591-600.

HILEMAN L. C. 2014. Bilateral flower symmetry - how, when and why? *Current Opinion in Plant Biology* 17: 146-152.

HOWARTH D. G. AND M. J. DONOGHUE. 2005. Duplications in *CYC*-like genes from Dipsacales correlate with floral form. *International Journal of Plant Sciences* 166: 357-370.

HOWARTH D. G. AND M. J. DONOGHUE. 2006. Phylogenetic analysis of the "'ECE" (CYC/TB1) clade reveals duplications predating the core eudicots. *Proceedings of the National Academy of Sciences of the United States of America* 103: 9101-9106.

HOWARTH D. G., T. MARTINS, E. CHIMNEY, AND M. J. DONOGHUE. 2011. Diversification of *CYCLOIDEA* expression in the evolution of bilateral flower symmetry in Caprifoliaceae and *Lonicera* (Dipsacales). *Annals of Botany* 107: 1521-1532.

HOWE G., P. BUCCIAGLIA, W. HACKETT, G. FURNIER, M. CORDONNIER-PRATT, AND G. GARDNER. 1998. Evidence that the phytochrome gene family in black cottonwood has one *PHYA* locus and two *PHYB* loci but lacks members of the PHYC/F and PHYE subfamilies. *Molecular Biology and Evolution* 15: 160-175.

HUA S., I. H. SHAMSI, Y. GUO, H. PAK, M. CHEN, C. SHI, H. MENG, AND L. JIANG. 2009. Sequence, expression divergence, and complementation of homologous *ALCATRAZ* loci in *Brassica napus*. *Planta* 230: 493-503.

HUELSENBECK J. AND F. RONQUIST. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.

IBM Corp.: IBM SPSS statistics for Windows. Released 2012, Version 19.0.

ILTIS H. H., J. C. HALL, T. S. COCHRANE, AND K. J. SYTSMA. 2011. Studies in the Cleomaceae I. on the separate recognition of Capparaceae, Cleomaceae, and Brassicaceae. *Annals of the Missouri Botanical Garden* 98: 28-36.

ITKIN M., H. SEYBOLD, D. BREITEL, I. ROGACHEV, S. MEIR, AND A. AHARONI. 2009. *TOMATO AGAMOUS-LIKE 1* is a component of the fruit ripening regulatory network. *Plant Journal* 60: 1081-1095.

JAGER M., A. HASSANIN, M. MANUEL, H. LE GUYADER, AND J. DEUTSCH. 2003. MADS-box genes in *Ginkgo biloba* and the evolution of the AGAMOUS family. *Molecular Biology and Evolution* 20: 842-854.

JAILLON O., J. AURY, B. NOEL, A. POLICRITI, C. CLEPET, A. CASAGRANDE, N. CHOISNE, S. AUBOURG, N. VITULO, C. JUBIN, A. VEZZI, F. LEGEAI, P. HUGUENEY, C. DASILVA, D. HORNER, E. MICA, D. JUBLOT, J. POULAIN, C. BRUYERE, A. BILLAULT, B. SEGURENS, M. GOUYVENOUX, E. UGARTE, F. CATTONARO, V. ANTHOUARD, V. VICO, C. DEL FABBRO, M. ALAUX, G. DI GASPERO, V. DUMAS, N. FELICE, S. PAILLARD, I. JUMAN, M. MOROLDO, S. SCALABRIN, A. CANAGUIER, I. LE CLAINCHE, G. MALACRIDA, E. DURAND, G. PESOLE, V. LAUCOU, P. CHATELET, D. MERDINOGLU, M. DELLEDONNE, M. PEZZOTTI, A. LECHARNY, C. SCARPELLI, F. ARTIGUENAVE, M. E. PE, G. VALLE, M. MORGANTE, M.

CABOCHE, A. ADAM-BLONDON, J. WEISSENBACH, F. QUETIER, P. WINCKER, AND FRENCH-ITALIAN PUBLIC. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463-475.

JARAMILLO M. A. AND E. M. KRAMER. 2007. Molecular evolution of the petal and stamen identity genes, *APETALA3* and *PISTILLATA*, after petal loss in the Piperales. *Molecular Phylogenetics and Evolution* 44: 598-609.

JIAO Y., N. J. WICKETT, S. AYYAMPALAYAM, A. S. CHANDERBALI, L. LANDHERR, P. E. RALPH, L. P. TOMSHO, Y. HU, H. LIANG, P. S. SOLTIS, D. E. SOLTIS, S. W. CLIFTON, S. E. SCHLARBAUM, S. C. SCHUSTER, H. MA, J. LEEBENS-MACK, AND C. W. DEPAMPHILIS. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97-113.

JOHNSON S. D., H. P. LINDER, AND K. E. STEINER. 1998. Phylogeny and radiation of pollination systems in *Disa* (Orchidaceae). *American Journal of Botany* 85: 402-411.

KAMPNY C. M. 1995. Pollination and flower diversity in Scrophulariaceae. *Botanical Review* 61: 352-366.

KARLOVA R., F. M. ROSIN, J. BUSSCHER-LANGE, V. PARAPUNOVA, P. T. DO, A. R. FERNIE, P. D. FRASER, C. BAXTER, G. C. ANGENENT, AND R. A. DE MAAGD. 2011. Transcriptome and metabolite profiling show that *APETALA2a* is a major regulator of tomato fruit ripening. *Plant Cell* 23: 923-941.

KATOH K. AND D. M. STANDLEY. 2013. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772-780.

KERS L. E. 2003. Capparaceae. *In* Kubitzki K. and C. Bayer [eds.], *The Families and Genera of Vascular Plants*, 36-56. Springer-Verlag, Berlin.

KINJO H., N. SHITSUKAWA, S. TAKUMI, AND K. MURAI. 2012. Diversification of three *APETALA1/FRUITFULL*-like genes in wheat. *Molecular Genetics and Genomics* 287: 283-294.

KNOWLES D. G. AND A. MCLYSAGHT. 2006. High rate of recent intron gain and loss in simultaneously duplicated *Arabidopsis* genes. *Molecular Biology and Evolution* 23: 1548-1557.

KOCH M., I. AL-SHEHBAZ, AND K. MUMMENHOFF. 2003. Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae). *Annals of the Missouri Botanical Garden* 90: 151-171.

KOCH M., B. HAUBOLD, AND T. MITCHELL-OLDS. 2001. Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *Chs* sequences. *American Journal of Botany* 88: 534-544.

KOCH M. A. AND K. MUMMENHOFF. 2006. Evolution and phylogeny of the Brassicaceae. *Plant Systematics and Evolution* 259: 81-83.

KOSUGI S. AND Y. OHASHI. 1997. *PCF1* and *PCF2* specifically bind to *cis* elements in the rice proliferating cell nuclear antigen gene. *Plant Cell* 9: 1607-1619.

KOSUGI S. AND Y. OHASHI. 2002. DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant Journal* 30: 337-348.

KOYAMA T., F. SATO, AND M. OHME-TAKAGI. 2010. A role of *TCP1* in the longitudinal elongation of leaves in *Arabidopsis*. *Bioscience Biotechnology and Biochemistry* 74: 2145-2147.

KRAMER E. M., L. HOLAPPA, B. GOULD, M. A. JARAMILLO, D. SETNIKOV, AND P. M. SANTIAGO. 2007. Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia*. *Plant Cell* 19: 750-766.

KRAMER E., R. DORIT, AND V. IRISH. 1998. Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. *Genetics* 149: 765-783.

KRAMER E., M. JARAMILLO, AND V. DI STILIO. 2004. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* 166: 1011-1023.

LANFEAR R., B. CALCOTT, S. Y. W. HO, AND S. GUINDON. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695-1701.

LEE H. AND V. F. IRISH. 2011. Gene duplication and loss in a MADS-box gene transcription factor circuit. *Molecular Biology and Evolution* 28: 3367-3380.

LENSER T. AND G. THEISSEN. 2013. Conservation of fruit dehiscence pathways between *Lepidium campestre* and *Arabidopsis thaliana* sheds light on the regulation of *INDEHISCENT*. *Plant Journal* 76: 545-556.

LEWIS M. W., M. E. LESLIE, AND S. J. LILJEGREN. 2006. Plant separation: 50 ways to leave your mother. *Current Opinion in Plant Biology* 9: 59-65.

LI S., L. FENG, AND D. NIU. 2007. Selection for the miniaturization of highly expressed genes. *Biochemical and Biophysical Research Communications* 360: 586-592.

LI W. AND C. CHINNAPPA. 2003. The phytochrome gene family in the *Stellaria longipes* complex. *International Journal of Plant Sciences* 164: 657-673.

LIDGARD S. AND P. R. CRANE. 1990. Angiosperm diversification and Cretaceous floristic trends - a comparison of palynofloras and leaf macrofloras. *Paleobiology* 16: 77-93.

