Evolution of the interaction of floral homeotic proteins

Dissertation

Zur Erlangung des akademischen Grades
‘doctor rerum naturalium‘ (Dr. rer. nat.)

Vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät
der Friedrich-Schiller-Universität Jena

von Diplombiologe Florian Rümpler
geboren am 04. Dezember 1986 in Erfurt

Gutachter

1. Prof. Dr. Günter Theißen (Jena)
2. Asst. Prof. Dr. Rainer Melzer (Dublin)
3. Prof. Dr. Richard Immink (Wageningen)

Tag der öffentlichen Verteidigung

Mittwoch 15. November 2017

Table of contents

1 Introduction

Developmental processes are controlled by the dynamic interplay between transcription factors and their target genes. Thereby transcription factors often bind to DNA not as monomers but as higher order homo- and heteromeric complexes. The activation or repression of target genes therefore highly depends on the protein-protein interactions of the corresponding transcription factor. A good case in point are floral homeotic MADS-domain transcription factors (MADS-TFs) that bind as heterotetramers to *cis*-regulatory DNA elements of target genes to control floral organ identity determination of angiosperms (Theißen and Saedler, 2001). It is presumed that the composition of the heterotetramer determines the set of target genes and eventually defines the identity of the developing floral organ. The protein-protein interactions of floral homeotic MADS-TFs and the structure and evolution of the protein-protein interaction network (PPI network) controlling flower development are thus of great scientific interest. With this thesis I aim to deepen our understanding of the molecular mechanisms underlying the protein-protein interactions of floral homeotic MADS-TFs and to give insights into how certain interaction patterns of floral homeotic proteins changed in the course of angiosperm evolution.

1.1 MADS-domain transcription factors in plants - a K-domain blessing

MADS-box genes are a synapomorphy of eukaryotes and encode for MADS-TFs that control a variety of different developmental processes of animals, plants and fungi (Messenguy and Dubois, 2003). Whereas only few MADS-box genes are present in animals and fungi (Messenguy and Dubois, 2003), their number dramatically increased during plant evolution, with more than 100 gene family members being present in *Arabidopsis thaliana* (Parenicova et al., 2003; Gramzow and Theißen, 2010). All MADS-TFs share a highly conserved DNA-binding MADS-domain, with ‘MADS’ being an acronym for the four founding members of this protein family: MINICHROMOSOME MAINTENANCE FACTOR 1, AGAMOUS, DEFICIENS, and SERUM RESPONSE FACTOR (Schwarz-Sommer et al., 1990). An ancestral gene duplication in the stem group of extant eukaryotes gave rise to two clades of MADS-box genes, termed Type I and Type II genes (Alvarez-Buylla et al., 2000; Gramzow et al., 2010). Whereas relatively little is known about function and evolution of Type I MADS-box genes in plants, Type II genes constitute key regulators of many developmental processes in angiosperms and they are therefore

one of the best studied gene families within plants (Kaufmann et al., 2005a; Gramzow and Theißen, 2010; Smaczniak et al., 2012a; Gramzow and Theißen, 2013).

Plant Type II MADS-TFs are characterized by a conserved domain structure comprising the highly conserved MADS-domain (M), followed by an intervening (I), a keratin-like (K) and a C-terminal (C) domain (Theißen et al., 1996; Becker and Theißen, 2003; Kaufmann et al., 2005a). This conserved architecture is unique to plant Type II MADS-TFs and due to this domain structure plant Type II MADS-TFs are also referred to as MIKC-type proteins. Whereas MADS- and I-domain function in the formation of DNA-bound dimers, the K-domain enables (at least many) MIKC-type proteins to also tetramerize among each other (Melzer and Theißen, 2009; Melzer et al., 2009; Smaczniak et al., 2012b; Puranik et al., 2014). It is presumed that the emergence of the K-domain was a key event in the evolution of plant Type II MADS-TFs and that the conjunction of a DNA-binding MADS-domain with a protein-protein interaction domain was of importance for success of this transcription factor family (Kaufmann et al., 2005a; Manuscript IV: Theißen et al., 2016; Manuscript V: Rümpler et al., 2017).

MIKC-type genes are found in charophytes (freshwater green algae) but not in chlorophytes (Tanabe et al., 2005; Derelle et al., 2006). Thus it appears most likely that the recruitment of the K-domain took place in the stem group of extant streptophytes (charophytes and land plants) (Fig. 1). During streptophyte evolution MIKC-type genes further diverged into two groups: $MIKC^C$ and MIKC* (Fig. 1) (Henschel et al., 2002; Kaufmann et al., 2005a; Gramzow and Theißen, 2010). In angiosperms $MIKC^C$-type genes predominantly function in sporophyte development, whereas the expression of MIKC*-type genes is mainly restricted to the male gametophyte (Zobell et al., 2010; Kwantes et al., 2012; Smaczniak et al., 2012a). Structurally $MIKC^C$- and MIKC*-type genes differ in length and number of the exons that encode for the K-domain (Henschel et al., 2002; Kwantes et al., 2012). It is not entirely clear as to exactly when the duplication of an ancestral MIKC-type gene gave rise to $MIKC^C$ and MIKC*. However, sequence features of MIKC-type gene from charophyte algae suggest that they represent direct descendants of a MIKC-type gene that was ancestral to both, $MIKC^C$- and MIKC*-type genes (Tanabe et al., 2005; Kwantes et al., 2012).

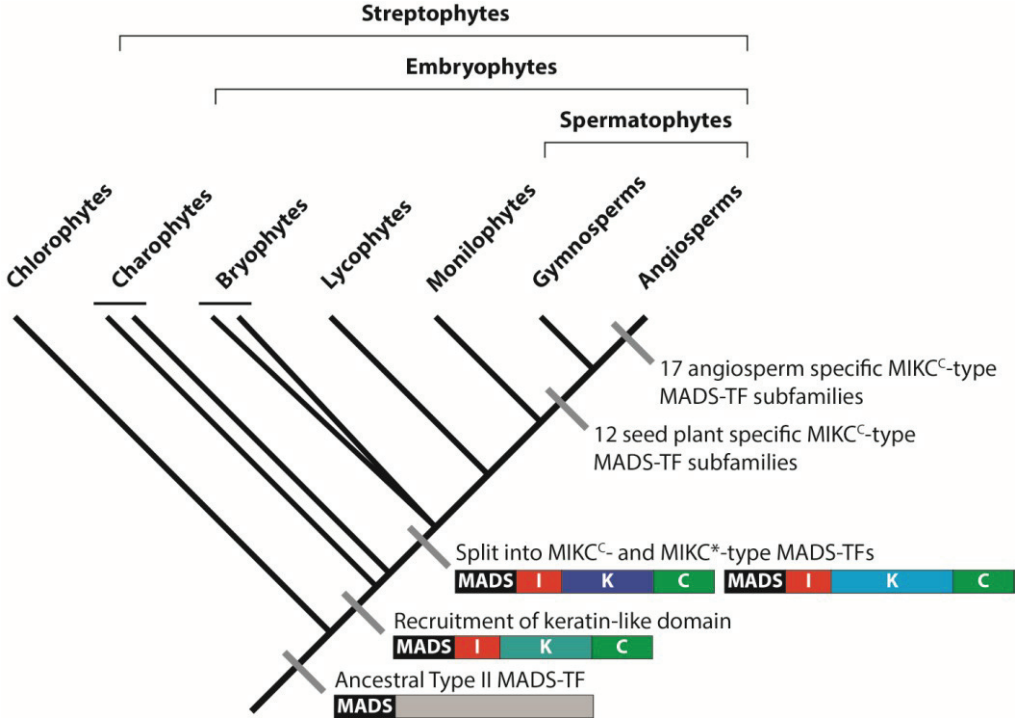


Figure 1: The evolution of plant Type II MADS-TFs. It is presumed that the K-domain coding K-box joined an ancestral Type II MADS-box gene in the stem group of extant streptophytes giving rise to an ancestral MIKC-type gene. Following a gene duplication in the MRCA of extant land plants, plant Type II MADS-box genes diverged into MIKC^C- and MIKC*-type genes. During land plant evolution the family of MIKC^C-type genes considerably expanded leading to 12 seed plant specific subfamilies and 17 angiosperm specific subfamilies of MIKC^C-type genes.

During land plant evolution the group of MIKC^C-type genes considerably expanded giving rise to 17 subfamilies that had already been established in the most recent common ancestor (MRCA) of extant angiosperms (Gramzow and Theißen, 2010; Gramzow and Theissen, 2015). In various angiosperm species members of the different subfamilies often have very similar or even identical functions (Smaczniak et al., 2012a). The best studied MIKC^C-type genes belong to the seven subfamilies of *APETALA1*- (*AP1*), *APETALA3*- (*AP3*), *PISTILLATA*- (*PI*), *AGAMOUS*- (*AG*), *SEEDSTICK*- (*STK*), *SEPALLATA1*- (*SEP1*), and *SEPALLATA3* (*SEP3*)-like genes which comprise key regulators for floral meristem and floral organ identity determination and all of which encode for floral homeotic proteins. In addition to the central role of some MIKC^C-type genes for flower development, members of less intensively studied subfamilies have been shown to fulfill diverse other functions ranging from root development and nutrient response to initiation of flowering and seed development (Gramzow and Theißen, 2010; Smaczniak et al., 2012a). Interestingly, analyses on whole genome data of 27 angiosperm species have shown that among

the 17 angiosperm-specific subfamilies of MIKC^C-type genes 15 have either never or only extremely rarely been completely lost in any of the examined species (Gramzow and Theissen, 2015). This suggests that also MIKC^C-type gene subfamilies of so far undetermined function may control important developmental or physiological processes in angiosperms.

1.2 MIKC^C-type MADS-TFs controlling flower development of angiosperms

Due to their conspicuous mutant phenotypes MADS-box genes that control the flower development of angiosperms were among the first MADS-box genes that have been identified (Sommer et al., 1990; Yanofsky et al., 1990). Based on the phenotypes of different floral homeotic mutants the involved genes were allocated to five classes of partially overlapping floral homeotic functions: A, B, C, D, and E (Coen and Meyerowitz, 1991; Krizek and Meyerowitz, 1996a; Theißen, 2001; Favaro et al., 2003). According to the ABCDE-model class A genes together with class E genes determine the identity of first whorl sepals; the combined function of A, B and E class genes controls the development of second whorl petals; B, C and E class genes together specify third whorl stamens; C together with E class genes determine fourth whorl carpels and the combination of C, D and E class genes controls the development of ovules inside the carpels (Fig. 2 bottom part). The genes that fulfill the different floral homeotic functions each belong to one of the different subfamilies of MIKC^C-type genes. A-function is exerted by *AP1*-like genes; *AP3*- and *PI*-like genes together fulfill B-function; C-function is realized by *AG*-like genes; *STK*-like genes exert D-function and the subfamilies of *SEP1*- and *SEP3*-like genes function as E-class genes (Fig. 2 bottom part) (reviewed by Theißen, 2001; O'Maoileidigh et al., 2014).

At the molecular level the combined activity of the different floral homeotic genes is realized by highly specific protein-protein interactions of the encoded MIKC^C-type MADS-TFs. The floral quartet model (FQM) proposes that the floral homeotic proteins interact to form DNA-bound floral organ specific heterotetrameric complexes, so called floral quartets (Fig. 2 top part) (Theißen, 2001; Theißen and Saedler, 2001; reviewed in Manuscript IV: Theißen et al., 2016). More precisely for *A. thaliana* it is presumed that a tetramer of two AP1 and two SEP proteins (i.e. an AAEE quartet) determines sepal development, an AP1/AP3/PI/SEP (ABBE) complex specifies petals, an AP3/PI/AG/SEP (BBCE) tetramer controls stamen identity, a complex of two AG proteins and two SEP proteins (CCEE) determines carpel development and a CDDE tetramer

composed of one AG protein, one SEP protein and two of the three STK-like subfamily members present in *A. thaliana* STK, SHATTERPROOF1 (SHP1) and SHP2 is assumed to control ovule development inside the carpels (Fig. 2 top part) (Theißen, 2001; Theißen and Saedler, 2001; Melzer et al., 2006; Melzer and Theißen, 2009; Smaczniak et al., 2012b). Within angiosperms the floral homeotic functions and the corresponding subfamilies of MIKC^{C}-type genes are highly conserved (Becker and Theißen, 2003; Gramzow and Theissen, 2015) and it is believed that floral organ identity determination is controlled by floral quartet-like complexes (FQCs) of floral homeotic MIKC^{C}-type proteins throughout angiosperms.

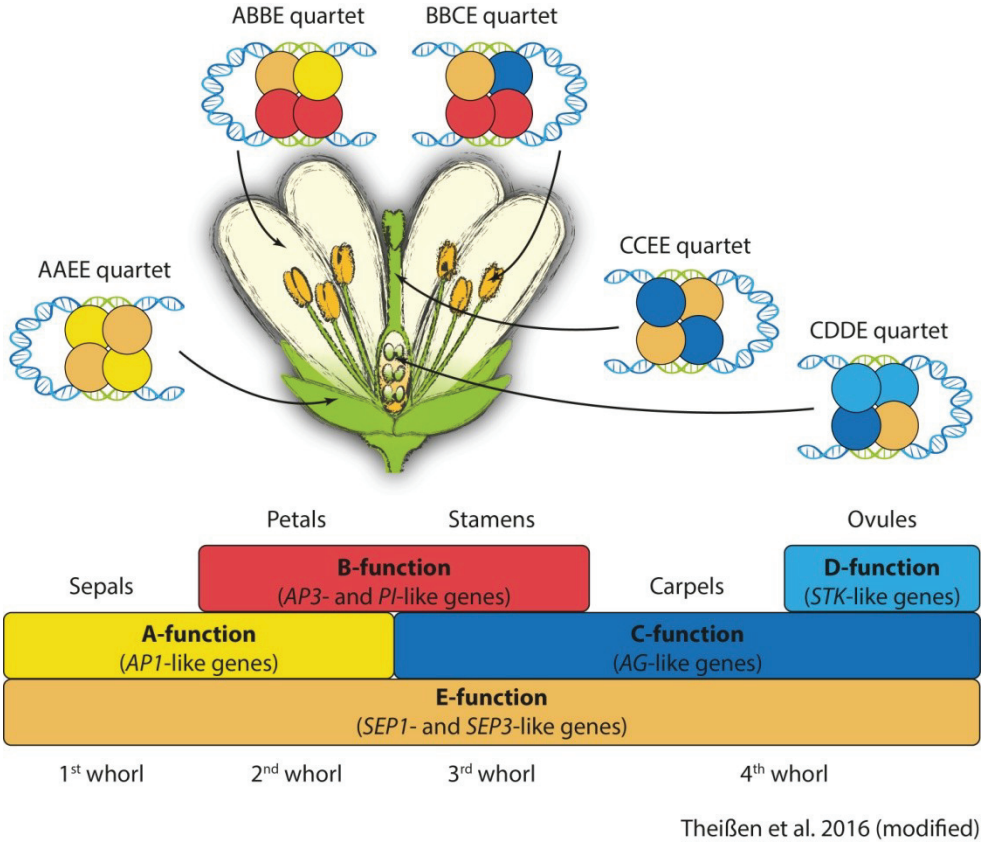


Figure 2: The ABCDE- and the floral quartet model of floral organ identity determination in angiosperms. According to the ABCDE-model (bottom part) the identity of each floral organ type (sepal, petal, stamen, carpel, and ovule) is controlled by 5 partially overlapping floral homeotic functions: A, B, C, D, and E. Each function is exerted by MIKC^{C}-type MADS-box genes belonging to seven different subfamilies: *AP1*-like genes (A-function), *AP3*- and *PI*-like genes (B-function), *AG*-like genes (C-function), *STK*-like genes (D-function), and *SEP1*- and *SEP3*-like genes (E-function). At the molecular level the combined activity of the floral homeotic functions is realized by the formation of floral organ specific DNA-bound heterotetramers of the encoded MIKC^{C}-type MADS-TFs (top part). Based on the generic floral quartet model shown here a tetramer of two A- and two E-function proteins (AAEE quartet) is assumed to control sepal development, an ABBE quartet specifies petals, a BBCE quartet determines stamens, a CCEE quartet specifies carpel identity and a CDDE quartet controls development of the ovules inside the carpel.

1.3 The protein-protein interaction network of floral homeotic proteins

As proposed by the FQM and meanwhile confirmed by numerous studies the functional unit of most if not all floral homeotic MIKC^C-type MADS-TFs is of tetrameric rather than dimeric state (Melzer and Theißen, 2009; Melzer et al., 2009; Smaczniak et al., 2012b; Mendes et al., 2013; Jetha et al., 2014; Ruelens et al., 2017). Furthermore recent *in silico* studies suggest that FQC-formation is not restricted to floral homeotic proteins but instead is a widespread property of MIKC^C-type proteins (Espinosa-Soto et al., 2014). For a detailed understanding of the combinatorial activity of MIKC^C-type proteins it is thus crucial to investigate the structure and evolution of the underlying PPI network. Although all MIKC^C-type proteins share the highly conserved protein-protein interacting K-domain their abilities to form DNA-bound homo- and heterodimers and -tetramers considerably differs. Most MIKC^C-type proteins display a quite restricted set of potential interaction partners whereas few members function as hubs that mediate interaction of numerous other MIKC^C-type proteins (de Folter et al., 2005; Immink et al., 2009; Al Hindi et al., 2017; Ruelens et al., 2017). This scale-free structure is characteristic for biological PPI networks as it makes them less vulnerable for the random removal of nodes (Barabasi and Oltvai, 2004; Kitano, 2004).

Within the PPI network controlling flower development of *A. thaliana* the E-class protein SEP3 constitutes a major hub as it incorporates other floral homeotic proteins into floral quartets that would otherwise not form (Favaro et al., 2003; Immink et al., 2009; Melzer and Theißen, 2009; Smaczniak et al., 2012b). However, the set of interaction partners of SEP3 is not restricted to floral homeotic proteins but also comprises MIKC^C-type proteins belonging to other subfamilies such as SUPPRESSOR OF OVEREXPRESSION OF CONSTANTS 1 (SOC1), SHORT VEGETATIVE PHASE (SVP), AGAMOUS-LIKE 6 (AGL6) and AGL24 which are involved in the initiation of flowering (Immink et al., 2009; Liu et al., 2009). Furthermore SEP3 has been shown to also mediate the interaction of ARABIDOPSIS B-SISTER (ABS) and STK that play an important role in seed development and fertilization (Kaufmann et al., 2005b; Mizzotti et al., 2012). The promiscuous interactions of SEP3 suggest that it fulfills a hub function in the PPI network of MIKC^C-type proteins also beyond floral organ identity determination. Interestingly SEP3 acts in a mostly redundant manner with the other E-class proteins SEP1, SEP2, and SEP4 (Pelaz et al., 2000; Ditta et al., 2004) which are also phylogenetically closely related to SEP3 (Zahn et al., 2005b). Although the other SEP proteins exhibit a much narrower set of interaction

partners compared to SEP3 (Immink et al., 2009), the functional redundancy suggests that the hub function of SEP3 can at least partially be undertaken by the other E-class proteins.

In contrast to the numerous interactions of SEP3 the B-class proteins AP3 and PI represent non-hubs in the PPI network controlling flower development of *A. thaliana* as they only form obligate heterodimers (Riechmann et al., 1996b; Immink et al., 2009) and require SEP proteins to be incorporated into tetrameric complexes (Melzer and Theißen, 2009; Smaczniak et al., 2012b). *AP3*- and *PI*-like genes are phylogenetically closely related and most likely originated by a duplication of an ancestral *AP3/PI*-like gene in the stem group of extant angiosperms (Winter et al., 2002a). Interestingly, orthologs of AP3/PI-like proteins from gymnosperms have been shown to homodimerize leading to the hypothesis that the obligate heterodimerization of AP3- and PI-like proteins originated from homodimerization (Winter et al., 2002b). It was proposed that the obligate heterodimerization of AP3- and PI-like proteins enhanced the developmental robustness during floral organ identity determination and thereby fostered the canalization of flower development during angiosperm evolution (Lenser et al., 2009). However it remained unknown whether the obligate heterodimerization of AP3- and PI-like proteins is an ancient feature that was already present in the MRCA of extant angiosperms or if it evolved during angiosperm evolution (addressed in Manuscript I: Melzer et al., 2014).

1.4 The keratin-like domain of MIKC^C-type proteins

Although for most MIKC^C-type proteins MADS- and I-domain together are sufficient for the formation of DNA-bound dimers (Huang et al., 1996), numerous studies illustrated that the K-domain makes important contributions to dimerization and that it is essential for tetramerization of MIKC^C-type proteins (Davies et al., 1996; Riechmann et al., 1996b; Fan et al., 1997; Moon et al., 1999; Yang et al., 2003b; Yang et al., 2003a; Yang and Jack, 2004; Immink et al., 2009; Melzer et al., 2009; van Dijk et al., 2010; Puranik et al., 2014; Bartlett et al., 2016; Silva et al., 2016). The amino acid sequence within the K-domain of most MIKC^C-type proteins shows three patterns of regularly spaced hydrophobic and charged residues based on which the K-domain was subdivided into three subdomains: K1-, K2- and K3-subdomain (Ma et al., 1991; Riechmann et al., 1996b) (Fig. 3A). The so called heptad repeat pattern of the form $[abcdefg]_n$ with hydrophobic residues at ‘a’ and ‘d’ positions and charged residues at ‘e’ and ‘g’ positions is characteristic for a widespread and intensively studies class of protein-protein interaction

domains termed coiled-coil (reviewed by Mason and Arndt, 2004; Lupas and Gruber, 2005). In a coiled-coil the amino acid α-helix itself is bent into a helical conformation in a way that the hydrophobic residues on heptad repeat 'a' and 'd' positions form a stripe that runs along the helix and facilitates hydrophobic interactions with a partner coiled-coil (Fig. 3B). The hydrophobic stripe is flanked by charged residues on heptad repeat 'e' and 'g' positions that allow for intermolecular electrostatic interactions (Mason et al., 2009). Coiled-coil interactions have been studied intensively in recent decades and much is known about energetic contributions of different amino acids on heptad repeat 'a', 'd', 'e', and 'g' positions (Zhu et al., 1993; Moitra et al., 1997; Mason et al., 2009; Azuma et al., 2014; Kükenshöner et al., 2014). Furthermore, computational models have been developed to predict coiled-coil interactions (Fong et al., 2004; Grigoryan and Keating, 2006; Potapov et al., 2015). However, due to a complex 'knobs-into-holes' side chain packing between interacting coiled-coils we do not yet fully understand the contribution of individual amino acid parings to the overall interaction strength and interaction specificity. Considering the abundance of data on attractive and repulsive forces that facilitate or impede the interaction of coiled-coil proteins, relatively little of this knowledge has yet been used to determine sequence determinants of the K-domain that are critical for dimerization and tetramerization of MIKC^{C}-type proteins (addressed in Manuscript V: Rümpler et al., 2017). Nevertheless, interaction studies with single amino acid substitution mutants of AP3, PI, SEP1 and SEP3 suggest that the physicochemical laws that underlie coiled-coil interactions are also applicable for the K-domain (Yang et al., 2003b; Yang et al., 2003a; Yang and Jack, 2004; Puranik et al., 2014; Silva et al., 2016).

For a long time all assumptions concerning the structure of the K-domain were based on sequence similarities to known coiled-coil proteins. However, in 2014 Puranik et al. determined the X-ray crystallographic structure of a K-domain tetramer of SEP3, according to which the K-domain forms two amphipathic α-helices separated by a kink region that prevents intramolecular association of both helices (Puranik et al., 2014). The first helix comprises the K1-subdomain heptad repeat and constitutes an interaction interface for dimerization of two SEP3 monomers. The second helix encompasses the K2-subdomain that further strengthens the interaction of two SEP3 monomers and the K3-subdomain that mediates interaction of two SEP3 dimers and thus facilitates tetramerization (Fig. 3C).

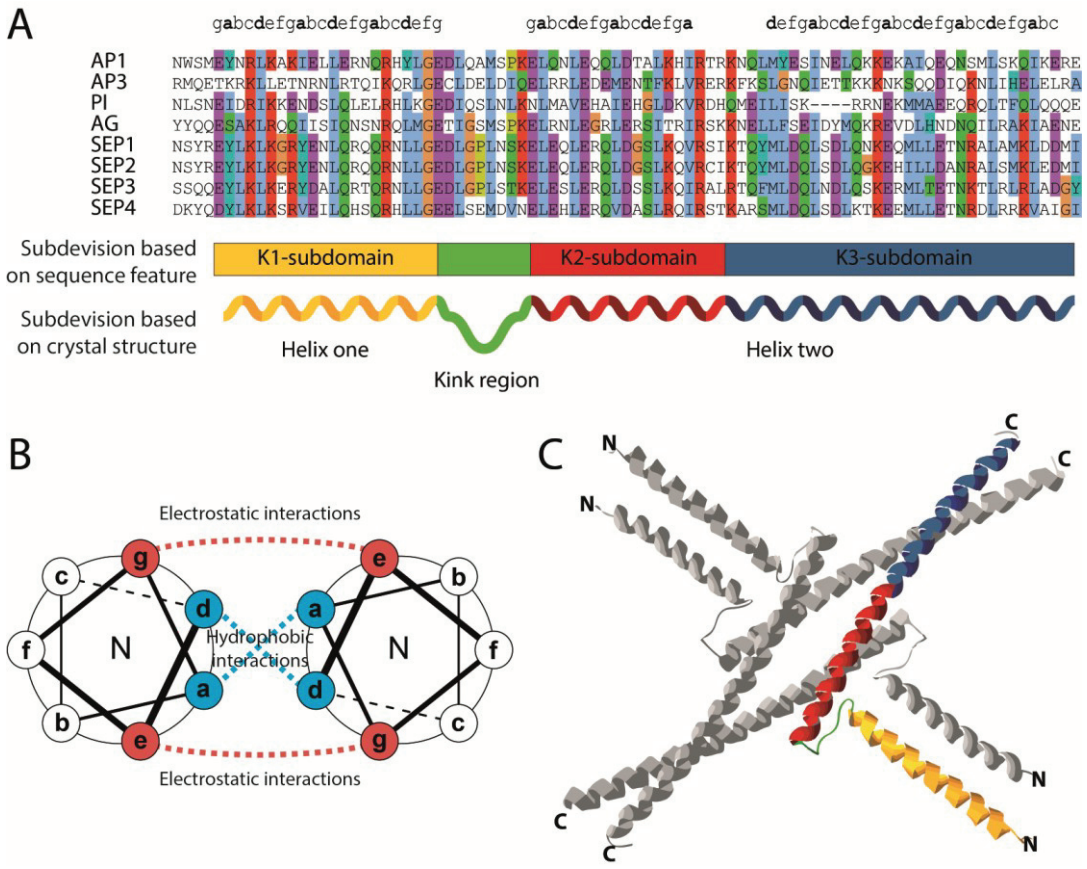


Figure 3: Sequence and structural features of the K-domain of MIKC^C-type MADS-TFs. (A) Aligned amino acid sequences of the K-domain of AP1, AP3, PI, AG, SEP1, SEP2, SEP3 and SEP4. The amino acid sequences follow a characteristic heptad repeat pattern of the form $[abcdefg]_n$ with hydrophobic residues at 'a' and 'd' positions and charged residues at 'e' and 'g' positions. This amino acid pattern is characteristic for a class of protein-protein interaction domains termed coiled-coil. Based on the presence of three separate heptad repeats the K-domain was subdivided into three K-subdomains: K1 (yellow), K2 (red) and K3 (blue). According to the crystal structure of the K-domain of SEP3 the K1-subdomain is located within the first K-domain α-helix and helix two spans K2- as well as K3-subdomain. (B) If an amino acid strand that follows a heptad repeat structure is wound up to a helix the hydrophobic residues at heptad repeat 'a' and 'd' positions are directed to the same site of the helix and facilitate hydrophobic interactions with a partner coiled-coil. In addition charged residues at heptad repeat 'e' and 'g' positions mediate attractive or repulsive electrostatic interactions. (C) Crystal structure of a K-domain homotetramer of SEP3 (Puranik et al., 2014). The color coding of the subdomains follows the color coding of panel A.

1.5 Floral homeotic proteins as targets of plant pathogen effectors

As key regulators of various plant developmental processes MADS-TFs also constitute targets of plant pathogens that aim to alter host development. A remarkable example are plant pathogenic bacteria termed phytoplasmas that are able to influence the development of their host plants in a number of impressive ways (reviewed by Lee et al., 2000; Christensen et al., 2005; Hogenhout et al., 2008). Characteristic symptoms of a phytoplasma infection include the clustering of branches

termed witches' broom, the green coloration of non-green floral organs (virescence), decline and stunting of plants and the development of leaf-like structures instead of floral organs known as phyllody (Lee et al., 2000; Maejima et al., 2014a; Marcone, 2014). Phytoplasmas are pathogens of numerous crop plants and the developmental alterations that come along with an infection cause devastating yield losses all over the world. In recent years an increasing number of studies investigated the molecular mechanisms that underlie different symptoms of a phytoplasma infection and it became apparent that at least many of these symptoms are induced by effector proteins secreted by the bacteria (Bai et al., 2009; Hogenhout et al., 2009; Sugio et al., 2011; MacLean et al., 2014; Sugio et al., 2014).

One especially fascinating mechanism disclosed how phytoplasmas induce phyllody. The phytoplasma effector proteins SAP54 (for **S**ECRETED **A**STER-YELLOWS WITCHES-BROOM **P**ROTEIN 54) was shown to bind to MIKC^C-type MADS-TFs and to destine them for degradation via the ubiquitin/26S proteasome pathway (Fig. 4A) (MacLean et al., 2014; Maejima et al., 2014b). SAP54 thereby specifically targets MIKC^C-type proteins of certain subfamilies, among others comprising the floral homeotic E-class proteins SEP1, SEP2, SEP3 and SEP4 (SEP-subfamily) and the A-class protein AP1 (SQUA/AP1-subfamily) (Fig. 4B) (MacLean et al., 2014). The depletion of the targeted transcription factors causes defects in floral meristem and floral organ identity determination that resemble knock-out phenotypes of the respective A- and E-function genes (Bowman et al., 1993; Pelaz et al., 2000; Ditta et al., 2004). Whether the altered phenotype of the host plant is beneficial for the bacteria or if the phenotypic alterations are an ancillary effect of another SAP54 mediated mechanism is not entirely clear yet. Phytoplasma infected plants as well as plants overexpressing *SAP54* have been shown to be more attractive to insect vectors that feed on the host plant and it was thus hypothesized that the bacteria induce leaf-like flowers as a means of increasing vegetative biomass to attract insect vectors (MacLean et al., 2014). However, in a recent study Orlovskis and Hogenhout (2016) could demonstrate that SAP54 mediates insect vector attraction independently of the altered floral phenotypes assuming another so far unexplored function of SAP54 (Orlovskis and Hogenhout, 2016).

Remarkably SAP54 was shown to specifically interact with the K-domain of the targeted floral homeotic proteins (Fig. 4C) (MacLean et al., 2014). The high specificity of the interaction and hints for amino acid sequence similarities between SAP54 and the K-domain thus led to the

hypothesis that the interaction between SAP54 and floral homeotic proteins is probably mediated by a mechanism that is similar to that mediating higher order complex formation among floral homeotic proteins (addressed in Manuscript III: Rümpler et al., 2015b).

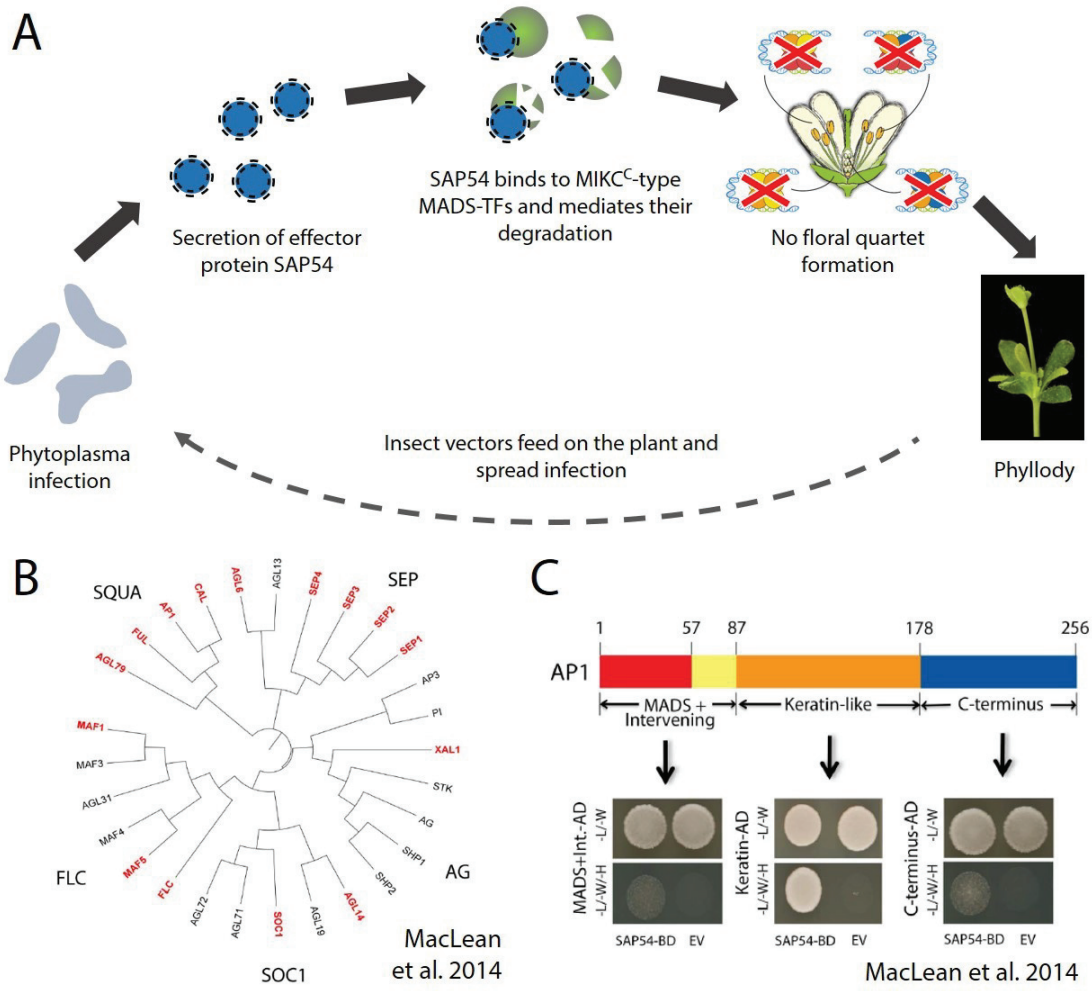


Figure 4: The interaction between SAP54 and MIKC^C-type MADS-TFs. (A) Following a phytoplasma infection the phytopathogenic bacteria secrete an effector protein termed SAP54 that specifically binds to MIKC^C-type MADS-TFs of certain subfamilies including floral homeotic A- and E-class proteins. After being bound by SAP54 the targeted transcription factors are destined for degradation via the ubiquitin/26S proteasome pathway. Depletion of the floral homeotic A- and E-class proteins prevents the formation of floral quartets and eventually causes the development of leaf-like structures instead of floral organs, a phytoplasma infection symptom known as phyllody. (B) In addition to floral homeotic A- and E-class proteins (i.e. SQUA/AP1-like and SEP-like proteins) SAP54 also targets MIKC^C-type MADS-TFs belonging to the subfamilies of FLC-, SOC1- and AG-like proteins (figure taken from MacLean et al., 2014). (C) Yeast-two-hybrid experiments with truncated MIKC^C-type MADS-TFs have shown that SAP54 specifically binds the K-domain of the targeted proteins (figure taken from MacLean et al., 2014).

1.6 Aims of this thesis

This thesis aims to deepen our understanding of the molecular mechanisms underlying the protein-protein interactions of floral homeotic proteins and to reveal how certain interaction patterns changed during angiosperm evolution. I focused my investigations on AP3-, PI- and SEP3-like MADS-TFs that play contrary roles within the PPI network controlling flower development, with SEP3-like proteins serving as major hubs and AP3- and PI-like proteins possessing highly restricted interaction capabilities. In the second part of this thesis I aim to provide insights into how the interaction between the phytoplasma effector protein SAP54 and the K-domain of floral homeotic proteins is realized at the molecular level and how this interaction may evolved.

In particular I answer the following questions:

- Had the protein-protein interactions governing flower development in core eudicots already been established at the base of extant angiosperms (Manuscript I)?
- When and how did the dimerization behavior of AP3- and PI-like proteins change during angiosperm evolution (Manuscript I)?
- Which amino acid residues within the K-domain are critical for the tetramerization of SEP3 and how conserved are these positions among SEP3-like proteins (Manuscript V)?
- Does the K-domain of other MIKC^C-type proteins fold into a structure similar to that determined for SEP3 and which amino acid preferences do non-hubs such as AP3- and PI-like proteins show at the amino acid sites that are critical for tetramerization of SEP3 (Manuscript V)?
- Does the phytoplasma effector protein SAP54 mimic the structure of the K-domain of MIKC^C-type MADS-TFs (Manuscript III)?
- Are the structural similarities between SAP54 and the K-domain of MIKC^C-type MADS-TFs a result of horizontal gene transfer or did both proteins evolve convergently (Manuscript III)?

In addition I describe the process of character state reconstruction as method to infer ancestral protein-protein interactions (Manuscript II) and discuss, based on current knowledge, when the tetramerization of MIKC^C-type MADS-TFs and thus the formation of FQCs may have originated during plant evolution (Manuscript IV).

2 Manuscripts

2.1 Manuscript overview

Manuscript I

Melzer, R., Härter, A., Rümpler, F., Kim, S., Soltis, P. S., Soltis, D. E., Theißen, G. (2014). DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. *Ann. Bot.* 114, 1431-1443.

In this manuscript we demonstrate that DEF- and GLO-like MADS-TFs from early diverging angiosperms display more diverse protein-protein interaction capabilities than their orthologs from core eudicots. We hypothesize that the more flexible interactions of DEF- and GLO-like proteins from early diverging angiosperms may account for the diverse flower morphology observed in these species. The reduction of interaction partners during angiosperm evolution probably fostered developmental robustness and thereby contributed to the canalization of flower development.

Author contributions:

Andrea Härter and Sangtae Kim amplified and cloned the cDNA sequences. Andrea Härter performed the EMSA and yeast-two-hybrid experiments. Rainer Melzer and Florian Rümpler compiled the collection of previously published interaction data. Florian Rümpler created the phylogenic trees, performed the ancestral character state reconstructions and prepared the manuscript figures. Rainer Melzer wrote the manuscript. Günter Theißen designed and supervised the project. Pamela Soltis, Douglas Soltis, Günter Theißen and Florian Rümpler contributed to improving the manuscript.

Overall contribution of Florian Rümpler: 20 %

I hereby certify the accuracy of the statements on the contributions of the authors.

Prof. Günter Theißen

Manuscript II

Rümpler, F., Theißen, G., Melzer, R. (2015). Character-state reconstruction to infer ancestral protein-protein interaction patterns. *Bio-protocol* 5, published online.

This manuscript constitutes an invited method protocol in which we provide a step by step description on how ancestral protein-protein interaction patterns can be inferred based on a set of known protein interactions. We provide links to suitable protein-protein interaction databases, programs for sequence alignments and phylogeny reconstructions and discuss difficulties and possible pitfalls during the process of ancestral character state reconstruction.

Author contributions:

Rainer Melzer and Florian Rümpler wrote the manuscript. Günter Theißen contributed to improving the manuscript.

Overall contribution of Florian Rümpler: 50 %

I hereby certify the accuracy of the statements on the contributions of the authors.

Prof. Günter Theißen

Manuscript III

Rümpler, F., Gramzow, L., Theißen, G., Melzer, R. (2015). Did convergent protein evolution enable phytoplasmas to generate 'zombie plants'? *Trends Plant Sci.* 20, 798-806.

In this opinion article we present preliminary evidence that the phytoplasma effector protein SAP54 is able to specifically target certain $MIKC^C$-type MADS-TFs because it folds into a structure similar to that of the K-domain. We hypothesize that SAP54 underwent convergent sequence and structural evolution to mimic the protein-protein interaction domain of its target proteins. Furthermore we discuss possible origins of SAP54 and outline the potential of SAP54-like proteins to serve as a molecular tool to study flower development in genetically intractable plant species.

Author contributions:

Florian Rümpler performed the structural predictions, sequence similarity analyses and remote homology searches. Lydia Gramzow introduced Florian Rümpler to remote homology search methods. Günter Theißen initiated the project, developed its major hypothesis and together with Rainer Melzer supervised the project. Rainer Melzer and Florian Rümpler wrote the manuscript. Günter Theißen and Lydia Gramzow contributed to improving the manuscript.

Overall contribution of Florian Rümpler: 65 %

I hereby certify the accuracy of the statements on the contributions of the authors.

Prof. Günter Theißen

Manuscript IV

Theißen, G., Melzer, R., Rümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development* 143, 3259-3271.

In this review article we give a short historical overview of the findings that led towards the development of the floral quartet model, review experimental evidences supporting it and give examples of how this model is used as basis for current research. Furthermore we discuss the direct linkage between K-domain emergence and the origin of floral quartet-like complexes. We point out that the emergence of the K-domain probably was an important preadaptation to the transition to land and that evolutionary changes of the tetramerization behavior of certain $MIKC^C$-type MADS-TFs may have played an essential role for the origin and evolution of the flower.

Author contributions:

Günter Theißen conceptualized the structure of the review and wrote most parts of the manuscript. Rainer Melzer wrote the text for text box 3 ('FQCs: beyond floral organ identity') and text box 4 ('Why quartets?'). Florian Rümpler wrote the paragraph 'On the origin of FQCs: a MIKC blessing' and prepared the manuscript figures. All authors contributed to improving the manuscript.

Overall contribution of Florian Rümpler: 20 %

I hereby certify the accuracy of the statements on the contributions of the authors.

Prof. Günter Theißen

Manuscript V

Rümpler, F., Theißen, G., Melzer, R. (2017). Sequence features of MADS-domain proteins that act as hubs in the protein-protein interaction network controlling flower development. *bioRxiv*,125294.

In this research article we demonstrate that leucine residues at intra- and intermolecular interaction interfaces within the K-domain are essential mediators of floral quartet-like complex formation of SEP3. SEP-subfamily proteins, which bear a hub-function within the PPI network controlling flower development, display an exceedingly high conservation of the identified leucine residues, whereas non-hub MADS-TFs exhibit preferences for other amino acids at homologous sites. In domain substitution experiments we could show that some of the leucine residues are not only essential but also sufficient for protein tetramerization. We hypothesize that the highly conserved leucine residues allow SEP-subfamily proteins to function as hubs and thereby contributed significantly to the present-day scale-free structure of the PPI network controlling flower development.

Author contributions:
Rainer Melzer and Günter Theißen initiated designed and supervised the project. Florian Rümpler created the substitution mutants of SEP3, performed the EMSA experiments, compiled the sequence collection and performed the *in silico* analyses. The cloning of most substitution mutants and some EMSA experiments had already been performed in the context of the diploma thesis of Florian Rümpler and are therefore not considered as contribution to this dissertation. Rainer Melzer and Florian Rümpler wrote the manuscript. Günter Theißen contributed to improving the manuscript.

Overall contribution of Florian Rümpler: 75 %
Overall contribution of Florian Rümpler during PhD period: 55 %

I hereby certify the accuracy of the statements on the contributions of the authors.

Prof. Günter Theißen

2.2 Manuscript I

Melzer, R., Härter, A., Rümpler, F., Kim, S., Soltis, P. S., Soltis, D. E., Theißen, G. (2014). DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. *Ann. Bot.* 114, 1431-1443.

Annals of Botany **114**: 1431–1443, 2014
doi:10.1093/aob/mcu094, available online at www.aob.oxfordjournals.org

ANNALS OF BOTANY
Founded 1887

PART OF A SPECIAL ISSUE ON FLOWER DEVELOPMENT

DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution

Rainer Melzer[1,2,†], Andrea Härter[1,†], Florian Rümpler[1], Sangtae Kim[3,‡], Pamela S. Soltis[4], Douglas E. Soltis[3,4] and Günter Theißen[1,*]

[1]*Department of Genetics, Friedrich Schiller University Jena, Philosophenweg 12, D-07743 Jena, Germany,* [2]*Department of Genetics, Institute of Biology, University of Leipzig, Talstraße 33, D-04103 Leipzig, Germany,* [3]*Department of Biology and* [4]*Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA*
** For correspondence. E-mail guenter.theissen@uni-jena.de*
[†]*These authors contributed equally to this work.*
[‡]*Present address: Department of Biology and Basic Science Institute, Sungshin Women's University, Seoul 142-732, Korea.*

Received: 3 December 2013 Returned for revision: 3 January 2014 Accepted: 28 March 2014 Published electronically: 5 June 2014

• **Background and Aims** DEFICIENS (DEF)- and GLOBOSA (GLO)-like proteins constitute two sister clades of floral homeotic transcription factors that were already present in the most recent common ancestor (MRCA) of extant angiosperms. Together they specify the identity of petals and stamens in flowering plants. In core eudicots, DEF- and GLO-like proteins are functional in the cell only as heterodimers with each other. There is evidence that this obligate heterodimerization contributed to the canalization of the flower structure of core eudicots during evolution. It remains unknown as to whether this strict heterodimerization is an ancient feature that can be traced back to the MRCA of extant flowering plants or if it evolved later during the evolution of the crown group angiosperms.
• **Methods** The interactions of DEF- and GLO-like proteins of the early-diverging angiosperms *Amborella trichopoda* and *Nuphar advena* and of the magnoliid *Liriodendron tulipifera* were analysed by employing yeast two-hybrid analysis and electrophoretic mobility shift assay (EMSA). Character-state reconstruction, including data from other species as well, was used to infer the ancestral interaction patterns of DEF- and GLO-like proteins.
• **Key Results** The yeast two-hybrid and EMSA data suggest that DEF- and GLO-like proteins from early-diverging angiosperms both homo- and heterodimerize. Character-state reconstruction suggests that the ability to form heterodimeric complexes already existed in the MRCA of extant angiosperms and that this property remained highly conserved throughout angiosperm evolution. Homodimerization of DEF- and GLO-like proteins also existed in the MRCA of all extant angiosperms. DEF-like protein homodimerization was probably lost very early in angiosperm evolution and was not present in the MRCA of eudicots and monocots. GLO-like protein homodimerization might have been lost later during evolution, but very probably was not present in the MRCA of eudicots.
• **Conclusions** The flexibility of DEF- and GLO-like protein interactions in early-diverging angiosperms may be one reason for the highly diverse flower morphologies observed in these species. The results strengthen the hypothesis that a reduction in the number of interaction partners of DEF- and GLO-like proteins, with DEF–GLO heterodimers remaining the only DNA-binding dimers in core eudicots, contributed to developmental robustness, canalization of flower development and the diversification of angiosperms.

Key words: Flower development, DEFICIENS, GLOBOSA, APETALA3, PISTILLATA, protein–protein interaction, yeast two-hybrid, EMSA, character-state evolution, MADS-domain protein, floral homeotic gene, early-diverging angiosperms, basal angiosperms.

INTRODUCTION

Depending on their partner proteins, transcription factors may affect the regulation of certain genes or developmental pathways in very different ways. MIKC-type MADS-domain proteins are a good case in point. In higher eudicots, virtually all MIKC-type MADS-domain proteins constitute dimers with several different partners (Immink *et al*., 2003; de Folter *et al*., 2005; Leseberg *et al*., 2008). These dimers bind to *cis*-regulatory DNA elements termed CArG-boxes [consensus 5′-CC(A/T)$_6$GG-3′]. Combinatorial dimer formation is assumed to be of vital importance for the ability of MADS-domain proteins to regulate a plethora of developmental processes (de Folter *et al*., 2005; Kaufmann *et al*., 2005, 2010). For example, the floral homeotic protein SEPALLATA3 (SEP3) from *Arabidopsis thaliana* may interact with APETALA1 (AP1), another floral homeotic protein, to control floral meristem identity (Gregis *et al*., 2009). Later during development, SEP3 interacts with the floral homeotic protein AGAMOUS (AG) to determine carpel identity and, even later, SEP3 forms complexes with SHATTERPROOF1 and SHATTERPROOF2 to control ovule development (de Folter *et al*., 2005; Immink *et al*., 2009).

Although combinatorial dimer formation emerges as a common property among MIKC-type MADS-domain proteins, the subfamily of DEFICIENS (DEF)- and GLOBOSA (GLO)-like proteins [also known as APETALA3 (AP3)- and PISTILLATA (PI)-like proteins, respectively] constitutes a remarkable exception from this rule. DEF- and GLO-like transcription factors are

highly conserved homeotic selector proteins that determine petal and stamen identity in probably all angiosperms (Kim *et al*., 2005; Zahn *et al*., 2005*b*; Litt and Kramer, 2010). In almost all core eudicots, DEF- and GLO-like proteins form DNA-binding dimers exclusively with each other, and do not form homodimers or DNA-binding heterodimers with other MADS-domain proteins (Riechmann *et al*., 1996*a*; Leseberg *et al*., 2008; Liu *et al*., 2010). This strict (or obligate) DEF–GLO heterodimerization is accompanied by a positive autoregulatory feedback loop in which DEF–GLO heterodimers foster the expression of their own transcripts (Schwarz-Sommer *et al*., 1992; Goto and Meyerowitz, 1994; McGonigle *et al*., 1996; Lenser *et al*., 2009). After the initial activation of *DEF*- and *GLO*-like genes (by factors that are not further considered here), there is an interdependence of DEF- and GLO-like protein expression in core eudicots; only if both of the partner proteins are expressed is the heterodimer formed and can in turn activate *DEF*- and *GLO*-like genes (Schwarz-Sommer *et al*., 1992; Goto and Meyerowitz, 1994).