LILJEGREN S., G. DITTA, H. ESHED, B. SAVIDGE, J. BOWMAN, AND M. YANOFSKY. 2000. *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404: 766-770.

LILJEGREN S., A. ROEDER, S. KEMPIN, K. GREMSKI, L. OSTERGAARD, S. GUIMIL, D. REYES, AND M. YANOFSKY. 2004. Control of fruit patterning in *Arabidopsis* by *INDEHISCENT*. *Cell* 116: 843-853.

LITT A. AND V. IRISH. 2003. Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: Implications for the evolution of floral development. *Genetics* 165: 821-833.

LITT A. 2007. An evaluation of A-function: evidence from the *APETALA1* and *APETALA2* gene lineages. *International Journal of Plant Sciences* 168: 73-91.

LIU D., D. WANG, Z. QIN, D. ZHANG, L. YIN, L. WU, J. COLASANTI, A. LI, AND L. MAO. 2014. The *SEPALLATA* MADS-box protein SLMBP21 forms protein complexes with *JOINTLESS* and *MACROCALYX* as a transcription activator for development of the tomato flower abscission zone. *The Plant Journal for Cell and Molecular Biology* 77: 284-96.

LORTS C. M., T. BRIGGEMAN, AND TAO SANG. 2008. Evolution of fruit types and seed dispersal: a phylogenetic and ecological snapshot. *Journal of Systematics and Evolution* 46: 396-404.

LUO D., R. CARPENTER, L. COPSEY, C. VINCENT, J. CLARK, AND E. COEN. 1999. Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* 99: 367-376.

LUO D., R. CARPENTER, C. VINCENT, L. COPSEY, AND E. COEN. 1996. Origin of floral asymmetry in *Antirrhinum*. *Nature* 383: 794-799.

LYNCH M. AND J. CONERY. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151-1155.

LYONS E. AND M. FREELING. 2008. How to usefully compare homologous plant genes and chromosomes as DNA sequences. *The Plant Journal* 53: 661-673.

LYSAK M., M. KOCH, A. PECINKA, AND I. SCHUBERT. 2005. Chromosome triplication found across the tribe Brassiceae. *Genome research* 15: 516-525.

LYSAK M. A., K. CHEUNG, M. KITSCHKE, AND P. BURES. 2007. Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiology* 145: 402-410.

MADDISON, W. P. AND D.R. MADDISON. 2009. Mesquite: a modular system for evolutionary analysis. Version 2.71.

MADDISON W. 1997. Gene trees in species trees. *Systematic Biology* 46: 523-536.

MARON L. G., C. T. GUIMARAES, M. KIRST, P. S. ALBERT, J. A. BIRCHLER, P. J. BRADBURY, E. S. BUCKLER, A. E. COLUCCIO, T. V. DANILOVA, D. KUDRNA, J. V. MAGALHAES, M. A. PINEROS, M. C. SCHATZ, R. A. WING, AND L. V. KOCHIAN. 2013. Aluminum tolerance in maize is associated with higher *MATE1* gene copy number. *Proceedings of the National Academy of Sciences of the United States of America* 110: 5241-5246.

MARTIN-TRILLO M. AND P. CUBAS. 2010. TCP genes: a family snapshot ten years later. *Trends in Plant Science* 15: 31-39.

MCKENNA D. D. 2011. Towards a temporal framework for "inordinate fondness": reconstructing the macroevolutionary history of beetles (Coleoptera). *Entomologica Americana* 117: 28-36.

MENA M., B. A. AMBROSE, R. B. MEELEY, S. P. BRIGGS, M. F. YANOFSKY, AND R. J. SCHMIDT. 1996. Diversification of C-function activity in maize flower development. *Science* 274: 1537-1540.

MING R., S. HOU, Y. FENG, Q. YU, A. DIONNE-LAPORTE, J. H. SAW, P. SENIN, W. WANG, B. V. LY, K. L. T. LEWIS, S. L. SALZBERG, L. FENG, M. R. JONES, R. L. SKELTON, J. E. MURRAY, C. CHEN, W. QIAN, J. SHEN, P. DU, M. EUSTICE, E. TONG, H. TANG, E. LYONS, R. E. PAULL, T. P. MICHAEL, K. WALL, D. W. RICE, H. ALBERT, M. WANG, Y. J. ZHU, M. SCHATZ, N. NAGARAJAN, R. A. ACOB, P. GUAN, A. BLAS, C. M. WAI, C. M. ACKERMAN, Y. REN, C. LIU, J. WANG, J. WANG, J. NA, E. V. SHAKIROV, B. HAAS, J. THIMMAPURAM, D. NELSON, X. WANG, J. E. BOWERS, A. R. GSCHWEND, A. L. DELCHER, R. SINGH, J. Y. SUZUKI, S. TRIPATHI, K. NEUPANE, H. WEI, B. IRIKURA, M. PAIDI, N. JIANG, W. ZHANG, G. PRESTING, A. WINDSOR, R. NAVAJAS-PEREZ, M. J. TORRES, F. A. FELTUS, B. PORTER, Y. LI, A. M. BURROUGHS, M. LUO, L. LIU, D. A. CHRISTOPHER, S. M. MOUNT, P. H. MOORE, T. SUGIMURA, J. JIANG, M. A. SCHULER, V. FRIEDMAN, T. MITCHELL-OLDS, D. E. SHIPPEN, C. W. DEPAMPHILIS, J. D. PALMER, M. FREELING, A. H. PATERSON, D. GONSALVES, L. WANG, AND M. ALAM. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452: 991-1007.

MOLLER A. P. AND M. ERIKSSON. 1995. Pollinator preference for symmetrical flowers and sexual selection in plants. *Oikos* 73: 15-22.

MOLLER M., M. CLOKIE, P. CUBAS, AND Q. C. B. CRONK. 1999. Integrating molecular phylogenies and developmental genetics: a Gesneriaceae case study. *Molecular Systematics and Plant Evolution* 57: 375-402.

MONDRAGON-PALOMINO M. AND G. THEISSEN. 2009. Why are orchid flowers so diverse? Reduction of evolutionary constraints by paralogues of class B floral homeotic genes. *Annals of Botany* 104: 583-594.

MOORE R. AND M. PURUGGANAN. 2005. The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology* 8: 122-128.

MOREAU C. S., C. D. BELL, R. VILA, S. B. ARCHIBALD, AND N. E. PIERCE. 2006. Phylogeny of the ants: diversification in the age of angiosperms. *Science* 312: 101-104.

MUEHLHAUSEN A., T. LENSER, K. MUMMENHOFF, AND G. THEISSEN. 2013. Evidence that an evolutionary transition from dehiscent to indehiscent fruits in *Lepidium* (Brassicaceae) was caused by a change in the control of valve margin identity genes. *Plant Journal* 73: 824-835.

MUMMENHOFF K., A. POLSTER, A. MUEHLHAUSEN, AND G. THEISSEN. 2009. *Lepidium* as a model system for studying the evolution of fruit development in Brassicaceae. *Journal of Experimental Botany* 60: 1503-1513.

NAKANO T., J. KIMBARA, M. FUJISAWA, M. KITAGAWA, N. IHASHI, H. MAEDA, T. KASUMI, AND Y. ITO. 2012. *MACROCALYX* and *JOINTLESS* interact in the transcriptional regulation of tomato fruit abscission zone development. *Plant Physiology* 158: 439-450.

NEI M. AND T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 3: 418-426.

NEI M. AND A. ROONEY. 2005. Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics* 39: 121-152.

NIELSEN C. B., B. FRIEDMAN, B. BIRREN, C. B. BURGE, AND J. E. GALAGAN. 2004. Patterns of intron gain and loss in fungi. *PLoS Biology* 2(12): e422.

OHNO S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, New York.

O'KANE S. AND I. AL-SHEHBAZ. 2003. Phylogenetic position and generic limits of *Arabidopsis* (Brassicaceae) based on sequences of nuclear ribosomal DNA. *Annals of the Missouri Botanical Garden* 90: 603-612.

OLIVEIRA A. E. A., D. BELLIENY-RABELO, AND T. M. VENANCIO. 2013. Impact of whole-genome and tandem duplications in the expansion and functional diversification of the F-box family in legumes (Fabaceae). *PLoS One* 8 (2): e55127.

OLMSTEAD R. G. 2013. Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: a comparison of continental and intercontinental diversification patterns. *Botanical Journal of the Linnean Society* 171: 80-102.

OSNATO M., M. R. STILE, Y. WANG, D. MEYNARD, S. CURIALE, E. GUIDERDONI, Y. LIU, D. S. HORNER, P. B. F. OUWERKERK, C. POZZI, K. J. MULLER, F. SALAMINI, AND L. ROSSINI. 2010. Cross talk between the *KNOX* and ethylene pathways is mediated by intron-binding transcription factors in barley. *Plant Physiology* 154: 1616-1632.

OTTO S. P. AND J. WHITTON. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401-437.

PABON-MORA N., B. A. AMBROSE, AND A. LITT. 2012. Poppy *APETALA1/FRUITFULL* orthologs control flowering time, branching, perianth identity, and fruit development. *Plant Physiology* 158: 1685-1704.

PABON-MORA N., O. HIDALGO, S. GLEISSBERG, AND A. LITT. 2013a. Assessing duplication and loss of *APETALA1/FRUITFULL* homologs in Ranunculales. *Frontiers in Plant Science* 4: 358-358.

PABON-MORA N., B. SHARMA, L. D. HOLAPPA, E. M. KRAMER, AND A. LITT. 2013b. The *Aquilegia FRUITFULL*-like genes play key roles in leaf morphogenesis and inflorescence development. *Plant Journal* 74: 197-212.

PALMER J. D., K. L. ADAMS, Y. CHO, C. L. PARKINSON, Y. QIU, AND K. SONG. 2000. Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences of the United States of America* 97: 6960-6966.

PARK K. I., J. D. CHOI, A. HOSHINO, Y. MORITA, AND S. IIDA. 2004. An intragenic tandem duplication in a transcriptional regulatory gene for anthocyanin biosynthesis confers pale-colored flowers and seeds with fine spots in Ipomoea tricolor. *Plant Journal* 38: 840-849.

PARKIN I. A., A. G. SHARPE, D. J. KEITH, AND D. J. LYDIATE. 1995. Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). *Genome* 38: 1122-1131.

PARKIN I., S. GULDEN, A. SHARPE, L. LUKENS, M. TRICK, T. OSBORN, AND D. LYDIATE. 2005. Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171: 765-781.

PARRA G., K. BRADNAM, A. B. ROSE, AND I. KORF. 2011. Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. *Nucleic Acids Research* 39: 5328-5337.

PATCHELL M. J. 2012. Connecting molecules to morphology: evolutionary developmental genetics of floral monosymmetry in *Cleome violacea* (Cleomaceae). MSc Thesis, University of Alberta, Edmonton.

PATCHELL M. J., M. C. BOLTON, P. MANKOWSKI, AND J. C. HALL. 2011. Comparative floral development in Cleomaceae reveals two distinct pathways leading to monosymmetry. *International Journal of Plant Sciences* 172: 352-365.

PATCHELL M. J., E. H. ROALSON, AND J. C. HALL. 2014. Resolved phylogeny of the Cleomaceae based on all three genomes. *Taxon* 63: 315-328.

PATERSON A. H., J. E. BOWERS, AND B. A. CHAPMAN. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences of the United States of America* 101: 9903-9908.

PHILLIPS M. J., F. DELSUC, AND D. PENNY. 2004. Genome-scale phylogeny and the detection of systematic biases. *Molecular Biology and Evolution* 21: 1455-1458.

PINYOPICH A., G. S. DITTA, B. SAVIDGE, S. J. LILJEGREN, E. BAUMANN, E. WISMAN, AND M. F. YANOFSKY. 2003. Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424: 85-88.

POSADA D. AND K. CRANDALL. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

PRESTON J. C. AND L. C. HILEMAN. 2012. Parallel evolution of TCP and B-class genes in Commelinaceae flower bilateral symmetry. *Evodevo* 3: 6.

PRESTON J. C., L. C. HILEMAN, AND P. CUBAS. 2011a. Reduce, reuse, and recycle: developmental evolution of trait diversification. *American Journal of Botany* 98: 397-403.

PRESTON J. C. AND E. A. KELLOGG. 2006. Reconstructing the evolutionary history of paralogous *APETALA1/FRUITFULL*-like genes in grasses (Poaceae). *Genetics* 174: 421-437.

PRESTON J. C., M. A. KOST, AND L. C. HILEMAN. 2009. Conservation and diversification of the symmetry developmental program among close relatives of snapdragon with divergent floral morphologies. *New Phytologist* 182: 751-762.

PRESTON J. C., C. C. MARTINEZ, AND L. C. HILEMAN. 2011b. Gradual disintegration of the floral symmetry gene network is implicated in the evolution of a wind-pollination syndrome. *Proceedings of the National Academy of Sciences of the United States of America* 108: 2343-2348.

PURUGGANAN M., S. ROUNSLEY, R. SCHMIDT, AND M. YANOFSKY. 1995. Molecular evolution of flower development - diversification of the plant MADS-box regulatory gene family. *Genetics* 140: 345-356.

RAMBAUT A. AND A.J. DRUMMOND. 2007. Tracer. Version 1.4.

RASTOGI P. A. 2000. MacVector. Integrated sequence analysis for the Macintosh. *Methods in Molecular Biology* 132: 47-69.

REEVES P. A. AND R. G. OLMSTEAD. 2003. Evolution of the TCP gene family in Asteridae: cladistic and network approaches to understanding regulatory gene family diversification and its impact on morphological evolution. *Molecular Biology and Evolution* 20: 1997-2009.

RIPOLL J. J., A. H. K. ROEDER, G. S. DITTA, AND M. F. YANOFSKY. 2011. A novel role for the floral homeotic gene *APETALA2* during *Arabidopsis* fruit development. *Development* 138: 5167-5176.

RODGERS-MELNICK E., S. P. MANE, P. DHARMAWARDHANA, G. T. SLAVOV, O. R. CRASTA, S. H. STRAUSS, A. M. BRUNNER, AND S. P. DIFAZIO. 2012. Contrasting patterns of evolution following whole genome versus tandem duplication events in Populus. *Genome Research* 22: 95-105.

RODRIGUEZ I., A. GUMBERT, N. H. DE IBARRA, J. KUNZE, AND M. GIURFA. 2004. Symmetry is in the eye of the 'beeholder': innate preference for bilateral symmetry in flower-naive bumblebees. *Naturwissenschaften* 91: 374-377.

ROEDER A., C. FERRANDIZ, AND M. YANOFSKY. 2003. The role of the *REPLUMLESS* homeodomain protein in patterning the *Arabidopsis* fruit. *Current Biology* 13: 1630-1635.

ROELANTS K., D. J. GOWER, M. WILKINSON, S. P. LOADER, S. D. BIJU, K. GUILLAUME, L. MORIAU, AND F. BOSSUYT. 2007. Global patterns of diversification in the history of modern amphibians. *Proceedings of the National Academy of Sciences of the United States of America* 104: 887-892.

ROSIN F. M. AND E. M. KRAMER. 2009. Old dogs, new tricks: regulatory evolution in conserved genetic modules leads to novel morphologies in plants. *Developmental Biology* 332: 25-35.

ROY S. W. AND D. PENNY. 2006. Smoke without fire: most reported cases of intron gain in nematodes instead reflect intron losses. *Molecular Biology and Evolution* 23: 2259-2262.

RUTLEDGE R., S. REGAN, O. NICOLAS, P. FOBERT, C. COTE, W. BOSNICH, C. KAUFFELDT, G. SUNOHARA, A. SEGUIN, AND D. STEWART. 1998. Characterization of an *AGAMOUS* homologue from the conifer black spruce (*Picea mariana*) that produces floral homeotic conversions when expressed in *Arabidopsis*. *Plant Journal* 15: 625-634.

SCHRANZ M. E. AND T. MITCHELL-OLDS. 2006. Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* 18: 1152-1165.

SCOTLAND R. W. 2011. What is parallelism? *Evolution & Development* 13: 214-227.

SEOIGHE C., C. GEHRING, AND L. HURST. 2005. Gametophytic selection in *Arabidopsis* thaliana supports the selective model of intron length reduction. *PLoS Genetics* 1: 154-158.

SEYMOUR G. B., L. OSTERGAARD, N. H. CHAPMAN, S. KNAPP, AND C. MARTIN. 2013. Fruit development and ripening. *Annual Review of Plant Biology* 64: 219-241.

SHAN H., N. ZHAN, C. LIU, G. XU, J. ZHANG, Z. CHEN, AND H. KONG. 2007. Patterns of gene duplication and functional diversification during the evolution of the AP1/SQUA subfamily of plant MADS-box genes. *Molecular Phylogenetics and Evolution* 44: 26-41.

SIMILLION C., K. VANDEPOELE, M. C. E. VAN MONTAGU, M. ZABEAU, AND Y. VAN DE PEER. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* 99: 13627-13632.

SKIPPER M., K. B. PEDERSEN, L. B. JOHANSEN, S. FREDERIKSEN, V. F. IRISH, AND B. B. JOHANSEN. 2005. Identification and quantification of expression levels of three *FRUITFULL*-like MADS-box genes from the orchid *Dendrobium thyrsiflorum* (Reichb. f.). *Plant Science* 169: 579-586.

SMITH J. F., M. M. FUNKE, AND V. L. WOO. 2006. A duplication of *GCYC* predates divergence within tribe Coronanthereae (Gesneriaceae): phylogenetic analysis and evolution. *Plant Systematics and Evolution* 261: 245-256.