DEF- and GLO-like proteins of core eudicots are usually expressed in the second and third whorl of the flower, in the primordia of which petals and stamens develop (for reviews, see Zahn *et al*., 2005*b*; Theißen and Melzer, 2007; Litt and Kramer, 2010). The interdependence in expression of DEF- and GLO-like proteins stabilizes this expression pattern; the misexpression of either a DEF- or a GLO-like protein alone would remain without consequences for floral organ identity as the appropriate partner would be missing (Winter *et al*., 2002*b*). It has therefore been proposed that the strict heterodimerization in conjunction with positive feedback regulation enhances developmental robustness of floral organ identity specification and contributed to the standardization of the floral structure (Winter *et al*., 2002*b*; Lenser *et al*., 2009). This raises the question as to when during evolution heterodimerization between DEF- and GLO-like proteins was established. *DEF*- and *GLO*-like genes originated by duplication in a common ancestor of all extant angiosperms (Kim *et al*., 2004). Extant gymnosperms, the closest living relatives of angiosperms, possess gene subfamilies (*GGM2*- and *DAL12*- and *CJMADS1*-like genes) that are ancestral to both *DEF*- and *GLO*-like genes (Winter *et al*., 2002*a*), whereas the sister to all other extant angiosperms, *Amborella trichopoda*, has both *DEF*- and *GLO*-like genes, indicating that the duplication that generated *DEF*- and *GLO*-like genes occurred in the lineage that led to extant angiosperms after the lineage that led to extant gymnosperms had already branched off (Aoki *et al*., 2004; Kim *et al*., 2004). In contrast to many angiosperm DEF- and GLO-like proteins, gymnosperm GGM2-like proteins form homodimers (Sundström and Engström, 2002; Winter *et al*., 2002*b*). It appears likely therefore that the heterodimerization between DEF- and GLO-like proteins originated from a homodimerizing ancestor (Winter *et al*., 2002*b*).

To understand better how DEF- and GLO-like proteins evolved towards the strictly heterodimerizing proteins in core eudicots, analysis of the orthologous proteins from early-diverging angiosperms is required. Here, we analyse the interaction of DEF- and GLO-like proteins from the early-diverging angiosperms *A. trichopoda* and *Nuphar advena* as well as from the magnoliid *Liriodendron tulipifera*. These branched off successively very early during angiosperm evolution, thus forming a grade in the phylogenetic tree (Fig. 1) (APG III, 2009). Our data suggest that DEF-like and GLO-like proteins from early diverging angiosperms heterodimerize with each other but also have the ability to homodimerize as well as to interact weakly with a number of other MADS-domain proteins. Character-state reconstruction revealed that DEF- and GLO-like proteins possessed the ability to form heterodimeric complexes with each other in the most recent common ancestor (MRCA) of extant angiosperms and that this property remained highly conserved throughout angiosperm evolution. In contrast, homodimerization and interactions with proteins from other subfamilies appear much less conserved.

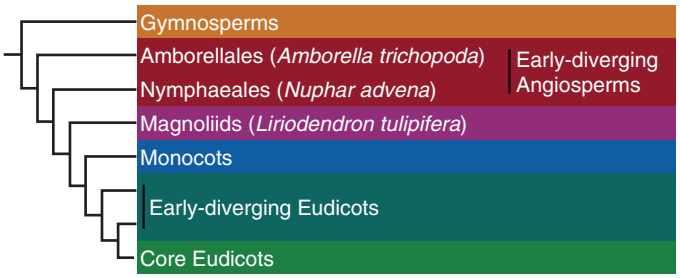


FIG. 1. Simplified seed plant phylogeny. The phylogeny is mainly based on analyses of the angiosperm phylogeny group (APG III, 2009). The phylogenetic position of *A.trichopoda*, *N. advena* and *L. tulipifera* is indicated. Major groups of seed plants are highlighted in different colours.

MATERIALS AND METHODS

Plant material

Flower buds from *Liriodendron tulipifera* were collected in the *Park an der Ilm*, Weimar, Germany. Flower buds of *Nuphar advena* were collected in the Old Botanical Garden of Göttingen, Germany. Male flower buds of *Amborella trichopoda* were collected in the Botanical Garden Bonn, Germany. The collected material was placed immediately into liquid nitrogen and stored at −80 °C until further use.

cDNA sequences used in this study

Partial coding sequences of *LtAP3*, *LtPI*, *Nu.ad.AP3·1*, *Nu.ad.AP3·2* and *Nu.ad.AGL2* have been published previously (Kramer *et al*., 1998; Kim *et al*., 2005; Zahn *et al*., 2005*a*). Full-length coding sequences of these cDNAs were obtained using 5′-RACE (rapid amplification of cDNA ends). The partial coding sequence of *AMtrAGL9* (Zahn *et al*., 2005*a*) was completed by alignment with an expressed sequence tag (EST) sequence (FD432914·1). The predicted transcript was subsequently PCR amplified.

LtPI2 was isolated using 3′- and 5′-RACE. *Nu.ad.PI1* and *Nu.ad.AGL6·1* were derived from ESTs. *Nu.ad.PI2* and *Nu.ad.AGL6·2* were isolated by PCR with primers derived from *Nu.ad.PI1* and *Nu.ad.AGL6·1*, respectively. *Nu.ad.AGL6·1* is, except for three nucleotide differences, identical to a previously published *AGL6*-like sequence from *N. advena* (accession no. GU048649) (Kim *et al*., 2013).

The *AmAP3* and *AmPI* cDNA sequences used here are very similar or identical to *Am.tr.AP3·1* and *Am.tr.PI1*, which were recently reported (Amborella Genome Project, 2013)

For a complete list of cDNA sequences used, see Supplementary Data Table S1. Protein and gene names as

assigned in previous publications were used throughout the manuscript.

cDNA sequences used for yeast two-hybrid and electrophoretic mobility shift assay (EMSA) experiments were provided by the Floral Genome Project (http://fgp.bio.psu.edu/) or by Seishiro Aoki (University of Tokyo), or were synthesized from RNA. For RNA isolation from *L. tulipifera*, total RNA isolation reagent (Biomol) was used. RNA from *N. advena* was extracted with the RNeasy Plant Mini Kit. For RNA extraction from *A. trichopoda*, we used a combination of a CTAB (cetyltrimethylammonium bromide) DNA extraction method and the RNeasy Plant Mini Kit (Kim *et al.*, 2004). cDNA was synthesized using an oligo(dT) primer with MuLV reverse transcriptase.

For yeast two-hybrid experiments, full-length cDNA sequences were cloned into pGADT7 and pGBKT7. Full-length and C-terminal deleted cDNA sequences were cloned into pSPUTK for EMSA experiments. *Nu.ad.PI2* was amplified with primers originally designed to amplify *Nu.ad.PI1*. As the primers used for amplification covered part of the coding sequence, the C-terminal end as well as the beginning of the MADS-domain were identical in the *Nu.ad.PI1* and *Nu.ad.PI2* clones used for EMSA and yeast two-hybrid analyses. This identity also applies to *Nu.ad.AGL6·1* and *Nu.ad.AGL6·2*.

Yeast two-hybrid studies

Yeast two-hybrid assays were carried out essentially as described (Wang *et al.*, 2010). For assaying an interaction, similar amounts of yeast cells were dissolved in water and 10-fold serially diluted up to 1:10 000 in water. Afterwards, the diluted yeast cells were spotted on selective medium lacking histidine, leucine and tryptophan containing 3 mM 3-amino-1,2,4-triazole. The plates were incubated for up to 14 d at 22 °C. All interactions were tested with at least two independent matings.

Electrophoretic mobility shift assay

The EMSA experiments were conducted essentially as described (Melzer *et al.*, 2009). Proteins were produced by *in vitro* translation using the SP6 TNT Quick Coupled Transcription/Translation mix (Promega). After *in vitro* translation, proteins were shock frozen in liquid nitrogen and stored at −80 °C until use.

cDNA sequences of the C-terminal deleted constructs used are listed in Supplementary Data Table S2.

The sequence of the DNA probe used was 5′-CGTTC CATAC TTTCC TTATT TGGAA TATAA TTAAA TTTCG-3′ (the CArG-box is underlined). The concentration of the labelled DNA probe was generally approx. 3 nM. Usually 4 μL of *in vitro* translated protein were used per binding reaction.

Phylogenetic analysis and ancestral character-state reconstructions

Phylogenetic trees for *DEF*- and *GLO*-like genes shown in Figs 5 and 6, and in Supplementary Data Figs S5, S6, S9–S12 were drawn manually according to the APG III (2009) tree topology with the following modifications. Gymnosperm sequences of *Gnetum gnemon* and *Picea abies* were included as outgroup representatives. Gymnosperm branching was implemented as described (Winter *et al.*, 2002*b*). Branching of the *DEF*-like genes within Asparagales was based on Mondragon-Palomino *et al.* (2009). The *TM6*/*euAP3* split was arranged according to phylogenies in Hernandez-Hernandez *et al.* (2007) and Lee and Irish (2011). The *GLO1*/*GLO2* split within the core lamiids was arranged according to Lee and Irish (2011).

In addition to the species-based phylogeny described above, ancestral character-state reconstruction was also done with phylogenetic trees inferred from the DEF- and GLO-like sequences under study (Supplementary Data Figs S1–S4, S7, S8). cDNA sequences used for this were either obtained in this study or downloaded from the NCBI nucleotide collection (http://www.ncbi.nlm.nih.gov/nuccore). To create a codon alignment, all cDNA sequences were first translated to amino acid sequences using ExPASy Translate (http://web.expasy.org/translate/). The amino acid sequences were aligned with MAFFT 7 applying the E-INS-i strategy (Katoh and Standley, 2013). Using the respective cDNA sequences, the resulting amino acid alignment was converted into a codon alignment with RevTrans 1·4 (http://www.cbs.dtu.dk/services/RevTrans/). The quality of the codon alignment was examined in Seaview 4 (Gouy *et al.*, 2010). Phylogenetic trees were calculated using the Bayesian inference method in MrBayes 3 (Ronquist and Huelsenbeck, 2003). Because of high sequence diversity that led to uncertain alignments of the C-terminal domain, only MADS-, I- and K-domains were considered for the calculation of the phylogenetic trees. The analyses were run for 4 000 000 generations applying the 4by4 nucleotide model. The first 25 % of the calculated trees were discarded. Gymnosperm relatives of *DEF*- and *GLO*-like genes were defined as the outgroup.

The interaction data used for the ancestral character-state reconstruction were based on yeast two-hybrid studies and EMSAs either obtained in this study or published previously. A complete list of publications from which interaction data were extracted can be found in Supplementary Data Table S3. All proteins included in the ancestral-state reconstruction were analysed for their homodimerization ability and their ability to form heterodimers with the respective DEF- or GLO-like partner proteins, as well as with proteins from a clade consisting of AGAMOUS-LIKE6-like (AGL6-like), LOFSEP- and SEP3-like proteins (AGL6/LOFSEP/SEP3 clade) (Malcomber and Kellogg, 2005; Zahn *et al.*, 2005*a*; Kim *et al.*, 2013). A DEF- or GLO-like protein was defined as 'interacting' with the other subfamily if interaction with at least one member of the other clade was detected, and as 'non-interacting' if none of the tested interactions showed a positive result. In general, these rules were also applied for interactions of DEF- and GLO-like proteins with AGL6/LOFSEP/SEP3-like proteins. However, in these cases, to be designated as 'not interacting', we required a protein to be negatively tested with members of at least two of the three sub-clades of AGL6/LOFSEP/SEP3-like proteins. This constraint was incorporated to minimize the risk of including false negatives. If different publications or methods yielded contradictory interaction data, the protein was defined as 'interacting' if at least one study showed the respective interaction.

The ancestral character-state reconstructions were performed in Mesquite 2·75 (Maddison and Maddison, 2011) following a maximum likelihood approach with a Markov one-parameter model (i.e. rates for gains and losses of interactions are identical).

We used the one-parameter model as this is usually preferred over a two-parameter model (in which different rates for gains vs. losses are allowed) for medium-sized data sets like those used here (Mooers and Schluter, 1999). Asymmetry likelihood ratio tests as implemented in Mesquite 2·75 also favoured a one-parameter model over a two-parameter model for the majority of the data sets ($P > 0{\cdot}05$). For the trees drawn manually and constrained to the species phylogeny, a branch length of one was uniformly assigned. For the trees based on the gene phylogeny, branch lengths as obtained from the phylogenetic reconstructions were taken.

Alignments and phylogenetic trees have been deposited at TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S15503).

RESULTS

Interactions among MADS-domain proteins from early-diverging angiosperms

To investigate protein–protein interactions of MADS-domain proteins from *A. trichopoda*, *N. advena* and *L. tulipifera*, a GAL-4-based yeast two-hybrid system was employed. DEF- and GLO-like proteins from each species were tested bidirectionally (i.e. proteins were tested as fusions with the GAL4 DNA-binding domain as well as with the GAL4 transcription activation domain) in an 'all against all' fashion. Orthologues of other floral homeotic proteins were also tested for comparison.

In addition to the yeast two-hybrid assays, EMSAs were conducted to characterize further the interactions between the MADS-domain proteins and to study their DNA-binding abilities. MADS-domain proteins are well known to bind as dimers to CArG-boxes (Schwarz-Sommer *et al*., 1992; Huang *et al*., 1996; Riechmann *et al*., 1996*b*). Using a CArG-box derived from the regulatory intron of *AG* from *A. thaliana*, we assayed the formation of protein–DNA complexes. In EMSAs, the formation of heteromeric protein complexes bound to DNA was inferred when the co-incubation of two different proteins with the DNA probe yielded a complex with an electrophoretic mobility different from that of the respective homomeric complexes (Huang *et al*., 1996; Melzer *et al*., 2009; Wang *et al*., 2010). If the homomeric complexes possessed similar electrophoretic mobilities, C-terminal deleted versions of one of the proteins were used to distinguish a homomeric from a heteromeric complex, as described previously (Wang *et al*., 2010).

From all three species, DEF-, GLO- and LOFSEP- or SEP3-like proteins were assayed using yeast two-hybrid analyses and EMSAs. From *A. trichopoda* and *N. advena*, AGL6- and AG-like proteins were also analysed. The results for *A. trichopoda* are shown in Figs 2 and 3, and those for *N. advena* and *L. tulipifera* in Supplementary Data Figs S13–S15. Results are summarized for all species analysed in Fig. 4. Homomeric interactions were detected in only a few instances in yeast two-hybrid assays. This is in contrast to the EMSA results, according to which probably all of the proteins except Nu.ad.AP3·2 formed homomeric complexes bound to DNA, although homomeric AmAP3 and Nu.ad.AP3·1 protein–DNA complexes were only very weakly and not consistently detected (Figs 2–4). With respect to heteromeric complexes, extensive interactions between LOFSEP-, SEP3-, AGL6- and AG-like proteins were detected. At least in EMSAs, but in many cases also in yeast two-hybrid assays, LOFSEP-, SEP3-, AGL6- and AG-like proteins interacted in all tested combinations with each other (Fig. 4). Also DEF- and GLO-like proteins reliably formed DNA-binding heterodimers with each other (Figs 3 and 4). This was in stark contrast to interactions of DEF- and GLO-like proteins with proteins from the LOFSEP-, SEP3-, AGL6- or AG-like clade. In these cases, interactions were often undetectable, were detected only in either yeast two-hybrid assays or EMSAs and/or were so weak that reliable detection was difficult (Fig. 4). For example, the interactions that were very weakly and/or not consistently observed in EMSAs all involved at least one DEF-like or one GLO-like protein (Fig. 4).

In some cases, co-translation of two proteins failed to detect one of the corresponding homomeric complexes but also failed to yield a strong heteromeric complex (compare AMtrAGL9 and AMtrAGL9/AmAP3ΔC, in Fig. 3, for example). We do not have a satisfactory explanation for this observation. However, it could be that heteromeric complexes that either

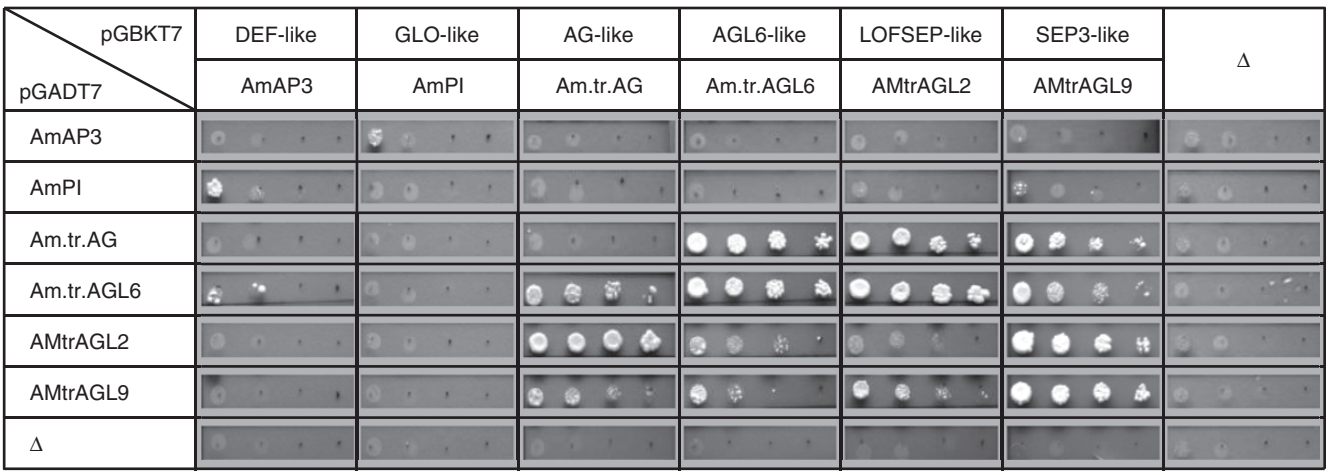


FIG. 2. Representative yeast two-hybrid results for MADS-domain proteins from *A. trichopoda*. Photographs show colony growth on selective –Leu/–Trp/–His media with yeasts grown at 22 °C. For each interaction tested, yeast cells were spotted in a 10-fold serial dilution (from left to right). Proteins that were expressed as fusions with the GAL4 DNA-binding domain (vector pGBKT7) are shown horizontally; proteins expressed as GAL4 activation domain fusions (vector pGADT7) are shown vertically. Δ indicates negative controls in which empty vectors that did not contain a MADS-box gene cDNA insert were used.

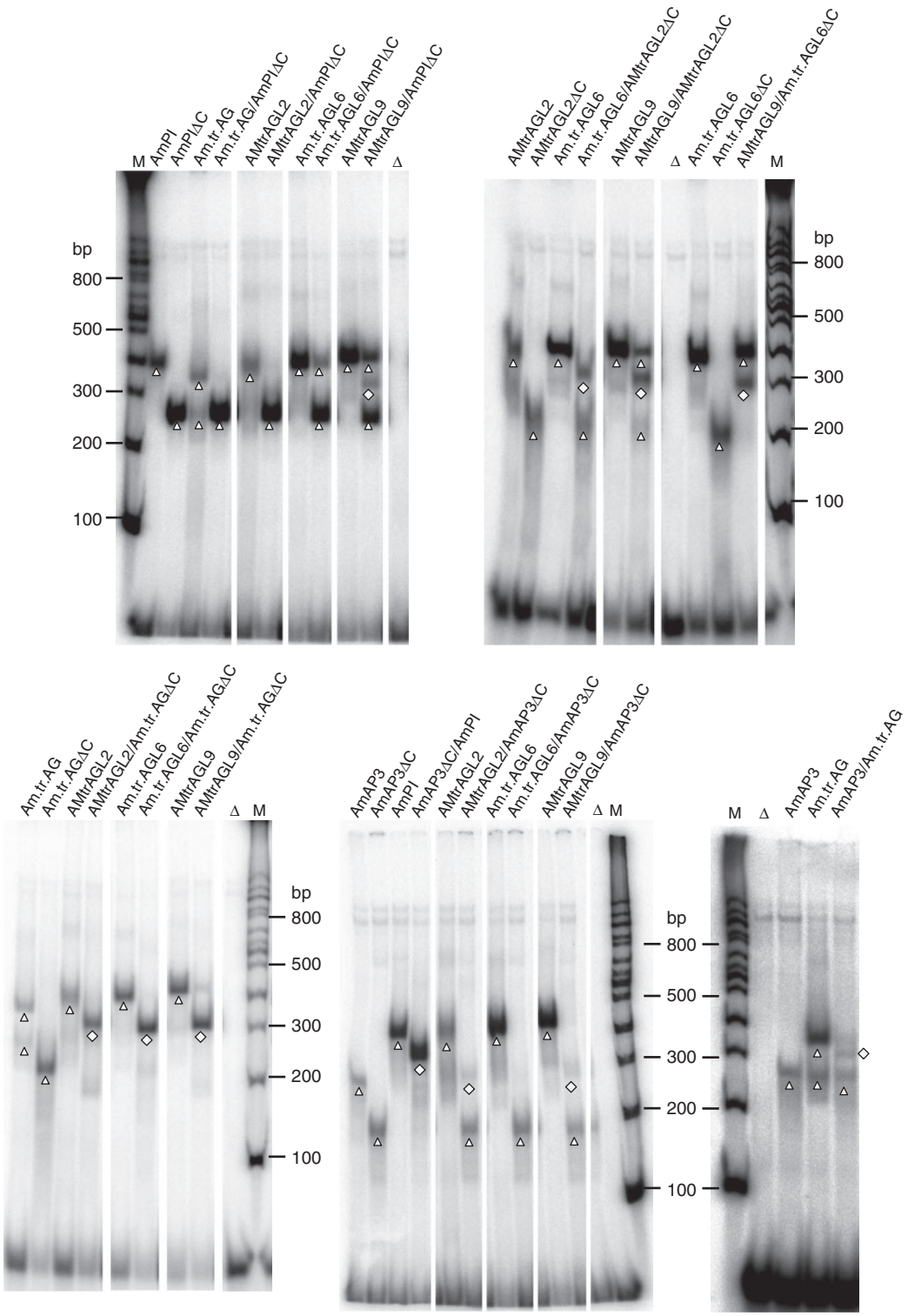


FIG. 3. EMSA results for MADS-domain proteins from *A. trichopoda*. *In vitro* translated proteins were incubated together with a radioactively labelled DNA probe which carried one CArG-box. Proteins applied are noted above the gel. 'ΔC' is used to indicate C-terminal deleted proteins. Triangles highlight homomeric DNA-bound complexes; squares highlight heteromeric DNA-bound complexes. Free DNA is seen at the bottom of the gels. Homomeric complexes were not always visible when heteromer formation was tested, possibly because the vast majority of protein was assembled into DNA-bound heteromeric complexes (AmAP3ΔC/AmPI, for example) or in heteromeric complexes not or only very weakly binding to DNA (AMtrAGL9/AmAP3ΔC, for example). Note that certain potential heteromeric complexes were very weak and sometimes difficult to distinguish from homomeric complexes (AMtrAGL2/AmAP3ΔC, for example). For unknown reasons, the homomeric AmAP3–DNA complex possessed an unsusually high electrophoretic mobility and some proteins (Am.tr.AG, for example) formed two distinct protein–DNA complexes even in the absence of a partner. 'Δ' indicates negative controls in which *in vitro* translation lysate programmed with a vector that did not contain a cDNA insert was added. 'M' denotes lanes in which a radioactively labelled DNA marker (NEB 100 bp DNA ladder) was applied (bp = base pairs).

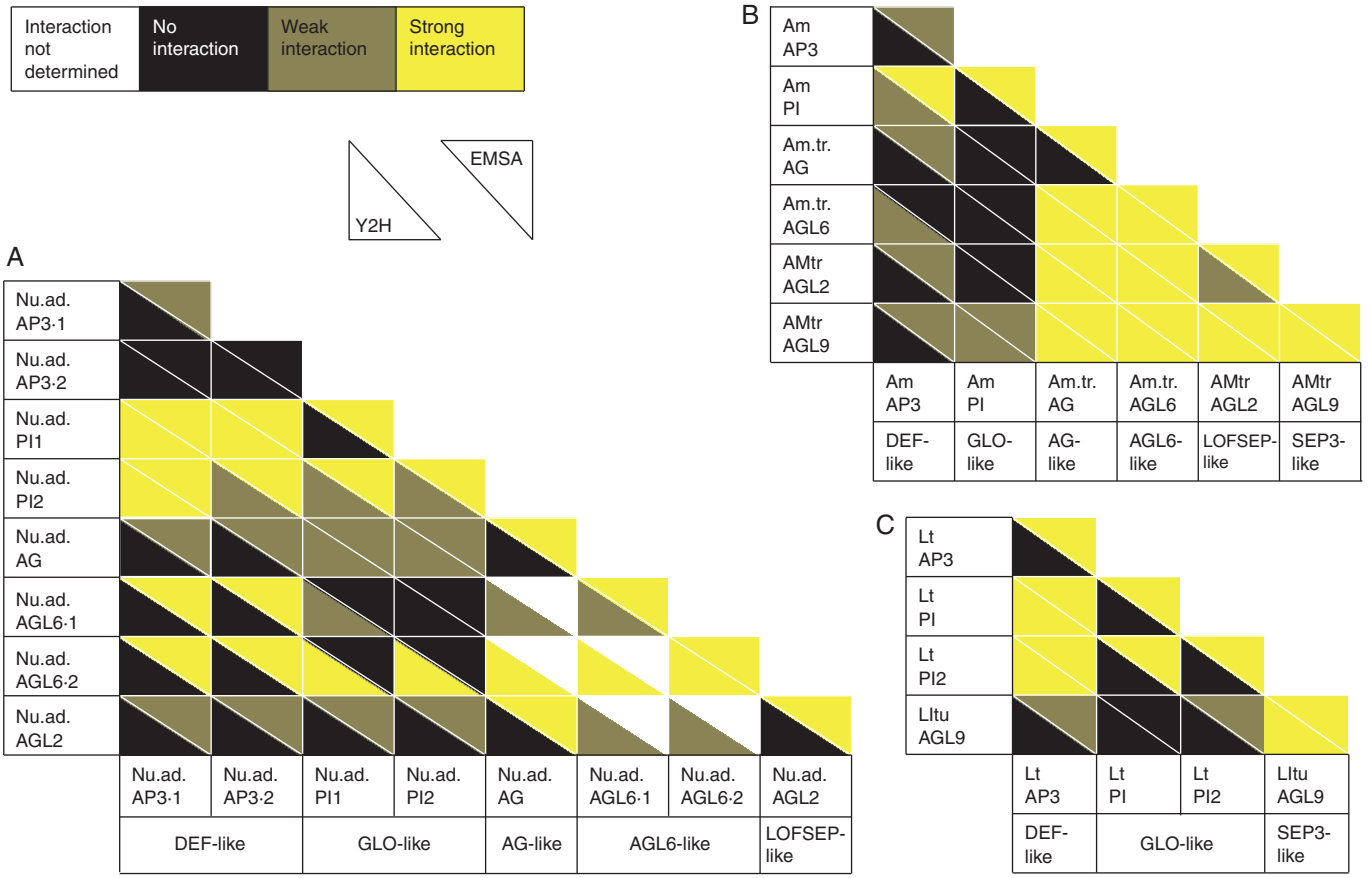


FIG. 4. Summary of the interaction patterns of MADS-domain proteins from (A) *N. advena*, (B) *A. trichopoda* and (C) *L. tulipifera*. Yeast two-hybrid interactions were scored as strong if yeast growth was observed in all dilutions in at least one direction, and weak if yeast growth was observed but not in all dilutions tested. EMSA results were designated as strong if a protein–DNA complex was reliably observed, and as weak if a protein–DNA complex was not reproducibly observed, and/or if the protein–DNA complex was detected as a faint band only (that was sometimes difficult to distinguish from homomeric complexes).

did not bind or only weakly bound to DNA were reconstituted in these cases. Furthermore, the protein–DNA complexes reconstituted in EMSAs span a considerable range of electrophoretic mobilities (compare homomeric AmAP3 with homomeric AmAG complexes in Fig. 3, for example), although the molecular masses and charges of the proteins are very similar. It is unclear whether the different mobilities resulted from different conformations of the proteins or of the protein–DNA complexes or from differences in the stoichiometry of binding.

Reconstruction of the ancestral DEF- and GLO-like protein interaction behaviour

Interaction patterns among DEF- and GLO-like proteins have previously been determined for several monocot and eudicot species (Davies *et al.*, 1996; Riechmann *et al.*, 1996*a*; Winter *et al.*, 2002*b*; Immink *et al.*, 2003; Kanno *et al.*, 2003; Yang *et al.*, 2003; Vandenbussche *et al.*, 2004; Whipple *et al.*, 2004; Kramer *et al.*, 2007; Yao *et al.*, 2008; Liu *et al.*, 2010, to mention but a few). Using published data along with the newly identified interactions from early-diverging angiosperms obtained in this study, we aimed at understanding the trajectories of interactions of DEF- and GLO-like proteins during angiosperm evolution employing character-state reconstruction. The EMSA and yeast two-hybrid data were combined for these analyses. Character-state reconstruction was conducted using (1) a phylogenetic tree drawn manually according to the species phylogeny as reported by APG III (Fig. 1) (APG III, 2009) and (2) phylogenetic trees based on the phylogenetic relationships among the genes. Tree topologies were similar for the two approaches, and the character-state reconstructions yielded essentially the same results (compare Figs 5, 6 and Supplementary Data Figs S5, S6 with Figs S1–S4, S7, S8). Interestingly, almost all of the DEF- and GLO-like proteins assayed so far are capable of interacting with a partner from the other subfamily, i.e. of forming DEF–GLO-like protein complexes (Figs 5 and 6; Supplementary Data Figs S1, S3). This result strongly indicates that heterodimerization between DEF- and GLO-like proteins was established at the base of extant angiosperms and remained highly conserved throughout angiosperm evolution (Figs 5 and 6).

Several DEF-like and GLO-like proteins not only constitute DEF–GLO heterodimers but also possess the ability to homodimerize (Winter *et al.*, 2002*a*; Liu *et al.*, 2010). Character-state reconstruction suggests that homodimerization of both DEF- and GLO-like proteins was probably present in the MRCA of extant angiosperms (Figs 5 and 6). The analyses further suggest that homodimerization of DEF-like proteins was lost relatively early during angiosperm evolution and regained in

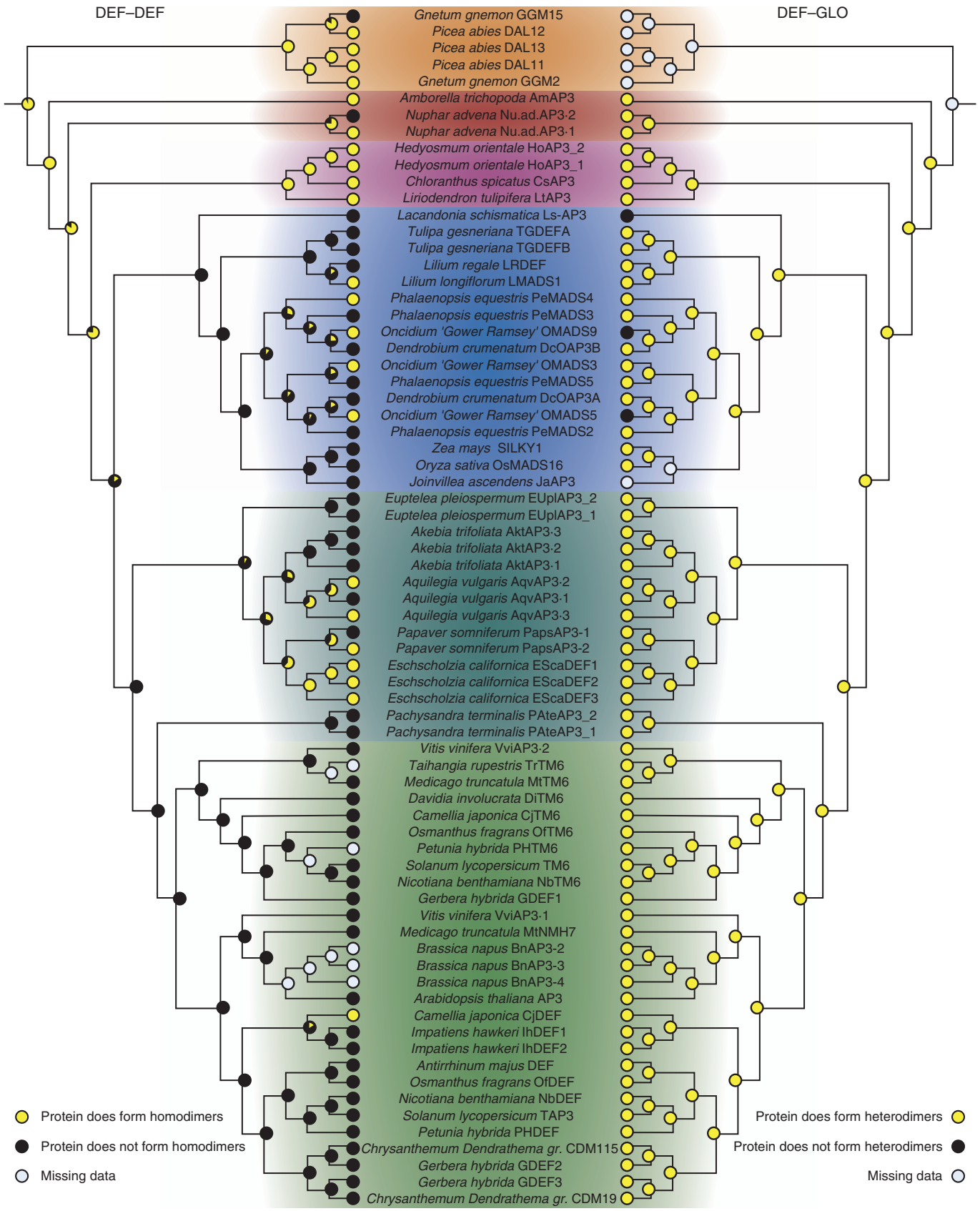


FIG. 5. Ancestral character-state reconstruction for the ability of DEF-like proteins to form homo- and heterodimeric complexes. Trees are in general based on the APG III phylogeny as described in the Materials and Methods. The tree on the left depicts character-state reconstruction for the homodimerization capability of

several eudicot and monocot lineages (Fig. 5). Character-state reconstruction using the gene phylogeny of *DEF*-like sequences indicates that DEF-like protein homodimerization was lost independently in eudicots and monocots. However, this result can very probably be attributed to the unsusual placement of magnoliid *DEF*-like genes as being closely related to eudicot *DEF*-like genes (Supplementary Data Fig. S2).

For GLO-like proteins, the situation was slightly different; GLO-like protein homodimerization prevails in the extant monocots analysed, although it is not clear whether homodimerization was preserved at the base of the monocots or lost and re-established later during monocot evolution (Fig. 6; Supplementary Data Fig. S4). In contrast, only a very limited number of eudicot species possess homodimerizing GLO-like proteins (Fig. 6; Supplementary Data Fig. S4), strongly indicating that the MRCA of eudicots did not possess a homodimerizing GLO-like protein.

We also conducted character-state reconstructions for interactions of DEF- and GLO-like proteins with proteins from the clade of AGL6/LOFSEP/SEP3-like proteins (Supplementary Data Figs S5–S8). These analyses were included because DEF- and GLO-like proteins act in tetrameric complexes with AGL6/LOFSEP/SEP3-like proteins to determine petal and stamen identity (Theißen, 2001; Theißen and Saedler, 2001; Wang *et al*., 2010). Interactions of DEF-like as well as of GLO-like proteins with AGL6/LOFSEP/SEP3-like proteins appear to be relatively scattered across the angiosperm phylogeny. For example, about half of the eudicot DEF-like proteins analysed interact with at least one AGL6/LOFSEP/SEP3-like protein (Supplementary Data Figs S5, S7). However, the pattern emerges that DEF–AGL6/LOFSEP/SEP3 as well as GLO–AGL6/LOFSEP/SEP3 interactions were already present early in angiosperm evolution. Also the MRCA of DEF- and GLO-like proteins at the base of the seed plants probably already possessed the respective interaction (Supplementary Data Figs S5–S8).

For the above-described character-state reconstruction, proteins were designated as interacting when a positive result had been reported either from yeast two-hybrid assays or from EMSAs. For 15 DEF-like and 12 GLO-like proteins from angiosperms, homodimerization data from both techniques (EMSA and yeast two-hybrid assays) were available (Supplementary Data Figs S9, S11). Of these 27 cases, 16 yielded identical results in yeast two-hybrid assays and EMSAs with respect to homodimerization (i.e. proteins did or did not form homodimers in both assays). For nine interactions, homodimerization was detected in EMSAs but not in yeast two-hybrid assays. The converse case, i.e. detection of homodimerization in yeast two-hybrid assays but not in EMSAs, was observed only for two proteins (Supplementary Data Figs S9, S11). In contrast, heterodimerization among DEF- and GLO-like proteins was, with one exception, consistently observed in EMSAs as well as yeast two-hybrid assays (Supplementary Data Figs S10, S12). This confirms previous observations that homodimerization, in particular of MADS-domain proteins, is more readily detected in EMSAs than in yeast two-hybrid assays (Wang *et al*., 2010). To account for the different results that these techniques yielded, ancestral character-state reconstruction for homo- and heterodimerization of DEF- and GLO-like proteins was conducted separately for data sets based on yeast two-hybrid and EMSA results (Supplementary Data Figs S9–S12). Heterodimerization between DEF- and GLO-like proteins was inferred to be ancestral and highly conserved for all data sets (Supplementary Figs S10, S12). Also homodimerization was still inferred to be ancestral for DEF- and GLO-like proteins from flowering plants when only EMSA data were considered. In contrast, when only yeast two-hybrid data were taken into account, homodimerization was inferred to be absent in the MRCA of extant flowering plants.

DISCUSSION

Conservation of the MADS-domain protein interaction pattern during angiosperm evolution

In core eudicots and monocots, certain dimers of MADS-domain proteins are involved in floral organ specification. Dimers of DEF- and GLO-like proteins function in petal and stamen specification, and dimers of an AGL6/LOFSEP/SEP3-like protein and an AG-like protein are involved in stamen and carpel development (Schwarz-Sommer *et al*., 1992; Huang *et al*., 1996; Riechmann *et al*., 1996*a*; Theißen, 2001; de Folter *et al*., 2005; Rijpkema *et al*., 2009; Thompson *et al*., 2009; Liu *et al*., 2010). Comparison of the interaction patterns obtained from *A. trichopoda*, *N. advena* and *L. tulipifera* with those from model eudicots and monocots (i.e. *A. thaliana*, *Petunia hybrida* and *Oryza sativa*) revealed that these heterodimeric interactions important for organ specification in derived angiosperms are also detected in early diverging angiosperms (Fig. 7), pointing towards a high degree of conservation of these interactions during flowering plant evolution. Recently published yeast two-hybrid data from *A. trichopoda* (Amborella Genome Project, 2013) largely agree with the interactions reported here. However, we detected some additional interactions (AMtrAGL9/AMtrAGL2, for example), probably because we used milder assay conditions (e.g. growing yeast cells at 22 °C instead of using higher temperatures).

Our data suggest that the interactions governing flower development in core eudicots were already established at the base of extant angiosperms and remained highly conserved since then. Specifically, our results indicate that the heterodimerization between DEF-like and GLO-like proteins was already present in the MRCA of extant angiosperms and was virtually never rewired. Alternative scenarios in which heterodimerization between DEF- and GLO-like proteins was established independently in eudicots and monocots would invoke multiple losses and gains of DEF–GLO heterodimerization and therefore appear less likely.

DEF-like proteins. The tree on the right has the same topology as the tree on the left but shows character-state reconstruction for heterodimerization of DEF-like proteins with GLO-like proteins. Yellow circles at the terminal positions indicate the presence of an interaction, and black circles indicate the absence of an interaction. Grey circles indicate that interaction data are not available for that particular protein. The likelihood of an interaction at internal nodes is indicated by pie charts. Proteins from different plant groups are highlighted in colours as in Fig. 1. Colour coding is as follows: gymnosperms, orange; early-diverging angiosperms, red; magnoliids, purple; monocots, blue; early-diverging eudicots, dark green; core eudicots, light green

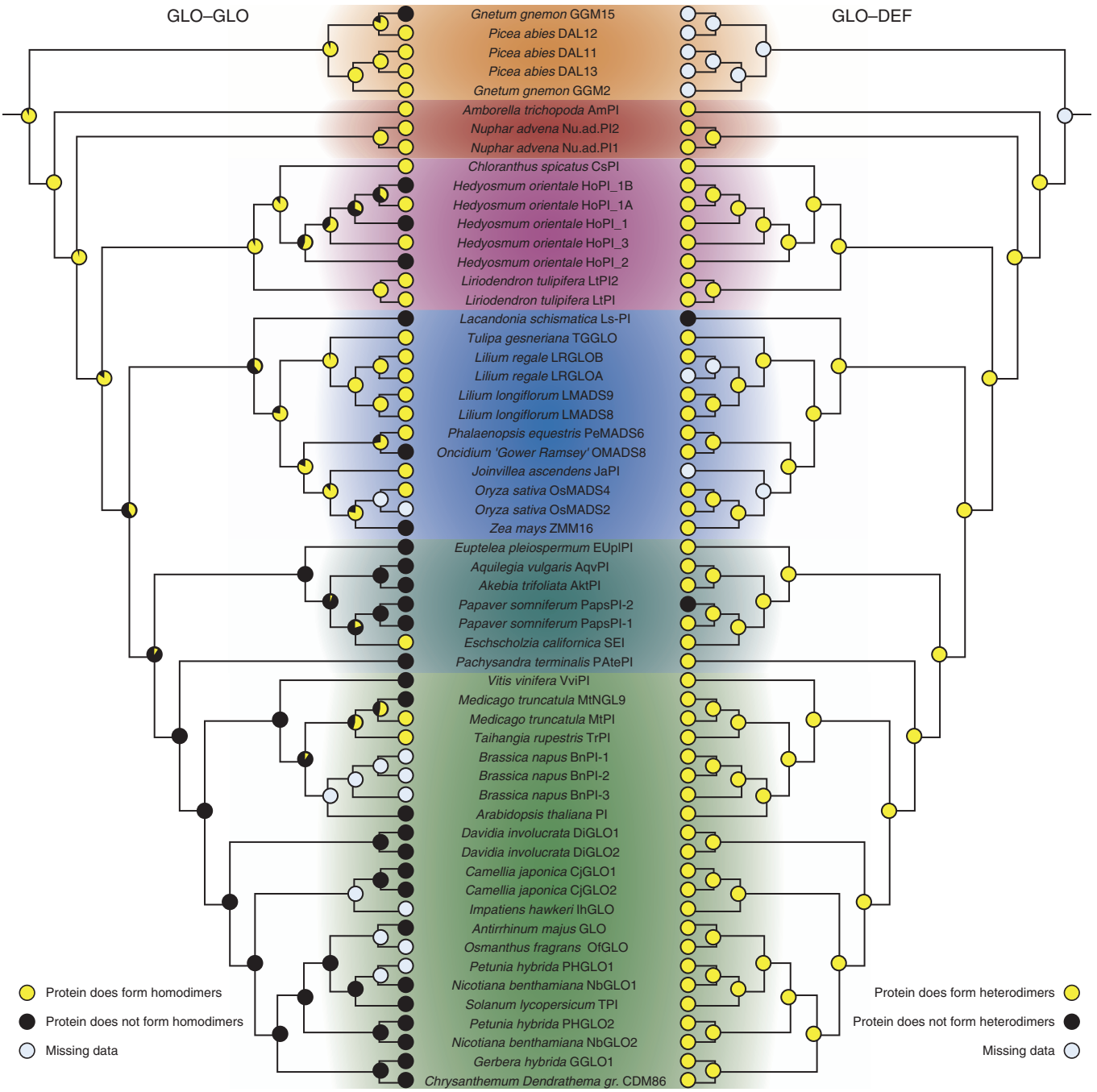


FIG. 6. Ancestral character-state reconstruction for the ability of GLO-like proteins to form homo- and heterodimeric complexes. The trees on the left and right depict character state reconstructions for homodimerization of GLO-like proteins and heterodimerization of GLO-like with DEF-like proteins, respectively. For details of the labelling, see Fig. 5.

We observed only three DEF-like (OMADS5 and OMADS9 from *Oncidium* 'Gower Ramsey' and Ls-AP3 from *Lacandonia schismatica*) (Alvarez-Buylla *et al.*, 2010; Chang *et al.*, 2010) and two GLO-like proteins (PapsPI-2 from *Papaver somniferum* and Ls-PI from *L. schismatica*) (Drea *et al.*, 2007; Alvarez-Buylla *et al.*, 2010) for which heterodimerization with a partner from the other subfamily was not detected (Figs 5 and 6). However, in *O.* 'Gower Ramsey' and *P. somniferum*, at least one additional DEF- or GLO-like protein exists that can constitute a DEF–GLO heterodimer and thus may compensate for the loss of heterodimerization in the respective paralogue. In contrast, Ls-AP3 and Ls-PI are the only DEF- and GLO orthologues that have been isolated from *L. schismatica*. Intriguingly, *L. schismatica* deviates from the basic floral 'bauplan' in that carpels surround stamens, which are positioned in the centre of the flower (Marquez-Guzman *et al.*, 1989; Alvarez-Buylla *et al.*, 2010), a feature that is otherwise only known from the genus *Trithuria* (Rudall *et al.*, 2009) and thus is extremely rare in angiosperms. It is interesting to note that the loss of DEF–GLO heterodimerization coincides with the modification of the floral bauplan

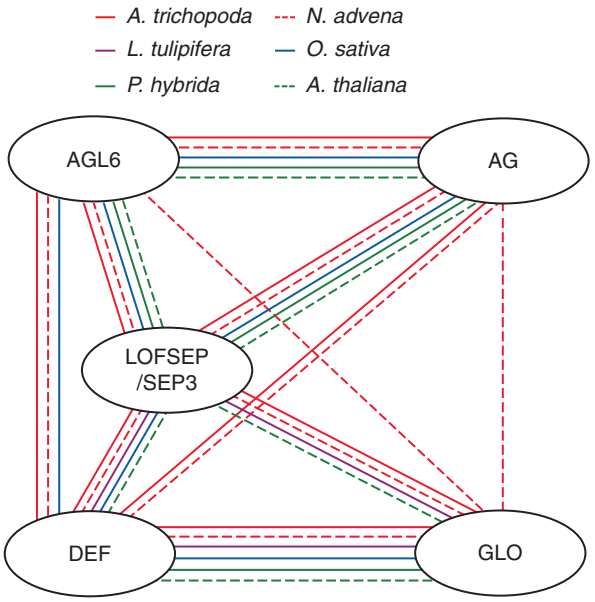


FIG. 7. Conservation of interactions among MADS-domain proteins. The different subfamilies are depicted by ovals; lines between the ovals indicate interactions in the different species as colour coded in the key. Only interactions between proteins of different subfamilies are depicted. Not all protein combinations have been tested in all species. The absence of an interaction here does not therefore necessarily mean that the proteins do not interact. Publications used to depict interactions among proteins from *A. thaliana*, *P.hybrida* and *O. sativa* are listed in Supplementary Data Table S3.

in *L. schismatica*, although it remains elusive whether there is a causal relationship between these two observations.

Use of the yeast two-hybrid system to detect homodimerization of proteins

The available data clearly suggest that homodimerization of DEF- and GLO-like proteins is under-represented in yeast two-hybrid assays as compared with EMSAs. Indeed, character-state reconstructions using only yeast two-hybrid data infer that homodimerization of DEF- or GLO-like proteins was not present in the MRCA of flowering plants, whereas the opposite is predicted when using only EMSA data or a combination of both data sets (Figs 5, 6; Supplementary Data Figs S9, S11). One explanation for the discrepancy between yeast two-hybrid and EMSA results might be that homodimerzation of DEF- and GLO-like proteins is stabilized by DNA binding, as has also been proposed for other MIKC-type MADS-domain proteins (Wang *et al*., 2010). In addition, it has been described that homodimerization is in general difficult to detect with the yeast two-hybrid system (Smirnova *et al*., 1999; Newman *et al*., 2000). One plausible explanation for that is that the GAL4 DNA-binding domain used in the yeast two-hybrid experiments is already capable of homodimerization. This homodimerization in turn elevates the local concentration of the proteins fused to the GAL4 DNA-binding domain (Hu, 2000; Newman *et al*., 2000), thereby favouring the formation of homodimers between hybrid proteins containing a GAL4 DNA-binding domain, at the expense of interactions with hybrid proteins containing the GAL4 activation domain. We therefore assume that EMSA data or a combination of yeast two-hybrid and EMSA data are better suited to trace the ancestral character state of homodimerizing proteins than yeast two-hybrid data alone.