SMITH J. F., L. C. HILEMAN, M. P. POWELL, AND D. A. BAUM. 2004. Evolution of *GCYC*, a Gesneriaceae homolog of *CYCLOIDEA*, within Gesnerioideae (Gesneriaceae). *Molecular Phylogenetics and Evolution* 31: 765-779.

SOLTIS D. E., V. A. ALBERT, J. LEEBENS-MACK, C. D. BELL, A. H. PATERSON, C. ZHENG, D. SANKOFF, C. W. DEPAMPHILIS, P. K. WALL, AND P. S. SOLTIS. 2009a. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336-348.

SOLTIS P. S., S. F. BROCKINGTON, M. YOO, A. PIEDRAHITA, M. LATVIS, M. J. MOORE, A. S. CHANDERBALI, AND D. E. SOLTIS. 2009b. Floral variation and floral genetics in basal angiosperms. *American Journal of Botany* 96: 110-128.

SONG C., Q. LIN, R. LIANG, AND Y. WANG. 2009. Expressions of ECE-CYC2 clade genes relating to abortion of both dorsal and ventral stamens in *Opithandra* (Gesneriaceae). *BMC Evolutionary Biology* 9: 244.

SOUTHERN, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98: 503–517

SPECHT C. D. AND M. E. BARTLETT. 2009. Flower evolution: the origin and subsequent diversification of the angiosperm flower. *Annual Review of Ecology Evolution and Systematics* 40: 217-243.

STEVENS P. F. 2012. Missouri Botanical Garden. http://www.mobot.org/MOBOT/research/APweb/ [Accessed April/01 2014].

STRATMANN J. W. AND S. R. HIND. 2011. Gene silencing goes viral and uncovers the private life of plants. *Entomologia Experimentalis et Applicata* 140: 91-102.

SUN W., W. HUANG, Z. LI, C. SONG, D. LIU, Y. LIU, A. HAYWARD, Y. LIU, H. HUANG, AND Y. WANG. 2014. Functional and evolutionary analysis of the *AP1/SEP/AGL6* superclade of MADS-box genes in the basal eudicot *Epimedium sagittatum*. *Annals of Botany* 113: 653-668.

SWARBRECK D., C. WILKS, P. LAMESCH, T. Z. BERARDINI, M. GARCIA-HERNANDEZ, H. FOERSTER, D. LI, T. MEYER, R. MULLER, L. PLOETZ, A. RADENBAUGH, S. SINGH, V. SWING, C. TISSIER, P. ZHANG, AND E. HUALA. 2008. The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research* 36: 1009-1014.

SWOFFORD D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0β10.

SZOELLOSI G. J. AND V. DAUBIN. 2012. Modeling gene family evolution and reconciling phylogenetic discord. *Evolutionary Genomics: Statistical and Computational Methods, Vol 2* 856: 29-51.

TAGHIZADEH M. S., S. CRAWFORD, M. E. NICOLAS, AND R. D. COUSENS. 2009. Water deficit changes the anatomy of the fruit abscission zone in *Raphanus raphanistrum* (Brassicaceae). *Australian Journal of Botany* 57: 708-714.

TAN X., Z. XIA, L. ZHANG, Z. ZHANG, Z. GUO, AND C. QI. 2009. Cloning and sequence analysis of oilseed rape (*Brassica napus*) *SHP2* gene. *Botanical Studies* 50: 403-412.

TANDRE K., V. A. ALBERT, A. SUNDAS, AND P. ENGSTROM. 1995. Conifer homologs to genes that control floral development in angiosperms. *Plant Molecular Biology* 27: 69-78.

TANI E., A. N. POLIDOROS, E. FLEMETAKIS, C. STEDEL, C. KALLONIATI, K. DEMETRIOU, P. KATINAKIS, AND A. S. TSAFTARIS. 2009. Characterization and expression analysis of *AGAMOUS*-like, *SEEDSTICK*-like, and *SEPALLATA*-like MADS-box genes in peach (*Prunus persica*) fruit. *Plant Physiology and Biochemistry* 47: 690-700.

THEISSEN G. 2001. Development of floral organ identity: stories from the MADS house. *Current Opinion in Plant Biology* 4: 75-85.

THOMAS B. C., B. PEDERSEN, AND M. FREELING. 2006. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homolog leaving clusters enriched in dose-sensitive genes. *Genome Research* 16: 934-946.

THOMPSON J., T. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. HIGGINS. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876-4882.

TONSOR S. J., C. ALONSO-BLANCO, AND M. KOORNNEEF. 2005. Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. *Plant Cell and Environment* 28: 2-20.

TSAI W. C., C. S. KUOH, M. H. CHUANG, W. H. CHEN, AND H. H. CHEN. 2004. Four *DEF*-Like MADS-box genes displayed distinct floral morphogenetic roles in *Phaldenopsis* orchid. *Plant and Cell Physiology* 45: 831-844.

TSAI W., Z. PAN, Y. HSIAO, M. JENG, T. WU, W. CHEN, AND H. CHEN. 2008. Interactions of B-class complex proteins involved in tepal development in *Phalaenopsis* orchid. *Plant and Cell Physiology* 49: 814-824.

USHIMARU A., I. DOHZONO, Y. TAKAMI, AND F. HYODO. 2009. Flower orientation enhances pollen transfer in bilaterally symmetrical flowers. *Oecologia* 160: 667-674.

VANDER WALL S. B. AND M. J. BECK. 2012. A comparison of frugivory and scatter-hoarding seed-dispersal syndromes. *Botanical Review* 78: 10-31.

VEIT B., E. VOLLBRECHT, J. MATHERN, AND S. HAKE. 1990. A tandem duplication causes the Kn1-0 allele of *KNOTTED*, a dominant morphological mutant of maize. *Genetics* 125: 623-631.

VEKEMANS D., S. PROOST, K. VANNESTE, H. COENEN, T. VIAENE, P. RUELENS, S. MAERE, Y. VAN DE PEER, AND K. GEUTEN. 2012. Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-box gene and species diversification. *Molecular Biology and Evolution* 29: 3793-3806.

VIAENE T., D. VEKEMANS, V. F. IRISH, A. GEERAERTS, S. HUYSMANS, S. JANSSENS, E. SMETS, AND K. GEUTEN. 2009. *Pistillata*-Duplications as a mode for floral diversification in (basal) Asterids. *Molecular Biology and Evolution* 26: 2627-2645.

VISION T. J., D. G. BROWN, AND S. D. TANKSLEY. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290: 2114-2117.

WANG X., H. WANG, J. WANG, R. SUN, J. WU, S. LIU, Y. BAI, J. MUN, I. BANCROFT, F. CHENG, S. HUANG, X. LI, W. HUA, J. WANG, X. WANG, M. FREELING, J. C. PIRES, A. H. PATERSON, B. CHALHOUB, B. WANG, A. HAYWARD, A. G. SHARPE, B. PARK, B. WEISSHAAR, B. LIU, B. LI, B. LIU, C. TONG, C. SONG, C. DURAN, C. PENG, C. GENG, C. KOH, C. LIN, D. EDWARDS, D. MU, D. SHEN, E. SOUMPOUROU, F. LI, F. FRASER, G. CONANT, G. LASSALLE, G. J. KING, G. BONNEMA, H. TANG, H. WANG, H. BELCRAM, H. ZHOU, H. HIRAKAWA, H. ABE, H. GUO, H. WANG, H. JIN, I. A. P. PARKIN, J. BATLEY, J. KIM, J. JUST, J. LI, J. XU, J. DENG, J. A. KIM, J. LI, J. YU, J. MENG, J. WANG, J. MIN, J. POULAIN, K. HATAKEYAMA, K. WU, L. WANG, L. FANG, M. TRICK, M. G. LINKS, M. ZHAO, M. JIN, N. RAMCHIARY, N. DROU, P. J. BERKMAN, Q. CAI, Q. HUANG, R. LI, S. TABATA, S. CHENG, S. ZHANG, S. ZHANG, S. HUANG, S. SATO, S. SUN, S. KWON, S. CHOI, T. LEE, W. FAN, X. ZHAO, X. TAN, X. XU, Y. WANG, Y. QIU, Y. YIN, Y. LI, Y. DU, Y. LIAO, Y. LIM, Y. NARUSAKA, Y. WANG, Z. WANG, Z. LI, Z. WANG, Z. XIONG, AND Z. ZHANG. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature genetics* 43: 1035-1040.

WANG X., D. C. TANK, AND T. SANG. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular Biology and Evolution* 17: 773-781.

WANG Z., Y. LUO, X. LI, L. WANG, S. XU, J. YANG, L. WENG, S. SATO, S. TABATA, M. AMBROSE, C. RAMEAU, X. FENG, X. HU, AND D. LUO. 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: 10414-10419.

WARWICK S. AND C. SAUDER. 2005. Phylogeny of tribe Brassiceae (Brassicaceae) based on chloroplast restriction site polymorphisms and nuclear ribosomal internal transcribed spacer and chloroplast *trnL* intron sequences. *Canadian Journal of Botany-Revue Canadienne De Botanique* 83: 467-483.