The developmental relevance of DEF- and GLO-like protein interactions beyond DEF–GLO heterodimerization

The functional relevance of DEF–GLO heterodimers for determining petal and stamen identity is well established (Schwarz-Sommer *et al*., 1992; Lenser *et al*., 2009). However, one may ask what the function of the occasionally observed DEF- and GLO-like protein homodimers or dimers of DEF- or GLO-like proteins with other floral homeotic proteins is. The frequent occurrence of GLO-like protein homodimers in particular has fostered speculation that these protein complexes are of developmental relevance (Winter *et al*., 2002*b*). This possibility is supported by expression studies showing that in monocots and early diverging angiosperms in particular, expression of *GLO*-like genes is sometimes not associated with *DEF*-like gene expression and is also observed in organs other than petals and stamens (see, for example, Münster *et al*., 2001; Kim *et al*., 2005).

However, to the best of our knowledge, a developmental function of GLO-like protein homodimers or DEF-like protein homodimers has not yet been described. A detailed analysis of these possibilities would require downregulation or knockout of *DEF*- and *GLO*-like genes in species in which DEF- or GLO-like proteins homodimerize. Respective data are only available from very few species, including *O. sativa* (Nagasawa *et al*., 2003; Ronai *et al*., 2003; Yao *et al*., 2008), *Aquilegia vulgaris* (Kramer *et al*., 2007) and *P. somniferum* (Drea *et al*., 2007). In these cases, downregulation of the complete set of *DEF*-like genes resulted in homeotic transformations that very much resembled or were identical to those obtained when the concomitant *GLO*-like genes were downregulated. These observations are in line with the assumption that DEF- and GLO-like proteins function as heterodimers during development and that potential DEF- or GLO-like protein homodimers cannot substitute for the developmental role of DEF–GLO heterodimers. However, it should be taken into consideration that the formation of GLO-like or DEF-like protein homodimers may lower the concentration of protein monomers available for heterodimer formation, thus perhaps indirectly affecting the activity of DEF–GLO heterodimers. Also, in several species, DEF- or GLO-like protein homodimers possess DNA-binding activity (Winter *et al*., 2002*b*; Kanno *et al*., 2003; Ronai *et al*., 2003; Whipple and Schmidt, 2006). These dimers may compete with DEF–GLO heterodimers for target gene occupancy. Mechanisms may thus have evolved to prevent the interference between homodimers and heterodimers of DEF- and GLO-like proteins. Further studies involving quantitative analyses of interaction strengths will reveal whether and how such mechanisms are employed.

The functional relevance of homodimers in determining floral organ identity notwithstanding, it remains to be noted that DEF- and GLO-like protein homodimerization is especially prevalent in early-diverging angiosperms. In these species, flower morphology is considerably less canalized than in core eudicots and monocots; very often the transition between different floral organ types is gradual and not as distinct as in more derived

angiosperms (Buzgo *et al*., 2004; Soltis *et al*., 2006, 2007, 2009). It has been proposed that this gradual transition is related to gradients of floral homeotic gene expression (Buzgo *et al*., 2004; Soltis *et al*., 2006, 2007, 2009). According to the 'fading border' model, weak expression of floral homeotic genes may lead to the establishment of organs of intermediate identity. For example, weak expression of *DEF*- and/or *GLO*-like genes in the outer floral organs of *A. trichopda* may give rise to sepaloid tepals. An increase of *DEF*- and *GLO*-like gene expression towards the centre of the flower may result in a more petaloid appearance of tepals. It could well be that GLO- or DEF-like protein homodimers contributed to the broadening of the expression domain and to the gradual transition between floral organs. The evolutionary establishment of obligate heterodimerization, possibly in conjunction with autoregulatory control, may have sharpened the expression boundaries of DEF- and GLO-like proteins and hence contributed to the origin of distinct organ types within the flower. Unfortunately, mutant analyses in the early-diverging angiosperms investigated here to test hypotheses on the functional relevance of homodimers of DEF- or GLO-like proteins are not yet possible.

It is also noteworthy that circumstantial evidence for a role for DEF-like protein homodimers during development was provided by the analysis of the orchid *DEF*-like gene *OMADS3* (Hsu and Yang, 2002). OMADS3 forms homodimers and does not interact with AP3 or PI from Arabidopsis. However, when ectopically expressed in Arabidopsis, precocious flowering is observed. This was taken as evidence that OMADS3 homodimers have a function in floral induction (Hsu and Yang, 2002). Future experiments may further substantiate whether such a novel function of DEF-like proteins in orchids does indeed exist.

Beyond homodimerization and interactions with each other, some DEF- and GLO-like proteins occasionally interact with AGL6/LOFSEP/SEP3-like proteins. These interactions may be considered in conjunction with the ability of floral homeotic MADS-domain proteins to form tetrameric complexes termed floral quartets (Theißen, 2001; Theißen and Saedler, 2001; Smaczniak *et al*., 2012). These tetrameric complexes are constituted by two dimers of MADS-domain proteins that are bound in the vicinity of each other on the DNA and interact with each other. Specific floral quartets are proposed to confer the identity of each of the different floral organ types. DNA-bound heterodimers of DEF- and GLO-like proteins are implicated to be part of the quartets determining petal and stamen identity. In *A. thaliana*, for example, these are AP3-PI/SEP3–AP1 and AP3-PI/SEP3–AG complexes, respectively (Honma and Goto, 2001; Theißen, 2001; Theißen and Saedler, 2001). Most of the heterodimeric interactions between DEF- or GLO-like and AGL6/LOFSEP/SEP3-like proteins may therefore reflect the interactions operating between DNA-bound dimers within the quartet. This hypothesis is supported by the notion that in the (admittedly few) eudicot species tested so far, DNA-binding activity of heterodimers consisting of a DEF-like or a GLO-like protein and an AGL6/LOFSEP/SEP3-like protein has not been detected (Davies *et al*., 1996; Geuten *et al*., 2006). Co-operative interactions between DNA-bound dimers are usually very weak in nature. Depending on the actual experimental set-up, they may escape detection with the yeast two-hybrid assay. That may explain why interactions between DEF- or GLO-like and AGL6/LOFSEP/SEP3-like proteins appear to be scattered across the angiosperm phylogeny. However, it remains to be noted that at the base of the angiosperms DNA-binding dimers between DEF- or GLO-like and AGL6/LOFSEP/SEP3-like proteins may have existed. Remnants of this property are detected in early-diverging angiosperms. Also, a gymnosperm orthologue of DEF- and GLO-like proteins, GGM2, forms DNA-binding dimers with AGL6-like proteins (Wang *et al*., 2010). We hypothesize that the ability of DEF- and GLO-like proteins to form DNA-binding dimers with proteins from other subfamilies was quickly lost during angiosperm evolution, as the respective complexes are difficult to detect in early-diverging angiosperms.

The evolutionary and developmental relevance of obligate heterodimerization

Taken together, our data indicate that a DEF–GLO heterodimer is the only functional DNA-binding dimer in the vast majority of angiosperms. As summarized elsewhere, this obligate heterodimerization may have contributed to the canalization of flower development as well as to developmental robustness (Winter *et al*., 2002*b*; Lenser *et al*., 2009). However, the question remains as to why specifically DEF- and GLO-like proteins and not other floral homeotic MADS-domain proteins have undergone such a drastic reduction of interaction partners.

The evolution of interactions among floral homeotic proteins can probably not be understood without taking tetramer formation into account. In eudicots, tetramer formation and hence floral organ development largely depends on AGL6/LOFSEP/SEP3-like proteins (or the closely related SQUA-like proteins) (Immink *et al*., 2009, 2010), and this dependence may effectively prevent the development of floral organs outside the floral context (Melzer *et al*., 2010). This dependence may in turn have contributed to the concerted development of petals, stamens and carpels in close proximity to each other and thus to the evolutionary origin and developmental stabilization of the flower as a reproductive entity (Melzer *et al*., 2010). However, if one protein of the AGL6/LOFSEP/SEP3 clade is always part of the tetramers determining organ identity, an obligate heterodimer beyond the DEF–GLO heterodimer cannot be constituted for stoichiometric reasons. For example, at the dimeric level, it is plausible to assume that a hypothetical obligate heterodimer of two different AG-like proteins could buffer developmental perturbations and contribute to the canalization of development within the flower in a similar way to as proposed for a DEF–GLO heterodimer. It may prevent, for example, misexpression of AG-like proteins outside the two inner floral whorls. In the context of tetramer formation, stamen development would in this hypothetical scenario be controlled by a tetramer constituted of a DEF–GLO heterodimer and a heterodimer of two AG-like proteins. However, AGL6/LOFSEP/SEP3-like proteins would not be part of such a complex. Therefore, selection may have acted against the formation of such complexes as these may have negatively affected the robustness of the flower as a single reproductive entity.

Concluding remarks

The evolution of DEF- and GLO-like interaction patterns is a fascinating example of how the reduction of molecular

interactions may have contributed to developmental robustness that in turn may have led to species diversity. It is clear, however, that much remains to be learned about the pattern of protein interactions of these two subfamilies of floral homeotic MADS-domain proteins. Of particular interest is the functional relevance of dimers other than DEF–GLO heterodimers in early-diverging angiosperms and gymnosperms. Eventually, analysis of loss-of-function mutants of *def*- and *glo*-like genes in these species may be necessary to substantiate the evolutionary dynamics of DEF- and GLO-like protein interactions.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. Table S1: list of genes used in this study. Table S2: cDNA sequences used to generate C-terminal deleted proteins for EMSAs. Table S3: list of publications from which protein–protein interaction data were extracted. Figs S1–12 give ancestral character-state reconstruction of DEF- and GLO-like protein interactions: S1-S4, S7 and S8 show phylogenetic trees based on the gene phylogenies, whilst S5, S6 and S9-S12 show phylogenetic trees manually drawn based on the species phylogeny. Fig. S1: heterodimerization of DEF- with GLO-like proteins. Fig. S2: homodimerization of DEF-like proteins. Fig. S3: heterodimerization of GLO- with DEF-like proteins. Fig. S4: homodimerization of GLO-like proteins. Figs S5, S7: heterodimerization of DEF-like with AGL6/LOFSEP/SEP3-like proteins. Figs S6, S8: heterodimerization of GLO-like with AGL6/LOFSEP/SEP3-like proteins. Fig. S9: homodimerization of DEF-like proteins, comparing yeast two-hybrid (Y2H) and EMSA data. Fig. S10: heterodimerization of DEF- with GLO-like proteins, comparing Y2H with EMSA data. Fig. S11: homodimerization of GLO-like proteins, comparing Y2H and EMSA data. Fig. S12: heterodimerization of GLO- with DEF-like proteins, comparing Y2H with EMSA data. Figs S13, S14: representative yeast two-hybrid results for MADS-domain proteins from (S13) *L. tulipifera* and (S14) *N. advena*. Fig. S15: EMSA results for MADS-domain proteins from *N. advena* and *L. tulipifera*.

ACKNOWLEDGEMENTS

We thank Seishiro Aoki (University of Tokyo) and Charlie Scutt (ENS Lyon) for providing cDNAs of *A. trichopoda* MADS-box genes, and Michael Schwerdtfeger (Botanical Garden Göttingen) and Wilhelm Barthlott (Botanical Garden Bonn) for providing plant material. Very special thanks to Chrissi Gafert for excellent technical support. We are grateful to Kerstin Kaufmann for valuable discussions during the planning phase of the project. R.M. is grateful to the University of Leipzig for general support. This research was supported by DFG grants to G.T. (TH417/5–1) and to G.T. and R.M. (TH417/5–2). R.M. was supported by a post-doctoral fellowship of the Carl-Zeiss-Foundation.

LITERATURE CITED

Alvarez-Buylla ER, Ambrose BA, Flores-Sandoval E, *et al.* 2010. B-function expression in the flower center underlies the homeotic phenotype of *Lacandonia schismatica* (Triuridaceae). *The Plant Cell* **22**: 3543–3559.

Amborella Genome Project. 2013. The *Amborella* genome and the evolution of flowering plants. *Science* **342**: 11.

Aoki S, Uehara K, Imafuku M, Hasebe M, Ito M. 2004. Phylogeny and divergence of basal angiosperms inferred from *APETALA3*- and *PISTILLATA*-like MADS-box genes. *Journal of Plant Research* **117**: 229–244.

APG III. 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.

Buzgo M, Soltis PS, Soltis DE. 2004. Floral developmental morphology of *Amborella trichopoda* (Amborellaceae). *International Journal of Plant Sciences* **165**: 925–947.

Chang YY, Kao NH, Li JY, *et al.* 2010. Characterization of the possible roles for B class MADS box genes in regulation of perianth formation in orchid. *Plant Physiology* **152**: 837–853.

Davies B, EgeaCortines M, Silva ED, Saedler H, Sommer H. 1996. Multiple interactions amongst floral homeotic MADS box proteins. *EMBO Journal* **15**: 4330–4343.

Drea S, Hileman LC, De Martino G, Irish VF. 2007. Functional analyses of genetic pathways controlling petal specification in poppy. *Development* **134**: 4157–4166.

de Folter S, Immink RGH, Kieffer M, *et al.* 2005. Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. *The Plant Cell* **17**: 1424–1433.

Geuten K, Becker A, Kaufmann K, *et al.* 2006. Petaloidy and petal identity MADS-box genes in the balsaminoid genera *Impatiens* and *Marcgravia*. *The Plant Journal* **47**: 501–518.

Goto K, Meyerowitz EM. 1994. Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes and Development* **8**: 1548–1560.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27**: 221–224.

Gregis V, Sessa A, Dorca-Fornell C, Kater MM. 2009. The *Arabidopsis* floral meristem identity genes *AP1*, *AGL24* and *SVP* directly repress class B and C floral homeotic genes. *The Plant Journal* **60**: 626–637.

Hernandez-Hernandez T, Martinez-Castilla LP, Alvarez-Buylla ER. 2007. Functional diversification of B MADS-box homeotic regulators of flower development: adaptive evolution in protein–protein interaction domains after major gene duplication events. *Molecular Biology and Evolution* **24**: 465–481.

Honma T, Goto K. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409**: 525–529.

Hsu HF, Yang CH. 2002. An orchid (*Oncidium* Gower Ramsey) *AP3*-like MADS gene regulates floral formation and initiation. *Plant and Cell Physiology* **43**: 1198–1209.

Hu JC. 2000. A guided tour in protein interaction space: coiled coils from the yeast proteome. *Proceedings of the National Academy of Sciences, USA* 97: 12935–12936.

Huang H, Tudor M, Su T, Zhang Y, Hu Y, Ma H. 1996. DNA binding properties of two *Arabidopsis* MADS domain proteins: binding consensus and dimer formation. *The Plant Cell* **8**: 81–94.

Immink RGH, Ferrario S, Busscher-Lange J, *et al.* 2003. Analysis of the petunia MADS-box transcription factor family. *Molecular Genetics and Genomics* **268**: 598–606.

Immink RG, Tonaco IA, de Folter S, *et al.* 2009. SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. *Genome Biology* **10**: R24.

Immink RG, Kaufmann K, Angenent GC. 2010. The 'ABC' of MADS domain protein behaviour and interactions. *Seminars in Cell and Developmental Biology* **21**: 87–93.

Kanno A, Saeki H, Kameya T, Saedler H, Theißen G. 2003. Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). *Plant Molecular Biology* **52**: 831–841.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kaufmann K, Melzer R, Theissen G. 2005. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* **347**: 183–198.

Kaufmann K, Pajoro A, Angenent GC. 2010. Regulation of transcription in plants: mechanisms controlling developmental switches. *Nature Reviews Genetics* **11**: 830–842.

Kim ST, Yoo MJ, Albert VA, Farris JS, Soltis PS, Soltis DE. 2004. Phylogeny and diversification of B-function MADS-box genes in angiosperms: evolutionary and functional implications of a 260-million-year-old duplication. *American Journal of Botany* **91**: 2102–2118.

Kim S, Koh J, Yoo MJ, *et al.* 2005. Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. *The Plant Journal* **43**: 724–744.

Kim S, Soltis PS, Soltis DE. 2013. *AGL6*-like MADS-box genes are sister to *AGL2*-like MADS-box genes. *Journal of Plant Biology* **56**: 315–325.

Kramer EM, Dorit RL, Irish VF. 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 EM, Holappa L, Gould B, Jaramillo MA, Setnikov D, Santiago PM. 2007. Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia*. *The Plant Cell* **19**: 750–766.

Lee HL, Irish VF. 2011. Gene duplication and loss in a MADS box gene transcription factor circuit. *Molecular Biology and Evolution* **28**: 3367–3380.

Lenser T, Theissen G, Dittrich P. 2009. Developmental robustness by obligate interaction of class B floral homeotic genes and proteins. *PLoS Computational Biology* **5**: e1000264.

Leseberg CH, Eissler CL, Wang X, Johns MA, Duvall MR, Mao L. 2008. Interaction study of MADS-domain proteins in tomato. *Journal of Experimental Botany* **59**: 2253–2265.

Litt A, Kramer EM. 2010. The ABC model and the diversification of floral organ identity. *Seminars in Cell and Developmental Biology* **21**: 129–137.

Liu CJ, Zhang JA, Zhang N, *et al.* 2010. Interactions among proteins of floral MADS-box genes in basal eudicots: implications for evolution of the regulatory network for flower development. *Molecular Biology and Evolution* **27**: 1598–1611.

Maddison WP, Maddison DR. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75 http://mesquiteproject.org.

Malcomber ST, Kellogg EA. 2005. *SEPALLATA* gene diversification: brave new whorls. *Trends in Plant Science* **10**: 427–4335.

Marquez-Guzman J, Engleman M, Martinez-Mena A, Martinez E, Ramos C. 1989. Reproductive anatomy of *Lacandonia schismatica* (Lacandoniaceae). *Annals of the Missouri Botanical Garden* **76**: 124–127.

McGonigle B, Bouhidel K, Irish VF. 1996. Nuclear localization of the Arabidopsis *APETALA3* and *PISTILLATA* homeotic gene products depends on their simultaneous expression. *Genes and Development* **10**: 1812–1821.

Melzer R, Verelst W, Theissen G. 2009. The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes *in vitro*. *Nucleic Acids Research* **37**: 144–157.

Melzer R, Wang YQ, Theissen G. 2010. The naked and the dead: the ABCs of gymnosperm reproduction and the origin of the angiosperm flower. *Seminars in Cell and Developmental Biology* **21**: 118–128.

Mondragon-Palomino M, Hiese L, Harter A, Koch MA, Theissen G. 2009. Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses. *BMC Evolutionary Biology* **9**: 81.

Mooers AO, Schluter D. 1999. Reconstructing ancestor states with maximum likelihood: support for one- and two-rate models. *Systematic Biology* **48**: 623–633.

Münster T, Wingen LU, Faigl W, Werth S, Saedler H, Theißen G. 2001. Characterization of three *GLOBOSA*-like MADS-box genes from maize: evidence for ancient paralogy in one class of floral homeotic B-function genes of grasses. *Gene* **262**: 1–13.

Nagasawa N, Miyoshi M, Sano Y, *et al.* 2003. *SUPERWOMAN1* and *DROOPING LEAF* genes control floral organ identity in rice. *Development* **130**: 705–718.

Newman JR, Wolf E, Kim PS. 2000. A computationally directed screen identifying interacting coiled coils from *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences, USA* 97: 13203–13208.

Riechmann JL, Krizek BA, Meyerowitz EM. 1996*a*. Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proceedings of the National Academy of Sciences, USA* 93: 4793–4798.

Riechmann JL, Wang MQ, Meyerowitz EM. 1996*b*. DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Research* **24**: 3134–3141.

Rijpkema AS, Zethof J, Gerats T, Vandenbussche M. 2009. The petunia *AGL6* gene has a *SEPALLATA*-like function in floral patterning. *The Plant Journal* **60**: 1–9.

Ronai Z, Wang Y, Khandurina J, *et al.* 2003. Transcription factor binding study by capillary zone electrophoretic mobility shift assay. *Electrophoresis* **24**: 96–100.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Rudall PJ, Remizowa MV, Prenner G, Prychid CJ, Tuckett RE, Sokoloff DD. 2009. Nonflowers near the base of extant angiosperms? Spatiotemporal arrangement of organs in reproductive units of Hydatellaceae and its bearing on the origin of the flower. *American Journal of Botany* **96**: 67–82.

Schwarz-Sommer Z, Hue I, Huijser P, *et al.* 1992. Characterization of the *Antirrhinum* floral homeotic MADS-box gene *deficiens* – evidence for DNA-binding and autoregulation of its persistent expression throughout flower development. *EMBO Journal* **11**: 251–263.

Smaczniak C, Immink RG, Angenent GC, Kaufmann K. 2012. Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* **139**: 3081–3098.

Smirnova E, Shurland DL, Newman-Smith ED, Pishvaee B, van der Bliek AM. 1999. A model for dynamin self-assembly based on binding between three different protein domains. *Journal of Biological Chemistry* **274**: 14942–14947.

Soltis PS, Soltis DE, Kim S, Chanderbali A, Buzgo M. 2006. Expression of floral regulators in basal angiosperms and the origin and evolution of ABC-function. *Advances in Botanical Research* **44**: 483–506.

Soltis DE, Chanderbali AS, Kim S, Buzgo M, Soltis PS. 2007. The ABC model and its applicability to basal angiosperms. *Annals of Botany* **100**: 155–163.

Soltis PS, Brockington SF, Yoo MJ, *et al.* 2009. Floral variation and floral genetics in basal angiosperms. *American Journal of Botany*, **96**: 110–128.

Sundström J, Engström P. 2002. Conifer reproductive development involves B-type MADS-box genes with distinct and different activities in male organ primordia. *The Plant Journal* **31**: 161–169.

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

Theißen G, Melzer R. 2007. Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Annals of Botany* **100**: 603–619.

Theißen G, Saedler H. 2001. Plant biology. Floral quartets. *Nature* **409**: 469–471.

Thompson BE, Bartling L, Whipple C, *et al.* 2009. *bearded-ear* encodes a MADS box transcription factor critical for maize floral development. *The Plant Cell* **21**: 2578–2590.

Vandenbussche M, Zethof J, Royaert S, Weterings K, Gerats T. 2004. The duplicated B-class heterodimer model: whorl-specific effects and complex genetic interactions in *Petunia hybrida* flower development. *The Plant Cell* **16**: 741–754.

Wang YQ, Melzer R, Theissen G. 2010. Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm Gnetum gnemon provide a clue to the evolutionary origin of 'floral quartets'. *The Plant Journal* **64**: 177–190.

Whipple CJ, Schmidt RJ. 2006. Genetics of grass flower development. *Advances in Botanical Research* **44**: 385–424.

Whipple CJ, Ciceri P, Padilla CM, Ambrose BA, Bandong SL, Schmidt RJ. 2004. Conservation of B-class floral homeotic gene function between maize and *Arabidopsis*. *Development* **131**: 6083–6091.

Winter KU, Saedler H, Theißen G. 2002*a*. On the origin of class B floral homeotic genes: functional substitution and dominant inhibition in *Arabidopsis* by expression of an orthologue from the gymnosperm *Gnetum*. *The Plant Journal* **31**: 457–475.

Winter KU, Weiser C, Kaufmann K, *et al.* 2002*b*. Evolution of class B floral homeotic proteins: obligate heterodimerization originated from homodimerization. *Molecular Biology and Evolution* **19**: 587–596.

Yang YZ, Fanning L, Jack T. 2003. The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. *The Plant Journal* **33**: 47–59.

Yao SG, Ohmori S, Kimizu M, Yoshida H. 2008. Unequal genetic redundancy of rice *PISTILLATA* orthologs, *OsMADS2* and *OsMADS4*, in lodicule and stamen development. *Plant and Cell Physiology* **49**: 853–857.

Zahn LM, King HZ, Leebens-Mack JH, *et al.* 2005*a*. The evolution of the *SEPALLATA* subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* **169**: 2209–2223.

Zahn LM, Leebens-Mack J, Depamphilis CW, Ma H, Theißen G. 2005*b*. To B or not to B a flower: the role of *DEFICIENS* and *GLOBOSA* orthologs in the evolution of the angiosperms. *Journal of Heredity* **96**: 225–240.

2.3 Manuscript II

Rümpler, F., Theißen, G., Melzer, R. (2015). Character-state reconstruction to infer ancestral protein-protein interaction patterns. *Bio-protocol* 5, published online.

Character-State Reconstruction to Infer Ancestral Protein-Protein Interaction Patterns

Florian Rümpler[1], Günter Theißen[1] and Rainer Melzer[1, 2*]

[1]Department of Genetics, Friedrich Schiller University Jena, Jena, Germany; [2]School of Biology and Environmental Science, University College Dublin, Dublin, Ireland
[*]For correspondence: rainer.melzer@ucd.ie

[Abstract] Protein-protein interactions are at the core of a plethora of developmental, physiological and biochemical processes. Consequently, insights into the origin and evolutionary dynamics of protein-protein interactions may provide information on the constraints and dynamics of specific biomolecular circuits and their impact on the organismal phenotype.

This protocol describes how ancestral protein-protein interaction patterns can be inferred using a set of known protein interactions from phylogenetically informative species. Although this protocol focuses on protein-protein interaction data, character-state reconstructions can in general be performed with other kinds of binary data in the same way.

Data

A. Protein-protein interaction data

A comprehensive list of interactions for the protein family under study should be compiled. As interaction data are typically generated only for proteins whose sequences have been deposited in databases, a recently published comprehensive phylogeny of the protein family under study may yield an upper estimate of the number and phylogenetic breadth of interaction data to be expected. In many cases recently published phylogenetic relationships need to be extracted from the publications itself, however a growing number of phylogenies are being uploaded in online databases such as TreeBASE (http://treebase.org/treebase-web/home.html) or Dryad (http://datadryad.org/).

1. For obtaining data on protein-protein interactions, databases might be used. Prominent examples of such databases include BioGRID (http://thebiogrid.org/) (Chatr-Aryamontri *et al.*, 2013), the Database of Interacting Proteins (http://dip.doe-mbi.ucla.edu/) (Salwinski *et al.*, 2004), IntAct (http://www.ebi.ac.uk/intact/) (Orchard *et al.*, 2014) and String (http://string-db.org/) (Franceschini *et al.*, 2013), to mention but a few. Above all, UniProt (http://www.uniprot.org/) (UniProt 2015) provides cross-references to a number of these database, thus facilitating searches for potential interaction partners.

2. Whereas database searches provide a good starting point, they very often do not capture all of the information available. It is therefore advisable to undertake a literature search. Special emphasize should be put on obtaining information from phylogenetically informative proteins, *i.e.* from proteins that occupy a position in the phylogeny that is critical for resolving the state of a particular trait (*i.e.* the character-state). Very often these are the early-diverging lineages, as their inclusion (together with more derived taxa) ensures that the whole phylogenetic breadth of a taxonomic group is captured. It might prove useful to obtain new experimental data for proteins that are phylogenetically especially informative. Indeed, generation of new protein-protein interaction data is often combined with character-state reconstruction to better understand the evolution of protein-protein interactions (Liu *et al.*, 2010; Melzer *et al.*, 2014; Li *et al.*, 2015).

B. Sequence retrieval

Protein or nucleotide sequences for phylogenetic reconstructions can be retrieved from the NCBI nucleotide collection (http://www.ncbi.nlm.nih.gov/nuccore) or the NCBI protein collection (http://www.ncbi.nlm.nih.gov/protein).

Software

A. For sequence alignment and subsequent phylogenetic reconstructions one or several of the following programs may be used:

Table 1. Programs for sequence alignments and phylogenetic reconstructions

Program	Purpose	Reference
ExPASy translate	Translation of nucleotide sequences into amino acid sequences.	(Artimo *et al.*, 2012) http://web.expasy.org/translate/
Clustal 2	Sequence alignment. Suited especially for closely related sequences.	(Larkin *et al.*, 2007) http://www.clustal.org/clustal2/
MAFFT7	Sequence alignment. Suited for closely as well as more distantly related sequences.	(Katoh and Standley, 2013) http://mafft.cbrc.jp/alignment/software/
RevTrans 1.4	Converting amino acid alignment into codon alignment.	(Wernersson and Pedersen, 2003) http://www.cbs.dtu.dk/services/RevTrans/
MEGA 6	Sequence alignment and phylogenetic reconstruction.	(Tamura *et al.*, 2013) http://www.megasoftware.net/
MrBayes 3	Phylogenetic reconstruction.	(Ronquist and Huelsenbeck, 2003) http://mrbayes.sourceforge.net/index.php

B. To collate the character matrix
   Microsoft Excel or a similar spreadsheet application
C. For character-state reconstruction:
   Mesquite 3.02 (Maddison and Maddison, 2015) (http://mesquiteproject.org/)
   Mesquite also provides extensive documentation: (http://mesquiteproject.wikispaces.com/)

Procedure

A. Compilation of the character matrix
   A character matrix is constructed that contains the names of the proteins and their interaction properties. This can be done using an Excel spreadsheet. Alternatively, data may be entered directly in Mesquite (Figure 1). It is possible to collate information for several interacting partners in separate columns. To conduct a likelihood character-state reconstruction with Mesquite (see below) data have to be coded categorically, *i.e.* '0' for no interaction and '1' for an interaction. Combinations for which the interactions status is unknown are left blank. Theoretically, one may also introduce three or more categories, *e.g.* '0': no interaction; '1': weak interaction; '2' strong interaction. However, one needs to be aware of the fact that the categories are still discrete and do not follow a hierarchy (*e.g.* there is no constraint such that evolution has to proceed from 'no' to 'weak' to 'strong' interactions).
   Coding of interactions can be complicated by the phylogenetic history of the interaction partner. Consider an example in which protein A interacts with protein B in a certain model organism. In another organism, one ortholog of A, termed A' here, may exist, but two co-orthologs of B, (B' and B") occur. If A' interacts with B' but not with B" it is difficult to assign an interaction status to A' (Figure 2). One compromise is to designate A' as interacting as long as an interaction with either B' or B" is observed (Melzer *et al.*, 2014). The situation gets more complicated if only incomplete data sets are available. Assume, for example, A' is not interacting with B", but information on the interaction between A' and B' is not available. In this case, one may designate the interaction status of A' as unknown to avoid the inclusion of false negatives in the dataset (Melzer *et al.*, 2014). It is difficult to estimate how frequently these problems will appear in a particular dataset. It is therefore important to consider the phylogenetic history of the interaction partner in character-state reconstructions.
   If the interaction data gathered rely on different methods it is helpful to also collect data for each method separately (Figure 1). This will later reveal whether the results of the character-state reconstruction depend on the method used to obtain the interaction data.

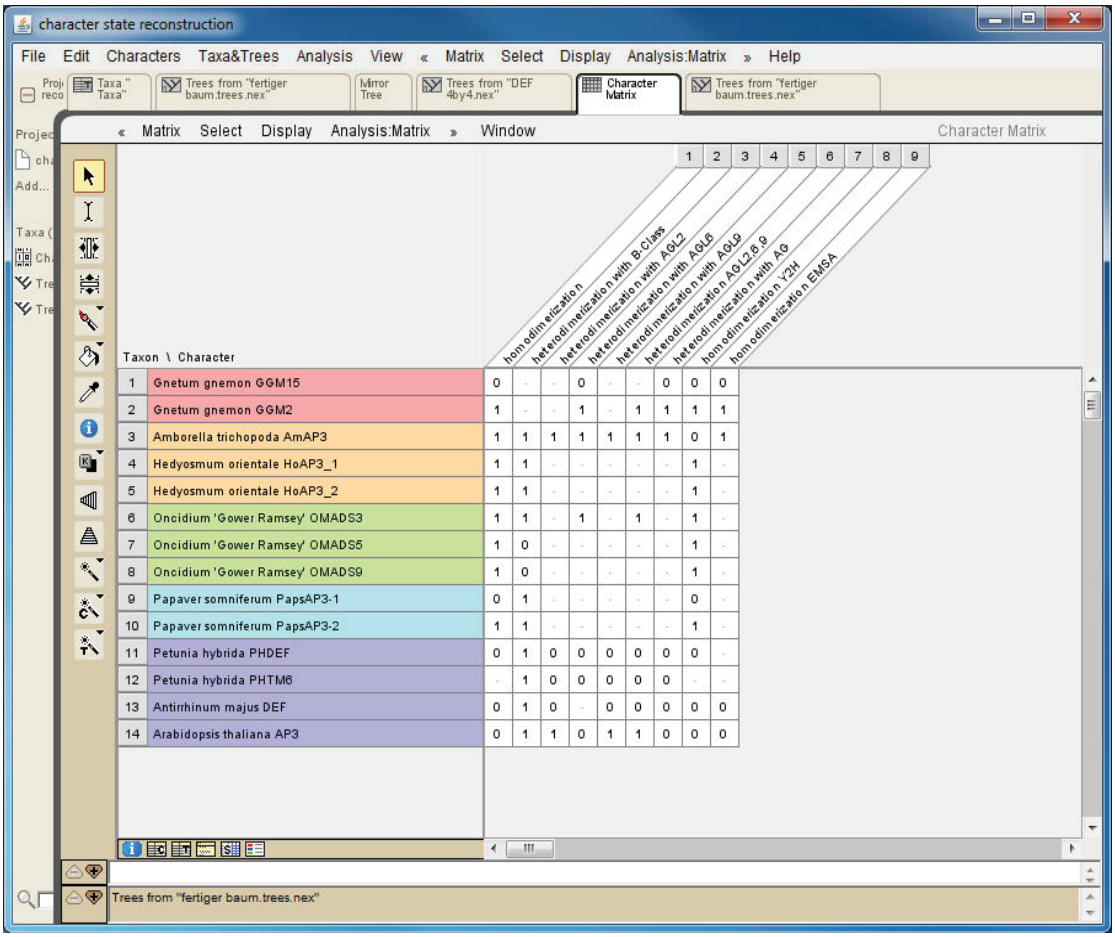


Figure 1. Screenshot from a character matrix in Mesquite. Protein names are listed in the second (coloured) column. Interaction characteristics are listed in subsequent columns. Data on homodimerization as well as on heterodimerization with other proteins and data obtained with different techniques (Y2H: yeast two-hybrid; EMSA: electrophoretic mobility shift assay) are collated in separate columns.

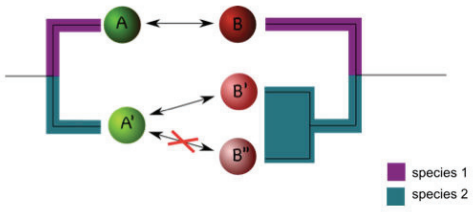


Figure 2. Duplications can complicate coding interactions. Proteins A and B interact in species 1 (as indicated by the double arrow). In species 2, A' interacts with B' but not with B''. This raises the question as how to code the interaction status of A'.

B. Phylogenetic reconstruction

A phylogeny covering all of the proteins under study needs to be constructed using one of the many software tools available (*e.g.* MrBayes, MEGA 6, Bali-Phy, PhyML, see also Table 1). For an overview of basic concepts and methods in phylogeny reconstruction see De Bruyn *et al.* (2014). The phylogeny can be constructed using the sequences of the proteins under study. However, in principle every tree can be used as long as each protein is assigned to a specific position in the tree. Protein names in the character matrix described above and in the phylogenetic tree have to be identical to be later able to connect the two datasets. Mesquite also offers the possibility to manually draw trees; this may be used for cases in which a computational phylogenetic reconstruction is not feasible.

The phylogeny may contain proteins for which interaction data are not available. These will later be ignored by the character-state analysis.

C. Character-state reconstruction

The character-state reconstruction is done using Mesquite. For general instructions on how to handle Mesquite one may visit the 'Mesquite ProjectTeam' YouTube channel (https://www.youtube.com/channel/UCfSmgC0O_dWLI0PEoXZbS4Q).

1. Import/generate tree:

   Mesquite allows to import trees from other files in several ways (http://mesquiteproject.wikispaces.com/Trees). If trees are read from NEXUS files note that Mesquite cannot handle some special characters (*e.g.* dash) if present in protein names. When importing a phylogenetic tree, the branch lengths will later be taken into consideration for the character-state reconstruction. If a manually drawn tree is used, all branch lengths will by default be set to 1. This may work well in a number of cases, but it should be kept in mind that proteins from early diverging taxa may possess artificially short branches under this setting (Figure 3). However, Mesquite also allows editing branch lengths (https://mesquiteproject.wikispaces.com/Trees).

A

Branches as displayed by default when tree is drawn

B

Branches displayed proportional to length

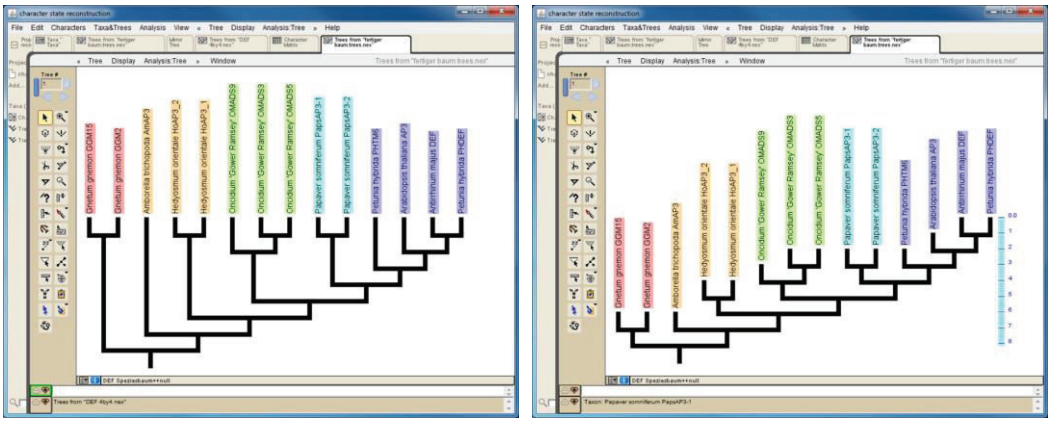


Figure 3. Screenshot of a manually drawn tree in Mesquite. A. Branches are by default displayed so that all tips reach the same level. However, branch lengths are by default set to one, as can be seen in B, where branches of the same tree are displayed proportional to length (see scaling on the right side of the tree). This reveals that some branches (*e.g.* that leading to Amborella trichopoda AmAP3) might be unreasonable short.

2. Import/generate data matrix:

   There are several ways to generate or import a data matrix implemented in Mesquite (http://mesquiteproject.wikispaces.com/Characters+%26+Matrices). A straightforward approach is to generate a new blank data matrix with the required number of characters and copy/paste the interaction data from the original data source (*i.e.* from the Excel spreadsheet). The matrix needs to be specified as categorical to be used for the character state reconstruction.

3. Model specification and character reconstruction:

   Mesquite provides an extensive documentation on the different settings for the character-state reconstruction: http://mesquiteproject.wikispaces.com/Ancestral+States. In our analyses we employed likelihood reconstruction methods (Melzer *et al.*, 2014), but parsimony reconstructions are also available (Li *et al.*, 2015). Two general models can be used for likelihood reconstructions: The 'Markov k-state 1 parameter model' (Mk1) and the 'Asymmetrical Markov k-state 2 parameter model' (AsymmMk). The principal difference between these two models is that the 2 parameter model allows 'forward' and 'backward' rates to be different, *i.e.* the probabilities for gaining and losing an interaction can be different. In the 1 parameter model, gaining and losing an interaction is equally probable. Biologically, it would in most cases make more sense

to apply the 2 parameter model, as one may assume that it is more likely to lose an interaction than gaining it. However, several reports have shown that 2 parameter models can lead to implausible results if small to medium sized datasets (data on less than 100 protein-protein interactions) are being used (Mooers and Schluter, 1999; Pagel, 1999). A likelihood ratio test can be used to infer whether the 2 parameter model significantly improves the fit of the model to the data as compared to the 1 parameter model (Pagel, 1999; Ree and Donoghue, 1999). This test is performed by subtracting the - log probability values derived from the two models and multiplying the absolute value of the result by 2 ($|(-\log L_{Mk1})-(-\log L_{AsymmMk})|\cdot 2$). The resulting number can be used as test statistic for a Chi-square test with one degree of freedom. The test is also integrated in Mesquite an can be conducted via Analysis: Tree > Values for Current Tree > Asymmetry Likelihood Ratio Test.

In the Mk1 and AsymmMk models, the rate of a character's evolution is estimated by Mesquite (http://mesquiteproject.wikispaces.com/Processes+of+Character+Evolution#param). However, it is also possible to create own models with specific parameters (http://mesquiteproject.wikispaces.com/Ancestral+States#editingModels). This can be useful if, for example, the probability of gaining vs. losing an interaction is known from prior experimental evidence.

4. Evaluation of the results:

   Results are best visualized using pie charts at the internal nodes of the tree (Figure 4). Mesquite offers the possibility to conduct the character-state reconstruction simultaneously over different phylogenetic trees. Also, several characters can be traced at once. This facilitates comparison of character-state reconstructions of one protein with different partners or comparison of character-state reconstructions based on different methods used to assay protein-protein interactions.

5. Export options:

   Mesquite can export trees and character matrices in numerous ways (http://mesquiteproject.wikispaces.com/Interactions+with+Other+Programs). For a graphical representation of the character state reconstruction results we recommend to export the tree as PDF and use this file for further post-processing with graphics software such as Adobe Illustrator. For a direct comparison of the character state evolution of two different traits one may utilize the mirror tree function (Figure 4).

Representative data

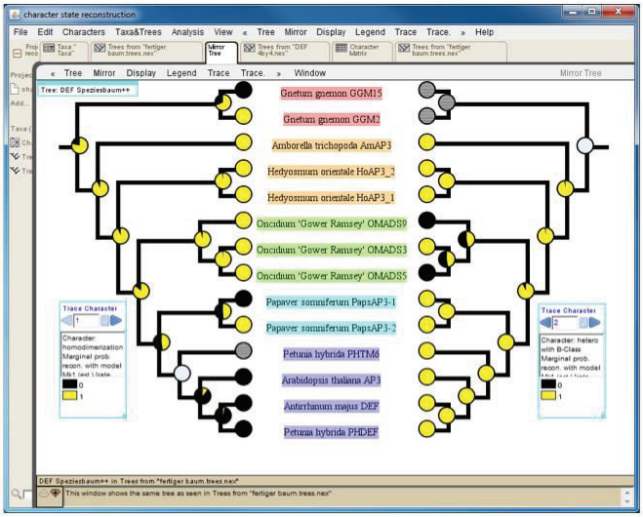


Figure 4. Mirror tree comparing results of character-state reconstruction for homodimerization (left) and heterodimerization (right) of a subfamily of plant transcription factors. Pie charts at internal nodes indicate the probability of the presence (yellow) or absence (black) of an interaction. Hatched circles at terminal positions (*e.g.* Petunia hybrida PHTM6 on the left tree) and grey circles at internal nodes designate an unknown interaction status.

Acknowledgements

This protocol was adapted from a previously published study (Melzer *et al.*, 2014). This research was supported by a DFG grant to G. T. and R. M. (TH417/5–2). R. M. was supported by a post-doctoral fellowship of the Carl-Zeiss-Foundation.

References

1. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I. and Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* 40(Web Server issue): W597-603.
2. Chatr-Aryamontri, A., Breitkreutz, B. J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., O'Donnell, L., Reguly, T., Breitkreutz, A., Sellam, A., Chen, D., Chang, C., Rust, J., Livstone, M., Oughtred, R., Dolinski, K. and Tyers, M.

(2013). The BioGRID interaction database: 2013 update. *Nucleic Acids Res* 41(Database issue): D816-823.

3. De Bruyn, A., Martin, D. P. and Lefeuvre, P. (2014). Phylogenetic reconstruction methods: an overview. *Methods Mol Biol* 1115: 257-277.
4. Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C. and Jensen, L. J. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41(Database issue): D808-815.
5. Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4): 772-780.
6. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21): 2947-2948.
7. Li, L., Yu, X. X., Guo, C. C., Duan, X. S., Shan, H. Y., Zhang, R., Xu, G. X. and Kong, H. Z. (2015). Interactions among proteins of floral MADS-box genes in *Nuphar pumila* (Nymphaeaceae) and the most recent common ancestor of extant angiosperms help understand the underlying mechanisms of the origin of the flower. *Journal of Systematics and Evolution*: n/a-n/a.
8. Liu, C., Zhang, J., Zhang, N., Shan, H., Su, K., Zhang, J., Meng, Z., Kong, H. and Chen, Z. (2010). Interactions among proteins of floral MADS-box genes in basal eudicots: implications for evolution of the regulatory network for flower development. *Mol Biol Evol* 27(7): 1598-1611.
9. Maddison, W. P. and Maddison, D. R. (2015). Mesquite: a modular system for evolutionary analysis. Version 3.02.
10. Melzer, R., Härter, A., Rumpler, F., Kim, S., Soltis, P. S., Soltis, D. E. and Theissen, G. (2014). DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. *Ann Bot* 114(7): 1431-1443.
11. Mooers, A. O. and Schluter, D. (1999). Reconstructing ancestor states with maximum likelihood: Support for one- and two-rate models. *Syst Biol* 48: 623-633.
12. Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N. H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R. C., Meldal, B., Melidoni, A. N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., van Roey, K., Cesareni, G. and Hermjakob, H. (2014). The

MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* 42(Database issue): D358-363.

13. Pagel, M. (1999). The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst Biol* 48: 612-622.
14. Ree, R. H. and Donoghue, M. J. (1999). Inferring rates of change in flower symmetry in asterid angiosperms. *Syst Biol* 48: 633-641.
15. Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12): 1572-1574.
16. Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U. and Eisenberg, D. (2004). The database of interacting proteins: 2004 update. *Nucleic Acids Res* 32(Database issue): D449-451.
17. Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30(12): 2725-2729.
18. UniProt, C. (2015). UniProt: a hub for protein information. *Nucleic Acids Res* 43(Database issue): D204-212.
19. Wernersson, R. and Pedersen, A. G. (2003). RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res* 31(13): 3537-3539.

2.4 Manuscript III

Rümpler, F., Gramzow, L., Theißen, G., Melzer, R. (2015). Did convergent protein evolution enable phytoplasmas to generate 'zombie plants'? *Trends Plant Sci.* 20, 798-806.

CellPress

Opinion

Did Convergent Protein Evolution Enable Phytoplasmas to Generate 'Zombie Plants'?

Florian Rümpler,[1] Lydia Gramzow,[1] Günter Theißen,[1] and Rainer Melzer[1,2,*]

Phytoplasmas are pathogenic bacteria that reprogram plant development such that leaf-like structures instead of floral organs develop. Infected plants are sterile and mainly serve to propagate phytoplasmas and thus have been termed 'zombie plants'. The developmental reprogramming relies on specific interactions of the phytoplasma protein SAP54 with a small subset of MADS-domain transcription factors. Here, we propose that SAP54 folds into a structure that is similar to that of the K-domain, a protein–protein interaction domain of MADS-domain proteins. We suggest that undergoing convergent structural and sequence evolution, SAP54 evolved to mimic the K-domain. Given the high specificity of resulting developmental alterations, phytoplasmas might be used to study flower development in genetically intractable plants.

Trends

The phytoplasma effector protein SAP54 induces the formation of leaf-like organs instead of flowers.

SAP54 specifically binds to the protein–protein interacting K-domain of MADS-domain proteins to reprogram flower development.

Based on its recently solved X-ray crystal structure, the K-domain forms two α-helices separated by a kink region.

In silico analysis of SAP54 indicates that it folds into a structure that is similar to that of the K-domain.

We suggest that SAP54 underwent convergent sequence evolution to mimic the K-domain of its targets. This enabled it to interact with its MADS-domain protein targets and thereby manipulate plant development to its own benefit.

Floral Homeotic Mutants versus Phytoplasma-Infected Zombies: United on Molecular Grounds

Daily life experience provides a number of compelling examples on the precision and conservation of developmental processes. Good cases in point are insects that almost invariably develop a pair of antennas and *Arabidopsis* (*Arabidopsis thaliana*) flowers that virtually always consist of sepals, petals, stamens, and carpels. Barely anything in developmental genetics is more fascinating and famous than **homeotic mutants** (see Glossary) in which these highly conserved developmental programs are confounded. In these types of mutants certain structures are 'changed into the likeness of something else' [1]. For example, the *Antennapedia* mutant from *Drosophila* (*Drosophila melanogaster*) develops legs instead of antennae on its head [2]. Likewise, the *sepallata* (*sep*) mutants from *Arabidopsis* develop a bunch of sepals or leaves instead of different floral organs as parts of flowers [3,4].

However, developmental processes cannot only be tampered by genetic changes as in the examples above. An especially exciting instance for another mechanism is provided by intracellular plant pathogens termed phytoplasmas. These bacteria reside mainly in the phloem tissue and secrete **effector proteins** that can reprogram plant development in a number of impressive ways (Box 1), one of the most striking developmental alterations being a change in floral structure as a means of attracting transmitting insect vectors [5–7]. Typical angiosperm flowers are terminate structures that consist (from outer to inner) of greenish sepals, showy petals, stamens, and carpels (Figure 1A). By contrast, phytoplasma-infected plants develop leaf-like structures instead of floral organs – a symptom termed **phyllody** (Figure 1B). Also

[1]Department of Genetics, Friedrich Schiller University Jena, Jena, Germany
[2]School of Biology and Environmental Science, University College Dublin, Dublin, Ireland

*Correspondence: rainer.melzer@ucd.ie (R. Melzer).

 http://dx.doi.org/10.1016/j.tplants.2015.08.004

Box 1. The Unfolding Universe of Phytoplasma Effector Proteins

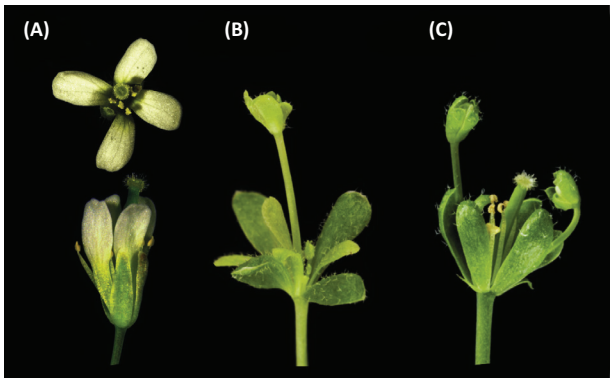
The morphological changes that a plant undergoes upon phytoplasma infection extend well beyond green coloration and organ identity alterations in flowers. Among others, a witches' broom phenotype that originates from the proliferation of shoot branches, resulting in a bushy look of the plant is observed. Also, rosette formation of leaves and dwarfism are typical symptoms of infection [5,6,8]. Many of these symptoms are probably induced by specific effector proteins produced by the phytoplasmas. Effectors are defined as 'all pathogen proteins and small molecules that alter host–cell structure and function' [37]. Based on the presence of an N-terminal signal peptide important for secretion, 56 candidate effector proteins could be identified [44]. Most of the putative effectors are smaller than 40 kDa, which is the upper size exclusion limit of sink tissue plasmodesmata [5,44]. This indicates that the majority of secreted proteins may be able to move out of the phloem and target host cellular processes in other plant tissues [44]. Expression studies revealed that about half of the predicted effector genes show higher transcript levels in plants, whereas the other half exhibits higher expression in insects, suggesting that plant hosts as well as insect vectors are targeted [25].

Beyond SAP54, the functions of two other effectors, SAP11 and TENGU, have been studied in detail [44,59,61–64]. In *Arabidopsis*, SAP11 locates to the plant nucleus and interacts with TCP TFs. The protein induces degradation of a specific subclass of TCP TFs, CIN–TCP proteins [44,59]. When ectopically expressed, SAP11 induces a witches' broom phenotype and the development of crinkly leaves in *Arabidopsis*. Furthermore, transgenic plants also show a decrease in jasmonic acid synthesis [59]. CIN–TCP proteins are implicated in leaf development and the control of jasmonic acid synthesis. Thus, the destabilization of CIN–TCPs is very likely at least partly responsible for the detrimental effects elicited by SAP11 [59].

Expression of TENGU in *Arabidopsis* induces dwarfism and witches' broom [61]. TENGU is processed into small peptides in the plant, the mechanism of how these interfere with plant development remains to be determined [62].

virescence, the green coloration of floral organs, is often observed (Figure 1B) [5,6]. In addition, flowers may lose their determinacy and secondary flowers may develop from within the flower (Figure 1B) [5,6]. The alterations in floral structure are so dramatic that infected plants are often sterile. Several phytoplasma strains possess a wide plant host range and are transmitted by leafhoppers, planthoppers, and related insects that suck phloem sap of infected plants. An infection can thus spread rapidly in a plant population, with devastating consequences for some economically important crops [5,6,8].

Being sterile, phytoplasma-infected plants approach an evolutionary dead end. They mainly propagate and reproduce phytoplasmas and have consequently been termed 'zombie plants' [7,9]. The alterations in floral structure that are induced by phytoplasmas are remarkably reminiscent of phenotypes of some floral homeotic mutants in *Arabidopsis*. In particular, certain *apetala1* (*ap1*) and certain *sep* mutants have strongly similar phenotypes to the ones induced by phytoplasmas as they also possess greenish leaf-like structures within the flower and indeterminate floral growth (compare Figure 1B with 1C) [3,4,7,10,11]. *AP1* and the four *SEP* genes *SEP1*, *SEP2*, *SEP3*, and *SEP4* encode **MADS-domain transcription factors** (TFs) that



Trends in Plant Science

Figure 1. Floral Phenotypes Caused by Phytoplasma Infections Resemble Phenotypes of Floral Homeotic *ap1* Mutants. (A) Healthy *Arabidopsis thaliana* wild-type flower; (B) *A. thaliana* flower from a phytoplasma-infected plant; (C) flower of an *ap1–12* mutant. (B) and (C) are from [7] published under the Creative Commons Attribution (CC BY) license, with kind permission from Saskia Hogenhout.

Glossary

Coiled-coil: a frequent 3D structure in proteins in which the axis of the amino acid α-helix (coil) itself is bent into a helical (coiled) conformation; two or more α-helices can thereby wind around one another in a way that specific hydrophobic and electrostatic side chain interactions are facilitated.
Effector protein: pathogen proteins that are secreted into host cells and alter host–cell structure and function.
Heptad repeat pattern: amino acid sequence feature of the form $[abcdefg]_n$, where most *a* and *d* positions are occupied by highly hydrophobic residues (e.g., leucine, isoleucine, methionine, valine, and alanine) and most *e* and *g* positions are occupied by charged residues (arginine, lysine, aspartic acid, and glutamic acid).
Homeotic mutants: mutants in which one structure (e.g., an organ) develops in a place where normally another structure develops.
Homology: similarity due to common ancestry.
Horizontal gene transfer: transfer of genetic material other than that between parents and offspring. Horizontal gene transfer can occur even between distantly related species, for example, between prokaryotes and eukaryotes.
MADS-domain transcription factors: a family of transcription factors found in almost all eukaryotes that possess the name-giving MADS-domain as DNA-binding domain. The term MADS is an acronym for the founding members of the family: *MCM1* from *Saccharomyces cerevisiae*, *AGAMOUS* from *Arabidopsis thaliana*, *DEFICIENS* from *Antirrhinum majus*, and *SRF* from *Homo sapiens*.
Molecular mimicry: in pathogen–host interactions, the display of any structure by the pathogen that resembles structures of the host at the molecular level and confers a benefit to the pathogen because of this resemblance.
Phyllody: metamorphosis of floral organs into leaf-like structures.
SAP54: secreted AY-WB protein 54; with AY-WB referring to the specific phytoplasma strain.
Sequence convergence: independent emergence of similar features in two unrelated sequences.

govern flower development and floral organ identity [12–14]. The *Arabidopsis* genome encodes some 40 paralogs of the *SEP* genes and *AP1*, collectively belonging to the subfamily of MIKC-type MADS-domain TFs [12,15]. MIKC-type MADS-domain proteins form an intricate network of protein–protein interactions by establishing different dimeric and tetrameric complexes [14,16–21]. These protein complexes can be considered as the functional unit of MIKC-type MADS-domain TFs. Specifically, it has been proposed that DNA-bound tetramers termed 'floral quartets' determine floral organ identity and thus control the development of reproductive tissues in flowering plants [12,22,23]. The four SEP proteins and AP1 are assumed to be essential components of the floral quartets and it is probably the disintegration of these complexes that leads to the severe defects in floral organ development of *sep* and *ap1* mutants [11,14,16,22,24].

The molecular genetic underpinnings of floral organ determination are relatively well understood, whereas light was shed only recently on the molecular mechanism by which phytoplasmas manipulate the flower development of their host plants [7,11,24–26]. The main determinant of altered flower development is a secreted protein termed **SAP54** in one study [7,25]. A close homolog of SAP54 was characterized in parallel in another investigation and termed PHYLLOGEN1 (PHYL1) [11]. Ectopic expression of SAP54/PHYL1 (termed SAP54 henceforth for simplicity) in *Arabidopsis* causes floral defects very similar to those observed upon phytoplasma infection [11,25]. The developmental reprogramming relies on specific interactions of SAP54 with a relatively small subset of MIKC-type MADS-domain TFs [7,11]. Thereby, SAP54 destines the MADS-domain TFs for degradation via the ubiquitin/26S proteasome pathway [7,11,24]. Intriguingly, SAP54 interacts, among others, with AP1 and SEP proteins [7,11]. The phytoplasma infected 'zombie plants' and the floral homeotic mutants are thus not only morphologically similar but their phenotypes may also be brought about by convergent molecular mechanisms that all eventually lead to the depletion of floral homeotic protein complexes that determine organ identity [7,11,24].

Is it the Mimic that Makes the Zombie?

MIKC-type MADS-domain TFs consist of a DNA-binding MADS-domain (M), an Intervening domain (I), a Keratin-like domain (K), and a C-terminal domain (C) [15]. It is this characteristic domain structure from which the term MIKC was derived. The K-domain possesses the second highest conservation (being topped only by the MADS-domain) and is of special importance for protein–protein interactions [12,15]. K-domain sequences exhibit a **heptad repeat pattern** of the form $[abcdefg]_n$ where most *a* and *d* positions are occupied by hydrophobic amino acids and most *e* and *g* positions by charged amino acids (Figure 2, Key Figure) [27–29]. This pattern is typical for α-helices capable of interacting with each other by forming **coiled-coils** [30–32]. The regular spacing of chemically similar amino acids generates a stripe of hydrophobic residues that runs along the α-helix and is flanked by charged residues. Two helices can interact with each other via these stripes, thus constituting coiled-coils [27–29,31,32]. Our views on the structural characteristics of the K-domain were long mainly based just on sequence analyses and computational predictions [15,27,28]. Recently, however, the X-ray crystal structure of the K-domain of SEP3 was determined, showing that indeed intricate coiled-coils are constituted [29]. The K-domain forms two α-helices separated by a kink region (Figure 2) [29]. The α-helices provide a platform for interaction with other SEP3 molecules, whereas the kink region prevents intramolecular association of both helices [29]. Eventually, this enables the formation of a homotetrameric complex consisting of four SEP3 proteins [29]. Given the high degree of sequence conservation of the K-domain [12,29], it is most likely that interactions among other plant MIKC-type MADS-domain TFs, including heteromeric interactions, take place via similar coiled-coils formations. Nevertheless, interactions among MIKC-type MADS-domain TFs are highly specific [16,21]. Very likely, this specificity is at least partly determined by a characteristic pattern of evolutionary conserved hydrophobic and charged amino acids within the K-domain

CellPress

Key Figure

The Secreted Part of SAP54 Shows Similarity to the K-Domain of MIKC-type MADS-Domain Transcription Factors (TFs)

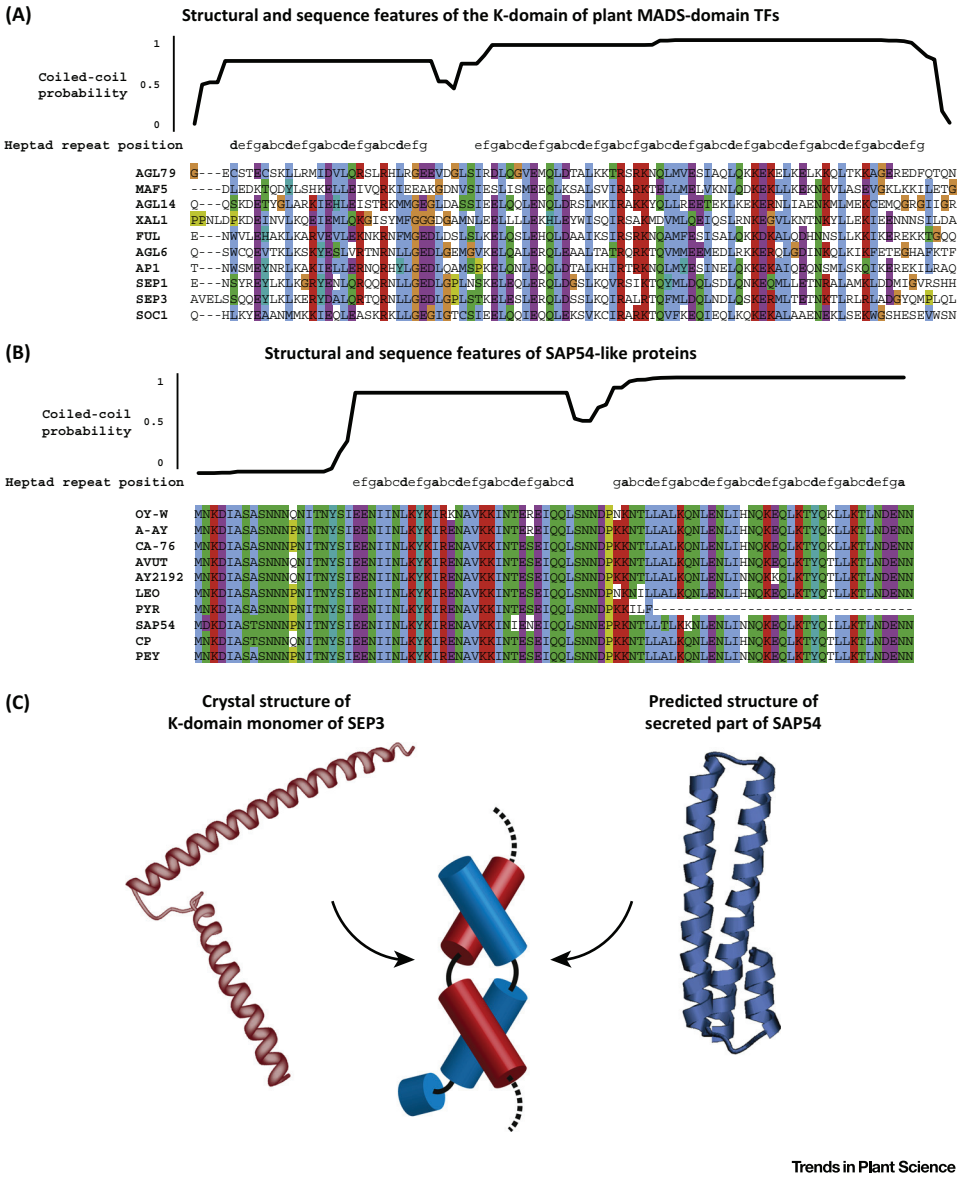


Figure 2. Coiled-coil probability, heptad repeat pattern, and amino acid alignment for (A) the K-domain of MIKC-type MADS-domain TF representatives and (B) the secreted part of SAP54 and its orthologs. Amino acid sequences were aligned with MAFFT applying the G-INS-i mode [65] and visualized in Jalview using the ClustalX coloring scheme[i]. Coiled-coil probability values and heptad repeat positions were calculated with PCOILS[ii] [30] for a sliding window size of 28 amino acids and plotted against the amino acid position. (C) Crystal structure of a K-domain monomer of SEP3 [29] (left side), predicted structure of the secreted part of SAP54 (right side, protein structure prediction was performed with Quark Online[iii] [66]), and hypothesized mode of interaction between SAP54 and its MIKC-type MADS-domain targets (center). For additional alignments and sequence identities between SAP54 and the K-domain of MADS-domain TFs, see Figure S1 and Table S1 in the supplemental information online.

that provide unique interaction interfaces for the different MIKC-type MADS-domain TFs [27–29].

Intriguingly, the K-domain also constitutes the interaction interface for SAP54 [7]. Yeast two-hybrid experiments show that the K-domain of AP1 alone is capable of interacting with SAP54 [7]. This is especially remarkable as MIKC-type MADS-domain TFs are not generally 'sticky' proteins in that they do not interact with all types of other proteins, but almost exclusively with other MIKC-type proteins [14,33]. Thus, the question arises as to how SAP54 evolved towards a protein that can manipulate flower development in such an intricate way. We performed different *in silico* analyses to better understand the primary and 3D structure of SAP54 (Figure 2B,C). These revealed that SAP54 possesses a heptad repeat pattern typical for coiled-coil proteins (Figure 2B). Furthermore, coiled-coil and structural predictions of SAP54 indicate that the protein most likely can constitute two relatively long α-helices separated by a short interhelical region in which a conserved proline 'breaks' the α-helix (Figure 2B,C). This structure is strikingly similar to the X-ray crystal structure of the K-domain of SEP3 (Figure 2) [29]. Based on the similarities between SAP54 and the K-domain of SEP3 found on the level of both primary and higher order structures, we hypothesize that the interaction between SAP54 and SEP3 is probably mediated by a mechanism similar to the one mediating protein interactions among MIKC-type MADS-domain TFs. Since MADS-domain proteins including MIKC-type proteins appear not to exist in prokaryotes [34], we thus propose that **molecular mimicry** of the K-domain enables SAP54 to interact with MIKC-type MADS-domain TFs (Figure 2C).

Molecular mimicry is a widespread phenomenon found in numerous pathogen–host interactions and can evolve either by **horizontal gene transfer** or convergent molecular evolution [35–37]. The secreted part of different SAP54-like proteins and the K-domains of its MADS-domain TF targets show pairwise sequence identities of up to 21% (see Table S1 in the supplemental information online), which is at the lower bound of the 'twilight zone' at which **homology** (i.e., similarity due to common ancestry and thus horizontal gene transfer) can be inferred [38]. However, since coiled-coil domains do, in general, underlie similar sequence constraints, *a priori* similarities for unrelated coiled-coils may be higher than for random sequences [39–41]. Owing to the relatively low sequence identity between SAP54 and the K-domain and to the repetitive nature of coiled-coils, we also found that different alignment algorithms produce different results, further complicating potential homology assignments (see Figure S1 in the supplemental information online). Probabilistic profile hidden Markov models (HMMs) are more sensitive for detecting remote homologies than sequence alignments [42]. We therefore created a sequence profile for the secreted part of SAP54-like proteins based on an alignment of different homologs and searched different databases. These HMM profile searches did not detect sequence profile similarities between the K-domain and SAP54 (Box 2, and see Table S2 in the supplemental information online). Taken together, these considerations favor the hypothesis that SAP54 originated via convergent evolution at the structural and sequence (**sequence convergence**) levels rather than by horizontal gene transfer to mimic the K-domain of MIKC-type MADS-domain TFs.

It remains speculative why plants did not evolve mechanisms to escape the fatal consequences of SAP54 interactions. Hosts do, in general, employ two complementary strategies to combat pathogenic proteins: the immune system may recognize the pathogenic protein and mitigate its detrimental effect and the host target may evolve to escape an interaction with the pathogenic protein [36]. However, with SAP54 mimicking native plant proteins it might be very difficult for the plant immune system to distinguish between friend and foe. At the same time, AP1 and the SEP proteins are major hubs in the protein interaction network controlling flower development [14,16,21]. Any evolutionary change in amino acid sequence of these proteins that weakens the interactions with SAP54 may also compromise essential interactions with other floral

Box 2. The Origin of SAP54: A Jumping Gene from Jumping Insects?

To study the origin of SAP54, we employed HMM searches based on a sequence alignment of SAP54-like sequences against the Conserved Domains Database (CDD) [67] and the Protein families (PFAM) [68] database using the HHPred web server [69]. Surprisingly, not the K-domain of plant MIKC-type MADS-domain TFs but a putative coiled-coil domain of an insect P element transposase possessed the most reliable profile similarity to the secreted part of SAP54-like proteins (see Table S2 in the supplemental information online). The potentially homologous region is very short, resulting in rather low probability values for the detected homology. Furthermore, other potentially homologous proteins with only slightly lower probability values and different functions were found. We thus aimed to evaluate the reliability of our findings by aligning all members of the transposase PFAM entry to the SAP54 sequence profile using HMMER [70]. This revealed that in addition to the regions identified by the HHPred web server, several other regions of SAP54 possess putative homology to the transposases (Figure I). The putative homologous regions are not organized consecutively when comparing the proteins, yielding a complex pattern of overlapping and reversed homology assignments (Figure I).

The identified transposase is also present in true bugs (hemiptera). This order contains leafhoppers and planthoppers, species of which serve as vectors for transmitting phytoplasmas between different host plants. Even though the sequence identity between the secreted part of SAP54 and the insect transposase is in the 'twilight zone' of homology assignments, the possibility exists that the secreted part of SAP54 originated from a coiled-coil domain of an insect transposase via horizontal gene transfer. However, although this hypothesis may constitute a plausible scenario, it is clear that more comprehensive and sophisticated methods for remote homology detection need to be applied in the future to shed light on the origin of SAP54.

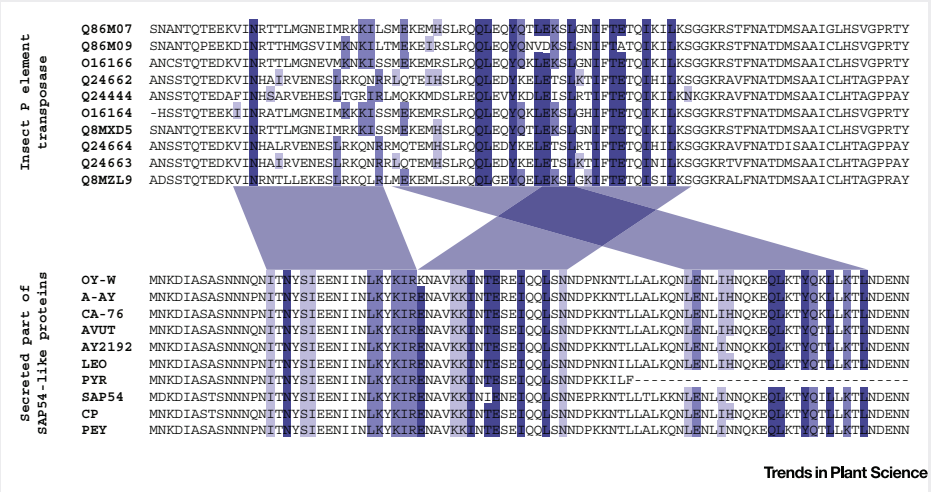


Figure I. Alignments of Some Selected Transposases Encoded by Insect P Element Transposons (top) and SAP54 Orthologs (bottom). Colored bands connect potentially homologous sequence stretches. Amino acid positions are colored when identical in more than 50% of the sequences, deeper blue shadings indicate higher percentages of identity. Only amino acids in the putatively homologous regions were shaded. For more details, see also Tables S2 and S3 in the supplemental information online.

homeotic proteins. The fact that SAP54 interacts with several MADS-domain TFs further complicates the evasion of the host machinery from pathogen attack. The different SEP proteins, AP1, and probably other MADS-domain TFs would have to evolve in parallel to avoid the SAP54 interaction. Thus, SAP54 may provide an impressive example as to how the mimicry of several essential host proteins can be employed to evade host defense mechanisms [36,43].

Our hypothesis does not explain how SAP54 interacts with the ubiquitin–proteasome machinery to eventually induce degradation of the MADS-domain TFs. However, truncation experiments with the SAP54 ortholog PHYL1 have shown that the N-terminal region of PHYL1, that shows no or very low coiled-coil formation probabilities and hence no structural similarities to the K-domain (Figure 2B), is not required for the interaction of PHYL1 and its MADS-domain protein targets [11]. Instead, this region seems to be involved in the recruitment of the ubiquitin–proteasome machinery, possibly by interacting with the ubiquitin receptor RAD23 [7,11].

Where did SAP54 Originate?

Assuming that SAP54 evolved via convergent evolution, the question remains as to from which ancestral protein sequence it originated. SAP54 possesses a short signal peptide, which mediates secretion into host plant cells, and that is also found in a number of other phytoplasma effector proteins [7,11,44]. However, our analyses indicated that the secreted part of SAP54 that interacts with the K-domain shows no sequence similarity to effector proteins other than putative orthologs of SAP54. We found neither in phytoplasmas nor in more distantly related bacteria sequences that could explain the molecular origin of SAP54 (see Table S3 in the supplemental information online). Surprisingly, the HMM profile searches mentioned above indicated that SAP54 may be homologous to an insect transposase, opening the possibility that a horizontal gene transfer from a transmitting insect may have been the initial step in the origin of SAP54 (Box 2).

It should also be noted that phytoplasmas are not only remarkable for reprograming plant development but they also bear one of the smallest bacterial genomes known to date [45]. Despite being so small, the genome harbors a significant amount of repetitive sequences that can make up some 20% of the genomic DNA [45,46]. Most of this repetitive DNA is organized in few relatively large 'potential mobile units' (PMUs) that possess features of transposons [46,47]. Phytoplasmas display a high degree of genome plasticity and PMUs probably play a vital role in generating and maintaining this plasticity [46,48]. Intriguingly, the gene encoding SAP54 is part of a PMU [44,46,48]. It is thus tempting to speculate that the evolution of SAP54 and its location in a PMU are causally linked. However, the molecular mechanisms that contributed to the origin of SAP54 remain to be determined.

One Protein to Rule Them All?

Phytoplasmas have a remarkably wide range of host plants, in many of which they induce the development of greenish flowers [6–8]. At the same time, as far as we know, all flowering plants possess SEP and AP1 orthologs, and the function of these proteins in flower development is highly conserved [13,49]. Thus, SAP54-mediated degradation of floral homeotic proteins may contribute to the similarities in phenotypes seen in a variety of infected hosts [7,24].

Beyond important crop and model plants, phytoplasma infections have also been described for early diverging angiosperms and even for some gymnosperm species [8,50–53]. The genetics of flower and cone development in these evolutionary interesting taxa is not well understood, primarily because these species are genetically intractable [54–58]. It will be very interesting to see whether phytoplasmas manipulate flower and cone development in early diverging angiosperms and gymnosperms, respectively, to the same extent as is observed in monocots and eudicots, and whether SAP54 interacts with orthologs of AP1 and SEP proteins in a similar manner as in *Arabidopsis*. A detailed account of the morphological and molecular consequences of phytoplasma infections in phylogenetically informative species could provide us with important information on the conservation of the gene regulatory program controlling reproductive development and thus eventually lead to a better understanding of the origin of the flower during evolution.

Concluding Remarks and Future Perspectives

Based on the presented data, we propose that SAP54 folds into a structure that is similar to that of the K-domain of MIKC-type MADS-domain TFs. As a result of convergent structural and sequence evolution, SAP54 may have evolved to mimic the K-domain of its targets and thus to reprogram plant development such that leaf-like structures instead of floral organs develop.

However, degradation of some floral homeotic proteins may not be the only mechanism by which phytoplasmas bring about aberrant floral phenotypes, even though SAP54 orthologs might be involved [26]. For example, a recent study provided evidence that a SAP54 ortholog inhibits the expression of a microRNA (miR396), which inhibits the translation of SHORT

Outstanding Questions

Do SAP54-like proteins indeed adopt a structure similar to that of the K-domain of plant MADS-domain TFs, which enables them to interact with some of these TFs? The *in silico* predictions presented in this opinion article strongly suggest structural similarities of SAP54 and its target proteins. However, experimental methods such as X-ray crystal structure determination of a SAP54/MADS-domain–TF complex will be needed to test our hypothesis about the structure and the mode of interaction.

Is molecular mimicry a widespread phenomenon among phytoplasma effector proteins? The vast majority of predicted secreted proteins may be able to move out of the phloem and target host cellular processes. The high sequence variability of their secreted part indicates that effector proteins fold into diverse tertiary structures that may mimic host cellular components.

Did *SAP54*-like genes enter phytoplasmas by horizontal gene transfer? No hints for the molecular origin of *SAP54* were found in phytoplasmas or in more distantly related bacteria. The life cycle of phytoplasmas as intracellular pathogens in their host plants and insect vectors may have facilitated the exchange of genetic material, opening the possibility that *SAP54* (and maybe other genes encoding effector proteins) originated through horizontal gene transfer.

Could SAP54-like proteins serve as molecular tools to study flower development in genetically intractable species? Phytoplasmas have been shown to infect some evolutionary interesting taxa that are genetically intractable. Studying the effect of phytoplasma infections in these species may yield information on the evolution of genes controlling reproductive development that is otherwise difficult to obtain.

VEGETATIVE PHASE (SVP), yet another MADS-domain protein [26]. Therefore, in phytoplasma-infected plants SVP is upregulated, which may also contribute to abnormal flower development [26].

Furthermore, phytoplasma infections do not only cause phyllody but also induce dwarfism and witches' broom phenotypes [5,8]. Additionally, the bacteria also influence insect host behavior and longevity [5,8]. Many of the developmental alterations induced by phytoplasmas are probably corroborated by effector proteins that interfere with the host cellular machinery in a very specific manner (Box 1). However, despite the considerable progress that has been made in recent years, the origin and evolutionary dynamics of most effector proteins remains largely elusive, and the ecological advantages of evolving such intricate proteins are only beginning to be explored (see Outstanding Questions) [7,48,59].

The study of host–pathogen interactions has repeatedly provided important insights into plant physiology and developmental biology [60]. The broad host range of some phytoplasmas and the specific effects caused by the effector proteins makes them ideally suited to better understand the evolutionary conservation of plant developmental processes.

Clearly, we are only beginning to explore the many facets of phytoplasma biology. The most exciting years in 'zombie research' might be still ahead of us.

Acknowledgments

R.M. received a post-doctoral fellowship from the Carl Zeiss Foundation. This research was supported by a DFG (Deutsche Forschungsgemeinschaft) grant to G.T. and R.M. (TH417/5-2).

Supplemental Information

Supplemental information associated with this article can be found online at http://dx.doi.org/10.1016/j.tplants.2015.08.004.

Resources

[i] www.jalview.org

[ii] http://toolkit.tuebingen.mpg.de/pcoils

[iii] http://zhanglab.ccmb.med.umich.edu/QUARK/

References

1. Bateson, W. (1894) *Materials for the Study of Variation Treated with Especial regard to Discontinuity in the Origin of Species,* Macmillan
2. Lindsley, D.L. and Grell, E.H. (1968) *Genetic Variations of Drosophila melanogaster,* Carnegie Institution of Washington
3. Ditta, G. *et al.* (2004) The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr. Biol.* 14, 1935–1940
4. Pelaz, S. *et al.* (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* 405, 200–203
5. Hogenhout, S.A. *et al.* (2008) Phytoplasmas: bacteria that manipulate plants and insects. *Mol. Plant Pathol.* 9, 403–423
6. Maejima, K. *et al.* (2014) Exploring the phytoplasmas, plant pathogenic bacteria. *J. Gen. Plant Pathol.* 80, 210–221
7. MacLean, A.M. *et al.* (2014) Phytoplasma effector SAP54 hijacks plant reproduction by degrading MADS-box proteins and promotes insect colonization in a RAD23-dependent manner. *PLoS Biol.* 12, e1001835
8. Marcone, C. (2014) Molecular biology and pathogenicity of phytoplasmas. *Ann. Appl. Biol.* 165, 199–221
9. Du Toit, A. (2014) Phytoplasma converts plants into zombies. *Nat. Rev. Microbiol.* 12, 393
10. Bowman, J.L. *et al.* (1993) Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development* 119, 721–743
11. Maejima, K. *et al.* (2014) Recognition of floral homeotic MADS domain transcription factors by a phytoplasmal effector, phyllogen, induces phyllody. *Plant J.* 78, 541–554
12. Gramzow, L. and Theissen, G. (2010) A hitchhiker's guide to the MADS world of plants. *Genome Biol.* 11, 214
13. Theißen, G. and Melzer, R. (2007) Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann. Bot.* 100, 603–619
14. Immink, R.G. *et al.* (2010) The 'ABC' of MADS domain protein behaviour and interactions. *Semin. Cell Dev. Biol.* 21, 87–93
15. Kaufmann, K. *et al.* (2005) MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347, 183–198
16. Immink, R.G. *et al.* (2009) SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. *Genome Biol.* 10, R24
17. Jetha, K. *et al.* (2014) *Arabidopsis* SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. *Nucleic Acids Res.* 42, 10927–10942
18. Melzer, R. and Theißen, G. (2009) Reconstitution of 'floral quartets' in vitro involving class B and class E floral homeotic proteins. *Nucleic Acids Res.* 37, 2723–2736
19. Melzer, R. *et al.* (2009) The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro. *Nucleic Acids Res.* 37, 144–157

20. Smaczniak, C. *et al.* (2012) Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1560–1565
21. de Folter, S. *et al.* (2005) Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. *Plant Cell* 17, 1424–1433
22. Theißen, G. (2001) Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* 4, 75–85
23. Theißen, G. and Saedler, H. (2001) Plant biology. Floral quartets. *Nature* 409, 469–471
24. Maejima, K. *et al.* (2015) Degradation of class E MADS-domain transcription factors in *Arabidopsis* by a phytoplasmal effector, phyllogen. *Plant Signal. Behav* Published online July 15, 2015. http://dx.doi.org/10.1080/15592324.2015.1042635
25. MacLean, A.M. *et al.* (2011) Phytoplasma effector SAP54 induces indeterminate leaf-like flower development in *Arabidopsis* plants. *Plant Physiol.* 157, 831–841
26. Yang, C-Y. *et al.* (2015) MicroRNA396-targeted *SHORT VEGETATIVE PHASE* is required to repress flowering and is related to the development of abnormal flower symptoms by the Phyllody Symptoms1 effector. *Plant Physiol.* 168, 1702–1716
27. Yang, Y.Z. *et al.* (2003) The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. *Plant J.* 33, 47–59
28. Yang, Y.Z. and Jack, T. (2004) Defining subdomains of the K domain important for protein–protein interactions of plant MADS proteins. *Plant Mol. Biol.* 55, 45–59
29. Puranik, S. *et al.* (2014) Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in *Arabidopsis*. *Plant Cell* 26, 3603–3615
30. Lupas, A. *et al.* (1991) Predicting coiled coils from protein sequences. *Science* 252, 1162–1164
31. Mason, J.M. and Arndt, K.M. (2004) Coiled coil domains: stability, specificity, and biological implications. *Chembiochem* 5, 170–176
32. Parry, D.A. *et al.* (2008) Fifty years of coiled-coils and α-helical bundles: a close relationship between sequence and structure. *J. Struct. Biol.* 163, 258–269
33. Melzer, R. *et al.* (2006) Missing links: DNA-binding and target gene specificity of floral homeotic proteins. *Adv. Bot. Res.* 44, 209–236
34. Gramzow, L. *et al.* (2010) On the origin of MADS-domain transcription factors. *Trends Genet.* 26, 149–153
35. Stebbins, C.E. and Galan, J.E. (2001) Structural mimicry in bacterial virulence. *Nature* 412, 701–705
36. Elde, N.C. and Malik, H.S. (2009) The evolutionary conundrum of pathogen mimicry. *Nat. Rev. Microbiol.* 7, 787–797
37. Hogenhout, S.A. *et al.* (2009) Emerging concepts in effector biology of plant-associated organisms. *Mol. Plant Microbe Interact.* 22, 115–122
38. Rost, B. (1999) Twilight zone of protein sequence alignments. *Protein Eng.* 12, 85–94
39. Mistry, J. *et al.* (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41, e121
40. Surkont, J. and Pereira-Leal, J.B. (2015) Evolutionary patterns in coiled-coils. *Genome Biol. Evol.* 7, 545–556
41. Kuhn, M. *et al.* (2014) Coiled-coil proteins facilitated the functional expansion of the centrosome. *PLoS Comput. Biol.* 10, e1003657
42. Fariselli, P. *et al.* (2007) The WWWH of remote homolog detection: the state of the art. *Brief. Bioinform.* 8, 78–87
43. Deslandes, L. and Rivas, S. (2012) Catch me if you can: bacterial effectors and plant targets. *Trends Plant Sci.* 17, 644–655
44. Bai, X. *et al.* (2009) AY-WB phytoplasma secretes a protein that targets plant cell nuclei. *Mol. Plant Microbe Interact.* 22, 18–30
45. Oshima, K. *et al.* (2004) Reductive evolution suggested from the complete genome sequence of a plant–pathogenic phytoplasma. *Nat. Genet.* 36, 27–29
46. Bai, X.D. *et al.* (2006) Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. *J. Bacteriol.* 188, 3682–3696
47. Toruno, T.Y. *et al.* (2010) Phytoplasma PMU1 exists as linear chromosomal and circular extrachromosomal elements and has enhanced expression in insect vectors compared with plant hosts. *Mol. Microbiol.* 77, 1406–1415
48. Sugio, A. and Hogenhout, S.A. (2012) The genome biology of phytoplasma: modulators of plants and insects. *Curr. Opin. Microbiol.* 15, 247–254
49. Gramzow, L. and Theissen, G. (2015) Phylogenomics reveals surprising sets of essential and dispensable clades of MIKC-group MADS-box genes in flowering plants. *J. Exp. Zool. B: Mol. Dev. Evol.* 324, 353–362
50. Kaminska, M. *et al.* (2001) The association of phytoplasma with stunting, leaf necrosis and witches' broom symptoms in *Magnolia* plants. *J. Phytopathol.* 149, 719–724
51. Li, Z.Y. *et al.* (2012) First report of elm yellows subgroup 16SrV-B *Phytoplasma* infecting Chinese tulip tree in China. *Plant Dis.* 96, 1064
52. Schneider, B. *et al.* (2005) '*Candidatus* Phytoplasma pini', a novel taxon from *Pinus silvestris* and *Pinus halepensis*. *Int. J. System. Evol. Microbiol.* 55, 303–307
53. Jezic, M. *et al.* (2013) '*Candidatus* Phytoplasma pini' in pine species in Croatia. *J. Plant Dis. Protect.* 120, 160–163
54. Wang, Y.Q. *et al.* (2010) Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm *Gnetum gnemon* provide a clue to the evolutionary origin of 'floral quartets'. *Plant J.* 64, 177–190
55. Melzer, R. *et al.* (2010) The naked and the dead: the ABCs of gymnosperm reproduction and the origin of the angiosperm flower. *Semin. Cell Dev. Biol.* 21, 118–128
56. Melzer, R. *et al.* (2014) DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. *Ann. Bot.* 114, 1431–1443
57. Amborella Genome Project (2013) The *Amborella* genome and the evolution of flowering plants. *Science* 342, 1241089
58. Frohlich, M.W. and Chase, M.W. (2007) After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450, 1184–1189
59. Sugio, A. *et al.* (2011) Phytoplasma protein effector SAP11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 108, E1254–E1263
60. Win, J. *et al.* (2012) Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb. Symp. Quant. Biol.* 77, 235–247
61. Hoshi, A. *et al.* (2009) A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6416–6421
62. Sugawara, K. *et al.* (2013) The alteration of plant morphology by small peptides released from the proteolytic processing of the bacterial peptide TENGU. *Plant Physiol.* 162, 2005–2014
63. Minato, N. *et al.* (2014) The phytoplasmal virulence factor TENGU causes plant sterility by downregulating of the jasmonic acid and auxin pathways. *Sci. Rep.* 4, 7399
64. Sugio, A. *et al.* (2014) The small phytoplasma virulence effector SAP11 contains distinct domains required for nuclear targeting and CIN-TCP binding and destabilization. *New Phytol.* 202, 838–848
65. Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780
66. Xu, D. and Zhang, Y. (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins Struct. Funct. Bioinform.* 80, 1715–1735
67. Sayers, E.W. *et al.* (2012) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 40, D13–D25
68. Finn, R.D. *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230
69. Soding, J. *et al.* (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248
70. Eddy, S.R. (2009) A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 23, 205–211

2.5 Manuscript IV

Theißen, G., Melzer, R., Rümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development* 143, 3259-3271.

© 2016. Published by The Company of Biologists Ltd | Development (2016) 143, 3259-3271 doi:10.1242/dev.134080

REVIEW

MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution

Günter Theißen[1,*], Rainer Melzer[2] and Florian Rümpler[1]

ABSTRACT

The floral quartet model of floral organ specification poses that different tetramers of MIKC-type MADS-domain transcription factors control gene expression and hence the identity of floral organs during development. Here, we provide a brief history of the floral quartet model and review several lines of recent evidence that support the model. We also describe how the model has been used in contemporary developmental and evolutionary biology to shed light on enigmatic topics such as the origin of land and flowering plants. Finally, we suggest a novel hypothesis describing how floral quartet-like complexes may interact with chromatin during target gene activation and repression.

KEY WORDS: ABC model, Floral homeotic gene, MIKC-type MADS-domain transcription factor, Nucleosome, Pioneer transcription factor, Plant evolution

Introduction

Flowers are frequently composed of four different classes of organs arranged in whorls, with sepals in the first floral whorl, petals in the second whorl, stamens (male reproductive organs) in the third whorl and carpels (female organs) in the fourth whorl. Understanding how these distinct floral organs are specified has been a long-standing challenge in plant developmental genetics (Meyerowitz et al., 1989; Schwarz-Sommer et al., 1990; Coen and Meyerowitz, 1991; Irish, 2010). According to the floral quartet model (FQM), which was proposed in 2001, the identity of the different floral organs is specified during development by quaternary (tetrameric) protein complexes composed of MIKC-type MADS-domain proteins (see Glossary, Box 1; Theißen, 2001). These quartets are assumed to function as transcription factors by binding to the DNA of their target genes, which they either activate or repress to control the development of the respective floral organs (Theißen, 2001).

MIKC-type MADS-box genes (see Glossary, Box 1; Münster et al., 1997; Kaufmann et al., 2005b) encode proteins that exhibit a characteristic domain organization that includes (from N- to C-terminus): a MADS (M) domain, an intervening (I) domain, a keratin-like (K) domain, and a C-terminal (C) domain (Theißen et al., 1996; Kaufmann et al., 2005b). The MADS domain is by far the most highly conserved region of all kinds of MADS-domain proteins, including MIKC-type proteins. It represents a DNA-binding domain but is also important for the dimerization and nuclear localization of MADS-domain transcription factors (Gramzow and Theißen, 2010). The I domain, by contrast, is only relatively weakly conserved and contributes to the selective formation of DNA-binding dimers (Kaufmann et al., 2005b). The K domain is characterized by a conserved, regular spacing of hydrophobic and charged residues, which allows the formation of amphipathic helices involved in protein dimerization and multimeric complex formation (Yang et al., 2003; Puranik et al., 2014). Finally, the C domain is quite variable and, in some MADS-domain proteins, is involved in transcriptional activation or multimeric complex formation (for a review on structural and phylogenetic aspects of MIKC-type proteins, see Kaufmann et al., 2005b; Theißen and Gramzow, 2016).

MADS-domain proteins bind as dimers to DNA sequences termed 'CArG-boxes' (see Glossary, Box 1; reviewed by Theißen and Gramzow, 2016). According to the FQM, two protein dimers of each tetramer recognize two different CArG-boxes and bring them into close vicinity by looping the DNA between the CArG-boxes (Theißen, 2001; Theißen and Saedler, 2001). In recent years, the remarkable capacity of MIKC-type proteins to constitute multimeric transcription factor complexes, together with the importance of these complexes in plant development and evolution, has been increasingly recognized. However, the heuristic value of the FQM in plant developmental and evolutionary biology has not yet been fully explored. To stimulate further research, we revisit the FQM and review the current status of the field. We first provide a short history of the FQM, summarize its recent experimental support, and outline its use in current research. We also propose a simplified and generic version of the FQM that helps to harmonize genetic and molecular models of floral organ identity specification. Finally, we discuss major open questions regarding floral quartet-like protein complexes (FQCs; see Glossary, Box 1), concerning their molecular mode of action during the activation or repression of target genes.

A brief history of the floral quartet model

The scientific journey that eventually led to the development of the FQM started with the analysis of mutants in which all or some organs of the flower had been replaced with organs of another identity, a phenomenon known as 'homeosis'. Mutational changes in floral organ identity have been known from many species and have fascinated humans for over centuries (Meyerowitz et al., 1989). It turned out that many of the respective mutants, termed floral homeotic mutants, including those of *Arabidopsis thaliana* (thale cress) and *Antirrhinum majus* (snapdragon), fall into three classes, termed A, B and C (Bowman et al., 1991; Coen and Meyerowitz, 1991). In ideal class A mutants, sepals are replaced by carpels and petals are substituted by stamens. In class B mutants, sepals instead of petals and carpels instead of stamens develop. In class C mutants, stamens are replaced by petals and carpels are substituted by sepals. The typical determinate growth of flowers is also often abolished in class C mutants, so that a potentially unlimited series of additional mutant flowers develops inside the primary mutant flower.

[1]Department of Genetics, Friedrich Schiller University Jena, 07743 Jena, Germany. [2]School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland.

*Author for correspondence (guenter.theissen@uni-jena.de)

G.T., 0000-0003-4854-8692

Development (2016) 143, 3259-3271 doi:10.1242/dev.134080

Box 1. Glossary

Angiosperms. Flowering plants *sensu stricto*. They produce seeds from ovules contained in ovaries ('vessel seeds') that develop into fruits.

CArG-box. 'C-Arich-G-box': a DNA-sequence motif bound by MADS-domain proteins, with the consensus sequence 5′-CC(A/T)$_6$GG-3′ or a similar sequence.

Floral quartet-like complex (FQC). A complex of four MIKC-type proteins that binds to two CArG-boxes involving looping of the DNA connecting the CArG-boxes.

Gymnosperms. Seed-bearing plants with ovules that are not contained in ovaries and hence develop as 'naked seeds'. Angiosperms very likely evolved from some (unknown) group of gymnosperms.

MADS-box gene. A gene containing a MADS box, which encodes the DNA-binding and nuclear-localization domain of the respective MADS-domain transcription factors. The acronym 'MADS' refers to the four founder genes *MINICHROMOSOME MAINTENANCE FACTOR1* (*MCM1*; from *Saccharomyces cerevisiae*), *AGAMOUS* (*AG*; from *Arabidopsis thaliana*), *DEFICIENS* (*DEF*; from *Antirrhinum majus*) and *SERUM RESPONSE FACTOR* (*SRF*; from *Homo sapiens*).

MIKC-type MADS-domain protein. A MADS-domain protein that exhibits a characteristic domain structure including a DNA-binding MADS (M) domain, an Intervening (I) domain, a keratin-like (K) domain and a C-terminal (C) domain.

Pioneer transcription factor (PTF). A transcription factor that can bind to nucleosome-associated DNA sites, possibly by evicting nucleosomes.

Based on these frequently found classes of homeotic mutants, simple genetic models were proposed and successfully tested by analysing double and triple mutants (for a historical perspective, see Theißen, 2001; Theißen and Melzer, 2006; Causier et al., 2010; Irish, 2010; Bowman et al., 2012). Arguably the most well-known of these models is the 'ABC model' as outlined by Bowman et al. (1991) and Coen and Meyerowitz (1991). It maintains that organ identity in each whorl is specified by a unique combination of three homeotic functions, termed A, B and C, which are accomplished by floral organ identity genes. Expression of the A function alone specifies sepal formation. The combination AB specifies the development of petals, while the combination BC specifies the formation of stamens. Expression of C alone determines the development of carpels. In order to explain the three classes of floral homeotic mutants, the ABC model proposes that the A- and C-function genes negatively regulate each other, so that the C function becomes expressed throughout the flower when the A function is mutated and vice versa (for reviews of the ABC model, see Theißen, 2001; Krizek and Fletcher, 2005; Bowman et al., 2012; Wellmer et al., 2014).

Subsequent genetic analyses identified five different genes that provide floral homeotic functions in *A. thaliana*. The A function is mediated by *APETALA1* (*AP1*) and *APETALA2* (*AP2*), the B function by *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), and the C function by *AGAMOUS* (*AG*). All of these genes encode putative transcription factors (Yanofsky et al., 1990; Jack et al., 1992; Mandel et al., 1992; Goto and Meyerowitz, 1994; Jofuku et al., 1994; for a review, see Theißen, 2001; Ó'Maoiléidigh et al., 2014), suggesting that ABC genes may control the transcription of other genes ('target genes') whose products are directly or indirectly involved in the formation or function of floral organs. Except for *AP2*, all ABC genes encode MIKC-type MADS-domain proteins (Irish, 2010).

The ABC model was attractively simple, but it soon revealed important shortcomings. For example, mutant and transgenic studies indicated that the ABC genes are required but usually not sufficient for the specification of floral organ identity, i.e. when the ABC genes were expressed outside the floral context they could not, in most cases, induce floral organ development from leaf primordia (Krizek and Meyerowitz, 1996a,b; Mizukami and Ma, 1992; Pelaz et al., 2001). It turned out that additional homeotic functions had escaped forward genetic approaches. Indeed, based on studies in petunia (*Petunia hybrida*), the ABC model was extended to an 'ABCD' model by addition of a D function specifying ovule identity (Angenent and Colombo, 1996). In *A. thaliana*, three genes closely related to *AG*, namely *SEEDSTICK* (*STK*; formerly known as *AGL11*), *SHATTERPROOF1* (*SHP1*; formerly known as *AGL1*) and *SHATTERPROOF2* (*SHP2*; formerly known as *AGL5*) (Favaro et al., 2003; Pinyopich et al., 2003) were identified as D-function genes; *stk shp1 shp2* triple mutants are characterized by conversion of ovules into carpel-like or leaf-like structures (Pinyopich et al., 2003). The C-function gene *AGAMOUS* was also considered as an additional class D gene (e.g. Theißen and Melzer, 2006), but reconciliation of the FQM with the genetic models suggests a more elegant solution (discussed below).