YANG Z., R. NIELSEN, N. GOLDMAN, AND A. PEDERSEN. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431-449.

YOON H. S. AND D. A. BAUM. 2004. Transgenic study of parallelism in plant morphological evolution. *Proceedings of the National Academy of Sciences of the United States of America* 101: 6524-6529.

ZHANG P. Y., H. T. W. TAN, K. H. PWEE, AND P. P. KUMAR. 2004. Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. *Plant Journal* 37: 566-577.

ZHANG W., E. M. KRAMER, AND C. 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: 6388-6393.

ZHANG W., Q. XIANG, D. T. THOMAS, B. M. WIEGMANN, M. W. FROHLICH, AND D. E. SOLTIS. 2008. Molecular evolution of *PISTILLATA*-like genes in the dogwood genus *Cornus* (Cornaceae). *Molecular Phylogenetics and Evolution* 47: 175-195.

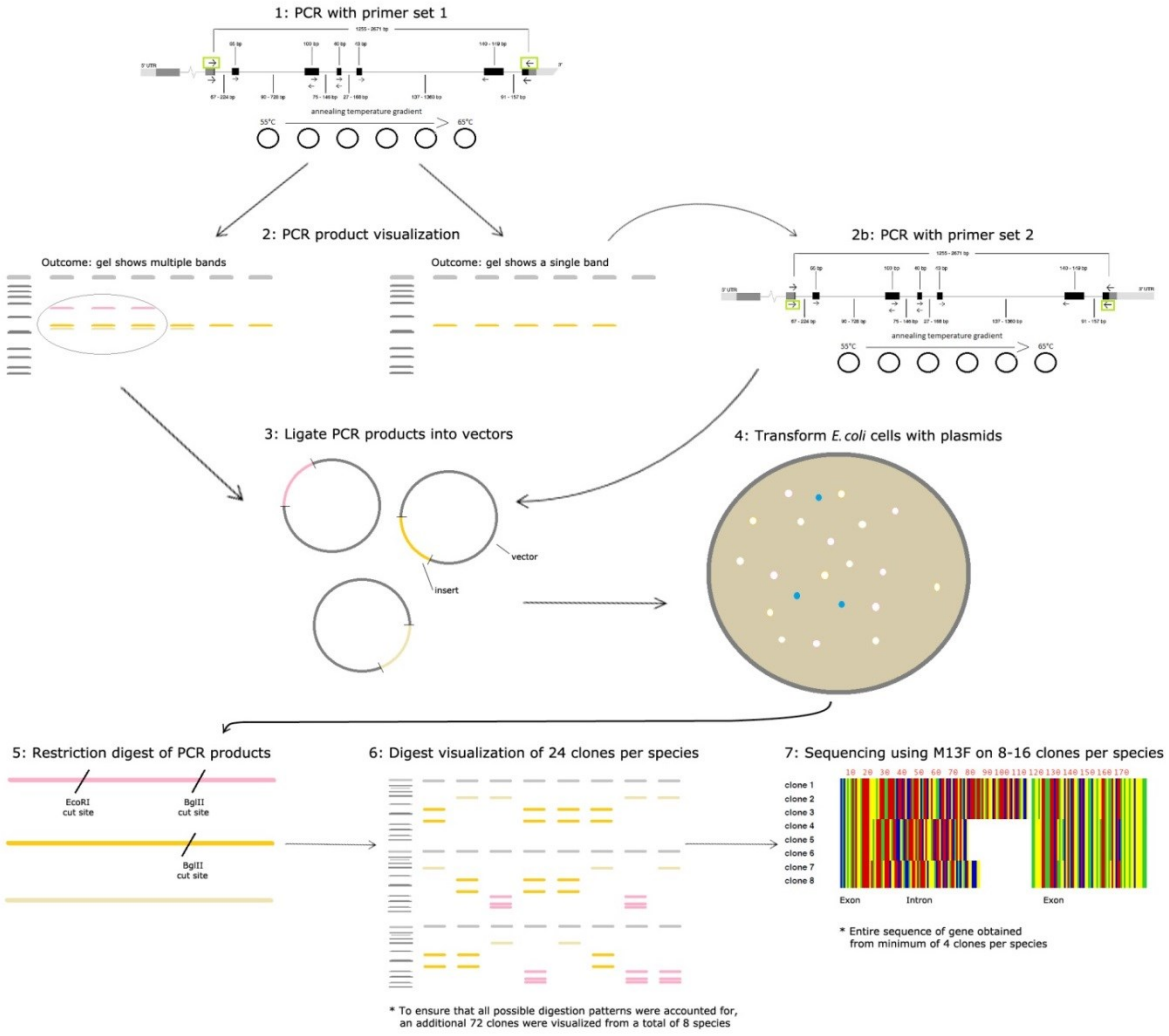
ZWICKL D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D Dissertation. University of Texas at Austin.

Appendices

Appendix 2-1. GenBank accession numbers and sampling information for *FUL* sequences. Voucher and herbarium information including seed stock number, collection information or population designation is included for data gathered from this study. DNA was extracted from greenhouse grown material and deposited at the Harvard University Herbaria (HUH) (Hall et al., 2011).

Species	Molecule type	Source of DNA sequence	GenBank accession No.	Voucher and herbarium information
Outgroup				
Cleome violacea L.	cDNA			Hall and Bolton, 20 Feb 2008 (ALTA)
Aethionema carneum (Banks & Sol.) B. Fedtsch	cDNA	GenBank	FR727244.1	-
Lepidium appelianum Al-Shehbaz	cDNA	GenBank	FR727231.1	-
Lepidium campestre (L.) W.T. Aiton	cDNA	GenBank	FR727237.1	-
Arabidopsis thaliana (L.) Heynh	cDNA	GenBank	AB008269.1	-
Arabidopsis lyrata (L.) O'Kane & Al-Shehbaz	cDNA	GenBank	XM_002866356.1	-
Brassiceae				
Brassica napus L.	cDNA	GenBank	DQ414534.1	-
Brassica napus L.	gDNA	This study		Excel canola; Ex-19
Brassica nigra (L.) W.D.J. Koch	gDNA	This study		0049-67, (HUH)
Brassica oleracea var. *acephala* DC.	gDNA	This study		2861-77, (HUH)
Brassica oleracea var. *botrytis* L.	cDNA	GenBank	AJ505841.1-44.1	-
Brassica spinescens Pomel	gDNA	This study		1800-70, (HUH)
Cakile lanceolata subsp. *fusiformis* (Greene) Rodman	gDNA	This study		GK1, (HUH)
Coincya monensis (L.) Greuter & Burdet	gDNA	This study		4429-75, (HUH)
Cordylocarpus muricatus Desf.	gDNA	This study		1137-68, (HUH)
Crambe orientalis L.	gDNA	This study		3696-75, (HUH)
Crambella teretifolia (Batt. & Trab.) Maire	gDNA	This study		1971, (HUH)
Didesmus bipinnatus Desf.	gDNA	This study		1853-70, (HUH)
Diplotaxis assurgens (Delile) Gren.	gDNA	This study		1120-67, (HUH)
Enarthrocarpus lyratus (Forssk.) D.C.	gDNA	This study		1206-68, (HUH)

Eruca vesicaria (L.) Cav. subsp *sativa* (Mill.) Thell.	gDNA	This study		3750-77,(HUH)
Erucaria erucarioides (Coss. & Durieu) Müll. & Berol.	gDNA	This study		1944-71, (HUH)
Erucastrum gallicum (Willd.) O.E. Schulz	gDNA	This study		1209-69, (HUH)
Guiraoa arvensis Coss.	gDNA	This study		1550-68, (HUH)
Hemicrambe fruticulosa Webb	gDNA	This study		2232-73, (HUH)
Henophyton deserti (Coss. & Durieu) Coss. & Durieu	gDNA	This study		1945-71, (HUH)
Hirschfeldia incana (L.) Lagr.-Foss.	gDNA	This study		2024-71, (HUH)
Moricandia arvensis (L.) D.C.	gDNA	This study		0863-66, (HUH)
Muricaria prostrata Desf.	gDNA	This study		1855-70, (HUH)
Pseuderucaria teretifolia (Desf.) O.E. Schulz	gDNA	This study		1844-70, (HUH)
Psychine stylosa Desf.	gDNA	This study		1458-68, (HUH)
Raffenaldia primuloides Godr.	gDNA	This study		4386-76, (HUH)
Rytidocarpus moricandiodes Coss.	gDNA	This study		0708-67, (HUH)
Schouwia thebaica Webb	gDNA	This study		5780-81, (HUH)
Sinapidendron angustifolium (DC.) Lowe	gDNA	This study		3620-75, (HUH)
Sinapis alba L.	cDNA	GenBank	U25695.1	-
Vella spinosa Boiss.	gDNA	This study		2007-71, (HUH)
Zilla spinosa (L.) Prantl	gDNA	This study		0731-67, (HUH)



Appendix 2-2. Overview of methods from initial PCR amplification to sequencing for each taxon.