Knocking out another class of MIKC-type MADS-box genes, initially known as *AGL2*-like genes, but later termed *SEPALLATA*-like genes, revealed additional floral organ identity genes (Pelaz et al., 2000; Ditta et al., 2004). Owing to functional redundancy, single and double mutants of *SEPALLATA1* (*SEP1*, formerly known as *AGL2*), *SEP2* (*AGL4*), *SEP3* (*AGL9*) or *SEP4* (*AGL3*) exhibit only weak mutant phenotypes, if any (Pelaz et al., 2000; Ditta et al., 2004). However, in *sep1 sep2 sep3* triple mutants, the organs in all whorls of the flower develop into sepals, and flower development becomes indeterminate (Pelaz et al., 2000); in *sep1 sep2 sep3 sep4* quadruple mutants, vegetative leaves rather than sepals develop in all whorls of indeterminate 'flowers' (Ditta et al., 2004). The function provided by the *SEP* genes was initially considered as a combined B/C function (Pelaz et al., 2000). However, since it had been shown that the initial expression patterns of class B and C genes are not altered in the *sep1 sep2 sep3* triple mutant (Pelaz et al., 2000) and to avoid confusion with the previously defined D function specifying ovule identity, it was proposed that *SEP* genes, rather than acting upstream or downstream of the floral homeotic genes, could constitute yet another class of redundant floral organ identity genes, for which the term 'class E genes' was suggested (Theißen, 2001). The corresponding 'ABCDE' model maintains that class A+E genes specify sepals, A+B+E specify petals, B+C+E specify stamens, C+E specify carpels and C+D+E specify ovules (Fig. 1; Theißen, 2001; Ditta et al., 2004; note that, in case of ovules, we deviate from previous views that considered *AG* to be a C+D gene and now we classify it only as a class C gene, but also consider the C function to be involved in ovule specification). Importantly, several lines of data (Honma and Goto, 2001; Pelaz et al., 2001) strongly suggested that the ABCDE genes are not only necessary, but also sufficient to superimpose floral organ identity upon vegetative developmental programs of angiosperms (see Glossary, Box 1), even though it has remained unclear up to now as to whether C+E genes suffice to generate carpels (Battaglia et al., 2006).

Like the ABC model, the ABCDE model relied mainly on genetic data. This raised questions with regards to the molecular mechanism by which the different floral homeotic genes interact. For example, how B and C class proteins interact to specify stamen identity remained elusive (Riechmann et al., 1996), and all attempts to fully explain the interactions of the floral homeotic genes and functions just by the dimerization of floral homeotic proteins were not successful. The inability to answer these questions was seen as another major shortcoming of the ABC model and its derivatives (Theißen, 2001). Overcoming this limitation required a switch from

Development (2016) 143, 3259-3271 doi:10.1242/dev.134080

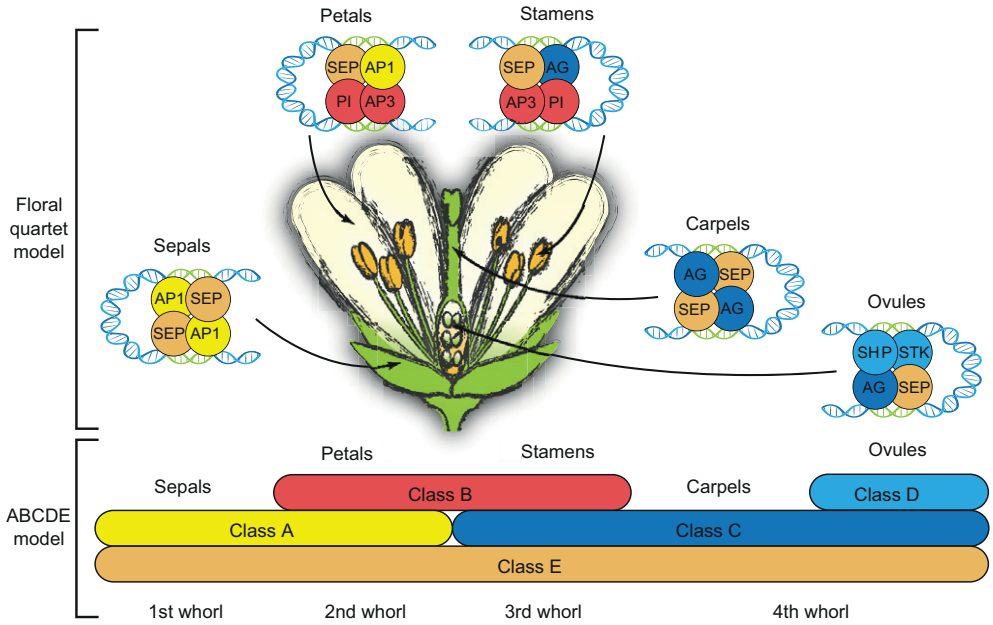


Fig. 1. The floral quartet model and the underlying ABCDE model of organ identity determination in *Arabidopsis thaliana*. The top part of the figure depicts a version of the floral quartet model, which maintains that the five floral organ identities (sepals, petals, stamens, carpels and ovules) are specified by the formation of floral organ-specific tetrameric complexes of MIKC-type MADS-domain transcription factors that bind to two adjacent *cis*-regulatory DNA binding sites (CArG-boxes, green) and loop the DNA (blue) in between. A complex of two APETALA1 (AP1) class A proteins and two SEPALLATA (SEP) class E proteins determines sepal identity, a complex of one AP1 and one SEP protein together with one of each of the class B proteins APETALA3 (AP3) and PISTILLATA (PI) determines petal identity, a complex of one SEP, one AP3, one PI protein and the class C protein AGAMOUS (AG) determines stamen identity, a complex of two SEP proteins together with two AG proteins determines carpel identity, and a complex of one SEP and one AG protein together with one of each of the class D proteins SHATTERPROOF (SHP) and/or SEEDSTICK (STK) controls ovule identity. The bottom part of the figure illustrates the genetic ABCDE model. According to this model, organ identity during flower development in *A. thaliana* is controlled by five sets of floral homeotic genes providing overlapping floral homeotic functions: A, B, C, D and E. Class A genes are expressed in the organ primordia of the 1st and the 2nd whorl of the flower, class B genes in the 2nd and 3rd whorl, class C genes in whorls 3 and 4, class D genes in parts of the 4th whorl (ovule primordia), and class E genes are expressed throughout all four whorls. Class A and E genes specify first whorl sepals, class A, B and E genes specify second whorl petals, class B, C, and E genes specify third whorl stamens, class C and E genes specify fourth whorl carpels, and class C, D and E genes control the development of the ovules within the fourth whorl carpels.

considerations at the gene level to the level of the encoded proteins and eventually led to a new model: the FQM.

The FQM suggests that tetrameric complexes of floral homeotic proteins, rather than individual dimers, control floral organ identity. An important clue that led to the proposition of the FQM was provided when Egea-Cortines et al. (1999) reported that the AP3, PI and AP1 orthologues DEFICIENS (DEF), GLOBOSA (GLO) and SQUAMOSA (SQUA) from *A. majus* form multimeric complexes in electrophoretic mobility shift and yeast three-hybrid assays. Interestingly, the multimeric complex appeared to have a higher DNA-binding affinity than the individual dimers. The authors suggested a model in which the protein complex is actually a protein tetramer, composed of a DEF-GLO heterodimer and a SQUA-SQUA homodimer, in which the DEF-GLO and SQUA-SQUA dimers recognize different CArG-boxes (Egea-Cortines et al., 1999). It remained unclear, however, whether the formation of multimeric protein-DNA complexes was just an idiosyncrasy of some MIKC-type proteins from snapdragon with limited functional relevance, or whether this observation revealed a general principle of MIKC-type protein interactions. Soon after this discovery, however, Pelaz et al. (2000) reported that not only the ABC genes, but also the *SEP* genes are required for the formation of petals, stamens and carpels. All available evidence, including some previous findings about protein dimerization specificities, were subsequently pulled together in the FQM (Theißen, 2001). According to the original 'quartet model', there is at least one unique quaternary complex for each type of the floral organs sepals, petals, stamens and carpels (Theißen, 2001). Based on the ABCDE model and considering carpels, which are unique to angiosperms, and ovules, which are present in all seed plants including gymnosperms (see Glossary, Box 1) as different organs, one may propose a more elaborate FQM (Fig. 1; Theißen and Melzer, 2006).

While the manuscript describing the FQM was in press but not yet available in print, Honma and Goto (2001) demonstrated the formation of the protein complexes postulated for stamens and petals, namely AP3-PI/AG-SEP and AP3-PI/AP1-AP1 (or AP3-PI/SEP-SEP), respectively, thus providing support for the FQM (Theißen and Saedler, 2001). Shortly thereafter, it was shown that partial loss of *SEP* gene (class E) activity leads to similar defects in ovule development as observed in *stk shp1 shp2* (class D gene) triple mutants, and that class D proteins form multimeric complexes together with the SEP3 protein in yeast three-hybrid assays (Favaro et al., 2003), strongly suggesting that floral quartets including class D and E proteins control ovule development (Fig. 1; Theißen and Melzer, 2006).

Recent experimental evidence supporting the floral quartet model

The FQM was rapidly accepted in the literature (see, e.g. Jack, 2001; Eckardt, 2003; Ferrario et al., 2004; Jack, 2004; Krizek and Fletcher, 2005; Baum and Hileman, 2006), suggesting that it was plausible and not in conflict with major evidence at the time of its inception. In addition, a number of protein interaction studies in yeast using proteins from different flowering plant species, such as

Development (2016) 143, 3259-3271 doi:10.1242/dev.134080

tomato, petunia, chrysanthemum, gerbera and rice, demonstrated that floral homeotic proteins could form multimers (e.g. Ferrario et al., 2003; Favaro et al., 2003; Shchennikova et al., 2004; Yang and Jack, 2004; Kaufmann et al., 2005a; Leseberg et al., 2008; Ruokolainen et al., 2010; Seok et al., 2010). Additional experimental evidence supporting the FQM, however, remained scarce for a while. In recent years, this has changed considerably. Diverse experimental approaches comprising analyses *in vitro*, *in vivo*, *in planta* and *in silico* have contributed to the view that floral quartets really exist and play an important role in controlling plant development.

An early experiment that provided evidence for the formation of multimeric complexes of MIKC-type MADS-domain proteins in plant cells employed different fusions between petunia MIKC-type proteins and yellow fluorescent protein (YFP) or cyan fluorescent protein (CFP). When two fusion proteins that dimerize only weakly were coexpressed in petunia protoplasts with a third, unlabelled MIKC-type protein, strong fluorescence resonance energy transfer (FRET) was observed, suggesting that a higher-order complex, as predicted by the FQM, had been formed (Nougalli-Tonaco et al., 2006). Next, in a series of experiments employing electrophoretic mobility shift assays (EMSA) and DNase I footprint assays, it was demonstrated that FQCs can be reconstituted from a limited number of components *in vitro*. Initial experiments revealed that not even a combination of different MIKC-type proteins is required to obtain FQCs; the class E floral homeotic protein SEP3 from *A. thaliana* shows an intrinsic capacity to cooperatively bind as a tetramer to two CArG-boxes (Melzer et al., 2009). The spacing and phasing of CArG-boxes influence the efficiency of FQC binding: binding occurs better if the CArG-boxes are separated by an integer number of helical turns (Melzer and Theißen, 2009; Melzer et al., 2009). In this context, the two CArG-boxes are in the same orientation, so that bending and looping, but not twisting, of the DNA is required when a MIKC-type protein tetramer binds. The ability of a SEP3 homotetramer to loop the DNA sequence separating the two binding sites supports some of the major tenets of the FQM. In follow-up experiments, it was shown that the other three SEP proteins (SEP1, SEP2 and SEP4) of *A. thaliana* also constitute FQCs involving protein homotetramers under suitable conditions *in vitro* (Jetha et al., 2014). All four SEP proteins bind to CArG-boxes in a similar way, and yet they also show subtly distinct DNA-binding properties. For example, the cooperativity of DNA binding differs among the different SEP proteins, with SEP3 often showing the least cooperativity (Jetha et al., 2014). It was also shown that all SEP proteins prefer surprisingly short distances of 4-6 helical turns (~42-63 nucleotides) between the CArG-boxes (Jetha et al., 2014). Remarkably, the optimal distance was shown to differ in *in vitro* experiments, with SEP2 preferring relatively large distances and SEP4 preferring small distances; SEP1 binds well to CArG-box pairs separated by a relatively broad range of distances (Jetha et al., 2014). It is conceivable that FQCs involving SEP proteins alone have a function in the development of flowering plants, but conclusive evidence for that is missing so far (Melzer et al., 2009; Melzer and Theißen, 2009) and other studies instead suggest that SEP proteins act as a kind of 'glue' in interactions of MIKC-type proteins (Immink et al., 2009). It was further shown that complexes composed of SEP3, AP3 and PI form preferentially over SEP3 homotetramers, suggesting a mechanism that would allow different target genes to be activated at different developmental stages (Melzer and Theißen, 2009). In addition, the ectopic expression of SEP3, together with the class B proteins AP3 and PI, is sufficient to induce the development of petals from primordia that would normally develop into vegetative leaves (Honma and Goto, 2001), highlighting that FQCs involving SEP3, AP3 and PI represent a minimal set of master control elements governing floral organ (petal) identity (Melzer and Theißen, 2009).

Data supporting FQC formation *in planta* has also been published. Identifying protein complexes isolated from transgenic plants by affinity purification followed by mass spectrometry and label-free quantification, Smaczniak et al. (2012b) collected data strongly suggesting that the five major floral homeotic MIKC-type proteins that were tested as baits – AP1 (A function), AP3 and PI (B function), AG (C function) and SEP3 (E function) – interact in floral tissues as proposed by the FQM, even though the data do not provide unequivocal evidence that exactly tetramers form *in planta*. Moreover, some tetramers of MIKC-type proteins appear to be able to bind to single CArG-boxes (see, e.g. Melzer et al., 2009; Smaczniak et al., 2012b). As such, an important aspect of the FQM – the looping of regulatory DNA of target genes bound by tetramers of MIKC-type proteins (Theißen, 2001; Theißen and Saedler, 2001) – remained untested *in planta*. Not much later, however, Mendes et al. (2013) reported a series of experiments in favour of FQC formation involving DNA looping. Employing the single-molecule *in vitro* method of tethered particle motion (TPM), the authors studied binding of the floral homeotic proteins STK (class D) and SEP3 (class E) to a fragment of the promoter region of *VERDANDI* (*VDD*), which is a direct target gene of STK that contains three CArG-boxes up to 444 bp apart. The data strongly suggested that loop formation indeed occurs and that FQC formation clearly favours one pair of CArG-boxes (CArG-box 1+CArG-box 3) over alternative combinatorial possibilities for protein binding (Mendes et al., 2013). Using promoter-reporter gene fusions, the authors also studied the functional importance of different CArG-boxes in transgenic *A. thaliana* plants, demonstrating that single CArG-boxes are not sufficient to drive *VDD* gene expression *in planta*, and that both CArG-boxes 1 and 3 are required to establish the typical *VDD* gene expression pattern. Together with chromatin immunoprecipitation (ChIP) studies demonstrating that STK and SEP3 preferentially bind to CArG-boxes 1 and 3 in the *VDD* promoter region, these findings suggest that FQCs involving STK, SEP3 and CArG-boxes 1 and 3 assemble in the *VDD* promoter region and are involved in controlling gene expression *in planta*. These findings provide remarkable *in vivo* evidence for the FQM, even though alternative scenarios have not been completely ruled out so far.

Additional support for the FQM has been provided by structural biology studies. Some EMSA experiments had demonstrated that the C-terminal half of the K domain, which was assumed to form an amphipathic α-helix involved in the formation of a coiled-coil, is of crucial importance for MIKC-type protein tetramerization (Melzer et al., 2009; Melzer and Theißen, 2009). Recent X-ray crystallography studies of the K domain of *A. thaliana* SEP3 revealed that the K domain forms two amphipathic α-helices separated by a rigid kink, which prevents intramolecular association (Puranik et al., 2014). The K domain thus provides two separate interaction interfaces to facilitate dimerization and tetramerization with other K domains (Puranik et al., 2014). Atomic force microscopy (AFM) further demonstrated the looping of target DNA by SEP3 and even allowed FQCs to be 'seen' for the first time (Puranik et al., 2014).

Last, but not least, recent *in silico* analyses have provided support for the FQM. Network-based analyses of the known physical interactions between MADS-domain proteins from *A. thaliana* (as

Development (2016) 143, 3259-3271 doi:10.1242/dev.134080

revealed by yeast two-hybrid and three-hybrid assays) indicated that the formation of functional tetramers is a widespread property of *A. thaliana* MIKC-type proteins, but not of non-MIKC-type MADS-domain proteins i.e. those that lack a K domain (Espinosa-Soto et al., 2014). Given that all floral organ identity proteins (ABCDE proteins) of the MADS-domain family are MIKC-type proteins, and that MIKC-type proteins have a tendency to tetramerize (even though not all of them may actually do so), it appears even more likely that the combinatorial interactions of the different homeotic genes predicted by the ABCDE model are indeed realized by the tetramerization of MIKC-type floral homeotic proteins.

The findings reviewed above, however, do not imply that all MIKC-type proteins exert their function only as constituents of tetrameric complexes. Several lines of evidence, such as ChIP-seq data of protein binding *in vivo* (Kaufmann et al., 2009; Kaufmann et al., 2010a,b), suggest that dimers of MIKC-type proteins are also of functional importance, and it also appears likely that at least some dimers and tetramers exist in dynamic equilibria.

The FQM as guiding model in current research

The heuristic value of the FQM is revealed by its use as a guiding model in current research. For example, the destruction of floral quartets has been proposed to cause the development of the often bizarre symptoms observed in plants infected by the bacterial pathogen phytoplasma (Maejima et al., 2014). One characteristic phenotype ('phyllody') of phytoplasma-infected plants from diverse species, including *A. thaliana*, resembles the phenotype of class E floral homeotic mutants, with floral organs unable to develop proper floral organ identity. This was recently shown to be due to proteasome-mediated degradation of the class A and E floral homeotic proteins AP1, CAULIFLOWER (CAL) and SEP3, which is initiated by interaction of the floral homeotic proteins with phytoplasma-secreted effector proteins termed SAP54 or PHYL1 (Maejima et al., 2014; MacLean et al., 2014). An *in silico* study suggests that the PHYL1 structure resembles that of the K domain, thus facilitating dimerization between some floral homeotic proteins and PHYL1 (Rümpler et al., 2015). The authors hypothesized that the similarity between PHYL1 and the K domain represents a case of convergent evolution ('molecular mimicry') that evolved to enable phytoplasmas to manipulate their host plants according to their needs. Maejima et al. (2014) noted that the strength of phenotype in the different floral whorls of plants expressing PHYL1 (severe in the first whorl, medium in whorl 2, weak in whorl 3 and again medium in the fourth whorl) correlates perfectly with the number of class A and E proteins (4, 2, 1, 2) in the floral quartets of whorls 1, 2, 3, and 4, respectively; they thus argue that the floral quartet model provides the basis for an explanation of the whorl-specific differences in the strength of phenotype in affected plants.

Diversification of the floral quartet that specifies petal identity has been used to explain the differences in the petaloid organs of orchids. Orchids typically have four paralogous *AP3*-like genes, in contrast to the one *AP3*-like class B gene found in *A. thaliana* and *A. majus*. According to the 'orchid code hypothesis', sub- and neo-functionalization involving differential expression of these genes led to a combinatorial system that specifies the identity of the different petaloid perianth organs, i.e. outer tepals (also called 'sepals') in the first floral whorl, and inner lateral tepals ('petals') and the labellum ('lip') in the second whorl (Mondragón-Palomino and Theißen, 2008, 2011). Recently, Hsu et al. (2015) reported several lines of evidence suggesting that competition between two floral quartets decides whether outer and inner lateral tepals ('sepals and petals'), or lips develop. Both floral quartets contain one protein encoded by the single *PI*-like (class B) gene, but different paralogs of *AP3*-like (class B) and *AGL6*-like genes (that may function in orchids as class E genes).

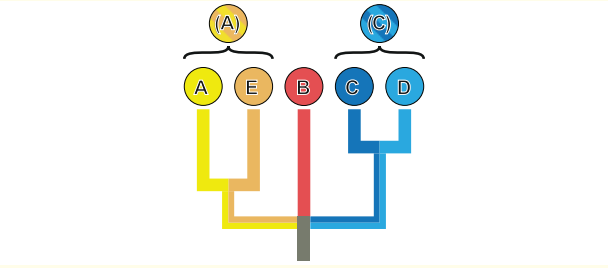
Floral quartets have also been used to explain how the interesting phenomenon of paralog interference can affect the evolutionary dynamics of genes after duplication (Kaltenegger and Ober, 2015). When proteins function in obligate homomeric complexes of identical subunits, duplication of the gene encoding these proteins generates paralogous genes whose gene products may cross-interact when co-expressed, thus resulting in paralog interference. Since independent mutations in the different gene copies may interfere with protein interaction and function, and hence may bring about a dominant negative effect, both copies are expected to remain under purifying selection during a prolonged time window. This increases the chance that they accumulate mutations that lead to novel properties of the different paralogous proteins. In line with this, positive selection may occur, creating asymmetric protein dimers or multimers that may contribute to evolutionary novelties or innovations. While Kaltenegger and Ober (2015) focused their discussion on the obligate heterodimerization of class B proteins within some floral quartets of angiosperms, it is tempting to speculate that paralog interference played an important role during the expansion and diversification of all kinds of MIKC-type genes and FQCs throughout the evolution of land plants.

The (A)B(C)s of floral quartets

When the simple and elegant ABC model developed into the more elaborate ABCDE model, the FQM was proposed to explain the interactions between floral homeotic genes and proteins, but it also intended to resimplify matters (Theißen, 2001). Recent improvements to the ABCDE model enable the FQM to further harmonize the genetic and the molecular models.

In contrast to the genetically, phylogenetically and developmentally quite well-defined B and C floral homeotic functions, the concept of A function has been considered controversial for almost as long as the ABC model itself (Theißen et al., 2000; Litt, 2007; Causier et al., 2010). One reason is that in almost all plants that have been investigated so far, with *A. thaliana*

Box 2. The phylogeny of floral homeotic genes and functions



The highly simplified phylogenetic tree depicts the relationships between floral homeotic genes, proteins and functions as defined in the ABCDE model (Fig. 1; Gramzow and Theißen, 2010). While the deep branching of the tree is still largely unknown (indicated by the basal trifurcation) there is strong support for a close relationship between class A and E genes, and class C and D genes, constituting the clades of (A) and (C) genes, respectively, as indicated.

Development (2016) 143, 3259-3271 doi:10.1242/dev.134080

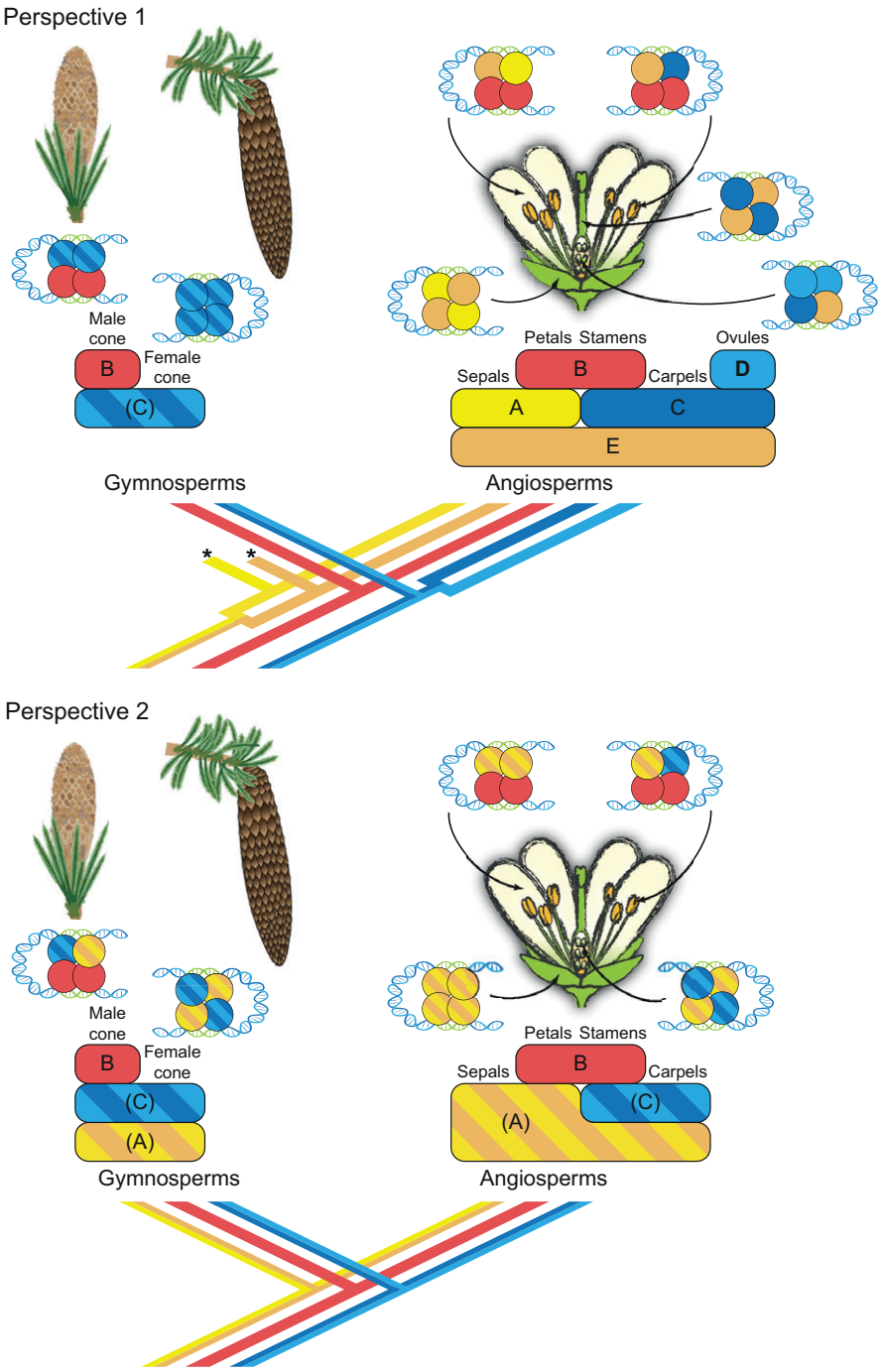


Fig. 2. Perspectives on floral homeotic functions in gymnosperms and angiosperms. Two perspectives (1 and 2) of how floral homeotic genes might function in gymnosperms and angiosperms are shown. The trees at the bottom of each panel illustrate proposed phylogenetic relationships between the five floral homeotic gene functions: A (yellow), B (red), C (blue), D (cyan) and E (brown). Potential combined/ancestral functions (A), which represents class A+E genes (bicoloured in yellow and brown) and (C), which represents class C+D genes (bicoloured in blue and cyan) are also shown. The coloured ovals illustrate the spatial expression patterns of floral homeotic genes within an angiosperm flower and for male and female cones of gymnosperms (applying the same colour code). The assumed composition of the FQC operating in each context is shown, with individual proteins colour coded according to their floral homeotic gene function. (Perspective 1) The generally well-accepted model of floral organ development assumes five classes of floral homeotic genes (A,B,C, D,E) with A and E descending from the duplication of a common ancestor (A) prior to the split of gymnosperms and angiosperms, and C and D resulting from a duplication of an ancestral (C)-function gene during early angiosperm evolution. Orthologues of class A (*AP1*-like) and some class E (*SEP*-like) genes had until very recently not been identified in gymnosperms and were therefore assumed to have been lost in the lineage that led to extant gymnosperms (*). The development of male and female cones is thus assumed to be controlled by tetramers of class B and (C) proteins only. (Perspective 2) We argue that the floral homeotic class A and E genes may be reconsidered as a combined (A) gene function that is present in angiosperms as well as gymnosperms, while the so far separated class C and D genes of angiosperms may be coalesced into a combined (C) gene function. Note that 'Carpels' in perspective 2 includes ovules. Based on the resulting (A)B(C) model, we hypothesize that the FQCs that determine male and female reproductive organs, respectively, are in principle the same for angiosperms and gymnosperms.

being a remarkable exception, one does not find recessive mutants in which the identity of both types of perianth organs is affected (Litt, 2007). But even in *A. thaliana*, the A function appears ill-defined (Litt, 2007; Causier et al., 2010). For example, an A function in specifying perianth (sepal and petal) organ identity and antagonizing the C function is difficult to separate genetically from a more fundamental function in specifying floral meristem identity. In fact, an early alternative to the ABC model that was focused on *A. majus* proposed two 'developmental pathways' named 'A' and 'B' in combination with a 'floral ground state' (Schwarz-Sommer et al., 1990), with A and B being equivalent to the class B and C function, respectively, of the ABC model (Causier et al., 2010). In this alternative model, sepal development represents the 'default state' of floral organ development and hence does not require a specific floral homeotic function (Schwarz-Sommer et al., 1990).

To resolve controversies surrounding A function, Causier et al. (2010) suggested an (A)BC model with (A) function controlling both floral meristem identity (the 'floral ground state') and floral organ identity in the first two floral whorls. According to the (A)BC model, (A) function also comprises the E function of the ABCDE model i.e. (A)=A+E. According to Causier et al., (A) function is provided by a group of genes, but if one focusses on the MADS-box genes involved – in the case of *A. thaliana* the class A gene *AP1* and the class E genes (*sensu lato*), i.e. the *SEP* genes and the *AGL6*-like genes (Mandel et al., 1992; Pelaz et al., 2000; Ditta et al., 2004; Rijpkema et al., 2009; Hsu et al., 2014) – one finds some support for the new (A) function in gene phylogeny. All of these genes are

Development (2016) 143, 3259-3271 doi:10.1242/dev.134080

Box 3. FQCs: beyond floral organ identity

Given that the formation of functional tetramers is a widespread property of *A. thaliana* MIKC-type proteins (Puranik et al., 2014; Espinosa-Soto et al., 2014), we hypothesize that FQCs play important roles beyond floral organ identity specification. Indeed, several studies have suggested that MIKC-type proteins other than the canonical class A-E floral homeotic proteins of the FQM can form FQCs. For example, members of the B_{sister} subfamily are involved in specifying the endothelium, and in case of *A. thaliana*, there is evidence that SHP proteins and/or STK, SEP3 and the B_{sister} protein ARABIDOPSIS BSISTER (ABS, also known as TT16 and AGL32) are components of a FQC that specifies this identity (Becker et al., 2002; Nesi et al., 2002; Kaufmann et al., 2005a; Mizzotti et al., 2012; Theißen and Gramzow, 2016). Several MIKC-type genes have also been implicated in fruit development, and it is reasonable to assume that FQCs also play a role in this context. For example, FQCs are involved in the development of the fleshy fruits of tomato (*Solanum lycopersicum*); such proposed FQCs were shown to contain MIKC-type proteins, such as the StMADS11-like protein JOINTLESS, that are part of clades not belonging to those containing floral homeotic proteins (Liu et al., 2014; Fujisawa et al., 2014). Smaczniak et al. (2012b) also identified complexes of several other MIKC-type proteins, in line with the hypothesis that they too are involved in protein tetramerization and FQC formation. For example, complexes of AP1 and the TM3-like protein SUPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), and of SOC1 and FUL were detected; both complexes might be part of FQCs involved in the transition to flower (Smaczniak et al., 2012b). These findings support the view that FQCs with protein compositions other than those described by the FQM play a role in processes other than floral organ identity specification.

relatively closely related members of a gene superclade (Gramzow and Theißen, 2010, 2013, 2015; Ruelens et al., 2013), and it is thus conceivable that the A and E functions known from flowering plants trace back to an ancestral function in specifying reproductive meristem identity (Box 2). Similarly, there is evidence that the C and D functions of angiosperms trace back to a combined C/D function provided by *AG*-like genes in extant gymnosperms and stem group seed plants (Box 2; Gramzow et al., 2014). Hence in analogy to (A) function, one may define (C) function, with (C)=C+D, yielding an (A)B(C) model for the angiosperm flower. The (C) function may specify reproductive organ identity, and its expression may distinguish reproductive from non-reproductive organs. Based on these considerations, one can transform a generalized ABCDE model into a more simple (A)B(C) model (Fig. 2). Note that the model shown is a generic model, and that the genes contributing to these functions may have been differentially sub- and neo-functionalized in different species of angiosperms. This hampers interspecific comparisons and might be one reason for some controversies about A/E and C/D functions in the literature (see, e.g. Litt, 2007).

'Translating' the (A)B(C) model into a model based on FQCs, one gets a generic floral quartet model (Fig. 2, perspective 2) with four (A) proteins specifying floral meristem identity and sepals, two (A)+two B proteins specifying petals, one (A)+two B+one (C) proteins specifying male reproductive organs (stamens), and two (A)+two (C) proteins specifying female reproductive organs (carpels including ovules). Thus, after somewhat of a detour, the ABC model regains its simplicity as an (A)B(C) model, and the FQM has also been generalized and simplified (e.g. Fig. 2, compare 1 and 2). Given that (A), B and (C) genes probably already existed in the most recent common ancestor (MRCA) of extant seed plants, the generic FQM has obvious consequences for understanding the origin of the angiosperm flower.

On the origin of floral quartets: towards solving the 'abominable mystery'

Floral organ identity does not develop without the proper activity of floral homeotic genes. It appears reasonable, therefore, that understanding the evolution of floral quartets is key to understanding the origin of the angiosperm flower – a scientific problem closely related to the origin of the angiosperms, which has been popularized as Darwin's 'abominable mystery' (Theißen and Saedler, 2001; Friedman, 2009). So how did floral homeotic genes of the MIKC type originate, and how did they start to constitute floral quartets?

Early studies had already documented a strong correlation between the evolution of MIKC-type genes and the origin of evolutionary novelties, including floral organs, in land plants (Theißen and Saedler, 1995; Purugganan et al., 1995; Theißen et al., 1996, 2000; Becker and Theißen, 2003). The phylogeny of MIKC-type genes is characterized by the formation of ancient paralogs, many of which originated by whole genome duplications, preferential gene retention after duplication, and sequence divergence resulting in sub- and neo-functionalization (Gramzow and Theißen, 2013, 2015; Theißen and Gramzow, 2016). Radiations of genes occurred independently in different groups of land plants. Even though diverse MIKC-type MADS-box genes are involved in the control of many developmental processes in angiosperms, and probably also in all other land plants (Smaczniak et al., 2012a; Gramzow and Theißen, 2010), the floral homeotic genes are all members of gene clades that are seed plant- or flowering plant-specific.

Recent phylogeny reconstructions involving the first whole-genome sequence data from conifers (gymnosperms) suggest that MIKC-type genes of seed plants are all members of 11 seed plant-specific superclades that were present in the MRCA of extant seed plants about 300 million years ago (MYA), but that did not yet exist in the MRCA of monilophytes (ferns and their allies such as horsetails) and seed plants (gymnosperms and angiosperms) about 400 MYA (Nystedt et al., 2013; Gramzow et al., 2014). Among

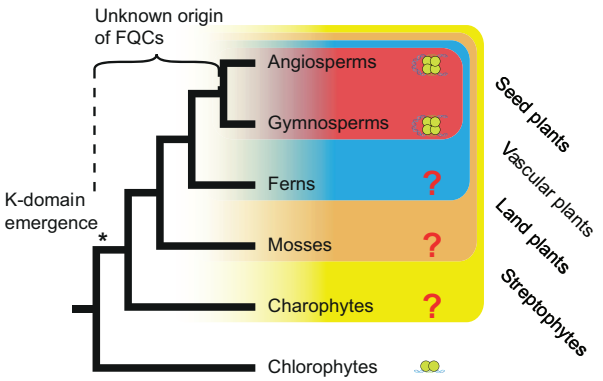


Fig. 3. The emergence of FQC formation during green plant evolution. The tree illustrates a simplified green plant phylogeny. Within the green plant phylogeny, seed plants (angiosperms and gymnosperms) together with ferns represent the clade of vascular plants; vascular plants and mosses represent the clade of land plants; and all land plants together with charophytes (a division of freshwater green algae) build up the clade of streptophytes. The K domain most likely emerged (indicated by *) prior to the split between extant charophytes and land plants and is thus a synapomorphy of streptophytes. There is experimental evidence that at least some MIKC-type proteins of angiosperms and gymnosperms can form FQCs, whereas MADS-domain proteins from chlorophytes lack a K domain and thus may bind to DNA only as dimers, as indicated by the icons next to the names of the different plant groups. It is not yet known if FQCs exist in charophytes, mosses or ferns (indicated by '?'). Whether FQC formation is directly linked to the emergence of the K domain and is a synapomorphy of streptophytes, or if it occurred later during land plant evolution, is therefore also unknown.

Development (2016) 143, 3259-3271 doi:10.1242/dev.134080

these superclades are those containing, besides other genes, genes providing floral homeotic A function (*FLC*/*SQUA*-like, or *FLC*/*AP1*-like genes), class B genes (*DEF*/*GLO*/*OsMADS32*-like, or *AP3*/*PI*/*OsMADS32*-like genes), class C genes (*AG*-like genes) and class E genes (*SEP/AGL6*-like genes) (Box 2). Based on gene or even whole genome duplications in the stem group of angiosperms, the 11 superclades evolved into 17 clades that had already been established in the MRCA of extant angiosperms, including distinct *DEF* (*AP3*)- and *GLO* (*PI*)-like genes (class B), the *AG*-like and *STK*-like genes (classes C and D), the *AGL2*-like (*SEP*-like) and *AGL6*-like genes (class E), and the *SQUA* (*AP1*)-like genes (class A) (Theißen et al., 1996, 2000; Becker and Theißen, 2003; Ruelens et al., 2013; Gramzow et al., 2014).

It has been shown that some putative DEF/GLO-like (class B) and AG-like (class C/D) MIKC-type proteins from gymnosperms can alone constitute FQCs that may specify male and female reproductive cone development (Wang et al., 2010). Moreover, early phylogeny reconstructions suggested that combined *DEF*/*GLO*-like (class B) and *AG*-like (class C), but no *SQUA*-like (class A) and *SEP*-like (class E) genes, existed in the MRCA of extant seed plants. Even though *AGL6*-like genes had been found in diverse extant gymnosperms (Winter et al., 1999), the function of these genes was, at that time, unknown even in angiosperms. These findings led to the view that the origin of *SEP*-like genes and the incorporation of SEP-like proteins into FQCs have been important steps during the origin of floral quartets, and hence floral organ identity and flower development (Fig. 2, perspective 1; Zahn et al., 2005; Baum and Hileman, 2006; Silva et al., 2016). However, recent experimental data from different species suggest that not only *SEP*-like but also *AGL6*-like genes can exert the E function (Thompson et al., 2009; Rijpkema et al., 2009; Hsu et al., 2014) and phylogeny reconstructions suggest that the genomes of extant conifers and the MRCA of extant seed plants contain(ed) orthologs of floral homeotic class A and E genes (Gramzow et al., 2014). It is conceivable, therefore, that FQCs quite similar to those of extant floral quartets also exist in extant gymnosperms and were already established in the MRCA of extant seed plants (Fig. 2, perspective 2). Specifically, and in contrast to previous views (Fig. 2, perspective 1; Zahn et al., 2005; Theißen and Melzer, 2007) that proposed that the incorporation of SEP-like proteins into FQCs played an important role during the origin of the flower, we consider it more likely now that the FQCs specifying male and female reproductive cone identity in ancestral and extant gymnosperms very much resemble(d) those of angiosperms, in that they contain(ed) AG-like proteins [(C) function] (female cones) or AG-like and DEF/GLO-like proteins [(B) function] (male cones) as well as SEP/AGL6-like and/or FLC/AP1-like proteins [(A) function] (Fig. 2, perspective 2).

The differences between the two hypotheses on the origin of floral quartets are obviously of heuristic relevance. Assuming that changes in the composition of FQCs played an essential role during the evolution of the flower may inspire investigations into the evolution of MIKC-type proteins interactions in seed plants (e.g. Wang et al., 2010; Melzer et al., 2014). However, if one hypothesizes that the FQCs specifying male and female reproductive organ identity in gymnosperms and angiosperms did not change substantially during the origin of the flower, one may conclude that changes in the interactions between the FQCs specifying reproductive organ identity and their target genes have been of special importance during the origin of the flower. If so, comparison of the target genes of the MIKC-type proteins in FQCs in extant gymnosperms and angiosperms would be most revealing. While target genes for several *A. thaliana* floral homeotic proteins have already been determined, for example by employing techniques such as ChIP-seq (Kaufmann et al., 2009, 2010b; Ó'Maoilléidigh et al., 2013; Wuest et al., 2012), respective data for gymnosperms are still missing.

Box 4. Why quartets?

Why do many MIKC-type transcription factors bind to the DNA of their target genes as tetramers (quartets) rather than as independent dimers, as is the case for many other MADS-domain proteins? One important difference between tetramers and two dimers binding to DNA is the increased cooperativity in DNA binding. This cooperativity creates a sharp transcriptional response, i.e. even small increases in protein concentration can lead to drastic changes in regulatory output (Georges et al., 2010). Floral homeotic proteins as well as many other MIKC-type proteins act as genetic switches that control discrete developmental stages, and cooperative DNA binding might be one important mechanism that translates the quantitative nature of biomolecular interactions into discrete phenotypic outputs (Theißen and Melzer, 2007; Kaufmann et al., 2010a). Tetramer formation could also potentially incorporate different signals and thereby increase the robustness of the system. If one protein component of the tetramer is missing, the entire complex will not form or will be greatly destabilized and the developmental switch will not occur (Whitty, 2008). The formation of tetramers might also, in principle, contribute to an increase in target gene specificity. It was previously shown that different tetramers have different DNA-binding affinities, and that different tetramers may prefer different CArG-box distances for maximum binding (Melzer and Theißen, 2009; Jetha et al., 2014). This offers the possibility to differentially regulate target genes even in the absence of differential DNA-binding of MIKC-type protein dimers (Georges et al., 2010). We are still in the early days of exploring the developmental and evolutionary relevance of cooperative DNA binding and FQC formation. Plants expressing mutant proteins defective specifically in cooperative DNA binding will hopefully yield additionally insights into how and why quartet formation is essential for MIKC-type proteins to act as developmental switches.

On the origin of FQCs: a MIKC blessing

The floral quartet model describes the interaction of the floral homeotic proteins at the molecular level. However, floral homeotic proteins represent only a minor fraction of the MIKC-type protein family, which comprises 45 members in *A. thaliana* alone (Becker and Theißen, 2003; Parenicová et al., 2003). Therefore, the question arises as to whether multimerization is restricted to the floral homeotic proteins of eudicots (an extreme hypothesis) or is a common feature of all MIKC-type proteins in all kinds of land plants (another extreme hypothesis) (Kaufmann et al., 2005b). Based on rapidly growing empirical evidence, we hypothesize that FQCs play an important role far beyond floral organ identity specification in *A. thaliana* (see Box 3). This raises the question as to when and where during evolution FQC formation began. Intriguingly, in contrast to many other multimeric complexes of transcription factors, the key protein constituents of floral quartets are all encoded by paralogous MIKC-type genes. This corroborates the view that duplications of ancestral MIKC-type genes are intimately interlinked with the evolution of developmental complexity in plants. Thus, the question arises as to how the origin of MIKC-type proteins and of their ability to constitute FQCs are linked.

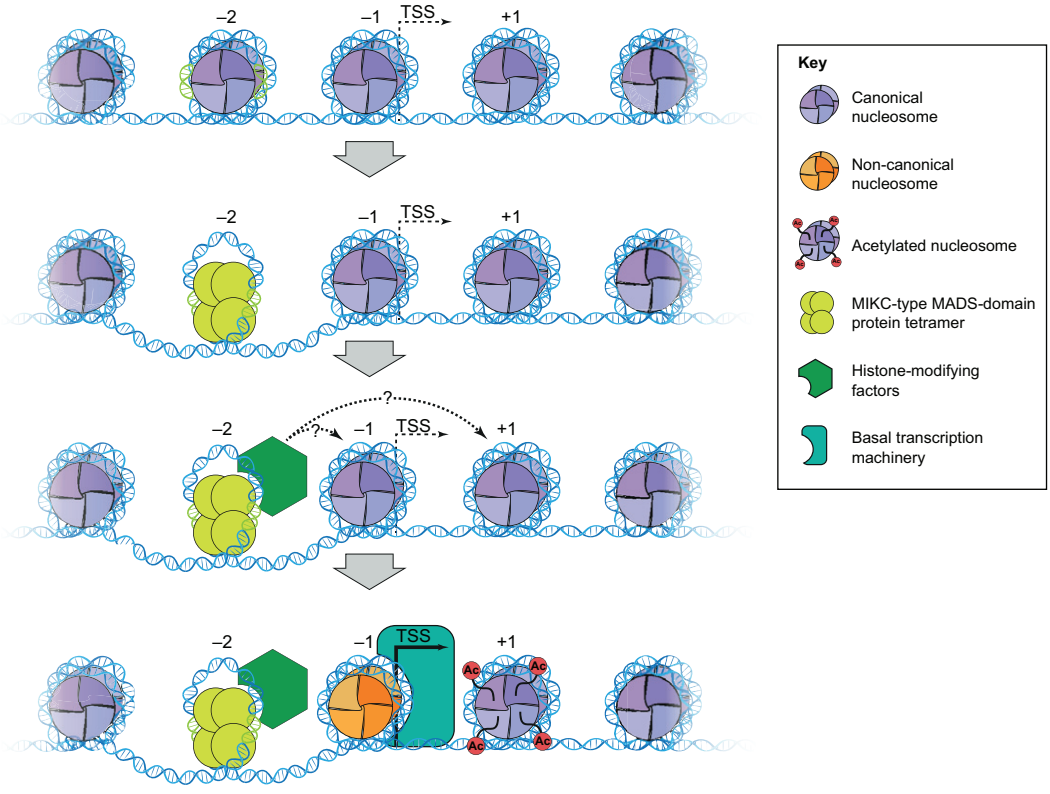
A variety of *in vitro* experimental data has demonstrated that the K domain is essential for mediating the interactions that are necessary for FQC formation (Yang et al., 2003; Yang and Jack, 2004; Melzer and Theißen, 2009; Melzer et al., 2009). As explained above, the K domain provides the structural basis on which FQC formation takes place (Puranik et al., 2014). It thus appears that the emergence of the K domain – with two distinct interaction interfaces that facilitate both

Development (2016) 143, 3259-3271 doi:10.1242/dev.134080

dimerization and tetramerization – constitutes an important precondition for the origin and evolution of FQCs. But when did such a K domain emerge? Even though MADS-box genes are present in almost all eukaryotes (Gramzow and Theißen, 2010; Gramzow et al., 2010), the most early diverging species in which MIKC-type genes were identified belong to the charophytes (Fig. 3; Tanabe et al., 2005); it is therefore presumed that the K domain is a synapomorphy of streptophytes (charophytes and land plants) and emerged more than 700 MYA (Kaufmann et al., 2005b; Gramzow and Theißen, 2010). How MIKC-type proteins from charophytes and early diverging land plants (such as liverworts, mosses and ferns) interact has not yet been investigated, and whether the ability to form FQCs was already present when the K domain emerged in the MRCA of extant streptophytes, or whether structural changes within the K domain that occurred during early land plant evolution were required still remains unresolved (Fig. 3). In any case, it appears reasonable that the emergence of a DNA-binding MADS domain with a dimerization and tetramerization enabling K domain was a key event in plant evolution. It provided the common ancestor of streptophytes or a major clade of land plants with the capacity to evolve efficient developmental switches (see Box 4) and to dramatically diversify these switches simply by gene duplications followed by mutations. It is tempting to speculate, therefore, that the origin of MIKC-type proteins and FQC formation have been important preadaptations to the transition to land, or remarkable prerequisites for the evolution of the complex body plans of land plants.

Box 5. The 'nucleosome mimicry' model of FQC action

We hypothesize that FQCs represent sequence-specific transcription factors with (half-) nucleosome-like properties that help to establish permissive or repressive chromatin modifications at CArG-box-containing promoters. A permissive, gene-activating case is illustrated below. In the first step, a nucleosome in inactive chromatin near to a transcription start site (TSS) is substituted by a FQC, resulting in a poised state of the chromatin. The FQC can then recruit histone-modifying factors such as acetylases and methylases, leading eventually to recruitment of the basal transcriptional machinery. The FQC and its co-factors may also be involved in substitution of a canonical nucleosome immediately upstream of the TSS (–1 position) by a labile, non-canonical one with modified histones (such as H2A.Z and H3.3). For simplicity, only histone acetylation is shown as symbol of gene activation here.



Our model is based on similarities between FQCs on the one hand, and (half-) nucleosomes and the transcription factor NF-Y, which mimics H2A/H2B-DNA nucleosome assembly (Nardini et al., 2013), on the other hand. Both FQCs and half-nucleosomes are composed of tetramers of similar proteins. Moreover, DNA might be wrapped around FQCs in a similar way as in nucleosomes, including similar loop sizes [about 42-94 base pairs in the case of FQCs and 86 (147:1.7) base pairs in the case of nucleosomes]. Like NF-Y, MADS-domain proteins insert a stretch of their sequence into the minor groove, and they bind to remarkably similar DNA sequences (note that a CCAAT box, to which NF-Y binds, is one half of a perfect CArG-box). Also, DNA containing short AT-rich sequences spaced by an integral number of DNA turns is easiest to bend around the nucleosome, and the same criterion is fulfilled by two CArG-boxes separated by an integer number of helical turns, an arrangement known to facilitate FQC formation (Jetha et al., 2014). In fact, the central region of the CArG-box largely resembles an 'A-tract' (sequence motif A_nT_m with n+m>3) and periodically spaced A-tracts outside the CArG-box have also been detected (Muino et al., 2014). Thus, the DNA binding of FQCs and nucleosomes is facilitated by similar structural motifs.

Development (2016) 143, 3259-3271 doi:10.1242/dev.134080

Conclusions

Much has been learned about FQCs and their role in plant development in recent years. However, two major questions that were not addressed by the original FQM remain largely unanswered. First, how do FQCs acquire target gene specificity? Second, by what molecular mechanisms do they activate or repress the expression of their target genes? These topics are highly inter-related, with chromatin structure and nucleosome activities providing an obvious link.

As is the case for many transcription factors, how MIKC-type proteins achieve target gene specificity still represents a major conundrum. The problem actually has at least two layers of complexity. First, DNA-sequence elements similar to CArG-boxes occur thousands of times in the *A. thaliana* genome, so that almost every gene possesses a potential binding site for MIKC-type transcription factors (de Folter and Angenent, 2006). This strongly indicates that the CArG-box motif alone is not sufficient to explain the target gene specificity of MIKC-type proteins. Second, all of the at least 45 different MIKC-type proteins encoded in the *A. thaliana* genome share the highly conserved DNA-binding MADS domain (Parenicová et al., 2003) and studies indicate that, for many of these proteins, DNA-binding specificity might be quite similar, although subtle differences can be detected (Huang et al., 1996; Riechmann et al., 1996). Yet, mutant phenotypes of different floral homeotic genes (and other MIKC-type proteins) differ drastically, suggesting a considerable level of target gene specificity among different floral homeotic proteins. Indeed, recent investigations suggest a complex picture in which the CArG-box sequence, structural features of the CArG-box (e.g. a narrow minor groove, the number, distance and orientation of CArG-boxes), sequences beyond the CArG-box and transcriptional cofactors all play a role in FQC target gene recognition (Melzer et al., 2006; Ó'Maoiléidigh et al., 2013; Jetha et al., 2014; Muino et al., 2014; Yan et al., 2016). Chromatin structure may also play a role in target site specificity. In line with this, chromatin-remodelling and -modifying factors were identified as interactors of MIKC-type proteins (Smaczniak et al., 2012b). For example, *A. thaliana* AP1 was suggested to recruit the H3K27 demethylase RELATIVE OF EARLY FLOWERING 6 (REF6) to the promoter of *SEP3*. This may explain the observed removal of the H3K27me3 inhibitory histone mark and, consequently, activation of SEP3, possibly by antagonizing Polycomb Group (PcG)-mediated transcriptional repression (Smaczniak et al., 2012b). It was also shown that AP1 and SEP3 bind to enhancer sites very early during flower development and that chromatin accessibility changes only subsequently, suggesting that SEP3 acts as a pioneer transcription factor (PTF, see Glossary, Box 1) that modifies chromatin accessibility (Pajoro et al., 2014). PTFs are by definition able to bind to inaccessible, nucleosome-associated DNA sites, thus creating an open chromatin environment that is permissive for the binding of non-pioneer factors that can only bind to accessible sites (termed 'settlers' if they almost always bind to sites matching their DNA-binding motif, and 'migrants' if they are more selective, e.g. because their binding requires co-factors) (Slattery et al., 2014; Todeschini et al., 2014). This raises the question as to what enables AP1 and SEP3 to function as PTFs. Jetha et al. (2014) calculated that the ability of cooperative DNA binding of SEP proteins during FQC formation could facilitate their invasion of nucleosomal DNA and thus their activity as PTFs. It is also known that nucleosomes are most efficiently ejected by DNA-binding proteins whose binding sites are spaced by up to 74 bp from each other (Polach and Widom, 1996; Moyle-Heyrman et al., 2011); this distance is close to the CArG-box distances for which the highest cooperativity was observed by Jetha et al. (2014).

The analysis of nucleosome-mediated control of gene expression has also provided clues into how FQCs might function. Nucleosomes are composed of an octamer of H2A, H2B, H3 and H4 histones, all of which are present in two copies, wrapped around with DNA almost exactly 147 base pairs long. However, nucleosomes are all but static systems, and chromatin is frequently reorganized at multiple levels (Henikoff, 2008). For example, nucleosomes near transcription start sites may continuously cycle between a repressed canonical form and an unstable, noncanonical form that contains histone variants such as H2A.Z and H3.3 substituting the standard histones H2A and H3, respectively (Soboleva et al., 2014). There is also experimental evidence for the existence of subnucleosomal particles such as half-nucleosomes that contain just one copy of H2A, H2B, H3 and H4. Again, especially at the 5′ end of genes, such dynamic nucleosomes may increase accessibility to transcription start sites and transcription factor binding sites (Rhee et al., 2014). Such dynamic half-nucleosomes (or even full nucleosomes) bear similarities to FQCs and, based on these similarities, we suggest a 'nucleosome mimicry' model of FQC action. Specifically, we hypothesize that FQCs represent sequence-specific transcription factors with (half-) nucleosome-like properties that help to establish permissive or repressive chromatin modifications at CArG-box-containing promoters (see Box 5 for details). This molecular mimicry might enable FQCs to evict nucleosomes from positions at which they are already quite labile, e.g. promoter regions with A-tracts (Henikoff, 2008) and hence to act as PTFs.

We hope that the 'nucleosome mimicry' model that we propose here will be rigorously tested in the near future. We have the same hope for the FQM itself and more general functions of FQCs. We feel that FQCs provide a useful framework for studying many more processes in plant development and evolution than just the specification of floral organ identity.

Acknowledgements

We are are grateful to three anonymous reviewers for their helpful comments on a previous version of the manuscript. G.T. thanks Mirna and Lydia Gramzow for their patience. We apologize to all authors whose publications could not be cited owing to space constraints.

Competing interests

The authors declare no competing or financial interests.

Funding

G.T. and R.M. received funding from the Deutsche Forschungsgemeinschaft (DFG) [TH417/5-3].

References

Angenent, G. C. and Colombo, L. (1996). Molecular control of ovule development. *Trends Plant Sci.* **1**, 228-232.

Battaglia, R., Brambilla, V., Colombo, L., Stuitje, A. R. and Kater, M. M. (2006). Functional analysis of MADS-box genes controlling ovule development in Arabidopsis using the ethanol-inducible alc gene-expression system. *Mech. Dev.* **123**, 267-276.

Baum, D. A. and Hileman, L. C. (2006). A developmental genetic model for the origin of the flower. In *Flowering and its Manipulation* (ed. C. Ainsworth) pp. 3-27. Sheffield, UK: Blackwell Publishing.

Becker, A. and Theißen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* **29**, 464-489.

Becker, A., Kaufmann, K., Freialdenhoven, A., Vincent, C., Li, M.-A., Saedler, H. and Theißen, G. (2002). A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Mol. Genet. Genomics* **266**, 942-950.

Bowman, J. L., Smyth, D. R. and Meyerowitz, E. M. (1991). Genetic interactions among floral homeotic genes. *Development* **112**, 1-20.

Bowman, J. L., Smyth, D. R. and Meyerowitz, E. M. (2012). The ABC model of flower development: then and now. *Development* **139**, 4095-4098.

Causier, B., Schwarz-Sommer, Z. and Davies, B. (2010). Floral organ identity: 20 years of ABCs. *Semin. Cell Dev. Biol.* **21**, 73-79.
Coen, E. S. and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* **353**, 31-37.
de Folter, S. and Angenent, G. C. (2006). Trans meets cis in MADS science. *Trends Plant Sci.* **11**, 224-231.
Ditta, G., Pinyopich, A., Robles, P., Pelaz, S. and Yanofsky, M. F. (2004). The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr. Biol.* **14**, 1935-1940.
Eckardt, N. A. (2003). MADS monsters: controlling floral organ identity. *Plant Cell* **15**, 803-805.
Egea-Cortines, M., Saedler, H. and Sommer, H. (1999). Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *EMBO J.* **18**, 5370-5379.
Espinosa-Soto, C., Immink, R. G. H., Angenent, G. C., Alvarez-Buylla, E. R. and de Folter, S. (2014). Tetramer formation in *Arabidopsis* MADS domain proteins: analysis of a protein-protein interaction network. *BMC Syst. Biol.* **8**, 9.
Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M. F., Kater, M. M. and Colombo, L. (2003). MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell* **15**, 2603-2611.
Ferrario, S., Immink, R. G. H., Shchennikova, A., Busscher-Lange, J. and Angenent, G. C. (2003). The MADS box gene *FBP2* is required for *SEPALLATA* function in petunia. *Plant Cell* **15**, 914-925.
Ferrario, S., Immink, R. G. H. and Angenent, G. C. (2004). Conservation and diversity in flower land. *Curr. Opin. Plant Biol.* **7**, 84-91.
Friedman, W. E. (2009). The meaning of Darwin's 'Abominable Mystery'. *Am. J. Bot.* **96**, 5-21.
Fujisawa, M., Shima, Y., Nakagawa, H., Kitagawa, M., Kimbara, J., Nakano, T., Kasumi, T. and Ito, Y. (2014). Transcriptional regulation of fruit ripening by tomato FRUITFULL homologs and associated MADS box proteins. *Plant Cell* **26**, 89-101.
Georges, A. B., Benayoun, B. A., Caburet, S. and Veitia, R. A. (2010). Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from? *FASEB J.* **24**, 346-356.
Goto, K. and Meyerowitz, E. M. (1994). Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes Dev.* **8**, 1548-1560.
Gramzow, L. and Theißen, G. (2010). A hitchhiker's guide to the MADS world of plants. *Genome Biol.* **11**, 214.
Gramzow, L. and Theißen, G. (2013). Phylogenomics of MADS-box genes in plants — two opposing life styles in one gene family. *Biology* **2**, 1150-1164.
Gramzow, L. and Theißen, G. (2015). Phylogenomics reveals surprising sets of essential and dispensable clades of MIKC^{c}-group MADS-box genes in flowering plants. *J. Exp. Zool. B Mol. Dev. Evol.* **324**, 353-362.
Gramzow, L., Ritz, M. S. and Theißen, G. (2010). On the origin of MADS-domain transcription factors. *Trends Genet.* **26**, 149-153.
Gramzow, L., Weilandt, L. and Theißen, G. (2014). MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Ann. Bot.* **114**, 1407-1429.
Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat. Rev. Genet.* **9**, 15-26.
Honma, T. and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409**, 525-529.
Hsu, W.-H., Yeh, T.-J., Huang, K.-Y., Li, J.-Y., Chen, H.-Y. and Yang, C.-H. (2014). AGAMOUS-LIKE13, a putative ancestor for the E functional genes, specifies male and female gametophyte morphogenesis. *Plant J.* **77**, 1-15.
Hsu, H.-F., Hsu, W.-H., Lee, Y.-I., Mao, W.-T., Yang, J.-Y., Li, Y.-Y. and Yang, C.-H. (2015). Model for perianth formation in orchids. *Nat. Plants* **1**, 15046.
Huang, H., Tudor, M., Su, T., Zhang, Y., Hu, Y. and Ma, H. (1996). DNA binding properties of two *Arabidopsis* MADS domain proteins: Binding consensus and dimer formation. *Plant Cell* **8**, 81-94.
Immink, R. G. H., Tonaco, I. A. N., de Folter, S., Shchennikova, A., van Dijk, A. D. J., Busscher-Lange, J., Borst, J. W. and Angenent, G. C. (2009). SEPALLATA3: The 'glue for MADS box transcription factor complex formation. *Genome Biol.* **10**, R24.
Irish, V. F. (2010). The flowering of Arabidopsis flower development. *Plant J.* **61**, 1014-1028.
Jack, T. (2001). Relearning our ABCs: new twists on an old model. *Trends Plant Sci.* **6**, 310-316.
Jack, T. (2004). Molecular and genetic mechanisms of floral control. *Plant Cell* **16**, S1-S17.
Jack, T., Brockman, L. L. and Meyerowitz, E. M. (1992). The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* **68**, 683-697.
Jetha, K., Theißen, G. and Melzer, R. (2014). Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. *Nucleic Acids Res.* **42**, 10927-10942.
Jofuku, K. D., den Boer, B. G. W., Van Montagu, M. and Okamuro, J. K. (1994). Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. *Plant Cell* **6**, 1211-1225.
Kaltenegger, E. and Ober, D. (2015). Paralogue interference affects the dynamics after gene duplication. *Trends Plant Sci.* **20**, 814-821.
Kaufmann, K., Anfang, N., Saedler, H. and Theißen, G. (2005a). Mutant analysis, protein–protein interactions and subcellular localization of the Arabidopsis B-sister (ABS) protein. *Mol. Genet. Genomics* **274**, 103-118.
Kaufmann, K., Melzer, R. and Theißen, G. (2005b). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* **347**, 183-198.
Kaufmann, K., Muiño, J. M., Jauregui, R., Airoldi, C. A., Smaczniak, C., Krajewski, P. and Angenent, G. C. (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS Biol.* **7**, e1000090.
Kaufmann, K., Pajoro, A. and Angenent, G. C. (2010a). Regulation of transcription in plants: mechanisms controlling developmental switches. *Nat. Rev. Genet.* **11**, 830-842.
Kaufmann, K., Wellmer, F., Muino, J. M., Ferrier, T., Wuest, S. E., Kumar, V., Serrano-Mislata, A., Madueno, F., Krajweski, P., Meyerowitz, E. M. et al. (2010b). Orchestration of floral initiation by APETALA1. *Science* **328**, 85-89.
Krizek, B. A. and Fletcher, J. C. (2005). Molecular mechanisms of flower development: An armchair guide. *Nat. Rev. Genet.* **6**, 688-698.
Krizek, B. A. and Meyerowitz, E. M. (1996a). The *Arabidopsis* homeotic genes *APETALA3* and *PISTILLATA* are sufficient to provide the B class organ identity function. *Development* **122**, 11-22.
Krizek, B. A. and Meyerowitz, E. M. (1996b). Mapping the protein regions responsible for the functional specificities of the Arabidopsis MADS domain organ-identity proteins. *Proc. Natl. Acad. Sci. USA* **93**, 4063-4070.
Leseberg, C. H., Eissler, C. L., Wang, X., Johns, M. A., Duvall, M. R. and Mao, L. (2008). Interaction study of MADS-domain proteins in tomato. *J. Exp. Bot.* **59**, 2253-2265.
Litt, A. (2007). An evaluation of A-function: Evidence from the *APETALA1* and *APETALA2* gene lineages. *Int. J. Plant Sci.* **168**, 73-91.
Liu, D., Wang, D., Qin, Z., Zhang, D., Yin, L., Wu, L., Colasanti, J., Li, A. and Mao, L. (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. *Plant J.* **77**, 284-296.
MacLean, A. M., Orlovskis, Z., Kowitwanich, K., Zdziarska, A. M., Angenent, G. C., Immink, R. G. H. and Hogenhout, S. A. (2014). Phytoplasma effector SAP54 hijacks plant reproduction by degrading MADS-box proteins and promotes insect colonization in a RAD23-dependent manner. *PLoS Biol.* **12**, e1001835.
Maejima, K., Iwai, R., Himeno, M., Komatsu, K., Kitazawa, Y., Fujita, N., Ishikawa, K., Fukuoka, M., Minato, N., Yamaji, Y. et al. (2014). Recognition of floral homeotic MADS domain transcription factors by a phytoplasmal effector, phyllogen, induces phyllody. *Plant J.* **78**, 541-554.
Mandel, M. A., Gustafson-Brown, C., Savidge, B. and Yanofsky, M. F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* **360**, 273-277.
Melzer, R. and Theißen, G. (2009). Reconstitution of 'floral quartets' *in vitro* involving class B and class E floral homeotic proteins. *Nucleic Acids Res.* **37**, 2723-2736.
Melzer, R., Kaufmann, K. and Theißen, G. (2006). Missing links: DNA-binding and target gene specificity of floral homeotic proteins. *Adv. Bot. Res.* **44**, 209-236.
Melzer, R., Verelst, W. and Theißen, G. (2009). The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes *in vitro*. *Nucleic Acids Res.* **37**, 144-157.
Melzer, R., Härter, A., Rümpler, F., Kim, S., Soltis, P. S., Soltis, D. E. and Theißen, G. (2014). DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. *Ann. Bot.* **114**, 1431-1443.
Mendes, M. A., Guerra, R. F., Berns, M. C., Manzo, C., Masiero, S., Finziu, L., Kater, M. M. and Colombo, L. (2013). MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in *Arabidopsis*. *Plant Cell* **25**, 2560-2572.
Meyerowitz, E. M., Smyth, D. R. and Bowman, J. L. (1989). Abnormal flowers and pattern formation in floral development. *Development* **106**, 209-217.
Mizukami, Y. and Ma, H. (1992). Ectopic expression of the floral homeotic gene AGAMOUS in transgenic Arabidopsis plants alters floral organ identity. *Cell* **71**, 119-131.
Mizzotti, C., Mendes, M. A., Caporali, E., Schnittger, A., Kater, M. M., Battaglia, R. and Colombo, L. (2012). The MADS box genes *SEEDSTICK* and *ARABIDOPSIS Bsister* play a maternal role in fertilization and seed development. *Plant J.* **70**, 409-420.
Mondragón-Palomino, M. and Theißen, G. (2008). MADS about the evolution of orchid flowers. *Trends Plant Sci.* **13**, 51-59.
Mondragón-Palomino, M. and Theißen, G. (2011). Conserved differential expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the flowers of Orchidaceae: refining the 'orchid code'. *Plant J.* **66**, 1008-1019.
Moyle-Heyrman, G., Tims, H. S. and Widom, J. (2011). Structural constraints in collaborative competition of transcription factors against the nucleosome. *J. Mol. Biol.* **412**, 634-646.

Muino, J. M., Smaczniak, C., Angenent, G. C., Kaufmann, K. and van Dijk, A. D. J. (2014). Structural determinants of DNA recognition by plant MADS-domain transcription factors. *Nucleic Acids Res.* **42**, 2138-2146.
Münster, T., Pahnke, J., DiRosa, A., Kim, J. T., Martin, W., Saedler, H. and Theissen, G. (1997). Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proc. Natl. Acad. Sci. USA* **94**, 2415-2420.
Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., Vonrhein, C., Moras, D., Romier, C., Bolognesi, M. et al. (2013). Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell* **152**, 132-143.
Nesi, N., Debeaujon, I., Jond, C., Stewart, A. J., Jenkins, G. I., Caboche, M. and Lepiniec, L. (2002). The *TRANSPARENT TESTA16* locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* **14**, 2463-2479.
Nougalli-Tonaco, I. A., Borst, J. W., de Vries, S. C., Angenent, G. C. and Immink, R. G. H. (2006). In vivo imaging of MADS-box transcription factor interactions. *J. Exp. Bot.* **57**, 33-42.
Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A. et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579-584.
Ó'Maoiléidigh, D. S., Wuest, S. E., Rae, L., Raganelli, A., Ryan, P. T., Kwaśniewska, K., Das, P., Lohan, A. J., Loftus, B., Garciet, E. et al. (2013). Control of reproductive floral organ identity specification in *Arabidopsis* by the C function regulator AGAMOUS. *Plant Cell* **25**, 2482-2503.
Ó'Maoiléidigh, D. S., Graciet, E. and Wellmer, F. (2014). Gene networks controlling *Arabidopsis thaliana* flower development. *New Phytol.* **201**, 16-30.
Pajoro, A., Madrigal, P., Muiño, J. M., Matus, J. T., Jin, J., Mecchia, M. A., Debernardi, J. M., Palatnik, J. F., Balazadeh, S., Arif, M. et al. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol.* **15**, R41.
Parenicová, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., Cook, H. E., Ingram, R. M., Kater, M. M., Davies, B. et al. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* **15**, 1538-1551.
Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E. and Yanofsky, M. F. (2000). B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* **405**, 200-203.
Pelaz, S., Tapia-López, R., Alvarez-Buylla, E. R. and Yanofsky, M. F. (2001). Conversion of leaves into petals in *Arabidopsis*. *Curr. Biol.* **11**, 182-184.
Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E. and Yanofsky, M. F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**, 85-88.
Polach, K. J. and Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J. Mol. Biol.* **258**, 800-812.
Puranik, S., Acajjaoui, S., Conn, S., Costa, L., Conn, V., Vial, A., Marcellin, R., Melzer, R., Brown, E., Hart, D. et al. (2014). Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in *Arabidopsis*. *Plant Cell* **26**, 3603-3615.
Purugganan, M. D., Rounsley, S. D., Schmidt, R. J. and Yanofsky, M. F. (1995). Molecular evolution of flower development - Diversification of the plant MADS-box regulatory gene family. *Genetics* **140**, 345-356.
Rhee, H. S., Bataille, A. R., Zhang, L. and Pugh, B. F. (2014). Subnucleosomal structures and nucleosome asymmetry across a genome. *Cell* **159**, 1377-1388.
Riechmann, J. L., Wang, M. Q. and Meyerowitz, E. M. (1996). DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Res.* **24**, 3134-3141.
Rijpkema, A. S., Zethof, J., Gerats, T. and Vandenbussche, M. (2009). The petunia *AGL6* gene has a *SEPALLATA*-like function in floral patterning. *Plant J.* **60**, 1-9.
Ruelens, P., de Maagd, R. A., Proost, S., Theißen, G., Geuten, K. and Kaufmann, K. (2013). FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. *Nat. Commun.* **4**, 2280.
Rümpler, F., Gramzow, L., Theißen, G. and Melzer, R. (2015). Did convergent protein evolution enable phytoplasmas to generate 'zombie plants'? *Trends Plant Sci.* **20**, 798-806.
Ruokolainen, S., Ng, Y. P., Albert, V. A., Elomaa, P. and Teeri, T. H. (2010). Large scale interaction analysis predicts that the Gerbera hybrida floral E function is provided both by general and specialized proteins. *BMC Plant Biol.* **10**, 129.
Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* **250**, 931-936.
Seok, H.-Y., Park, H.-Y., Park, J.-I., Lee, Y.-M., Lee, S.-Y., An, G. and Moon, Y.-H. (2010). Rice ternary MADS protein complexes containing class B MADS heterodimer. *Biochem. Biophys. Res. Commun.* **401**, 598-604.
Shchennikova, A. V., Shulga, O. A., Immink, R., Skryabin, K. G. and Angenent, G. C. (2004). Identification and characterization of four chrysanthemum MADS-box genes, belonging to the *APETALA1*/*FRUITFULL* and *SEPALLATA3* subfamilies. *Plant Physiol.* **134**, 1632-1641.
Silva, C. S., Puranik, S., Round, A., Brennich, M., Jourdain, A., Parcy, F., Hugouvieux, V. and Zubieta, C. (2016). Evolution of the plant reproduction master regulators LFY and the MADS transcription factors: the role of protein structure in the evolutionary development of the flower. *Front. Plant Sci.* **6**, 1193.
Slattery, M., Zhou, T., Yang, L., Dantas Machado, A. C., Gordân, R. and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* **39**, 381-399.
Smaczniak, C., Immink, R. G. H., Angenent, G. C. and Kaufmann, K. (2012a). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* **139**, 3081-3098.
Smaczniak, C., Immink, R. G. H., Muino, J. M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q. D., Liu, S., Westphal, A. H., Boeren, S. et al. (2012b). Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci. USA* **109**, 1560-1565.
Soboleva, T. A., Nekrasov, M., Ryan, D. P. and Tremethick, D. J. (2014). Histone variants at the transcription start-site. *Trends Genet.* **30**, 199-209.
Tanabe, Y., Hasebe, M., Sekimoto, H., Nishiyama, T., Kitani, M., Henschel, K., Münster, T., Theißen, G., Nozaki, H. and Ito, M. (2005). Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. *Proc. Natl. Acad. Sci. USA* **102**, 2436-2441.
Theißen, G. (2001). Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* **4**, 75-85.
Theißen, G. and Gramzow, L. (2016). Structure and evolution of plant MADS-domain transcription factors. In *Plant Transcription Factors: Evolutionary, Structural and Functional Aspects* (ed. D. H. Gonzalez), pp. 127-138. Philadelphia: Elsevier.
Theißen, G. and Melzer, R. (2006). Combinatorial control of floral organ identity by MADS-domain transcription factors. In *Ann. Plant Rev., Volume 29 - Regulation of Transcription in Plants* (ed. K. Grasser), pp. 253-265. New York: Blackwell Publishing.
Theißen, G. and Melzer, R. (2007). Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann. Bot.* **100**, 603-619.
Theißen, G. and Saedler, H. (1995). MADS-box genes in plant ontogeny and phylogeny: Haeckel's 'biogenetic law' revisited. *Curr. Opin. Genet. Dev.* **5**, 628-639.
Theißen, G. and Saedler, H. (2001). Floral quartets. *Nature* **409**, 469-471.
Theißen, G., Kim, J. T. and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* **43**, 484-516.
Theißen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J. T., Münster, T., Winter, K.-U. and Saedler, H. (2000). A short history of MADS-box genes in plants. *Plant Mol. Biol.* **42**, 115-149.
Thompson, B. E., Bartling, L., Whipple, C., Hall, D. H., Sakai, H., Schmidt, R. and Hake, S. (2009). *bearded-ear* encodes a MADS box transcription factor critical for maize floral development. *Plant Cell* **21**, 2578-2590.
Todeschini, A.-L., Georges, A. and Veitia, R. A. (2014). Transcription factors: specific DNA binding and specific gene regulation. *Trends Genet.* **30**, 211-219.
Wang, Y.-Q., Melzer, R. and Theißen, G. (2010). Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm *Gnetum gnemon* provide a clue to the evolutionary origin of 'floral quartets'. *Plant J.* **64**, 177-190.
Wellmer, F., Graciet, E. and Riechmann, J. L. (2014). Specification of floral organs in *Arabidopsis*. *J. Exp. Bot.* **65**, 1-9.
Whitty, A. (2008). Cooperativity and biological complexity. *Nat. Chem. Biol.* **4**, 435-439.
Winter, K.-U., Becker, A., Münster, T., Kim, J. T., Saedler, H. and Theißen, G. (1999). MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. USA* **96**, 7342-7347.
Wuest, S. E., O'Maoileidigh, D. S., Rae, L., Kwasniewska, K., Raganelli, A., Hanczaryk, K., Lohan, A. J., Loftus, B., Graciet, E. and Wellmer, F. (2012). Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. *Proc. Natl. Acad. Sci. USA* **109**, 13452-13457.
Yan, W., Chen, D. and Kaufmann, K. (2016). Molecular mechanisms of floral organ specification by MADS domain proteins. *Curr. Opin. Plant Biol.* **29**, 154-162.
Yang, Y. Z. and Jack, T. (2004). Defining subdomains of the K domain important for protein–protein interactions of plant MADS proteins. *Plant Mol. Biol.* **55**, 45-59.
Yang, Y. Z., Fanning, L. and Jack, T. (2003). The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. *Plant J.* **33**, 47-59.

Development (2016) 143, 3259-3271 doi:10.1242/dev.134080

Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A. and Meyerowitz, E. M. (1990). The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. *Nature* **346**, 35-39.

Zahn, L. M., Kong, H., Leebens-Mack, J. H., Kim, S., Soltis, P. S., Landherr, L. L., Soltis, D. E., dePamphilis, C. W. and Ma, H. (2005). The evolution of the SEPALLATA subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* **169**, 2209-2223.

2.6 Manuscript V

Rümpler, F., Theißen, G., Melzer, R. (2017). Sequence features of MADS-domain proteins that act as hubs in the protein-protein interaction network controlling flower development. *bioRxiv*,125294.

Sequence features of MADS-domain proteins that act as hubs in the protein-protein interaction network controlling flower development

Florian Rümpler[a], Günter Theißen[a,1,*] and Rainer Melzer[a,b,1,*]

[a] *Department of Genetics, Friedrich Schiller University Jena, Philosophenweg 12, D-07743 Jena, Germany*

[b] *School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland*

[1] *Contributed equally to this project*

[*] *For correspondence (e-mail: guenter.theissen@uni-jena.de; rainer.melzer@ucd.ie)*

Key Words

Amborella trichopoda, Arabidopsis thaliana, Coiled-coil; Floral quartet; Keratin-like domain; Leucine-zipper; MIKC-type protein; SEPALLATA3; Transcription factor

ABSTRACT

Protein-protein interaction networks (PPIs) are usually scale-free networks that contain a small number of highly connected nodes (hubs) and many poorly connected nodes. However, the molecular mechanisms that underlie the promiscuous interactions of hub proteins remain largely unknown. Here, we show that the floral homeotic MADS-domain transcription factor SEPALLATA3 from *Arabidopsis thaliana* can act as a hub in the PPI controlling flower development because it contains leucine residues at inter- and intramolecular interaction interfaces. Comprehensive sequence analyses of diverse MADS-domain proteins indicate exceedingly high conservation of the identified leucine residues within SEPALLATA-subfamily proteins, whereas non-hub MADS-domain proteins exhibit preferences for other amino acids at homologous sites. Our results indicate that the conservation of leucine residues at positions critical for protein-protein interactions contributed significantly to the present-day structure of the PPI and may have facilitated the evolution of the flower.

Complexity of biological systems is often achieved by the combinatorial activity of a small number of factors [1]. One important example are protein-protein interaction networks (PPIs) that are based on transcription factors (TFs) that act in a combinatorial manner to accomplish the required degree of e.g. morphological complexity. PPIs are often scale-free networks. They contain a small number of hub proteins with many interaction partners and a large number of poorly connected nodes. Though combinatorial control is of eminent importance for almost all developmental processes, the molecular determinants that are underlying the specific combinatorial interactions remain poorly understood. This is especially true for protein-protein interactions among TFs belonging to the same family. The respective TFs are often very similar in terms of sequence and biochemical properties yet fulfill highly distinct and specific functions which are at least partially determined by distinct protein-protein interactions. The PPI controlling flower development in angiosperms is a good case in point. Floral organ specification is regulated by so called floral quartets which are organ specific tetrameric complexes of MIKC-type MADS-domain TFs bound to two adjacent DNA-binding sites while looping the DNA to regulate target genes [2, 3, 4, 5, 6, 7]. In the model plant species *Arabidopsis thaliana* the floral homeotic protein SEPALLATA3 (SEP3) together with its paralogs SEP1, SEP2 and SEP4 from the closely related LOFSEP-subfamily bears a central role by forming tetrameric complexes with numerous other MIKC-type MADS-domain TFs [5, 8, 9, 10]. The four SEP proteins act in a largely redundant manner but in agreement with their central position in the PPI controlling flower development *sep* multiple mutants show severe developmental defects [3, 4]. *sep1 sep2 sep3* triple mutant plants develop sepals from primordia that would normally develop into petals, stamens and carpels and *sep1 sep2 sep3 sep4* quadruple mutants develop vegetative leaves instead of floral organs [3, 4].

Among the four SEP genes, SEP3 has been studied best [5, 6, 8, 9, 11, 12]. Beyond the formation of complexes that determine floral organ identity SEP3 is also involved in controlling flowering time, floral transition and ovule development [8, 11, 13, 14]. It does therefore constitute one of the major hub proteins within the PPI controlling reproductive development [8, 9, 11, 13]. However, it is unclear which biochemical and biophysical properties enable SEP3 to interact with numerous partners whereas other MIKC-type MADS-domain TFs show a much narrower interaction spectrum. For example, the floral homeotic proteins APETALA3 (AP3) and PISTILLATA (PI) from *A. thaliana* that are involved in the developmental specification of petals and stamens do only form obligate heterodimers and require SEP proteins for tetramer formation [5, 8, 15].

The protein-protein interactions that allow for tetramer formation are mainly mediated by the about 80 amino acids long keratin-like domain (K-domain), which is shared by all MIKC-type MADS-domain TFs [5, 16, 17]. The amino acid sequence within the K-domain of most MADS-domain proteins shows three characteristic heptad repeat patterns (K1-; K2-; K3-subdomain repeat) of the form $[abcdefg]_n$, where most 'a' and 'd' positions are occupied by highly hydrophobic residues [16, 17, 18]. This sequence feature is typical for coiled-coils, a common and intensively studied type of protein-protein interaction domains [19, 20, 21, 22] (Fig. 1). Within a coiled-coil, an α-helix is formed and the amino acids on heptad repeat 'a' and 'd' positions form a stripe of hydrophobic residues that runs along the α-helix and facilitates hydrophobic interaction with a partner coiled-coil [20, 21].

Recently the crystal structure of the complete K-domain of SEP3 was reported [23]. Based on the crystal structure, the K-domain forms two amphipathic α-helices separated by a kink region which prevents intramolecular association of both helices. Helix one comprises the first heptad repeat (K1-subdomain) and is involved in dimerisation of two SEP3 monomers. Helix two spans heptad repeat two (K2-subdomain) that further stabilizes the interaction of two SEP3 monomers and heptad repeat three (K3-subdomain) which constitutes an interface for the interaction of two SEP3 dimers i.e. tetramerisation (Fig. 1). In this study, we determine sequence features that enable SEP3 to form tetrameric complexes and identify the amino acid patterns that distinguish members of the SEP3- and LOFSEP-protein subfamilies (termed SEP subfamily henceforth for simplicity) from other MIKC-type MADS-domain TFs with more restricted interaction capabilities. Our data suggest that leucine residues at intramolecular contact points and at the interaction interface of the K3-subdomain are indispensable for tetrameric complex formation. Those leucines are highly conserved in the SEP subfamily but much less frequent in e.g. AP3- and PI-subfamily MIKC-type MADS-domain TFs. They may thus be a critical denominator that determines the ability of SEP-subfamily proteins to act as a hub protein in the scale free PPI controlling flower development.

RESULTS

Leucine residues in the K-domain strongly influence cooperative DNA-binding of SEP3

To investigate the relevance of the different K-subdomains for cooperative DNA-binding and tetramer formation, single and double amino acid substitutions to proline were performed. Proline was chosen because it is known to possess helix-breaking properties [24, 25]. For each of the three K-subdomains two substitution mutants were created (Fig. 2a, Supplementary Fig. 1). Based on coiled-coil predictions one substitution mutant was supposed to destroy the K-subdomain coiled-coil (L115P for K1-, L131P-L135P for K2- and L164P for K3-subdomain, respectively) whereas the other one was expected to not alter the formation of the respective coiled-coil (S94P (K1); L145P (K2); G178P (K3)). Beyond the three K-subdomains, we also introduced proline substitutions at positions occupied by two conserved hydrophobic amino acids in the interhelical region between the K1- and the K2-subdomain (L120P and L123P, Fig. 2a) because positions homologous to L120 and L123 have been shown to be important for the interaction of MADS-domain proteins [17].

We used electrophoretic mobility shift assays (EMSAs) to study the DNA-binding and tetramerisation behavior of the mutant SEP3 proteins. Based on previous studies it is known that SEP3 binds as homodimer to a DNA-element termed CArG-box (for CCArichGG; consensus sequence 5'-CC(A/T)$_6$GG-3') and that four SEP3 proteins bind to a DNA probe containing two CArG-boxes [6]. To first investigate whether DNA-binding affinities of individual dimers were affected by the different amino acid substitutions, we performed saturation binding EMSA experiments using increasing amounts of a DNA probe containing only one CArG-box together with constant amounts of protein as previously described [12]. The estimated affinities for binding of the altered SEP3 proteins to a single DNA-binding site varied slightly but did not considerably differ from the values obtained for SEP3 wild type protein (Supplementary Fig. 2, Supplementary Table 1), indicating that the different amino acid substitutions did not or only marginally affect DNA-binding of individual dimers.

If increasing amounts of SEP3 were incubated together with constant amounts of a DNA probe containing two CArG-boxes, three bands of different electrophoretic mobility were observed (Fig. 2b left side). As determined previously [6] the band of high electrophoretic mobility constitutes unbound DNA (indicated with '0' in Fig. 2b), the band of intermediate electrophoretic mobility constitutes a DNA probe bound by two SEP3 proteins ('2') and the band of low electrophoretic mobility constitutes a DNA probe bound by four SEP3

proteins ('4'). By analyzing the signal intensities of the three different fractions the ability of SEP3 to form DNA-bound tetrameric complexes can be quantified and expressed via the cooperativity constant k_{coop} (equation (4) in Methods). k_{coop} equals 1 for non-cooperative binding and increases with increasing tetramer formation capabilities of the examined protein. SEP3 wild type protein always showed a highly cooperative DNA-binding although the degree of cooperativity varied between different experiments and was slightly higher than previously estimated [6, 12], probably owing to difficulties to precisely determine high k_{coop} values (Fig. 2b and d, Supplementary Table 1).

In contrast to the wild type protein, all of the leucine-to-proline substitution mutants of SEP3 (L115P; L120P-L123P; L131P-L135P; L145P; L164P) showed a considerably reduced ability to bind cooperatively to DNA *in vitro*, independent of whether the formation of coiled-coils was predicted to be affected or not (Fig. 2c and d, Supplementary Table 1). Only the two proline substitutions S94P and G178P, located at the N- and C-terminal borders of the K-domain, respectively, did not strongly reduce cooperative binding of SEP3.

To test the effect of amino acid substitutions that are supposed to have a less severe effect on helix formation than proline, we substituted a subset of the previously selected leucines (L115; L145; L164) by alanine. Surprisingly, of these 3 substitutions only L145A showed a cooperative DNA-binding ability comparable to that of SEP3 wild type protein, whereas substitutions L115A and L164A caused an almost complete loss of cooperative DNA-binding, comparable to the proline substitutions at the respective positions (Fig. 2d, Supplementary Table 1). We further substituted position L164 by three additional amino acids (L164E; L164W; L164I) comprising glutamate and tryptophan which occur at position 164 in several members of the SEP subfamily and isoleucine which has very similar physicochemical properties to leucine. However, none of the resulting mutants was able to approach SEP3 wild type cooperative binding strength (Fig. 2d, Supplementary Table 1). Our results indicate that the examined leucine residues are of critical importance for tetramer formation and cooperative binding of SEP3.

Within the $[abcdefg]_n$ heptad repeat of the K3-subdomain of SEP3 two neighboring 'a' positions (E161; N168) are not occupied by hydrophobic amino acids. Substituting these positions by leucine (E161L-N168L) resulted in a higher probability for the formation of the K3-subdomain coiled-coil *in silico* (Supplementary Fig. 1). The respective mutant protein showed a cooperativity at least as high as the wild type protein in EMSAs. In contrast to the wild type protein, repeated measurements yielded k_{coop} values that consistently were above

200 (Fig. 2d, Supplementary Table 1). In fact, in none of the performed EMSAs a signal of a DNA probe bound by only one protein dimer was detected, an observation that was different from the other proteins for which high cooperativity in DNA-binding was detected (e.g. SEP3-WT and SEP3-L145A) indicating that cooperative binding was increased by the E161L-N168L substitutions (Supplementary Fig. 3). Surprisingly, when we performed saturation binding EMSA experiments using increasing amounts of a DNA probe containing only one CArG-box the mutant protein SEP3-E161L-N168L exhibited no binding of individual dimers. Instead a signal of low electrophoretic mobility occasionally occurred for high amounts of applied DNA probe that might constitutes a protein DNA complex consisting of more than two proteins (Supplementary Fig. 4).

Mutations in a distantly related ortholog of SEP3 have very similar effects on cooperative DNA-binding as in SEP3

The SEP3 ortholog AMtrAGL9 from the early diverging angiosperm *Amborella trichopoda* [15] forms homotetrameric protein-DNA complexes with a cooperative binding affinity comparable to SEP3 (Fig. 2e, Supplementary Fig. 5). AMtrAGL9 amino acid position I141 is homologous to SEP3 L145 and is thus located in the K2-subdomain heptad repeat of AMtrAGL9 (Fig. 2a). Substitution to alanine at that position interfered to some extend with cooperative binding capabilities whereas substitution to proline at position I141 results in an almost complete loss of cooperative binding (Fig. 2e, Supplementary Table 1). If amino acid position L160 of AMtrAGL9, which is homologous to position L164 in the center of the K3-subdomain of SEP3, is exchanged by proline or alanine, the ability of AMtrAGL9 to cooperatively bind to DNA is almost completely lost in either case, a behavior that is similar to that observed for SEP3 (compare Fig. 2d and e).

Interacting sites are more often occupied by leucine in SEP-subfamily proteins than in proteins of other MIKC-type subfamilies

The importance of leucine residues for the tetramerisation ability of SEP3 and AMtrAGL9 raised the question as to which extent these positions are conserved within the SEP subfamily and which amino acid preferences members of other MIKC-type protein subfamilies show at homologous sites. We therefore created a multiple sequence alignment based on 1,325 sequences of MIKC-type MADS-domain proteins belonging to 14 subfamilies and comprising sequences from a diverse array of seed plants. Despite the high evolutionary distance of the sampled taxa, the sequences aligned almost without gaps throughout the

187 complete K-domain (i.e. without potential insertions or deletions). The only exception were
188 PI-subfamily protein sequences, among which a deletion of four amino acids within the C-
189 terminal half of the K-domain was very common. This deletion within the PI-linage most
190 likely occurred after early diverging angiosperms branched off, as most of the sampled PI-
191 subfamily sequences from early diverging angiosperms still possess those four amino acids.

192 We first compared the conservation of sites that are homologous to the 15 residues that (based
193 on the crystal structure of SEP3) mediate the hydrophobic intra- and intermolecular
194 interactions in the SEP3 homotetramer [23] to the overall conservation of the K-domain. We
195 found that within the SEP3 subfamily, sites that are homologous to interacting sites in the
196 SEP3 homotetramer are significantly less variable than the remaining residues of the K-
197 domain (Fig. 3a). This conservation pattern also holds true for sequences of all other 13
198 subfamilies of MIKC-type MADS-domain proteins (Fig. 3b, Supplementary Fig. 6a) as well
199 as for sequences from gymnosperms to core eudicots (Supplementary Fig. 6b). Beyond this
200 similar pattern of conserved positions also the amino acid properties in terms of
201 hydrophobicity at homologous sites appear highly similar among all examined subfamilies
202 (Supplementary Fig. 7), suggesting that the overall structure of the K-domain as determined
203 for SEP3 is conserved among MIKC-type proteins of most if not all subfamilies and
204 throughout seed plants.

205 Next we analyzed the amino acid distribution at sites homologous to the 12 leucine residues
206 (L101, L108, L115, L120, L123, L128, L131, L135, L154, L157, L164 and L171) that
207 contribute to inter- and intramolecular interactions in a SEP3 homotetramer (Fig. 4a) [23]. All
208 these residues were found to be highly conserved within the 78 examined SEP3-subfamily
209 sequences; 8 out of 12 positions were completely invariable (Fig. 4b). In contrast to this,
210 members of other subfamilies (e.g. AP3- and PI-subfamily proteins) often show preferences
211 for other amino acids on equivalent sites (Fig. 4b, Supplementary Fig. 8). Especially positions
212 equivalent to L154, L157 and L164 of SEP3 that are located within the center of the
213 tetramerisation interface are often not occupied by leucines in AP3- and PI-subfamily
214 proteins. The high conservation of leucines also becomes apparent within LOFSEP-subfamily
215 proteins (comprising SEP1, SEP2 and SEP4 from *A. thaliana*) which form the sister group of
216 SEP3-subfamily proteins and that are assumed to function in a mostly redundant manner with
217 SEP3 during flower development (Fig. 4c) [3, 4, 10]. The closest relatives of the SEP subfamily
218 are AGL6-subfamily proteins followed by AP1-subfamily proteins [26]. However, despite the
219 close relationship AGL6- as well as AP1-subfamily proteins display a considerably lower

leucine frequency especially on sites within the tetramerisation interface (Fig. 4c, Supplementary Fig. 8). Instead, these positions are more frequently occupied by other hydrophobic amino acids such as isoleucine and methionine. It has previously been shown that within a coiled-coil, leucine packs very well at heptad repeat 'd' positions and enables the formation of a tight dimeric coiled-coil as it becomes apparent in a leucine-zipper [22, 27]. In contrast other hydrophobic amino acids such as isoleucine or valine lead to steric hindrance at heptad repeat 'd' positions [22] (Fig. 4d).

Insertion of leucine residues into the K3-subdomain of AP3 facilitates homotetramerisation of the chimeric protein SEP3$_{AP3chim}$

Based on our data we hypothesized that the overall structure of the K-domain is conserved throughout most if not all subfamilies of MIKC-type MADS-domain proteins and that preferences for different amino acids on interacting sites account for subfamily specific interaction patterns. We aimed to test our hypothesis with help of the chimeric protein SEP3$_{AP3chim}$, in which we substituted the K3-subdomain (i.e. tetramersiation interface) of SEP3 (residues 150-181) by the homologous sites of AP3 (Fig. 5a and b). AP3 is known to form obligate heterodimers with PI and is thus not able to form DNA-binding homodimers or homotetramers [28, 29]. As expected, the chimeric protein SEP3$_{AP3chim}$ showed a complete loss of homotetramerisation capabilities compared to SEP3 wild type protein in EMSA experiments (Fig. 5c and d right side, Supplementary Table 1). Although the K3-subdomains of SEP3 and AP3 share only four identical residues at homologous sites the sequence similarity in terms of hydrophobicity on most heptad repeat 'a' and 'd' positions is comparatively high (Fig. 5a). However, two heptad repeat 'd' positions occupied by leucine in SEP3 (L157 and L164) are occupied by threonine and glutamine in AP3, respectively (Fig. 5a). Both leucines are highly conserved throughout SEP3-subfamily proteins whereas homologous sites in AP3-subfamily proteins are almost exclusively occupied by residues other than leucine (Fig. 4b). We thus substituted positions T157 and Q164 of the chimeric protein by leucine and tested the ability of the resulting mutants to form homotetramers. Both single amino acid substitutions could not improve tetramerisation ability of the chimeric protein (Supplementary Table 1). However, the insertion of both leucine residues into the K3-subdomain of SEP3$_{AP3chim}$ sufficed to fully 'restore' the ability to form DNA-binding homotetramers (Fig. 5e right side). Visualizing the amino acid sequence of the tetramerisation interface of SEP3 and AP3 in a helical wheel diagram illustrates how residues M150, L157, L164 and L171 form a strong hydrophobic stripe within the tetramerisation interface of SEP3, whereas the hydrophobic

253 stripe is interrupted by threonine and glutamine in AP3 (Fig. 5c and d left side). Substituting
254 both residues by leucine closes the gap within the hydrophobic stripe and most likely thereby
255 facilitates homotetramerisation (Fig. 5e left side).

DISCUSSION

Tetramer formation among MIKC-type MADS-domain transcription factors is of central importance for flower development [5, 7, 8, 9, 30]. However, knowledge about the molecular determinants facilitating tetramer formation remains scarce. Our data indicate that substitution of leucines in the K-domain of SEP3 did almost invariably lead to a strong reduction in tetramer formation abilities (Fig. 2). This was expected for leucine to proline substitutions within the helical regions of the K-domain as proline has helix-breaking properties. However, also the rather conservative substitution from leucine to alanine in the tetramerization interface (L164A) affected cooperative binding and tetramerization strongly. Similar results have been obtained for substituting other leucine residues in the tetramerization interface by alanine [23].

The question arises as to why specifically leucine residues are favoured over other hydrophobic amino acid residues. The tetramerization interface forms coiled-coils and it is well established that complex 'knobs-into-holes' side chain interactions within the hydrophobic core determine the strength of the interaction between coiled-coils [19]. Numerous studies on energetic contributions of different hydrophobic amino acids inside the hydrophobic core revealed that β-branched amino acids (e.g. isoleucine or valine) as well as amino acids with small side chains (e.g. alanine) in heptad repeat 'd' positions have a strong destabilizing effect on formation of parallel dimeric coiled-coils [27, 31]. The local stereochemical environment at heptad repeat 'd' positions instead strongly favours γ-branched amino acids for intermolecular interactions, making leucines uniquely suited at these sites [22, 27, 29, 31]. This is in line with the observation that L145, which is located at a heptad repeat 'd' position but according to structural data not involved in intermolecular interactions [23] can be mutated to alanine without a decrease in tetramer formation capabilities. In contrast, L164 (also at a heptad repeat 'd' position but involved in intermolecular interactions) mutation to alanine leads to a strong decrease in tetramerization. In addition, L145 is by far not as conserved as leucines involved in interactions (Supplementary Fig. 8).

A decrease in tetramer formation was also observed for substitution of leucines in the kink region between the two helices, where an effect on helix formation was not predicted (Supplementary Fig. 1). However, although the leucine residues in the kink are not directly involved in tetramer formation, they interact intramolecularly with each other to stabilize the kink and thus bring the tetramer interface in a favourable position for protein-protein interactions [23]. It is likely that substitutions to proline or alanine in the kink region altered or

destabilized the orientation of the tetramerisation interface and thus impeded tetramer formation indirectly. Similar to the leucines at interacting sites within the helical regions of the K-domain stereochemical restrictions may also in this case favour leucines over other hydrophobic amino acids. This may explain why the L115A mutation in the kink region, which presumably only affects intramolecular interactions, caused a decrease in tetramer formation capabilities.

Taken together, these findings indicate that inter- and intramolecular hydrophobic interactions specifically among leucines are of critical importance for SEP3 homotetramerization. This principle does very likely apply to the entire subfamily of SEP proteins, as leucines at interaction positions are evolutionarily highly conserved throughout this subfamily (Fig. 4). The evolutionary conserved and important role of leucines is further highlighted by the observation that in the SEP3 ortholog AMtrAGL9 from *A. trichopoda* leucines at positions homologous to those in SEP3 were also of critical importance for tetramer formation (Fig. 2).