Appendix 2-3. Taxa and genes selected for southern blot hybridizations

Taxa	Gene
Cakile lanceolata	*FUL*
Cleome violaceae	*TCP1*
Enarthrocarpus lyratus	*FUL*
Erucaria erucarioides	*FUL, TCP1*
Moricandia arvensis	*FUL*
Polanisia dodecandra	*TCP1*
Thlaspi arvense	*TCP1*

Appendix 2-4. Probe amplification primer sequences used in southern blot hybridizations

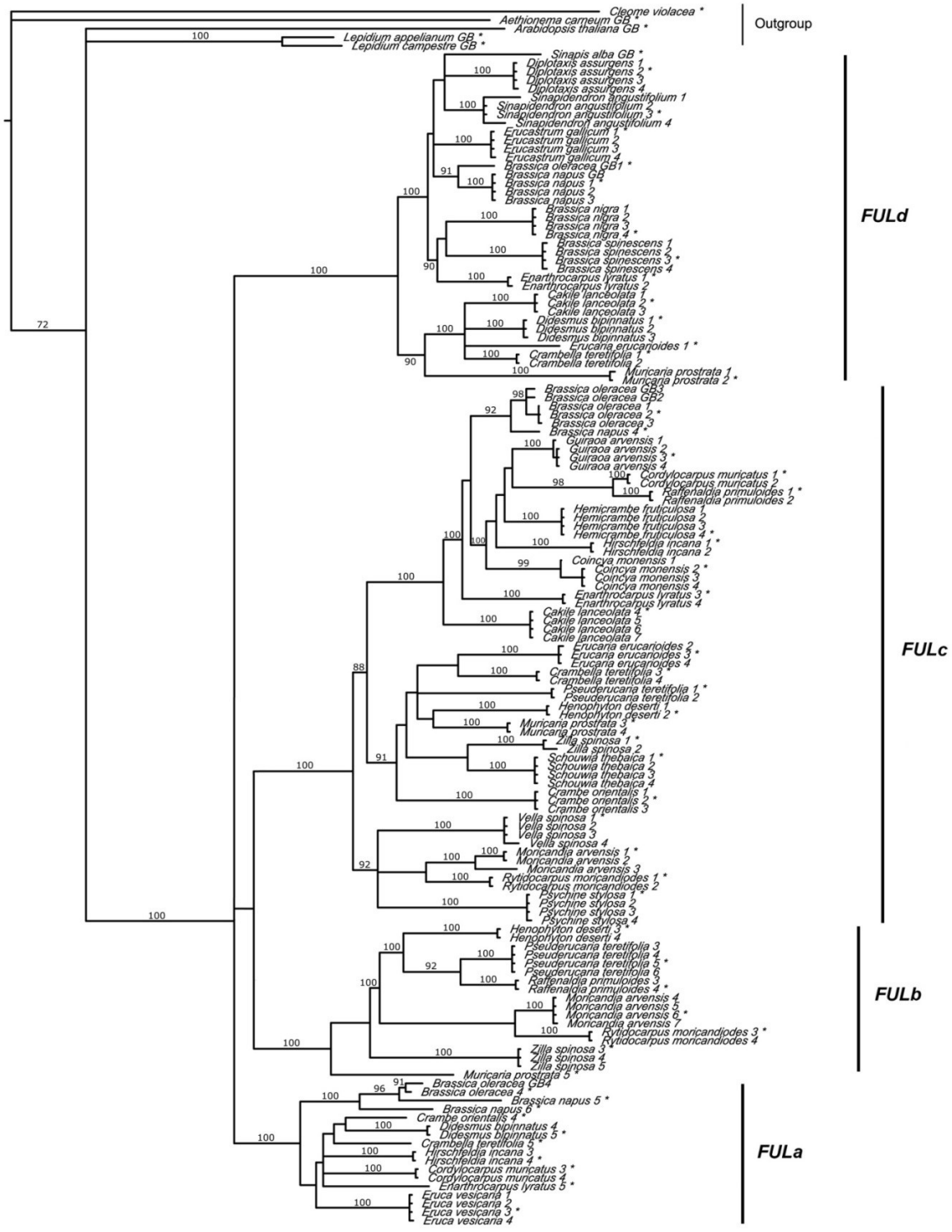
Primer name	Sequence
FULprobeF	AATTTTATGGGGGAAGATCTTGATTCC
FULprobeR	ACCTTTTTGAGAAGCGTATTGTTGTG
TCPprobeF	CCACAGCCTACAATCATCATCATCAT
TCPprobeR	GATGAGCTTCTTCTCGTGAACGAC

Appendix 2-5. Southern blot hybridization protocol

DAY 1		
Step 1	Amplify probe	a. Use any PCR protocol to amplify probe sequence b. Prepare a 20µl aliquot diluted to 10ng/µl c. Store diluted probe in -20° C
Step 2	Extract >10 µg DNA	Use Qiagen DNeasy Plant Maxi Kit, which results in high yields to avoid having to combine multiple extractions. Follow manufacturer's instructions.
Step 3	Digest 10 µg of DNA using restriction enzyme of your choice	a. Use spectrophotometer to quantify and assess quality of gDNA. b. Use 30 units of restriction enzyme (3 units restriction enzyme per µg DNA). If your DNA is concentrated, dilute your DNA such that the final reaction volume is at least 400 µl for best results. Use the appropriate buffer according to the restriction enzyme manufacturer's instructions. Let reaction continue overnight.
DAY 2		
Step 1	Use ethanol precipitation method to clean restriction enzyme and buffer away from digested DNA	a. Add 1/10 th volume of your final digest reaction of 3M sodium acetate to the DNA solution and mix gently. b. Add 2 volumes of -20° C 100% ethanol and vortex (med) for 10 seconds. Put the tube -80° C freezer for 1 hour. c. Spin in a centrifuge for 5 minutes. Pour out the ethanol, saving the pellet in the tube d. Wash the pellet with 500 µl of 4° C 70% EtOH, gently roll the tube, then pour out the EtOH. e. Dry using a speedvac or let tubes dry inverted for ~15 mins. Ensure they are dry or sample will "crawl" out of the wells of the gel. f. Resuspend DNA in 15µl of TE buffer. g. Add 3 µl of loading buffer.

Step 2	Prepare gel electrophoresis	a. Make a 0.8% gel by mixing 1.6 grams of agarose in 200 ml of 1X TAE. Add 2 µl of Ethidium bromide after it has sufficiently cooled. Pour into a 20 cm long gel rig using the largest, double wide gel combs available. b. Submerge gel TAE buffer in gel rig. c. Load ladder and samples carefully d. Run gel at 40 V for 6 -12 hours
Step 3	Prepare hybridization and wash buffers (they will be used in the upcoming next 2 days)	Reference GE Healthcare's Amersham Gene Images AlkPhos Direct Labelling and Detection System product booklet (Buckinghamshire, UK). Follow manufacturer's instructions to prepare the hybridization buffer, primary wash buffer, and secondary wash buffer (20X) stock.
Step 4	Photograph gel	Use photograph to assess whether the gDNA was properly digested. If not, do not proceed.
Step 5	Gel transfer to membrane	a. Assemble vacuum blotter. Place the membrane on the grating, and the plastic gasket over the membrane. Wet underside of gasket at the rubber seal interface of the vacuum blotting unit with left over TAE buffer from gel rig. b. Place gel on top of gasket so the edges of the gasket are completely covered by the gel. c. Turn on pump and slowly increase vacuum to ~70 in. H_2O. d. Cover top of gel with approximately 0.25N HCl. Vacuum 5 min at ~70 in. H_2O. Remove and discard excess solution. e. Add enough denaturation solution to just to cover gel. Vacuum 5 min at about ~70 in. H_2O. f. Remove denaturation solution. Add enough neutralization solution to cover gel. Vacuum 5 min at ~70 in. H_2O. Remove and discard excess solution. g. Transfer DNA to membrane for 1 hour under a constant vacuum (70-80 in. H_2O), ensuring the gel is constantly submerged in 10X SSPE (about 1-2 L).
Step 6	Post-transfer membrane storage	a. Discard gel and disassemble vacuum transfer apparatus. b. Affix DNA to membrane by placing membrane in cross linker over on "auto crosslink" setting. c. You may store the membrane overnight. Ensure it is dry.
Gel Electrophoresis and Transfer Buffers		