The K-domain is the second highest conserved domain of MIKC-type proteins (the most highly conserved domain is the MADS-domain) [32]. Previous structural predictions indicated that the K-domain is forming coiled-coils in most if not all MIKC-type proteins [18, 23, 33]. Our analyses indeed strongly support this view. The chemical properties of amino acids that are of importance for intra- and intermolecular interactions in SEP3 are conserved in MIKC-type proteins from all of the 14 subfamilies analysed here. This indicates that most K-domains fold in a structure similar to that determined for SEP3 and that residues that are homologous to interacting sites in the SEP3 homotetramer may also constitute intra- and intermolecular contact points in most other protein family members.

However, although the chemical properties of amino acids important for interactions were conserved in subfamilies other than SEP, their identity was not always. Whereas the vast majority of leucine residues important for intra- and intermolecular interactions is highly conserved within the SEP subfamily, leucine residues are observed at a clearly lower frequency in other subfamilies (Fig. 4, Supplementary Fig. 4). This indicates that, although the overall structure of the K-domain is conserved in all MIKC-type proteins, their tetramerization capabilities may vary depending on the presence of leucines on critical interaction sites. For example, AP3 and PI, who do not possess leucines on all inter- and intramolecular contact points, are unable to form tetramers not involving SEP3 [5, 9]. Indeed, the K3-subdomain of AP3, which is not capable of mediating homotetramer formation, gained this ability when placed in the SEP3 protein context and two leucines were introduced

(Fig. 4). Thus, we speculate that leucines at intra- and intermolecular contact points may not only be necessary but also sufficient for tetramer formation of MIKC-type proteins.

Intriguingly, the high conservation of leucines in the K-domain of SEP-subfamily proteins and their importance for homotetramer formation correlates very well with the crucial function of those proteins as hubs within the PPI controlling flower development. In addition, proteins like AP3 and PI that have less central positions within the interaction network also lack leucines at several positions critical for tetramerization. It thus appears plausible that leucines in SEP-subfamily proteins are not only important for homotetramer formation but also play a pivotal role in the formation of heterotetrameric complexes. For example, though a lack of leucines in the kink region of many MIKC-type proteins may destabilizes the orientation of the teramerization interface and prevents homotetramer formation, the high structural stability of the K-domain of SEP-subfamily proteins that is brought about by intramolecular leucine interactions may serve as a scaffold that helps to align the interaction interface of partner proteins and hence facilitate heterotetramer formation.

The pattern of leucines at the tetramerization interface may be explained in a similar manner. Though data on the interaction of leucines at heptad repeat ‘d’ positions with other amino acids at ‘d’ positions in a heteromeric coiled-coil are scarce, data from leucine zippers indicate that beyond leucine-leucine interactions also interactions of leucines with a number of other amino acids are more favourable than most other interactions not involving any leucine [34].

Taken together, we propose that the leucine residues in SEP-subfamily proteins serve to facilitate heterotetrameric interactions while at the same time the absence of leucines in the interaction partners prevents homotetramer formation or formation of heterotetramers not involving SEP-subfamily proteins. This way, SEP-subfamily proteins could act as hubs in the scale free PPI controlling flower development: tetramerization of many proteins depends on them and probably cannot occur in the absence of SEP-subfamily proteins.

We previously proposed that the dependence of other MIKC-type proteins on SEP-subfamily proteins for tetramer formation facilitated the concerted development of the different floral organs and the evolution of the flower as a single reproductive entity [15]. The evolutionary conservation of leucines in the SEP subfamily as opposed to most other subfamilies may be thus one important molecular mechanism that fostered the evolution of the flower.

353 Importantly, however, coiled-coil interactions are very complex, with amino acids occupying
354 heptad repeat 'a', 'd', 'e' and 'g' position playing key roles in determining the affinity and
355 specificity of an interaction [20, 21, 35] and we are far from completely understanding the
356 implications of sequence variations on the different positions for MIKC-type protein
357 interactions. For example, polar and charged residues are observed at heptad repeat 'd'
358 positions in a number of MIKC-type protein subfamilies and those would be expected to not
359 only hinder homotetramerization but also heterotetramerization with SEP-subfamily proteins.
360 Furthermore, subfamily specific patterns of charged residues at heptad repeat 'e' and 'g'
361 positions can be observed that may account for differences in interaction specificity. Although
362 our findings bring us one step closer towards solving the code for floral quartet-like complex
363 formation, additional structural and biophysical analyses are required to more completely
364 understand the molecular mechanisms and evolutionary patterns of MIKC-type protein
365 interactions. This will eventually also lead to a better understanding as to why this
366 transcription factor family expanded in seed plants and plays a role in virtually every
367 reproductive developmental process.

368

METHODS

Cloning procedures and site-directed mutagenesis.

The plasmids for *in vitro* transcription/translation of *SEP3*, *AP3*, *PI* and *AMtrAGL9* (pTNT-SEP3, pSPUTK-AP3, pSPUTK-PI and pSPUTK-AMtrAGL9) have been generated previously [6, 15]. The cDNA sequences for the single- and double amino acid substitution mutants of SEP3 were created by site-directed mutagenesis PCR according to the Q5 Site-Directed Mutagenesis Manual (New England Biolabs). The cDNA sequence for the chimeric protein $SEP3_{AP3chim}$ was created by megaprimer-mediated mutagenesis PCR for domain substitutions according to [36].

Design of DNA probes and radioactive labeling.

Design and preparation of DNA probes have been described previously [6]. The CArG-box sequence 5'-CCAAATAAGG-3' that was used for all DNA probes was derived from the regulatory intron of *AGAMOUS*. For studies on homotetramer formation a 151 nt long DNA probe was used that contained two CArG-boxes in a distance of 63 bp, i.e. 6 helical turns (sequence: 5'- TCGAG GTCGG AAATT TAATT ATATT CCAAA TAAGG AAAGT ATGGA ACGTT CGACG GTATC GATAA GCTTG ATGAA ATTTA ATTAT ATTCC AAATA AGGAA AGTAT GGAAC GTTAT CGAAT TCCTG CAGCC CGGGG GATCC ACTAG TTCTA G -3', CArG-box sequences are underlined). Saturation binding assays to quantify dimer binding affinities were performed with a 51 nt long DNA probe harboring a single CArG-box in the center (sequence: 3'- AATTC GAAAT TTAAT TATAT TCCAA ATAAG GAAAG TATGG AACGT TGAAT T- 5', CArG-box sequence is underlined). The DNA probes were radioactively labeled via Klenow fill-in reaction of 5'-overhangs with [α-^{32}P] dATP.

In vitro transcription/translation and electrophoretic mobility shift assay.

Proteins were produced *in vitro* using the TNT SP6 Quick Coupled Transcription/Translation System (Promega) according to the manufacturer's instructions and used directly without freezing and thawing. The composition of the protein-DNA binding reaction buffer was essentially as described [37], with final concentration of 1.6 mM EDTA, 10.3 mM HEPES, 1 mM DTT, 1.3 mM spermidine, 33.3 ng/µl Poly dI/dC, 2.5 % CHAPS, 4.3 % glycerol, and a minimum of 1.3 µg/µl BSA. The amounts of protein, DNA probe and BSA were varied according to the performed assay. For cooperative DNA-binding studies to infer tetramer

formation capabilities a constant amount of 0.1 ng of a DNA probe containing two CArG-boxes in a distance of six helical turns was co-incubated with variable amounts of *in vitro* translated protein ranging from 0.05 µl to 3 µl. Variable amounts of applied *in vitro* translated protein were compensated by adding appropriate volumes of BSA (10 µg/µl). For saturation binding assays to quantify dimer binding affinities a constant amount of 2 to 5 µl *in vitro* translated protein was co-incubated with variable amounts of a DNA probe containing one CArG-box in the center, ranging from 0.05 to 32 ng as previously described in [12]. Binding reactions had a total volume of 12 µl, were incubated overnight at 4°C and subsequently loaded on a polyacrylamide (5 % acrylamide, 0.1725 % bisacrylamid) 0.5x TBE gel that has been pre-run for 30 min. The gel was run with 0.5x TBE buffer for 2.5 h at 7.5 V/cm and afterwards dried and exposed onto a phosphorimaging screen to quantify signal intensities.

Quantification of cooperative DNA-binding.

For each lane of the EMSA gel relative signal intensities of all fractions were measured using Multi Gauge 3.1 (Fujifilm). The equations that were used to quantify the ability for cooperative DNA-binding of two dimers to a DNA probe carrying two CArG-boxes have been described previously [6, 38]. Briefly, if the relative concentration of unbound DNA probe [Y_0] (signal of high electrophoretic mobility), a DNA probe bound by two proteins [Y_2] (signal of intermediate electrophoretic mobility), and a DNA probe bound by four proteins [Y_4] (signal of low electrophoretic mobility) are described as a function of applied protein [P2],

$$[Y_0] = \frac{1}{1+\left(\frac{2}{k_{d1}}\right)\times[P2]+\left(\frac{1}{k_{d1}\times k_{d2}}\right)\times[P2]^2} \quad \textbf{(1)}$$

$$[Y_2] = \frac{\left(\frac{2}{k_{d1}}\right)\times[P2]}{1+\left(\frac{2}{k_{d1}}\right)\times[P2]+\left(\frac{1}{k_{d1}\times k_{d2}}\right)\times[P2]^2} \quad \textbf{(2)}$$

$$[Y_4] = \frac{\left(\frac{1}{k_{d1}\times k_{d2}}\right)\times[P2]^2}{1+\left(\frac{2}{k_{d1}}\right)\times[P2]+\left(\frac{1}{k_{d1}\times k_{d2}}\right)\times[P2]^2} \quad \textbf{(3)}$$

then k_{d1} is the dissociation constant for binding of a protein dimer to a DNA probe with two unoccupied binding sites and k_{d2} is the dissociation constant for binding of a second protein dimer to a DNA probe where one of the two binding sites is already occupied. By nonlinear regression of the measured signal intensities of the three fractions to equation (1) to (3), k_{d1}

and k_{d2} were estimated using GraphPad Prism 5 (GraphPad Software). As we used *in vitro* transcription/translation for protein production, the exact protein concentrations were unknown. Therefore the amount of applied *in vitro* transcription/translation mixture was used as proxy for [P2], as previously described [6]. As a result of the unknown protein concentrations the estimated values for k_{d1} and k_{d2} depend on the *in vitro* transcription/translation efficiency and can only be considered as relative values. However, estimating a cooperativity constant k_{coop} (defined as the ratio of k_{d1} and k_{d2}) is still possible:

$$k_{coop} = \frac{k_{d1}}{k_{d2}} \quad \textbf{(4)}$$

As described earlier, k_{coop} values of ≈ 200 were the upper limit that could be determined with our experimental setup [12].

Saturation binding assay. To estimate the dissociation constant for binding of a protein dimer to a single DNA-binding site k_d, saturation binding assays with a DNA probe carrying a single CArG-box were performed. The equation that was used to infer k_d has been described previously [12]. k_d can be defined as

$$k_d = \frac{([\mathrm{P_t}]-[\mathrm{PD}])\times[\mathrm{D}]}{[\mathrm{PD}]} \quad \textbf{(5)}$$

with [PD], [P_t], and [D] being the concentration of the protein-DNA complex, total protein, and unbound DNA probe, respectively. By expressing [PD] as a function of [D] for increasing concentrations of applied DNA probe, [P_t] and k_d were determined via nonlinear regression using GraphPad Prism 5.

Multiple sequence alignments and *in silico* sequence analysis.

For analyses on amino acid preferences of different MIKC-type MADS-domain protein subfamilies throughout the K-domain a comprehensive sequence collection was compiled. Via BLAST search [39] representatives of all 14 subfamilies [40, 41] of MIKC-type proteins present in *A. thaliana* (AP1-, AP3-, PI-, AG-, ABS-, SEP3-, LOFSEP-, AGL6-, AGL12-, AGL15-, AGL17-, FLC-, TM3-, and SVP-subfamily) were collected using the amino acid sequences of *A. thaliana* AP1, AP3, PI, AG, ABS, SEP3, SEP1, AGL6, AGL12, AGL15, AGL17, FLC, SOC1, and SVP, respectively, as query. To cover a broad set of species, six individual searches were performed for each subfamily. Each of those searches was restricted to a different group of seed plants: core eudicots, early diverging eudicots, monocots, magnoliids,

early diverging angiosperms, and gymnosperms. For sequences from core eudicots the search queries were restricted to asterids (BLAST tax-ID: 71274), Dilleniaceae (24942), Caryophyllidea (108240), Santalales (41947), Berberidopsidales (403664), Saxifragales (41946), rosids (71275), and Gunnerales (232382); for sequences from early diverging eudicots the search queries were restricted to Proteales (232378), Buxales (280577), and Ranunculales (41768); for sequences from monocots and magnoliids, respectively, the queries were restricted to the corresponding predefined organism groups implemented in BLAST (tax-ID: 4447 and 232347, respectively); for sequences from early diverging angiosperms the queries were restricted to Austrobaileyales (82956), Hydatellaceae (178426), Nymphaeales (261007), and Amborella (13332); and for sequences from gymnosperms the queries were restricted to Gnetales (3378), Pinaceae (3318), Taxaceae (25623), Cephalotaxus (50178), Cupressaceae (3367), Araucariaceae (25664), Podocarpaceae (3362), Ginkgoales (3308), and Cycadales (3297). For each of the 84 resulting BLAST searches the amino acid sequences of all hits were downloaded (if more than 100 sequences were found, only top 100 hits according to the total score calculated by BLAST were downloaded). The results of all BLAST searches were combined into a single data set, all completely redundant sequences as well as all sequences that did not constitute MIKC-type proteins were removed and the remaining sequences were aligned with Mafft applying E-INS-i mode using Jalview [42, 43]. The subfamily assignment of each sequence was performed according to its clustering within a phylogenetic tree calculated with MrBayes (based on MADS-, I- and K-domain sequences, applying mixed AA model with 20 million generations, 50% burn-in, and a sample frequency of 1000) [44]. All sequences with uncertain subfamily assignment were removed. To optimize the alignment quality of the K-domain 133 further sequences were removed that produced gaps and that did not appear to be representative for the respective subfamily. The final sequence collection comprised 1325 MIKC-type protein sequences.

Relative sequence similarities at homologous sites were calculated with R (https://www.R-project.org/). Each pair of amino acids at equivalent sites were assigned a similarity score based on BLOSUM40 values normalized to 1. Subsequently, all pairwise similarity scores were averaged to calculate the mean relative sequence similarity for all amino acid positions within the K-domain. BLOSUM40 was chosen because the average sequence identity within the K-domain of all examined sequences was about 40 %. Box-plots and line graphs of sequence similarity scores were created with SPSS (IBM). Statistical significance of sequence similarity differences were tested via Mann-Whitney U-tests implemented in SPSS.

Subfamily specific amino acid frequencies and mean hydrophobicity values for positions within the K-domain were calculated with R. SEP3 K-domain crystal structure pictures were created with Swiss-PdbViewer (SIB). Helical wheel diagrams were created with R. Coiled-coil predictions to preselect potential positions for single and double amino acid substitutions were performed with COILS [45].

ACKNOWLEDGEMENTS

We are grateful to Fredo-Torpedo (Fred Ferber), Chris-Master (Christian Gafert) and Tanja Schulze for valuable help with some experiments. This research was supported by the DFG (Deutsche Forschungsgemeinschaft) grant to G.T. and R.M. (TH417/5-3). R.M. received a post-doctoral fellowship from the Carl Zeiss Foundation.

AUTHOR CONTRIBUTIONS

G.T., R.M. and F.R. designed the study. F.R. performed the experiments. G.T., R.M. and F.R. wrote the manuscript.

COMPETING FINANCIAL INTEREST

The authors declare no competing financial interests.

REFERENCES

1. Remenyi A, Scholer HR, Wilmanns M. Combinatorial control of gene expression. *Nat Struct Mol Biol* **11**, 812-815 (2004).

2. Coen ES, Meyerowitz EM. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**, 31-37 (1991).

3. Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. *Curr Biol* **14**, 1935-1940 (2004).

4. Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **405**, 200-203 (2000).

5. Melzer R, Theißen G. Reconstitution of 'floral quartets' in vitro involving class B and class E floral homeotic proteins. *Nucleic Acids Res* **37**, 2723-2736 (2009).

6. Melzer R, Verelst W, Theißen G. The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro. *Nucleic Acids Res* **37**, 144-157 (2009).

7. Theißen G, Saedler H. Plant biology. Floral quartets. *Nature* **409**, 469-471 (2001).

8. Immink RGH*, et al.* SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. *Genome Biol* **10**, R24 (2009).

9. Smaczniak C*, et al.* Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc Natl Acad Sci USA* **109**, 1560-1565 (2012).

10. Zahn LM*, et al.* The evolution of the SEPALLATA subfamily of MADS-Box genes: A preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* **169**, 2209-2223 (2005).

11. Favaro R*, et al.* MADS-box protein complexes control carpel and ovule development in Arabidopsis. *Plant Cell* **15**, 2603-2611 (2003).

12. Jetha K, Theißen G, Melzer R. Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. *Nucleic Acids Res* **42**, 10927-10942 (2014).

13. Liu C, Xi W, Shen L, Tan C, Yu H. Regulation of floral patterning by flowering time genes. *Dev Cell* **16**, 711-722 (2009).

14. Lopez-Vernaza M, Yang SX, Muller R, Thorpe F, de Leau E, Goodrich J. Antagonistic roles of SEPALLATA3, FT and FLC genes as targets of the polycomb group gene CURLY LEAF. *Plos One* **7**, e30715 (2012).

15. Melzer R*, et al.* DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. *Ann Bot* **114**, 1431-1443 (2014).

16. Yang YZ, Fanning L, Jack T. The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. *Plant J* **33**, 47-59 (2003).

17. Yang YZ, Jack T. Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins. *Plant Mol Biol* **55**, 45-59 (2004).

18. Riechmann JL, Meyerowitz EM. MADS domain proteins in plant development. *Biol Chem* **378**, 1079-1101 (1997).

19. Parry DA, Fraser RD, Squire JM. Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. *J Struct Biol* **163**, 258-269 (2008).

20. Mason JM, Arndt KM. Coiled coil domains: Stability, specificity, and biological implications. *Chembiochem* **5**, 170-176 (2004).

21. Mason JM, Hagemann UB, Arndt KM. Role of hydrophobic and electrostatic interactions in coiled coil stability and specificity. *Biochemistry* **48**, 10380-10388 (2009).

22. Betz SF, Bryson JW, Degrado WF. Native-like and structurally characterized designed alpha-helical bundles. *Curr Opin Struc Biol* **5**, 457-463 (1995).

23. Puranik S*, et al.* Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in Arabidopsis. *Plant Cell* **26**, 3603-3615 (2014).

24. Nilsson I, Saaf A, Whitley P, Gafvelin G, Waller C, von Heijne G. Proline-induced disruption of a transmembrane alpha-helix in its natural environment. *J Mol Biol* **284**, 1165-1175 (1998).

25. Richardson JS. The anatomy and taxonomy of protein structure. *Adv Protein Chem* **34**, 167-339 (1981).

26. Kim S, Soltis PS, Soltis DE. AGL6-like MADS-box genes are sister to AGL2-like MADS-box genes. *J Plant Biol* **56**, 315-325 (2013).

27. Zhu BY, Zhou NE, Kay CM, Hodges RS. Packing and hydrophobicity effects on protein folding and stability - effects of beta-branched amino-acids, valine and isoleucine, on the formation

and stability of 2-stranded alpha-helical coiled coils leucine zippers. *Protein Sci* **2**, 383-394 (1993).

28. de Folter S*, et al.* Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. *Plant Cell* **17**, 1424-1433 (2005).

29. Moitra J, Szilak L, Krylov D, Vinson C. Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. *Biochemistry* **36**, 12567-12573 (1997).

30. Theißen G, Melzer R, Rümpler F. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development* **143**, 3259-3271 (2016).

31. Takei T*, et al.* The effects of the side chains of hydrophobic aliphatic amino acid residues in an amphipathic polypeptide on the formation of a helix and its association. *J Biochem* **139**, 271-278 (2006).

32. Kaufmann K, Melzer R, Theißen G. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* **347**, 183-198 (2005).

33. Ma H, Yanofsky MF, Meyerowitz EM. AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev* **5**, 484-495 (1991).

34. Fong JH, Keating AE, Singh M. Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biol* **5**, R11 (2004).

35. Potapov V, Kaplan JB, Keating AE. Data-driven prediction and design of bZIP coiled-coil interactions. *Plos Comput Biol* **11**, (2015).

36. Perez K, Yeam I, Jahn MM, Kang BC. Megaprimer-mediated domain swapping for construction of chimeric viruses. *J Virol Methods* **135**, 254-262 (2006).

37. Egea-Cortines M, Saedler H, Sommer H. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in Antirrhinum majus. *Embo J* **18**, 5370-5379 (1999).

38. Senear DF, Brenowitz M. Determination of binding constants for cooperative site-specific protein-DNA interactions using the gel mobility-shift assay. *J Biol Chem* **266**, 13661-13671 (1991).

39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).

630 40. Becker A, Theißen G. The major clades of MADS-box genes and their role in the development
631 and evolution of flowering plants. *Mol Phylogenet Evol* **29**, 464-489 (2003).

632
633 41. Gramzow L, Theißen G. Phylogenomics of MADS-Box genes in plants - two opposing life styles
634 in one gene family. *Biology (Basel)* **2**, 1150-1164 (2013).

635
636 42. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple
637 sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).

638
639 43. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
640 improvements in performance and usability. *Mol Biol Evol* **30**, 772-780 (2013).

641
642 44. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees.
643 *Bioinformatics* **17**, 754-755 (2001).

644
645 45. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science* **252**,
646 1162-1164 (1991).

FIGURES

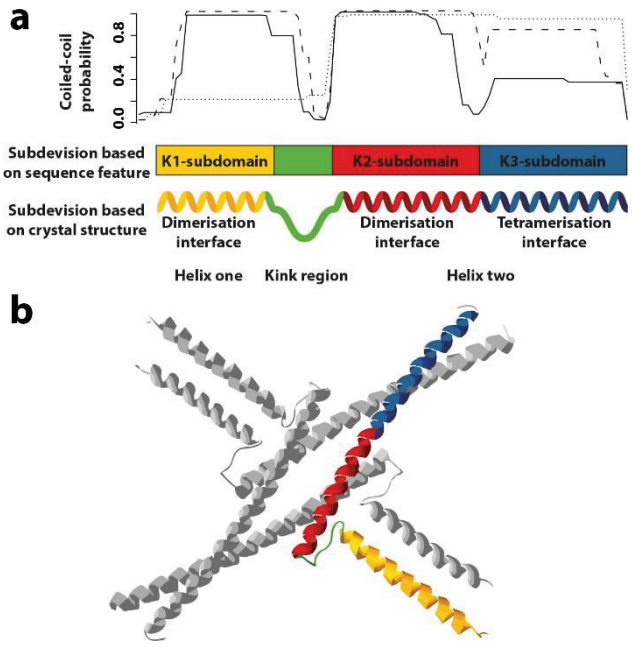


Figure 1. Domain architecture of the K-domain of SEP3 based on sequence and structural features. (**a**) Based on coiled-coil predictions (top) the K-domain was assumed to fold into three separate coiled-coils and was thus subdivided into three subdomains K1-, K2- and K3-subdomain (middle). The crystal structure of the K-domain of SEP3 revealed that it folds into two α-helices separated by a kink region (bottom). The first helix spans the K1-subdomain (color coded in yellow) and is involved in the dimerisation of two SEP3 monomers (i.e. dimerisation interface). The second helix spans the K2- and K3-subdomains and constitutes an N-terminal interaction interface that further stabilizes dimerisation of two SEP3 monomers (red) and a second C-terminal interaction interface that mediates the interaction of two SEP3 dimers (i.e. tetramerisation interface, blue). Coiled-coil predictions were performed with COILS [45]. The solid, dashed and dotted lines correspond to a sliding window size of 14, 21 and 28 amino acids used for the prediction, respectively. (**b**) Crystal structure of a SEP3 K-domain homotetramer [23]. The dimerisation interface of helix one, the kink region, the dimerisation interface of helix two and the tetramerisation interface of one K-domain are color coded in yellow, green, red and blue, respectively.

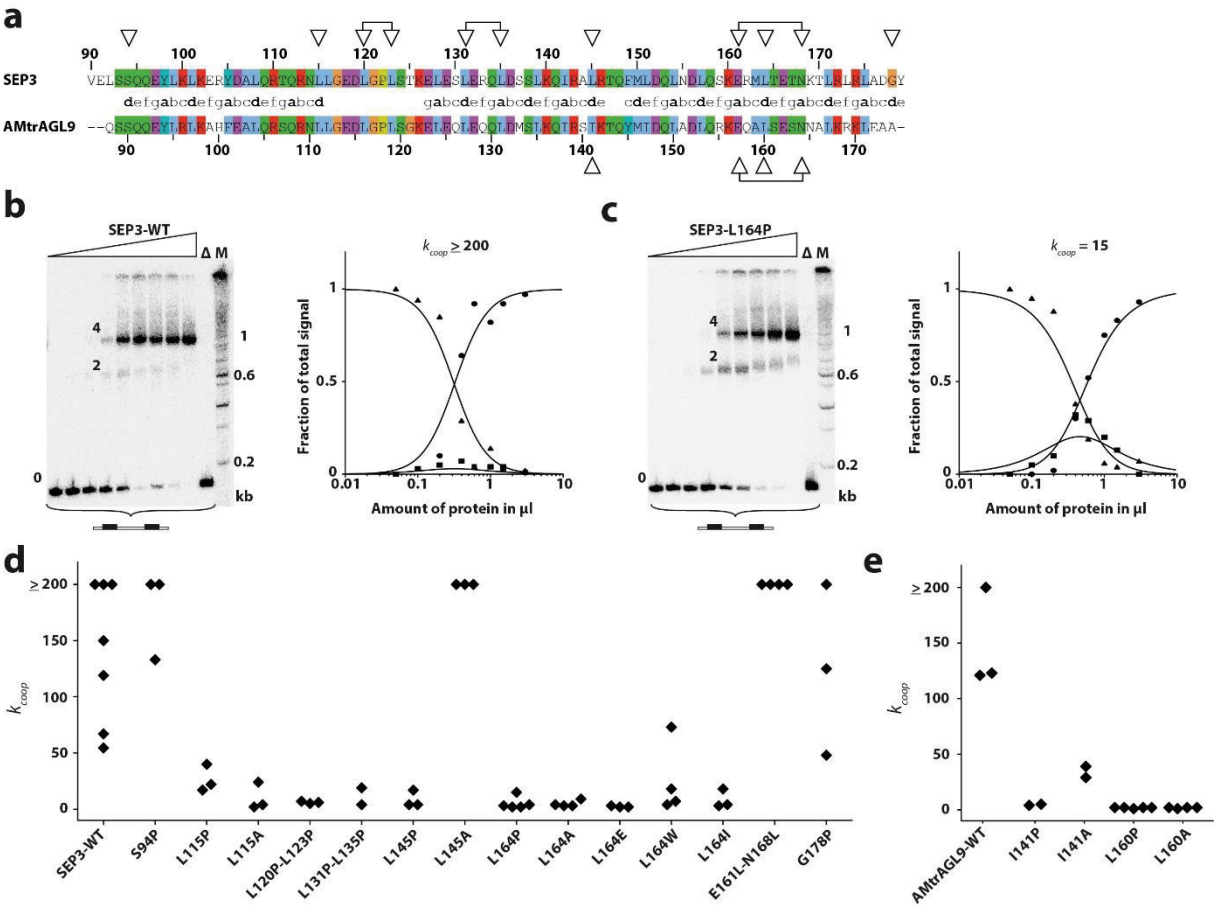


Figure 2. Ability of SEP3 and AMtrAGL9 wild type proteins and different amino acid substitution mutants to cooperatively bind to DNA. (**a**) Pairwise sequence alignment of the K-domains of SEP3 and AMtrAGL9. The heptad repeat pattern is depicted in the center. Positions at which amino acids were substituted are indicated by open triangles. (**b** and **c**) Binding of SEP3 wild type (**b**) and SEP3-L164P (**c**) to a DNA probe containing two CArG-boxes. Increasing amounts of *in vitro* translated protein were incubated with constant amounts of DNA probe. As negative control the empty pTNT vector without any cDNA insert was used as template DNA for the *in vitro* translation (lane Δ). For size comparison a radioactively labeled DNA ladder (100 bp Ladder, NEB) was applied (lane M). The labeling of the three different fractions '0', '2' and '4' corresponds to the number of proteins bound to one DNA molecule. Quantified signal intensities of the different fractions and graphs, fitted according to equation (1) to (3) described in Methods, are shown next to the gel pictures (▲ free DNA; ■ DNA probe bound by two proteins; ● DNA probe bound by four proteins). The k_{coop} value inferred from this particular measurement is depicted above the diagram. (**d** and **e**) k_{coop} values for the wild type protein and all examined single and double amino acid substitution mutants of SEP3 (**d**) and AMtrAGL9 (**e**). k_{coop} values above 200 could not be determined reliably (see Methods).

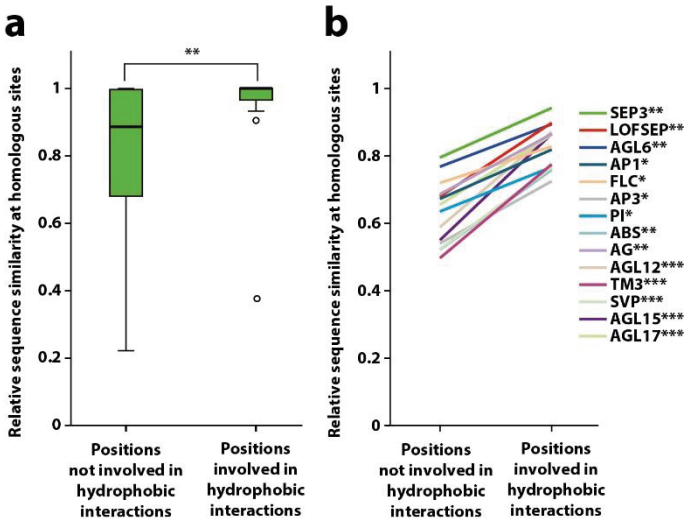


Figure 3. Sequence similarity analysis of SEP3-subfamily proteins and members of other MIKC-type MADS-domain protein subfamilies. (**a**) Box-plot showing relative sequence similarity at homologous sites of SEP3-subfamily proteins for positions that are involved in hydrophobic interactions within the SEP3 homotetramer and positions that are not involved in hydrophobic interactions. (**b**) Line graph showing the same analysis as in (**a**) but for all MIKC-type protein subfamilies. For all subfamilies amino acid positions that are homologous to sites involved in hydrophobic interactions are significantly less variable than positions that are homologous to non-interacting sites (Mann-Whitney-U-test; * $p = 0.01$-0.05; ** $p = 0.001$-0.01; *** $p < 0.001$).

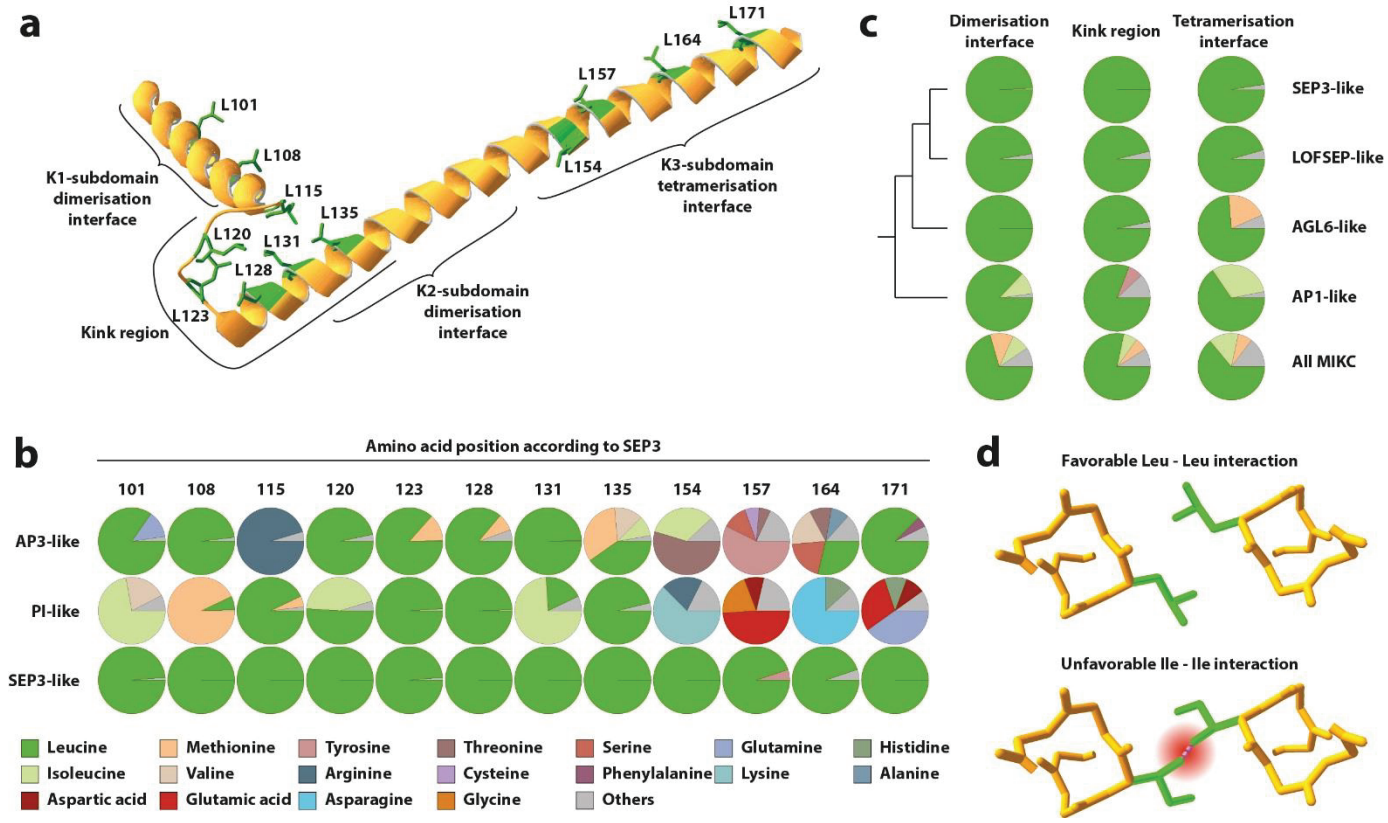


Figure 4. Amino acid preferences of SEP3-subfamily proteins and members of other MIKC-type MADS-domain protein subfamilies. (**a**) Picture of the crystal structure of a single K-domain of SEP3. Leucine side chains that are involved in inter- and intramolecular interactions are shown in green. (**b**) Amino acid frequencies at sites homologous to leucine residues that are involved in inter- and intramolecular interactions in the SEP3 homotetramer shown for SEP3-, AP3- and PI-subfamily proteins. Amino acids that occurred in less than 5 % of the examined subset of sequences were condensed as 'others'. The vast majority of the positions shown vertically are homologous to each other. The only exception are positions 154, 157, 164 and 171 of PI-like proteins. In this case, a gap was detected in the alignment but amino acids directly following the gap were included here. (**c**) Amino acid preferences at sites homologous to leucine residues that contribute to dimerisation interface (L101, L108), kink region (L115, L120, L123, L128, L131, L135) and tetramerisation interface (L154, L157, L164, L171) in the SEP3 homotetramer, shown for SEP3-, LOFSEP-, AGL6- and AP1-subfamily proteins and all MIKC-type proteins, respectively, following the color coding of panel B. (**d**) Part of the crystal structure of two interacting tetramerisation interfaces within a SEP3 homotetramer. The picture illustrates the favorable Leu-Leu interaction at heptad repeat 'd' positions as it becomes apparent several times within a SEP3 homotetramer (upper part). In contrast to the γ-branched leucine a β-branched amino acid such as isoleucine would potentially lead to steric hindrance at heptad repeat 'd' positions (lower part).

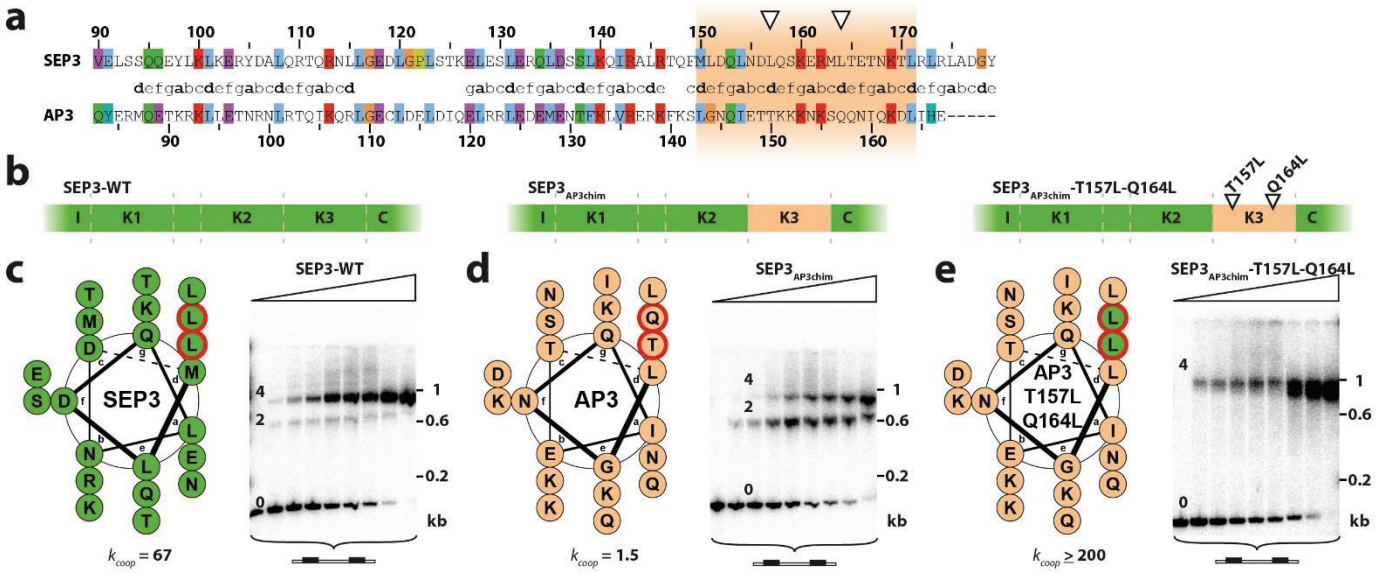


Figure 5. Design and cooperative DNA-binding capabilities of the chimeric protein SEP3$_{AP3chim}$. (**a**) Pairwise sequence alignment of the K-domains of SEP3 and AP3. The heptad repeat pattern is depicted in the center. Orange background marks the region that was substituted to create the chimeric protein SEP3$_{AP3chim}$. Open triangles mark the positions of the subsequently introduced amino acid substitutions. (**b**) Experimental setup to test for the ability of leucines to restore tetramerisation ability of SEP3$_{AP3chim}$. First the complete tetramerisation interface (i.e. the K3-subdomain) of SEP3 was substituted by the equivalent positions of AP3. Subsequently the two residues T157 and Q164 were substituted back to leucine to reestablish the hydrophobic stripe. (**c**-**e**, left) Helical wheel diagram of the tetramerisation interface of SEP3 wild type (**c**), SEP3$_{AP3chim}$ (**d**) and SEP3$_{AP3chim}$-T157L-Q164L (**e**), respectively, to illustrate the presumed position of amino acids 157 and 164 (framed in red) within the hydrophobic stripe of the K3-subdomain coiled-coil. (**c**-**e**, right) Binding of SEP3 wild type (**c**), SEP3$_{AP3chim}$ (**d**) and SEP3$_{AP3chim}$-T157L-Q164L (**e**) to a DNA probe containing two CArG-boxes. Increasing amounts of *in vitro* translated protein were incubated together with constant amounts of DNA probe. k_{coop} values inferred from this particular measurement are depicted below the helical wheel diagrams.

3 Discussion

Plant Type II MADS-TFs control a plethora of different developmental processes, most notably the flower development of angiosperms and are thus one of the most important and intensively studied transcription factor families of plants (Becker and Theißen, 2003; Kaufmann et al., 2005a; Smaczniak et al., 2012a). In contrast to MADS-TFs from animals and fungi, plant Type II MADS-TFs exhibit the highly conserved K-domain that mediates protein-protein interactions and enables members of the MIKC^C-type MADS-TF family to tetramerize among each other (Theißen et al., 1996; Kaufmann et al., 2005a; Melzer et al., 2009). The protein-protein interaction capabilities of plant Type II MADS-TFs are thus of utmost importance to understand the combinatorial action and evolution of this class of transcription factors. In recent years an increasing number of studies focused on the protein-protein interactions of floral homeotic MIKC^C-type proteins and the structure of the K-domain (Immink et al., 2009; Melzer and Theißen, 2009; Melzer et al., 2009; van Dijk et al., 2010; Espinosa-Soto et al., 2014; Puranik et al., 2014; Silva et al., 2016; Zhang et al., 2016; Al Hindi et al., 2017; Ruelens et al., 2017). The insights gained through this thesis together with recent findings of the scientific community allows us to draw a clearer picture of the evolutionary trajectories that shaped the present day structure of the PPI network controlling flower development and to narrow down which sequence determinants of the K-domain account for different interaction capabilities of floral homeotic proteins.

3.1 Evolution of the PPI network controlling flower development of angiosperms

In *A. thaliana* the PPI network of floral homeotic proteins shows a characteristic scale-free structure where most proteins display a highly limited set of interaction partners and only few members possess promiscuous interactions, most notably SEP3 (de Folter et al., 2005; Immink et al., 2009; Al Hindi et al., 2017). PPI networks of floral homeotic proteins have also been investigated for several other core eudicot species such as *Antirrhinum majus* (Causier et al., 2003), *Petunia x hybrida* (Immink et al., 2003), *Solanum lycopersicum* (Leseberg et al., 2008; Al Hindi et al., 2017) and *Gerbera hybrida* (Ruokolainen et al., 2010) as well as for the more distantly related early diverging eudicot species *Euptelea pleiospermum*, *Akebia trifoliata* and *Pachysandra terminalis* (Liu et al., 2010) and for the monocot *Oryza sativa* (Cooper et al., 2003). Comparison of the PPI networks reveals that the protein-protein interactions that are required for the formation of the different floral quartets are highly conserved throughout all investigated

networks pointing towards an essential role of these interactions for the development of the primary flower architecture (Liu et al., 2010). Beside the conserved interactions also several variable interactions were found that are presumed to account for deviant flower morphologies of the examined species (Liu et al., 2010). In addition and no less remarkable also the absence of numerous interactions appeared to be 'conserved'. For example, in none of the investigated eudicot and monocot species a direct interaction of AP3-like proteins with AP1-, AG- or STK-like proteins was detected (Liu et al., 2010; Ruokolainen et al., 2010). Furthermore, also no direct interaction between AG- and STK-like proteins was found (Liu et al., 2010). The absence of certain interaction patterns suggests that these interactions probably bring about detrimental changes in gene regulation that may cause severe malfunctions during floral organ identity determination.

In our study on the evolution of the obligate heterodimerization of AP3 and PI my colleagues and I investigated protein-protein interaction capabilities of floral homeotic proteins from the early diverging angiosperm species *Amborella trichopoda*, *Nuphar advena* and from the magnoliid *Liriodendron tulipifera* (Manuscript I: Melzer et al., 2014). We could show that even floral homeotic proteins of such distantly related species possess all protein-protein interactions that are necessary for the formation of the different floral quartets (Manuscript I: Melzer et al., 2014). However, beside these highly conserved interaction patterns we observed additional interactions among floral homeotic proteins of early diverging angiosperms that are not found among their orthologs from monocots and eudicots (Manuscript I: Melzer et al., 2014). For example AP3-like proteins from *A. trichopoda* and *N. advena* do not only interact with PI-like proteins but also with AG-like and AGL6-like proteins. This finding is consistent with other studies on the interactions of floral homeotic proteins from *A. trichopoda* and *N. pumila* that detected direct interactions of AP3- and PI-like proteins with AP1-, AGL6- and STK-like proteins, respectively (Amborella Genome Project, 2013; Li et al., 2015). The more promiscuous interactions of floral homeotic proteins from early diverging angiosperms suggest that ancestral precursors of floral homeotic proteins being present at the base of angiosperm evolution may also possessed a wide interaction spectrum, pointing towards a shift from promiscuity to specificity in the PPI network.

To better comprehend changes within the PPI network of floral homeotic proteins during angiosperm evolution two different approaches could be applied. The method of ancestral

character state reconstruction uses experimentally determined protein-protein interaction data of proteins from extant species, maps the interaction data onto the phylogeny of the examined proteins and eventually reconstructs the ancestral states at internal nodes (Manuscript II: Rümpler et al., 2015a). By contrast the technique of ancestral sequence reconstruction calculates the most likely amino acid sequence of ancestral proteins at certain time points during evolution. The reconstructed amino acid sequences can be used to synthesize the encoding nucleotide sequences and to subsequently examine the protein-protein interaction behavior of the corresponding proteins experimentally (Gumulya and Gillam, 2017). In recent years both approaches were applied to reconstruct the state of the PPI network of floral homeotic proteins in the MRCA of extant angiosperms (Li et al., 2015; Ruelens et al., 2017). The results consistently substantiate the assumption that ancestral floral homeotic proteins indeed possessed more promiscuous interaction capabilities compared to their orthologs of extant eudicots and monocots (Li et al., 2015; Ruelens et al., 2017).

In the context of this thesis protein-protein interaction data of floral homeotic proteins from phylogenetically informative angiosperms were examined to infer how the PPI network controlling flower development changed during early angiosperm evolution. However, to conceive the evolutionary trajectories that shaped the structure of the PPI network at the base of angiosperm evolution it is necessary to also consider the interaction capabilities of orthologs of floral homeotic proteins from gymnosperms, angiosperms closest extant relatives. The seven subfamilies of floral homeotic genes that are present in angiosperms (*AP1*-, *AP3*-, *PI*-, *AG*-, *STK*-, *SEP1*- and *SEP3*-like genes) most likely trace back to four ancestral gene families (ancestral *AP1/FLC*-, ancestral *AP3/PI*-, ancestral *AG/STK*- and ancestral *SEP*-like genes) that were present in the MRCA of extant seed plants i.e. angiosperms and gymnosperms (Gramzow and Theißen, 2010; Ruelens et al., 2013; Gramzow et al., 2014; Chen et al., 2017). In angiosperms ancestral *AP1/FLC*-like genes diverged into *AP1*- and *FLOWERING LOCUS C* (*FLC*)-like genes of which latter function as central repressors of flowering in *A. thaliana.* Ancestral *AP3/PI*-like genes diverged into *AP3*- and *PI*-like genes, a duplication of an ancestral *AG/STK*-like gene gave rise to *AG*- and *STK*-like genes and ancestral *SEP*-like genes diverged into *SEP1*- and *SEP3*-like genes (Ruelens et al., 2013; Gramzow et al., 2014; Chen et al., 2017). Phylogeny reconstructions of orthologs of floral homeotic genes from the gymnosperm *Gnetum gnemon* suggest that *GGM2* and *GGM3* constitute direct descendants of an ancestral *AP3/PI*-like

and an ancestral *AG/STK*-like gene, respectively (Becker et al., 2003; Wang et al., 2010; Gramzow et al., 2014). Two further genes of *G. gnemon GGM9* and *GGM11* were found to be phylogenetically closely related to *AP1*- and *SEP*-like genes of angiosperms, although their exact phylogenetic relationship is still discussed controversially in the literature (Kim et al., 2013; Ruelens et al., 2013; Gramzow et al., 2014). Studies on the protein-protein interactions of the encoded proteins revealed direct interaction of GGM2 with GGM3, GGM9 and GGM11 as well as direct interaction of GGM3 with GGM9 and GGM11 (Winter et al., 2002b; Wang et al., 2010). Thus similar to the direct interactions of AP3-, PI-, AG- and STK-like proteins observed in early diverging angiosperms (Amborella Genome Project, 2013; Melzer et al., 2014; Li et al., 2015) also the AP3/PI- and AG/STK-like proteins of the gymnosperm *G. gnemon* are capable to directly interact. Thus it appears most likely that also the ancestral AP1/FLC-, AP3/PI-, AG/STK- and SEP-like proteins being present at the base of seed plant evolution possessed promiscuous interaction capabilities. The angiosperm specific duplications of floral homeotic genes increased the number of floral homeotic proteins which probably still possessed promiscuous interaction capabilities resulting in a complex PPI network with highly connected nodes (Li et al., 2015; Ruelens et al., 2017). During early angiosperm evolution numerous protein-protein interactions were lost leading to a shift from promiscuous to more specific interactions within the PPI network of floral homeotic proteins (Ruelens et al., 2017). Remarkably the loss of interactions thereby was not random but rather predominantly involved interactions of AP3- and PI-like proteins and to a certain extent also AG- and STK-like proteins (Li et al., 2015; Ruelens et al., 2017). In contrast SEP-like proteins retained their promiscuous interaction behavior eventually leading towards the scale-free network structure found in extant eudicots and monocots with SEP-like proteins serving as hubs that mediate interaction of most other floral homeotic proteins (Ruelens et al., 2017).