1X TAE		Prepare stock 50X TAE a. Add 242 g of Tris to 750 ml of Milli-Q water. Mix until completely dissolved. b. Add 100 ml of 0.5 M EDTA c. Add 57 ml of glacial acetic acid. d. Add Milli-Q water until final volume is 1 L e. Mix and autoclave to sterilize. Prepare 1X TAE a. Add 20 ml of 50X TAE to 980 ml of milli-Q water
0.25 N HCl		a. Add 25 ml of concentrated HCl to 975 ml of milli-Q water
Denaturation Solution		b. Dissolve 43.83 g of NaCl in 300 ml of milli-Q water c. Add 50 ml of 10 N NaOH solution d. Add Milli-Q water until final volume is 500 ml
Neutralization Solution		a. Add 60.5 g of Tris to 850 ml of milli-Q water b. Add 87.45 g of NaCl c. Mix until dissolved d. Adjust to pH 7.5 using concentrated HCl. e. Add Milli-Q water until final volume is 1 L
10X SSPE		a. Add 175.5 g of NaCl to 1.8 L of milli-Q water b. Add 27.6 g of NaH_2PO_4 x H_2O c. Add 7.4 g of Na_2EDTA d. Adjust to pH 7.4 using 10M NaOH e. Add Milli-Q water until final volume is 2 L
DAY 3		
Step 1	Labelling the probe	Use GE Healthcare's Amersham Gene Images AlkPhos Direct Labelling and Detection System. Follow manufacturers' instructions. Note: 1.5 µl of labelling reagent may be used instead of 2 µl.
Step 2	Hybridization	Use GE Healthcare's Amersham Gene Images AlkPhos Direct Labelling and Detection System. Follow all manufacturers' instructions. Let the hybridization reaction run overnight.
DAY 4		
Step 1	Post-hybridization stringency washes	Use GE Healthcare's Amersham Gene Images AlkPhos Direct Labelling and Detection System. Follow all manufacturers' instructions. While pre-heating the primary wash solution, use this time to dilute the20X secondary wash buffer to 1 L of 1X concentration.
Step 2	Signal detection	Use GE Healthcare's Amersham Gene Images AlkPhos Direct Labelling and Detection System. Follow all manufacturers' instructions using the chemiflourescent detection option.



Appendix 2-6. Bayesian consensus including branch lengths of *FUL* sequences from the "all clones coding + intron" nucleotide dataset. Bayesian posterior probabilities above 80% are indicated above branches and "GB" indicates sequences that were retrieved from GenBank. Sequences that were selected for subsequent analyses are indicated by an asterisk.

Appendix 2-7. Intron and exon lengths of representative *FUL* copies from gDNA for each taxon.

Taxon	Intron 2	Exon 3	Intron 3	Exon 4	Intron 4	Exon 5	Intron 5	Exon 6	Intron 6	Exon 7	Intron 7
Arabidopsis thaliana	95	65	420	100	139	40	168	43	708	140	136
Brassica napus FULa.1	224	65	549	100	127	40	83	43	893	140	113
Brassica napus FULa.2	82	65	423	100	78	40	83	43	443	140	110
Brassica napus FULc	112	65	338	100	129	40	130	43	876	149	114
Brassica napus FULd	82	65	423	100	80	40	69	43	324	140	113
Brassica nigra FULd	78	65	429	100	86	40	69	43	766	140	103
Brassica oleracea var. *acephala FULa*	135	65	518	100	127	40	83	43	444	140	110
Brassica oleracea var. *acephala FULc*	112	65	330	100	129	40	130	43	809	149	104
Brassica spinescens FULd	67	65	636	100	79	40	70	43	1290	140	97
Cakile lanceolata subsp. *fusiformis FULd*	198	65	575	100	77	40	80	43	649	140	150
Cakile lanceolata subsp. *fusiformis FULc*	111	65	370	100	110	40	122	43	919	140	119
Coincya monensis FULc	112	65	341	100	116	40	156	43	705	140	117
Cordylocarpus muricatus FULc	108	65	328	100	119	40	162	43	675	140	107
Cordylocarpus muricatus FULa	118	65	382	100	127	40	84	43	1354	140	114
Crambe orientalis FULc	118	65	607	100	146	40	79	43	1036	140	116
Crambe orientalis FULa	110	65	466	100	92	40	84	43	949	140	111
Crambella teretifolia FULd	85	65	582	100	107	40	82	43	635	140	142
Crambella teretifolia FULa	122	65	467	100	123	40	84	43	936	140	112
Crambella teretifolia FULc	114	65	402	100	131	40	81	43	814	140	97
Didesmus bipinnatus FULa	105	65	467	100	129	40	77	43	731	140	115
Didesmus bipinnatus FULd	84	65	571	100	79	40	115	43	620	140	126
Diplotaxis assurgens FULd	77	65	447	100	80	40	69	43	925	140	113
Enarthrocarpus lyratus FULa	115	65	448	100	97	40	82	43	1065	140	
Enarthrocarpus lyratus FULc	110	65	335	100	122	40	141	43	1130	140	127
Enarthrocarpus lyratus FULd	77	65	495	100	81	40	69	43	741	140	115
Eruca vesicaria subsp *sativa FULa*	117	65	565	100	121	40	84	43	1099	140	105
Erucaria erucarioides FULc	125	65	687	100	132	40	81	43	812	140	91
Erucaria erucarioides FULd	80	65	632	100	80	40	77	43	598	140	108
Erucastrum gallicum FULd	76	65	404	100	79	40	68	43	1242	140	104
Guiraoa arvensis FULc	151	65	327	100	111	40	125	43	1360	140	115
Hemicrambe fruticulosa FULc	109	65	272	100	124	40	147	43	808	140	118
Henophyton deserti FULc	108	65	332	100	132	40	84	43	734	140	111
Henophyton deserti FULb	81	65	305	100	129	40	74	43	617	140	111
Hirschfeldia incana FULa	121	65	728	100	133	40	81	43	347	140	112
Hirschfeldia incana FULc	111	65	333	100	125	40	144	43	754	140	97
Moricandia arvensis FULc	113	65	288	100	97	40	82	43	683	140	115

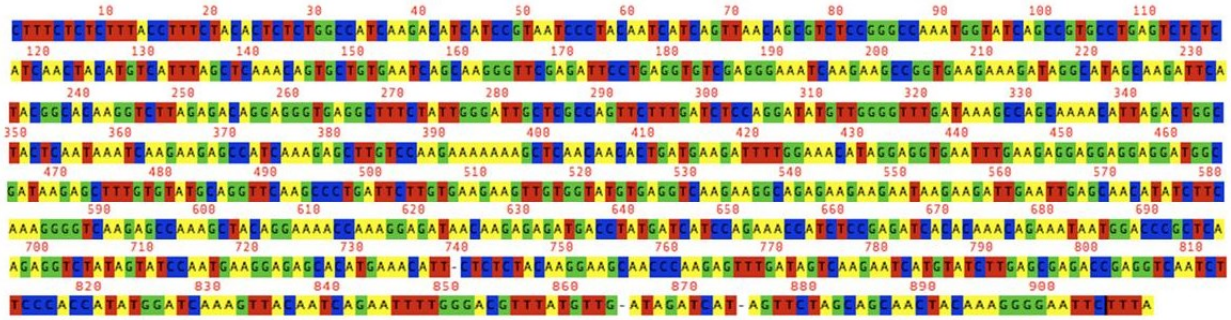
Moricandia arvensis FULb	107	65	346	100	130	40	94	43	598	140	105
Muricaria prostrata FULd	88	65	502	100	121	40	87	43	719	140	114
Muricaria prostrata FULb	103	65	90	100	128	40	82	43	703	140	131
Muricaria prostrata FULc	137	65	450	100	115	40	84	43	736	140	101
Pseuderucaria teretifolia FULb	82	65	286	100	132	40	80	43	137	140	106
Pseuderucaria teretifolia FULc	110	65	332	100	135	40	77	43	905	140	119
Psychine stylosa FULc	105	65	452	100	113	40	91	43	721	140	157
Raffenaldia primuloides FULb	78	65	470	100	79	40	70	43	1160	140	92
Raffenaldia primuloides FULc	108	65	328	100	82	40	27	43		140	
Rytidocarpus moricandiodes FULc	93	65	202	100	131	40	84	43	895	140	108
Rytidocarpus moricandiodes FULb	95	65	209	100	129	40	96	43	598	140	105
Schouwia thebaica FULc	106	65	420	100	118	40	76	43	672	140	91
Sinapidendron angustifolium FULd	77	65	449	100	79	40	69	43	1351	140	112
Vella spinosa FULc	112	65	398	100	131	40	35	43	611	140	115
Zilla spinosa FULb	111	65	669	100	127	40	82	43	652	140	149
Zilla spinosa FULc	113	65	605	100	117	40	76	43	628	140	91

Appendix 3-1. List of taxa where contamination prevented inclusion in the current phylogeny alongside original reasons for their retrieval.

Taxon name	Hall Lab DNA extraction #	Potential utility in *TCP1* phylogeny
Cardamine hirsuta L.	#230 – February 13, 2003	Supplement Brassicaceae sampling; compare copy # with Busch et al. 2012.
Cleome diandra (Burch.) T. Durand & Schinz	#362 – May 13, 2010	Necessary to represent the *"Angustifolia"* Cleomaceae lineage.
Tarenaya hassleriana (Chodat) Iltis	#354 – July 8, 2008	Useful to verify copy number retrieval methods with Genome.
Physostemon hemsleyanus (Bullock) R.C.Foster	#409 – June 9, 2010	Necessary to represent the *"Dactylaena"* Cleomaceae lineage.
Peritoma serrulata (Pursh) DC.	#448 – September 8, 2010	One of only two Cleomaceae representatives native to Alberta; Necessary to represent the North American Cleomoids
Cleome viridiflora Schreb.	#360 – December 4, 2008	Contamination and non-target amplicons prevented sufficient sampling of this taxa; needed to represent the *"Melidiscus" Cleomaceae* lineage
Arivela viscosa (L.) Raf.	#348 – July 8, 2008	Necessary to represent the Australian Cleomaceae lineage
Cleomella longipes Torr.	#464 – September 4, 2010	Necessary to represent the North American Cleomoids
Dactylaena micrantha Schrad. ex Schult. & Schult.f.	#291 – June 16, 2007	Necessary to represent the *"Dactylaena"* Cleomaceae lineage; only exceptional taxon displaying early monosymmetry despite belonging to the early dissymmetry clade (Patchell et al., 2011); the only Cleomaceae genus to have one stamen
Gynandropsis gynandra (L.) Briq.	#419 – June 22, 2010	Necessary to represent the "*Gynandropsis*" Cleomaceae lineage
Podandrogyne macrophylla (Turcz.) Woodson	#292 – June 16, 2007	Necessary to represent the Andean Cleomaceae lineage
Capparis flexuosa (L.) J. Presl	#346 – May 15, 2008	Outgoup sequence from Capparaceae

Appendix 3-2. Summary of attempts to find and exclude source of contamination.

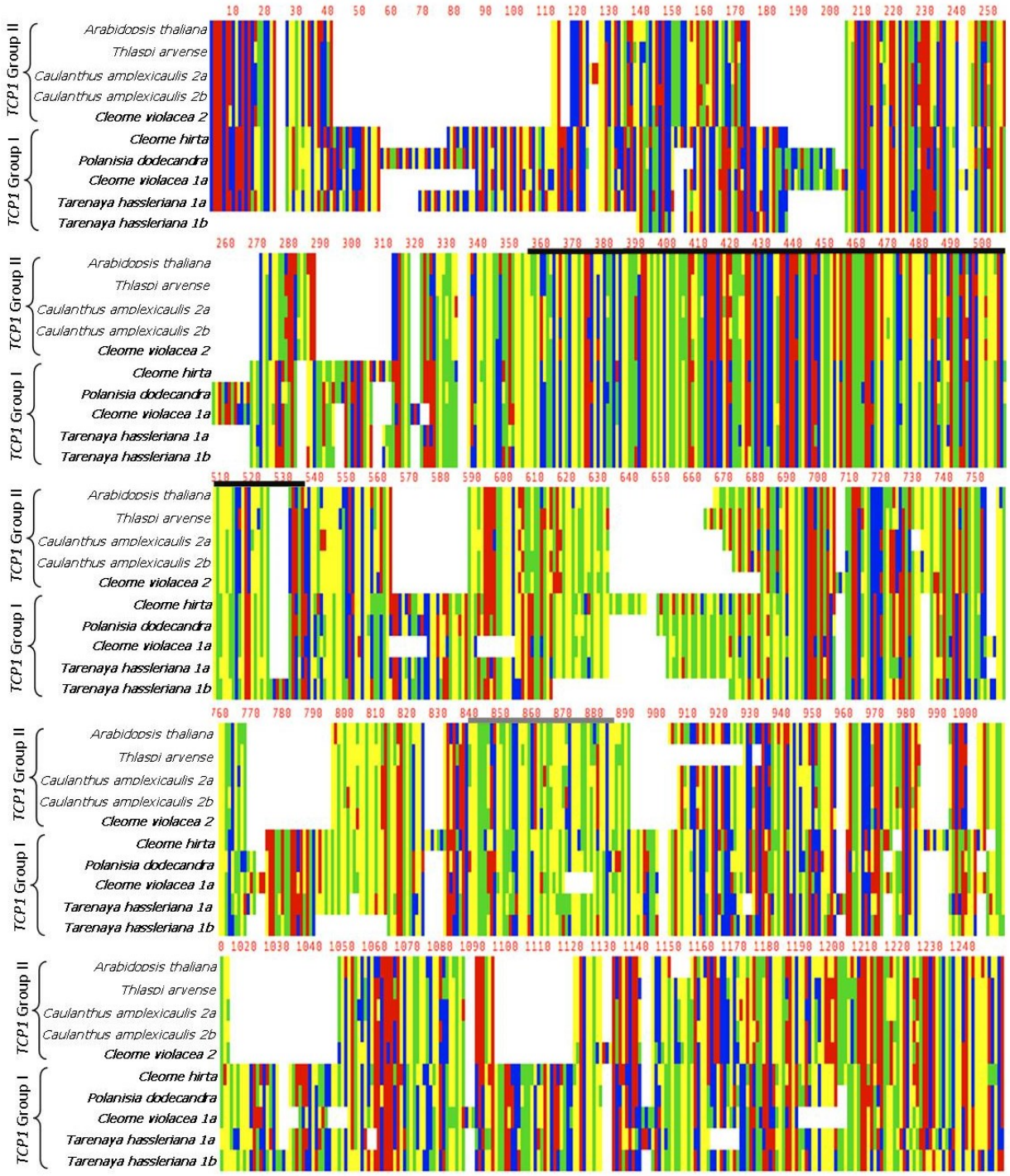
	List of anti-contamination efforts in order of attempt
1)	Used new autoclaved/milliQ water and newly diluted primers and dNTPs
2)	Used newly ordered stock primers and dNTPs
3)	Bleached all equipment and surrounding surfaces
4)	Used fresh genomic DNA extractions (extracted one at a time)
5)	Bleached equipment and surfaces, switched work area to be further away from amplification and gel-running area, conducted PCRs one at a time
6)	Had two additional trained people attempt amplification and sequencing



Appendix 3-3. Nucleotide sequence of suspected *TCP1* contaminant.

Appendix 3-4. PCR screening effort, copy number and sequencing of TCP1 in this study.

Species	Primers	clones PCR screened	# of clones sequenced	# of clones included	# of Group I copies	# of Group II copies
Aethionema grandiflorum	TCP1-fwd 5', TCP1-rev 5', TCP1 F degen.2, TCP1 R degen.2	48	16	2	0	1
Arabis lyrata	TCP1 F degen.2, TCP1 R degen.2	24	8	3	0	1
Brassica napus	TCP1 F degen.2, TCP1 R degen.2	24	8	7	0	2
Cakile lanceolata	TCP1-fwd 5', TCP1-rev 5', TCP1 F degen.2, TCP1 R degen.2	24	8	2	0	1
Caulanthus amplexicaulis	TCP1 F degen.2, TCP1 R degen.2	48	16	8	0	2
Cleome africana	TCP1-fwd 5', TCP1-rev 5', TCP1 F degen.2, TCP1 R degen.2	48	16	3	1	0
Cleome ambylocarpa	TCP1 F degen.2, TCP1 R degen.2	24	8	3	1	0
Cleome arabica	TCP1 F degen.2, TCP1 R degen.2	48	16	3	1	0
Cleome brachycarpa	TCP1 F degen.2, TCP1 R degen.2	24	8	6	1	0
Cleome droserifolia	TCP1 F degen.2, TCP1 R degen.2	48	16	2	1	0
Cleome hirta	TCP1-fwd 5', TCP1-rev 5', TCP1 F degen.2, TCP1 R degen.2	48	16	11	1	0
Cleome viridiflora	TCP1 F degen.2, TCP1 R degen.2	72	24	1	0	1
Cleome violacea	TCP1-fwd 5', TCP1-rev 5', TCP1 F degen.2, TCP1 R degen.2	48	16	7	1	1
Cleome spinosa	TCP1-fwd 5', TCP1-rev 5', TCP1 F degen.2, TCP1 R degen.2	48	16	7	1	1
Erucaria erucarioides	TCP1 F degen.2, TCP1 R degen.2	24	8	4	0	2
Iberis amara	TCP1 F degen.2, TCP1 R degen.2	24	8	4	0	1
Myagrum perfoliatum	TCP1 F degen.2, TCP1 R degen.2	24	8	4	0	1
Polanisia dodecandra	TCP1 F degen.2, TCP1 R degen.2	48	16	3	1	0
Stanleya pinnata	TCP1-fwd 5', TCP1-rev 5'	24	8	7	0	1
Thlaspi arvense	TCP1-fwd 5', TCP1-rev 5', TCP1 F degen.2, TCP1 R degen.2	48	16	10	0	1



Appendix 3-5. Alignment of 10 representative sequences from Group I and Group II. The TCP domain is indicated by a black line and the R domain is represented by the grey line. Cleomaceae taxa are bolded.