3.2 Origin and evolution of floral quartet-like complex formation

Most studies on protein-protein interactions of floral homeotic proteins are based on yeast-two-hybrid (Y2H) and Y3H experiments. Therefore relatively little is known about the capabilities of different floral homeotic proteins to form DNA-bound homo- and heterotetramers, i.e. the formation of floral quartet-like complexes (FQCs). For floral homeotic proteins of *A. thaliana* it has been shown that the B-class proteins AP3 and PI bind as heterodimer to a single CArG-box but they are unable to tetramerize among each other (Melzer and Theißen, 2009). Similarly also

AP1 and AG, respectively can bind as homodimers to a single DNA-binding site but they are incapable of forming DNA-bound homotetramers nor heterotetramers with one another (Smaczniak et al., 2012b). However, in combination with SEP3 AP3 and PI as well as AP1 and AG can be incorporated into FQCs (Melzer and Theißen, 2009; Smaczniak et al., 2012b). In contrast to this SEP3 as well as its closely related paralogs SEP1, SEP2 and SEP4 are able to form DNA-bound homotetramers with high affinity (Melzer et al., 2009; Jetha et al., 2014; Manuscript V: Rümpler et al., 2017). Thus among floral homeotic proteins of *A. thaliana* the ability to form DNA-bound homotetramers seems to be restricted to SEP-like proteins.

In this thesis I investigated the homotetramerization abilities of the distantly related SEP3 ortholog AMtrAGL9 from *A. trichopoda* and I could show that it is capable of forming DNA-bound homotetramers with an affinity similar to that of SEP3 (Manuscript V: Rümpler et al., 2017). Furthermore I could demonstrate that the amino acid residues that are essential for tetramerization of SEP3 are highly conserved among SEP-like proteins (Manuscript V: Rümpler et al., 2017). These findings strongly suggest that the ability of SEP-like proteins to form homotetramers and thus probably also their ability to incorporate other proteins into tetrameric complexes is highly conserved throughout angiosperms. Surprisingly, studies on homo- and heterotetramerization capabilities of AP3-, PI-, and AG-like proteins from *A. trichopoda* and *N. advena* suggest that the ability to form homotetramers is more widespread among floral homeotic proteins from early diverging angiosperm species (Härter, 2011). In a very recent study Ruelens et al. (2017) reconstructed and synthesized the ancestral floral homeotic proteins of the MRCA of extant angiosperms and tested their abilities to form DNA-bound homo- and heterotetramers. Their findings indicate that in contrast to AP3-, PI- and AG-like proteins from *A. thaliana* ancestral AP3/PI- and AG/STK-like proteins do not require ancestral SEP-like proteins to form FQCs. Instead the ancestral AP3/PI- and AG/STK-like proteins were shown to form heterotetramers among themselves and probably even homotetramers (Ruelens et al., 2017). This finding is consistent with previous studies on FQC formation capabilities of floral homeotic proteins from gymnosperms. GGM2 (AP3/PI-like) and GGM3 (AG/STK-like) from *G. gnemon* have been shown to bind to DNA as heterotetramer and GGM3 is even capable to form DNA-bound homotetramers (Wang et al., 2010). Thus also AP3/PI- and AG/STK-like proteins from gymnosperms do not require SEP-like proteins to form FQCs. Taken together these findings suggest that the ability to form DNA-bound homo- and heterotetramers is an ancestral feature of

floral homeotic proteins. During early angiosperm evolution most subfamilies of floral homeotic proteins lost the ability to from DNA-bound homotetramers and probably also the ability to incorporate other floral homeotic proteins into FQCs, whereas this feature was retained among SEP-like proteins.

If FQC formation is the ancestral rather than the derived state of floral homeotic proteins the question still remains as to when during plant evolution FQCs first occurred. Due to the importance of the K-domain for mediating the protein-protein interactions that are necessary for FQC formation it appears very likely that the ability to form DNA-bound tetramers is restricted to MIKC-type MADS-TFs (Manuscript IV: Theißen et al., 2016). All so far described FQCs are exclusively composed of MIKC^{C}-type MADS-TFs (Melzer and Theißen, 2009; Wang et al., 2010; Smaczniak et al., 2012b; Jetha et al., 2014; Ruelens et al., 2017; Manuscript V: Rümpler et al., 2017). However, as no MIKC*-type proteins have yet been tested for their ability to cooperatively bind to DNA it remains unresolved whether the presence of the K-domain per se or rather specific amino acid features within the K-domain of MIKC^{C}-type proteins facilitate tetramerization. MIKC^{C}- and MIKC*-type genes differ in length and number of the exons that encode for the K-domain (Henschel et al., 2002; Kwantes et al., 2012). This suggests that even though MIKC*-type proteins may form FQCs the structural basis upon which FQC formation of MIKC*-type proteins takes place is probably different from that facilitating tetramerization of MIKC^{C}-type proteins. This hypothesis is consistent with the observation that MIKC^{C}- and MIKC*-type MADS-TFs almost exclusively interact among each other but not with members of the other clade (de Folter et al., 2005; Immink et al., 2009). Clearly, studies on tetramerization capabilities of MIKC-type proteins from lycophytes, bryophytes and even from charophytes are needed to further narrow down the origin of FQCs and to clarify whether it was directly linked to the emergence of the K-domain or if it evolved later during land plant evolution.

3.3 Sequence features determining protein-protein interactions of floral homeotic proteins

The studies discussed above demonstrate that the structure of the PPI network as well as the homotetramerization capabilities of floral homeotic proteins changed during angiosperm evolution. Thus, the question emerges as to which changes at the molecular level account for subfamily specific interaction patterns and the evolutionary changes of the PPI network. Based on substitutions of amino acids within the K-domain of SEP3 I could demonstrate that smallest

changes in the amino acid composition (i.e. residue substitutions at a single interacting site) can have severe consequences on the overall interaction behavior of the respective protein (Manuscript V: Rümpler et al., 2017). The experimental investigations of this thesis mainly focused on the importance of hydrophobic residues for FQC formation. My findings demonstrate that highly conserved leucine residues at intra- and intermolecular interaction sites are indispensable for tetramerization of SEP3 and it is presumed that the absence of these leucines at least partially accounts for the less promiscuous interactions of non-hub proteins such as AP3 and PI (Manuscript V: Rümpler et al., 2017). Analyses on relative sequence similarity patterns within the K-domain have shown that throughout all subfamilies of MIKC^C-type MADS-TFs amino acid positions that are homologous to interacting sites of SEP3 are less variable than amino acid positions that constitute no direct interaction partners within a SEP3 homotetramer (Manuscript V: Rümpler et al., 2017). This finding strongly suggest that the structure of the K-domain as determined for SEP3 is highly conserved throughout all subfamilies of MIKC^C-type proteins. Furthermore it indicates that amino acid positions that are homologous to interacting sites within a SEP3 homotetramer also constitute direct contact points within the K-domains of other MIKC^C-type proteins. These insights together with the profound knowledge about favorable and unfavorable amino acid pairings of coiled-coil proteins (Mason and Arndt, 2004; Mason et al., 2009) can be used to reconsider previous findings on sequence determinants mediating protein-protein interactions of certain floral homeotic proteins. For example this knowledge could be applied to understand the molecular reasons for the obligate heterodimerization of AP3 and PI.

Almost 15 years ago Yang et al. (2003b) screened for single amino acid substitution mutants of AP3 and PI that were unable to heterodimerize with one another. Based on their findings they postulated a favorable electrostatic interaction between the positively charged R102 of AP3 and the negatively charged E97 of PI (Yang et al., 2003b). Based on the assumption that the K-domain of AP3 and PI folds into a structure similar to that determined for SEP3 (Manuscript V: Rümpler et al., 2017) the residues identified by Yang and colleagues indeed constitute direct interaction partners on opposing heptad repeat ‘e’ and ‘g’ positions in the center of helix one (K1-subdomain). As already suggested by the authors this favorable electrostatic interaction would not occur in a hypothetical AP3 homodimer nor in a PI homodimer because one of both charged residues would be missing in either case (Yang et al., 2003b). My analyses on subfamily specific amino acid distributions within the K-domain of MIKC^C-type proteins (Manuscript V:

Rümpler et al., 2017) reveal that both charged residues are exceedingly conserved among AP3- and PI-like proteins, respectively. In addition also the absence of charged residues at amino acid position 97 of AP3-like proteins and at position 102 of PI-like proteins is highly conserved (Manuscript V: Rümpler et al., 2017). This suggests that attractive electrostatic interactions between the first K-domain helices of AP3- and PI-like proteins facilitate heterodimerization and that the absence of this favorable interaction prevents homodimerization of AP3- and PI-like proteins, respectively. Interestingly, the gymnosperm AP3/PI-like protein GGM2 from *G. gnemon* that is capable of forming homodimers (Wang et al., 2010) possesses E97 as well as R102 suggesting that a favorable electrostatic interaction could take place and stabilize a GGM2 homodimer (Yang et al., 2003b).

Beside the single amino acid substitution mutants investigated by Yang et al. (2003b) also studies on the mutant *PI* allele *pi-5* (Yang et al., 2003a) help to understand the molecular reasons for the obligate heterodimerization of AP3- and PI-like proteins. The *pi-5* allele possesses a missense mutation in the K-box that causes a single amino acid substitution from the negatively charged glutamate to the positively charged lysine at amino acid position 125 of the encoded protein (Yang et al., 2003a). The resulting mutant PI^{E125K} is defective in dimerization with AP3 resulting in a partial loss of B-function in *pi-5* mutant plants. When Yang et al. (2003a) screened for compensatory AP3 mutant proteins they identified a single amino acid substitution mutant possessing a charge swap from lysine to glutamate at amino acid position 139 ($AP3^{K139E}$). Although amino acid position 125 of PI and 139 of AP3 appear to be relatively far away from each other, the compensatory charge substitutions suggest that both positions directly interact within an AP3-PI heterodimer. Intriguingly, in two interacting K-domains of SEP3 amino acid positions homologous to E125 of PI and K139 of AP3, respectively, indeed constitute direct interaction partners that are both located within the N-terminal half of the second K-domain helix (K2-subdomain). Both charged residues are highly conserved among AP3- and PI-like proteins, respectively, suggesting that also within the second K-domain helix a favorable electrostatic interaction stabilizes the AP3-PI heterodimer. The favorable interaction is not expected to occur in a hypothetical homodimer of PI as the positively charged residue at amino acid position 139 is absent among PI-like proteins. In accordance to the example discussed before, the gymnosperm AP3/PI-like protein GGM2 contains both E125 and K139 suggesting

that the favorable interaction between both charged residues could take place to stabilize the GGM2 homodimer.

Taken together these observations strongly support the assumption that the K-domains of AP3 and PI fold into a structure similar to that determined for SEP3. Furthermore it presumes that the heterodimerization of AP3 and PI is at least partially mediated by attractive electrostatic interactions and that the absence of these stabilizing interactions prevents homodimerization of AP3 and PI, respectively. In our study on the evolution of the obligate heterodimerization of AP3 and PI my colleagues and I demonstrated that the ability of AP3- and PI-like proteins to form homodimers was most likely present at the base of extant angiosperms and subsequently got lost during early angiosperm evolution (Manuscript I: Melzer et al., 2014). With these findings my colleagues and I supported the previous hypothesis that AP3- and PI-like proteins evolved from an ancestral AP3/PI-like protein that was able to form homodimers (Winter et al., 2002b; Lenser et al., 2009). The observations discussed above give a possible explanation as to which alterations at the molecular level account for the evolutionary changes in the protein-protein interactions of AP3- and PI-like proteins. It suggests that during early angiosperm evolution AP3- and PI-like proteins independently lost charged residues at different amino acid positions within the K-domain that are critical for intermolecular electrostatic interactions. As a result the favorable electrostatic interactions no longer take place in homodimers of AP3- and PI-like proteins, respectively, but still occur in a heterodimer eventually leading to an obligate heterodimerization of both proteins. In a very recent study on the evolution of the protein-protein interactions of AP3- and PI-like proteins in monocots Bartlett et al. (2016) demonstrated that homodimerization of PI-like proteins was regained several times independently within Poales. Interestingly they could show that a single amino acid change within the I-domain from the uncharged glycine to the negatively charged aspartic acid is responsible for a switch from obligate heterodimerization to homodimerization (Bartlett et al., 2016). This suggests that in addition to the K1- and K2-subdomains also the I-domain mediates attractive or repulsive electrostatic interactions that determine the interaction capabilities of AP3- and PI-like proteins.

The discussed amino acids within the K1- and K2-subdomains all occupy heptad repeat ‘e’ and ‘g’ positions, respectively. It is well known that whereas coiled-coil interaction stability is mainly achieved by hydrophobic amino acids at heptad repeat ‘a’ and ‘d’ positions, pairing specificity is

greatly influenced by charged amino acids at ‘e’ and ‘g’ positions (Mason and Arndt, 2004; Mason et al., 2009). My investigations on amino acid distributions within the K-domain of MIKC^C-type proteins belonging to different subfamilies indicate that also beyond AP3- and PI-like proteins subfamily specific charge distributions at heptad repeat ‘e’ and ‘g’ positions occur (Manuscript V: Rümpler et al., 2017). Thus it appears very likely that the different protein-protein interaction capabilities of MIKC^C-type MADS-TFs are to some extend mediated by the subfamily specific presence or absence of charged residues at certain ‘e’ and ‘g’ positions throughout the K-domain.

3.4 Implications of the K-domain mimicry of SAP54

The highly specific protein-protein interactions that are mediated by the K-domain of MIKC^C-type MADS-TFs are not only essential for dimerization and tetramerization of floral homeotic proteins. The interaction capabilities are also exploited by an effector protein of phytopathogenic bacteria. In this thesis I present preliminary evidences that the phytoplasma effector protein SAP54 folds into a structure similar to that of the K-domain of MIKC^C-type MADS-TFs (Manuscript III: Rümpler et al., 2015b). My colleagues and I hypothesize that SAP54 mimics the K-domain structure as a result of convergent sequence and structural evolution to specifically target certain MIKC^C-type MADS-TFs (Manuscript III: Rümpler et al., 2015b). As a consequence of the molecular mimicry the sequence determinants that enable the interaction between SAP54 and its targets are probably very similar to that mediating dimerization and tetramerization among MIKC^C-type MADS-TFs.

SAP54 has been shown to preferentially target MIKC^C-type proteins of certain subfamilies (MacLean et al., 2014). Among the 15 MIKC^C-type MADS-TFs of *A. thaliana* that are bound by SAP54, 12 belong to the subfamilies of AP1-, FLC-, AGL6- and SEP-like proteins, respectively (MacLean et al., 2014; Maejima et al., 2014b; Maejima et al., 2015). These subfamilies are phylogenetically closely related and most likely trace back to one ancestral gene that existed prior to the split of angiosperms and gymnosperms (Kim et al., 2013; Ruelens et al., 2013; Gramzow et al., 2014). In a very recent study Kitazawa et al. (2017) show that the SAP54 ortholog PHYL1 is able to bind to AP1- and SEP-like MADS-TFs of the eudicots *P. hybrida* and *Chrysanthemum morifolium* and the monocots *O. sativa* and *Lilium longiflorum*. The specific binding of PHYL1 to AP1- and SEP-like proteins of such distantly related angiosperm species

suggests that the effector protein probably exploits subfamily specific protein-protein interaction capabilities of MIKC^{C}-type MADS-TFs to specifically target members of the AP1/FLC/AGL6/SEP-superfamily.

In our opinion article on the convergent evolution of SAP54 (Manuscript III: Rümpler et al., 2015b) my colleagues and I hypothesized that the specific interactions of SAP54 could probably serve as a molecular tool to study the function of floral homeotic proteins in genetically intractable species. It is currently discussed controversially whether the incorporation of SEP-like proteins into FQCs was an important step in the evolution of the angiosperm flower or if AGL6-like proteins of gymnosperms may fulfill similar functions (Manuscript IV: Theißen et al., 2016; Ruelens et al., 2017). Thus phytoplasma infections could be used to create phenotypes that resemble knockouts of SEP- and AGL6-like genes in early diverging angiosperms and gymnosperms, respectively, to reveal the function of these genes for reproductive development. Remarkably, Kitazawa et al. (2017) could show that indeed the two gymnosperm AGL6-like proteins DAL1 from *Picea abies* and CjMADS14 from *Cryptomeria japonica* are bound by PHYL1.

4 Summary

4.1 Summary

The flower development of angiosperms is controlled by floral homeotic MIKC^C-type MADS-domain transcription factors (MADS-TFs) that activate or repress target genes by forming floral organ specific DNA-bound heterotetrameric complexes termed floral quartets. The ability to form floral quartets highly differs between floral homeotic MADS-TFs of certain subfamilies. However, to date relatively little is known about how these subfamily-specific interaction patterns of floral homeotic proteins evolved during angiosperm evolution and which sequence determinants account for the different interaction capabilities.

Based on interaction studies of floral homeotic proteins from early diverging angiosperms I could show that the interactions governing flower development in core eudicots are also present in these distantly related species. However, especially AP3- and PI-like proteins from early diverging angiosperms possess additional interactions compared to their orthologs from core eudicots which form obligate heterodimers only. The more diverse interactions among floral homeotic proteins from early diverging angiosperms suggest a shift from promiscuity to specificity in the protein-protein interaction network during early angiosperm evolution.

By comprehensive amino acid sequence analyses of MADS-TFs I demonstrated that the structure of the protein-protein interacting keratin-like domain (K-domain) is most likely highly similar among all subfamilies of floral homeotic proteins. Amino acid substitutions within the K-domain of the floral homeotic hub protein SEP3 revealed that highly conserved leucine residues at interacting sites are essential mediators of floral quartet-like complex formation. The absence of leucine residues at homologous amino acid positions in non-hubs such as AP3- and PI-like proteins probably accounts for their less promiscuous interactions.

Beside the highly specific protein-protein interactions among floral homeotic proteins I studied another K-domain mediated interaction. The phytoplasma effector protein SAP54 targets the K-domain to specifically bind MADS-TFs of certain subfamilies and destines them for degradation. By performing amino acid sequence analyses and structural predictions I provided preliminary evidence that SAP54 folds into a structure similar to that of the K-domain. Based on my findings I hypothesized that SAP54 evolved via convergent molecular evolution to mimic the K-domain of its MADS-TF targets.

4.2 Zusammenfassung

Die Blütenentwicklung von Angiospermen wird von floral homöotischen MIKC^C-Typ MADS-Domänen Transkriptionsfaktoren (MADS-TFs) kontrolliert. Diese aktivieren oder reprimieren Zielgene indem sie blütenorganspezifische DNA-gebundene Heterotetramere, sogenannte ‚florale Quartette', ausbilden. Die Fähigkeit zur Ausbildung floraler Quartette unterscheidet sich dabei stark zwischen floral homöotischen MADS-TFs verschiedener Subfamilien. Bisher ist allerdings nur sehr wenig darüber bekannt, wie diese subfamilien-spezifischen Interaktionsmuster im Laufe der Angiosperm-Evolution entstanden sind und durch welche Sequenzdeterminanten die verschiedenen Interaktionsfähigkeiten bestimmt werden.

Ausgehend von Interaktionsstudien floral homöotischer Proteine aus basalen Angiospermen, konnte ich zeigen, dass die Interaktionen, die die Blütenentwicklung von Kern-Eudikotylen steuern, auch in diesen entfernt verwandten Spezies zu finden sind. Darüber hinaus zeigten vor allem AP3- und PI-ähnliche Proteine aus basalen Angiospermen zusätzliche Interaktionen; im Gegensatz zu ihren Orthologen aus Kern-Eudikotylen die lediglich obligatorische Heterodimere ausbilden. Die vielfältigeren Interaktionen zwischen floral homöotischen Proteinen aus basalen Angiospermen weisen auf eine Zunahme der Spezifität innerhalb des Protein-Protein-Interaktionsnetzwerks während der frühen Angiosperm-Evolution hin.

Mit Hilfe von Aminosäuresequenzanalysen von MADS-TFs konnte ich zeigen, dass die Struktur der Keratin-ähnlichen Protein-Protein-Interaktionsdomäne (K-Domäne) subfamilienübergreifend höchstwahrscheinlich sehr ähnlich ist. Aminosäuresubstitutionen innerhalb der K-Domäne des floral homöotischen ‚Hub'-Proteins SEP3 zeigten, dass hoch konservierte Leucin-Seitengruppen an Interaktionspunkten essentiell für die Ausbildung floraler Quartette sind. Fehlende Leucin-Seitengruppen auf homologen Aminosäurepositionen von ‚Non-Hub'-Proteinen, wie AP3 und PI, sind möglicherweise für deren weniger vielfältige Interaktionen verantwortlich.

Neben hochspezifischen Protein-Protein-Interaktionen zwischen floral homöotischen Proteinen habe ich eine weitere Interaktion der K-Domäne studiert. Das Phytoplasma Effektorprotein SAP54 interagiert mit der K-Domäne, um gezielt MADS-TFs bestimmter Subfamilien zu binden und anschließend deren Abbau zu vermitteln. Aminosäuresequenzanalysen und Struktur-vorhersagen von SAP54 weisen darauf hin, dass die Struktur von SAP54 der Struktur der K-Domäne sehr ähnlich ist. Ausgehend davon habe ich die Hypothese aufgestellt, dass sich SAP54 über konvergente molekulare Evolution entwickelt hat, um die K-Domäne nachzuahmen.

5 Bibliography

Al Hindi, T., Zhang, Z., Ruelens, P., Coenen, H., Degroote, H., Iraci, N., and Geuten, K. (2017). Protein interaction evolution from promiscuity to specificity with reduced flexibility in an increasingly complex network. Sci Rep **7,** 44948.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol **215,** 403-410.

Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., Ribas de Pouplana, L., Martinez-Castilla, L., and Yanofsky, M.F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc Natl Acad Sci U S A **97,** 5328-5333.

Alvarez-Buylla, E.R., Ambrose, B.A., Flores-Sandoval, E., Englund, M., Garay-Arroyo, A., Garcia-Ponce, B., de la Torre-Barcena, E., Espinosa-Matias, S., Martinez, E., Pineyro-Nelson, A., Engstrom, P., and Meyerowitz, E.M. (2010). B-function expression in the flower center underlies the homeotic phenotype of Lacandonia schismatica (Triuridaceae). The Plant Cell Online **22,** 3543-3559.

Amborella Genome Project. (2013). The Amborella genome and the evolution of flowering plants. Science **342,** 1241089.

Angenent, G.C., and Colombo, L. (1996). Molecular control of ovule development. Trends in Plant Science **1,** 228-232.

Aoki, S., Uehara, K., Imafuku, M., Hasebe, M., and Ito, M. (2004). Phylogeny and divergence of basal angiosperms inferred from APETALA3- and PISTILLATA-like MADS-box genes. J Plant Res **117,** 229-244.

APGIII. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APGIII. Bot. J. Linn. Soc **161,** 105-121.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., and Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. Nucleic Acids Research **40,** W597-W603.

Azuma, Y., Kukenshoner, T., Ma, G.Y., Yasunaga, J., Imanishi, M., Tanaka, G., Nakase, I., Maruno, T., Kobayashi, Y., Arndt, K.M., Matsuoka, M., and Futaki, S. (2014). Controlling leucine-zipper partner recognition in cells through modification of a-g interactions. Chem Commun **50,** 6364-6367.

Bai, X., Correa, V.R., Toruno, T.Y., Ammar el, D., Kamoun, S., and Hogenhout, S.A. (2009). AY-WB phytoplasma secretes a protein that targets plant cell nuclei. Molecular plant-microbe interactions : MPMI **22,** 18-30.

Bai, X.D., Zhang, J.H., Ewing, A., Miller, S.A., Radek, A.J., Shevchenko, D.V., Tsukerman, K., Walunas, T., Lapidus, A., Campbell, J.W., and Hogenhout, S.A. (2006). Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. Journal of Bacteriology **188,** 3682-3696.

Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: Understanding the cell's functional organization. Nature Reviews Genetics **5,** 101-U115.

Bartlett, M., Thompson, B., Brabazon, H., Del Gizzi, R., Zhang, T., and Whipple, C. (2016). Evolutionary dynamics of floral homeotic transcription factor protein-protein interactions. Molecular Biology and Evolution **33,** 1486-1501.

Bateson, W. (1894). Materials for the study of variation treated with especial regard to discontinuity in the origin of species. (Macmillan).

Battaglia, R., Brambilla, V., Colombo, L., Stuitje, A.R., and Kater, M.M. (2006). Functional analysis of MADS-box genes controlling ovule development in Arabidopsis using the ethanol-inducible alc gene-expression system. Mech Develop **123,** 267-276.

Baum, D.A., and Hileman, L.C. (2006). A developmental genetic model for the origin of the flower. In Flowering and its Manipulation, C. Ainsworth, ed (Sheffield, UK: Blackwell Publishing), pp. 3-27.

Becker, A., and Theißen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Phylogenet Evol **29,** 464-489.

Becker, A., Saedler, H., and Theissen, G. (2003). Distinct MADS-box gene expression patterns in the reproductive cones of the gymnosperm Gnetum gnemon. Development Genes and Evolution **213,** 567-572.

Becker, A., Kaufmann, K., Freialdenhoven, A., Vincent, C., Li, M.A., Saedler, H., and Theissen, G. (2002). A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. Molecular Genetics and Genomics **266,** 942-950.

Betz, S.F., Bryson, J.W., and Degrado, W.F. (1995). Native-like and structurally characterized designed alpha-helical bundles. Curr Opin Struc Biol **5,** 457-463.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1991). Genetic Interactions among Floral Homeotic Genes of Arabidopsis. Development **112,** 1-20.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (2012). The ABC model of flower development: then and now. Development **139,** 4095-4098.

Bowman, J.L., Alvarez, J., Weigel, D., Meyerowitz, E.M., and Smyth, D.R. (1993). Control of flower development in Arabidopsis thaliana by Apetala1 and interacting genes. Development **119,** 721-743.

Buzgo, M., Soltis, P.S., and Soltis, D.E. (2004). Floral developmental morphology of Amborella trichopoda (Amborellaceae). Int J Plant Sci **165,** 925-947.

Causier, B., Cook, H., and Davies, B. (2003). An antirrhinum ternary complex factor specifically interacts with C-function and SEPALLATA-like MADS-box factors. Plant Mol Biol **52,** 1051-1062.

Causier, B., Schwarz-Sommer, Z., and Davies, B. (2010). Floral organ identity: 20 years of ABCs. Seminars in Cell & Developmental Biology **21,** 73-79.

Chang, Y.Y., Kao, N.H., Li, J.Y., Hsu, W.H., Liang, Y.L., Wu, J.W., and Yang, C.H. (2010). Characterization of the possible roles for B class MADS box genes in regulation of perianth formation in orchid. Plant Physiol **152,** 837-853.

Chatr-aryamontri, A., Breitkreutz, B.J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., O'Donnell, L., Reguly, T., Breitkreutz, A., Sellam, A., Chen, D.C., Chang, C., Rust, J., Livstone, M., Oughtred, R., Dolinski, K., and Tyers, M. (2013). The BioGRID interaction database: 2013 update. Nucleic Acids Research **41,** D816-D823.

Chen, F., Zhang, X.T., Liu, X., and Zhang, L.S. (2017). Evolutionary analysis of MIKCc-type MADS-box genes in gymnosperms and angiosperms. Front Plant Sci **8,** 895.

Christensen, N.M., Axelsen, K.B., Nicolaisen, M., and Schulz, A. (2005). Phytoplasmas and their interactions with hosts. Trends in Plant Science **10,** 526-535.

Coen, E.S., and Meyerowitz, E.M. (1991). The war of the whorls: genetic interactions controlling flower development. Nature **353,** 31-37.

Cooper, B., Clarke, J.D., Budworth, P., Kreps, J., Hutchison, D., Park, S., Guimil, S., Dunn, M., Luginbuhl, P., Ellero, C., Goff, S.A., and Glazebrook, J. (2003). A network of rice genes associated with stress response and seed development. P Natl Acad Sci USA **100,** 4945-4950.

Davies, B., Egea-Cortines, M., de Andrade Silva, E., Saedler, H., and Sommer, H. (1996). Multiple interactions amongst floral homeotic MADS box proteins. Embo J **15,** 4330-4343.

De Bruyn, A., Martin, D.P., and Lefeuvre, P. (2014). Phylogenetic reconstruction methods: an overview. Methods Mol Biol **1115,** 257-277.

de Folter, S., and Angenent, G.C. (2006). trans meets cis in MADS science. Trends in Plant Science **11,** 224-231.

de Folter, S., Immink, R.G.H., Kieffer, M., Parenicova, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M., Davies, B., and Angenent, G.C. (2005). Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. Plant Cell **17,** 1424-1433.

Derelle, E., Ferraz, C., Rombauts, S., Rouze, P., Worden, A.Z., Robbens, S., Partensky, F., Degroeve, S., Echeynie, S., Cooke, R., Saeys, Y., Wuyts, J., Jabbari, K., Bowler, C., Panaud, O., Piegu, B., Ball, S.G., Ral, J.P., Bouget, F.Y., Piganeau, G., De Baets, B., Picard, A., Delseny, M., Demaille, J., Van de Peer, Y., and Moreau, H. (2006). Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. P Natl Acad Sci USA **103,** 11647-11652.

Deslandes, L., and Rivas, S. (2012). Catch me if you can: bacterial effectors and plant targets. Trends in Plant Science **17,** 644-655.

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., and Yanofsky, M.F. (2004). The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. Curr Biol **14,** 1935-1940.

Drea, S., Hileman, L.C., de Martino, G., and Irish, V.F. (2007). Functional analyses of genetic pathways controlling petal specification in poppy. Development **134,** 4157-4166.

Du Toit, A. (2014). BACTERIAL PATHOGENICITY Phytoplasma converts plants into zombies. Nat Rev Microbiol **12,** 393-393.

Eckardt, N.A. (2003). MADS monsters: Controlling floral organ identity. Plant Cell **15,** 803-805.

Eddy, S.R. (2009). A new generation of homology search tools based on probabilistic inference. Genome Inform **23,** 205-211.

Egea-Cortines, M., Saedler, H., and Sommer, H. (1999). Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in Antirrhinum majus. Embo J **18,** 5370-5379.

Elde, N.C., and Malik, H.S. (2009). The evolutionary conundrum of pathogen mimicry. Nat Rev Microbiol **7,** 787-797.

Espinosa-Soto, C., Immink, R.G.H., Angenent, G.C., Alvarez-Buylla, E.R., and de Folter, S. (2014). Tetramer formation in Arabidopsis MADS domain proteins: analysis of a protein-protein interaction network. Bmc Syst Biol **8,** 9.

Fan, H.Y., Ma, H., Hu, Y., and Tudor, M. (1997). Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. Plant J **12,** 999-1010.

Fariselli, P., Rossi, I., Capriotti, E., and Casadio, R. (2007). The WWWH of remote homolog detection: The state of the art. Brief Bioinform **8,** 78-87.

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M.F., Kater, M.M., and Colombo, L. (2003). MADS-box protein complexes control carpel and ovule development in Arabidopsis. Plant Cell **15,** 2603-2611.

Ferrario, S., Immink, R.G., and Angenent, G.C. (2004). Conservation and diversity in flower land. Current Opinion in Plant Biology **7,** 84-91.

Ferrario, S., Immink, R.G.H., Shchennikova, A., Busscher-Lange, J., and Angenent, G.C. (2003). The MADS box gene FBP2 is required for SEPALLATA function in petunia. Plant Cell **15,** 914-925.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., and Punta, M. (2014). Pfam: the protein families database. Nucleic Acids Research **42,** D222-D230.

Fong, J.H., Keating, A.E., and Singh, M. (2004). Predicting specificity in bZIP coiled-coil protein interactions. Genome Biol **5,** R11.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J.Y., Minguez, P., Bork, P., von Mering, C., and Jensen, L.J. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Research **41,** D808-D815.

Friedman, W.E. (2009). The Meaning of Darwin's "Abominable Mystery". Am J Bot **96,** 5-21.

Frohlich, M.W., and Chase, M.W. (2007). After a dozen years of progress the origin of angiosperms is still a great mystery. Nature **450,** 1184-1189.

Fujisawa, M., Shima, Y., Nakagawa, H., Kitagawa, M., Kimbara, J., Nakano, T., Kasumi, T., and Ito, Y. (2014). Transcriptional Regulation of Fruit Ripening by Tomato FRUITFULL Homologs and Associated MADS Box Proteins. Plant Cell **26,** 89-101.

Georges, A.B., Benayoun, B.A., Caburet, S., and Veitia, R.A. (2010). Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from? Faseb J **24,** 346-356.

Geuten, K., Becker, A., Kaufmann, K., Caris, P., Janssens, S., Viaene, T., Theißen, G., and Smets, E. (2006). Petaloidy and petal identity MADS-box genes in the balsaminoid genera Impatiens and Marcgravia. The Plant Journal **47,** 501-518.

Goto, K., and Meyerowitz, E.M. (1994). Function and regulation of the Arabidopsis floral homeotic gene Pistillata. Gene Dev **8,** 1548-1560.

Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView Version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular Biology and Evolution **27,** 221-224.

Gramzow, L., and Theißen, G. (2010). A hitchhiker's guide to the MADS world of plants. Genome Biology **11,** 214.

Gramzow, L., and Theißen, G. (2013). Phylogenomics of MADS-Box genes in plants - two opposing life styles in one gene family. Biology (Basel) **2,** 1150-1164.

Gramzow, L., and Theissen, G. (2015). Phylogenomics reveals surprising sets of essential and dispensable clades of MIKCc-group MADS-box genes in flowering plants. J Exp Zool Part B **324,** 353-362.

Gramzow, L., Ritz, M.S., and Theißen, G. (2010). On the origin of MADS-domain transcription factors. Trends Genet **26,** 149-153.

Gramzow, L., Weilandt, L., and Theissen, G. (2014). MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. Annals of Botany **114,** 1407-1429.

Gregis, V., Sessa, A., Dorca-Fornell, C., and Kater, M.M. (2009). The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes. Plant Journal **60,** 626-637.

Grigoryan, G., and Keating, A.E. (2006). Structure-based prediction of bZIP partnering specificity. J Mol Biol **355,** 1125-1142.

Gumulya, Y., and Gillam, E.M.J. (2017). Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the ' retro' approach to protein engineering. Biochemical Journal **474,** 1-19.

Härter, A. (2011). Evolution der Interaktion floraler Proteine. (Dissertation) https://www.db-thueringen.de/receive/dbt_mods_00019604.

Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. Nature Reviews Genetics **9,** 15-26.

Henschel, K., Kofuji, R., Hasebe, M., Saedler, H., Münster, T., and Theißen, G. (2002). Two ancient classes of MIKC-type MADS-box genes are present in the moss Physcomitrella patens. Mol Biol Evol **19,** 801-814.

Hernandez-Hernandez, T., Martinez-Castilla, L.P., and Alvarez-Buylla, E.R. (2007). Functional diversification of B MADS-box homeotic regulators of flower development: Adaptive evolution in protein-protein interaction domains after major gene duplication events. Mol Biol Evol **24,** 465-481.

Hogenhout, S.A., Van der Hoorn, R.A.L., Terauchi, R., and Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. Mol Plant Microbe In **22,** 115-122.

Hogenhout, S.A., Oshima, K., Ammar, E.D., Kakizawa, S., Kingdom, H.N., and Namba, S. (2008). Phytoplasmas: bacteria that manipulate plants and insects. Mol Plant Pathol **9,** 403-423.

Honma, T., and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature **409,** 525-529.

Hoshi, A., Oshima, K., Kakizawa, S., Ishii, Y., Ozeki, J., Hashimoto, M., Komatsu, K., Kagiwada, S., Yamaji, Y., and Namba, S. (2009). A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. P Natl Acad Sci USA **106,** 6416-6421.

Hsu, H.F., and Yang, C.H. (2002). An orchid (Oncidium Gower Ramsey) AP3-like MADS gene regulates floral formation and initiation. Plant Cell Physiol **43,** 1198-1209.

Hsu, H.F., Hsu, W.H., Lee, Y.I., Mao, W.T., Yang, J.Y., Li, J.Y., and Yang, C.H. (2015). Model for perianth formation in orchids. Nat Plants **1,** 15046.

Hsu, W.H., Yeh, T.J., Huang, K.Y., Li, J.Y., Chen, H.Y., and Yang, C.H. (2014). AGAMOUS-LIKE13, a putative ancestor for the E functional genes, specifies male and female gametophyte morphogenesis. Plant Journal **77,** 1-15.

Hu, J.C. (2000). A guided tour in protein interaction space: Coiled coils from the yeast proteome. P Natl Acad Sci USA **97,** 12935-12936.

Huang, H., Tudor, M., Su, T., Zhang, Y., Hu, Y., and Ma, H. (1996). DNA binding properties of two Arabidopsis MADS domain proteins: binding consensus and dimer formation. Plant Cell **8,** 81-94.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17,** 754-755.

Immink, R.G.H., Kaufmann, K., and Angenent, G.C. (2010). The 'ABC' of MADS domain protein behaviour and interactions. Seminars in Cell & Developmental Biology **21,** 87-93.

Immink, R.G.H., Ferrario, S., Busscher-Lange, J., Kooiker, M., Busscher, M., and Angenent, G.C. (2003). Analysis of the petunia MADS-box transcription factor family. Mol Genet Genomics **268,** 598-606.

Immink, R.G.H., Tonaco, I.A., de Folter, S., Shchennikova, A., van Dijk, A.D., Busscher-Lange, J., Borst, J.W., and Angenent, G.C. (2009). SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. Genome Biol **10,** R24.

Irish, V.F. (2010). The flowering of Arabidopsis flower development. Plant Journal **61,** 1014-1028.

Jack, T. (2001). Relearning our ABCs: new twists on an old model. Trends in Plant Science **6,** 310-316.

Jack, T. (2004). Molecular and genetic mechanisms of floral control. Plant Cell **16,** S1-S17.

Jack, T., Brockman, L.L., and Meyerowitz, E.M. (1992). The homeotic gene Apetala3 of Arabidopsis thaliana encodes a Mads Box and is expressed in petals and stamens. Cell **68,** 683-697.

Jetha, K., Theißen, G., and Melzer, R. (2014). Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. Nucleic Acids Res **42,** 10927-10942.

Jezic, M., Poljak, I., Safaric, B., Idzojtic, M., and Curkovic-Perica, M. (2013). 'Candidatus Phytoplasma pini' in pine species in Croatia. J Plant Dis Protect **120,** 160-163.

Jofuku, K.D., Denboer, B.G.W., Vanmontagu, M., and Okamuro, J.K. (1994). Control of Arabidopsis flower and seed development by the homeotic gene Apetala2. Plant Cell **6,** 1211-1225.

Kaltenegger, E., and Ober, D. (2015). Paralogue interference affects the dynamics after gene duplication. Trends in Plant Science **20,** 814-821.

Kaminska, M., Sliwa, H., and Rudzinska-Langwald, A. (2001). The association of phytoplasma with stunting, leaf necrosis and witches' broom symptoms in magnolia plants. J Phytopathol **149,** 719-724.

Kanno, A., Saeki, H., Kameya, T., Saedler, H., and Theissen, G. (2003). Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (Tulipa gesneriana). Plant Mol Biol **52,** 831-841.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **30,** 772-780.

Kaufmann, K., Melzer, R., and Theißen, G. (2005a). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene **347,** 183-198.

Kaufmann, K., Pajoro, A., and Angenent, G.C. (2010a). Regulation of transcription in plants: mechanisms controlling developmental switches. Nature Reviews Genetics **11,** 830-842.

Kaufmann, K., Anfang, N., Saedler, H., and Theissen, G. (2005b). Mutant analysis, protein-protein interactions and subcellular localization of the Arabidopsis B-sister (ABS) protein. Molecular Genetics and Genomics **274,** 103-118.

Kaufmann, K., Muino, J.M., Jauregui, R., Airoldi, C.A., Smaczniak, C., Krajewski, P., and Angenent, G.C. (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. Plos Biol **7,** 854-875.

Kaufmann, K., Wellmer, F., Muino, J.M., Ferrier, T., Wuest, S.E., Kumar, V., Serrano-Mislata, A., Madueno, F., Krajewski, P., Meyerowitz, E.M., Angenent, G.C., and

Riechmann, J.L. (2010b). Orchestration of floral initiation by APETALA1. Science **328,** 85-89.

Kim, S., Soltis, P.S., and Soltis, D.E. (2013). AGL6-like MADS-box genes are sister to AGL2-like MADS-box genes. J Plant Biol **56,** 315-325.

Kim, S., Koh, J., Yoo, M.J., Kong, H., Hu, Y., Ma, H., Soltis, P.S., and Soltis, D.E. (2005). Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. Plant J **43,** 724-744.

Kim, S.T., Yoo, M.J., Albert, V.A., Farris, J.S., Soltis, P.S., and Soltis, D.E. (2004). Phylogeny and diversification of B-function MADS-box genes in angiosperms: Evolutionary and functional implications of a 260-million-year-old duplication. Am J Bot **91,** 2102-2118.

Kitano, H. (2004). Biological robustness. Nature Reviews Genetics **5,** 826-837.

Kitazawa, Y., Iwabuchi, N., Himeno, M., Sasano, M., Koinuma, H., Nijo, T., Tomomitsu, T., Yoshida, T., Okano, Y., Yoshikawa, N., Maejima, K., Oshima, K., and Namba, S. (2017). Phytoplasma-conserved phyllogen proteins induce phyllody across the Plantae by degrading floral MADS domain proteins. J Exp Bot**,** erx158.

Kramer, E.M., Dorit, R.L., and Irish, V.F. (1998). Molecular evolution of genes controlling petal and stamen development: Duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages. Genetics **149,** 765-783.

Kramer, E.M., Holappa, L., Gould, B., Jaramillo, M.A., Setnikov, D., and Santiago, P.M. (2007). Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot Aquilegia. The Plant Cell Online **19,** 750-766.

Krizek, B.A., and Meyerowitz, E.M. (1996a). The Arabidopsis homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. Development **122,** 11-22.

Krizek, B.A., and Meyerowitz, E.M. (1996b). Mapping the protein regions responsible for the functional specificities of the Arabidopsis MADS domain organ-identity proteins. Proc Natl Acad Sci U S A **93,** 4063-4070.

Krizek, B.A., and Fletcher, J.C. (2005). Molecular mechanisms of flower development: An armchair guide. Nature Reviews Genetics **6,** 688-698.

Kuhn, M., Hyman, A.A., and Beyer, A. (2014). Coiled-coil proteins facilitated the functional expansion of the centrosome. Plos Computational Biology **10,** e1003657.

Kükenshöner, T., Wohlwend, D., Niemoller, C., Dondapati, P., Speck, J., Adeniran, A.V., Nieth, A., Gerhardt, S., Einsle, O., Muller, K.M., and Arndt, K.M. (2014). Improving coiled coil stability while maintaining specificity by a bacterial hitchhiker selection system. J Struct Biol **186,** 335-348.

Kwantes, M., Liebsch, D., and Verelst, W. (2012). How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. Mol Biol Evol **29,** 293-302.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007). Clustal W and clustal X version 2.0. Bioinformatics **23,** 2947-2948.

Lee, H.L., and Irish, V.F. (2011). Gene duplication and loss in a MADS box gene transcription factor circuit. Mol Biol Evol **28,** 3367-3380.

Lee, I.M., Davis, R.E., and Gundersen-Rindal, D.E. (2000). Phytoplasma: Phytopathogenic mollicutes. Annu Rev Microbiol **54,** 221-255.

Lenser, T., Theissen, G., and Dittrich, P. (2009). Developmental robustness by obligate interaction of class B floral homeotic genes and proteins. PLoS Comput Biol **5,** e1000264.

Leseberg, C.H., Eissler, C.L., Wang, X., Johns, M.A., Duvall, M.R., and Mao, L. (2008). Interaction study of MADS-domain proteins in tomato. Journal of Experimental Botany **59,** 2253-2265.

Li, L., Yu, X.X., Guo, C.C., Duan, X.S., Shan, H.Y., Zhang, R., Xu, G.X., and Kong, H.Z. (2015). Interactions among proteins of floral MADS-box genes in Nuphar pumila (Nymphaeaceae) and the most recent common ancestor of extant angiosperms help understand the underlying mechanisms of the origin of the flower. J Syst Evol **53,** 285-296.

Li, Z.Y., Dong, Z.P., Hao, Z.M., and Dong, J.G. (2012). First report of elm yellows subgroup 16SrV-B phytoplasma infecting chinese tulip tree in China. Plant Dis **96,** 1064-1064.

Lindsley, D.L. (1968). Genetic variations of Drosophila melanogaster. (Carnegie Institution of Washington).

Litt, A. (2007). An evaluation of A-function: Evidence from the APETALA1 and APETALA2 gene lineages. Int J Plant Sci **168,** 73-91.

Litt, A., and Kramer, E.M. (2010). The ABC model and the diversification of floral organ identity. Seminars in Cell & Developmental Biology **21,** 129-137.

Liu, C., Xi, W., Shen, L., Tan, C., and Yu, H. (2009). Regulation of floral patterning by flowering time genes. Dev Cell **16,** 711-722.

Liu, C.J., Zhang, J.A., Zhang, N., Shan, H.Y., Su, K.M., Zhang, J.S., Meng, Z., Kong, H.Z., and Chen, Z.D. (2010). Interactions among proteins of floral MADS-box genes in basal eudicots: Implications for evolution of the regulatory network for flower development. Molecular Biology and Evolution **27,** 1598-1611.

Liu, D., Wang, D., Qin, Z., Zhang, D., Yin, L., Wu, L., Colasanti, J., Li, A., and Mao, L. (2013). The SEPALLATA MADS-box protein SLMBP21 forms protein complexes with JOINTLESS and MACROCALYX as a transcription activator for tomato flower abscission zone development. The Plant Journal **77,** 284-296.

Lopez-Vernaza, M., Yang, S.X., Muller, R., Thorpe, F., de Leau, E., and Goodrich, J. (2012). Antagonistic roles of SEPALLATA3, FT and FLC genes as targets of the polycomb group gene CURLY LEAF. Plos One **7,** e30715.

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. Science **252,** 1162-1164.

Lupas, A.N., and Gruber, M. (2005). The structure of alpha-helical coiled coils. Adv Protein Chem **70,** 37-78.

Ma, H., Yanofsky, M.F., and Meyerowitz, E.M. (1991). AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. Genes Dev **5,** 484-495.

MacLean, A.M., Orlovskis, Z., Kowitwanich, K., Zdziarska, A.M., Angenent, G.C., Immink, R.G., and Hogenhout, S.A. (2014). Phytoplasma effector SAP54 Hijacks plant reproduction by degrading MADS-box proteins and promotes insect colonization in a RAD23-dependent manner. Plos Biol **12,** e1001835.

MacLean, A.M., Sugio, A., Makarova, O.V., Findlay, K.C., Grieve, V.M., Toth, R., Nicolaisen, M., and Hogenhout, S.A. (2011). Phytoplasma effector SAP54 induces indeterminate leaf-like flower development in Arabidopsis plants. Plant Physiol **157,** 831-841.

Maddison, W.P., and Maddison, D.R. (2011). Mesquite: a modular system for evolutionary analysis. Version 2.75 http://mesquiteproject.org.

Maejima, K., Oshima, K., and Namba, S. (2014a). Exploring the phytoplasmas, plant pathogenic bacteria. J Gen Plant Pathol **80,** 210-221.

Maejima, K., Kitazawa, Y., Tomomitsu, T., Yusa, A., Neriya, Y., Himeno, M., Yamaji, Y., Oshima, K., and Namba, S. (2015). Degradation of class E MADS-domain transcription factors in Arabidopsis by a phytoplasmal effector, phyllogen. Plant Signaling & Behavior **10,** e1042635.

Maejima, K., Iwai, R., Himeno, M., Komatsu, K., Kitazawa, Y., Fujita, N., Ishikawa, K., Fukuoka, M., Minato, N., Yamaji, Y., Oshima, K., and Namba, S. (2014b). Recognition of floral homeotic MADS domain transcription factors by a phytoplasmal effector, phyllogen, induces phyllody. Plant Journal **78,** 541-554.

Malcomber, S.T., and Kellogg, E.A. (2005). SEPALLATA gene diversification: brave new whorls. Trends Plant Sci **10,** 427-435.

Mandel, M.A., Gustafsonbrown, C., Savidge, B., and Yanofsky, M.F. (1992). Molecular characterization of the Arabidopsis floral homeotic gene Apetala1. Nature **360,** 273-277.

Marcone, C. (2014). Molecular biology and pathogenicity of phytoplasmas. Ann Appl Biol **165,** 199-221.

Marquez-Guzman, J., Engleman, M., Martinezmena, A., Martinez, E., and Ramos, C. (1989). Reproductive anatomy of Lacandonia schismatica (Lacandoniaceae). Ann Mo Bot Gard **76,** 124-127.

Mason, J.M., and Arndt, K.M. (2004). Coiled coil domains: Stability, specificity, and biological implications. Chembiochem **5,** 170-176.

Mason, J.M., Hagemann, U.B., and Arndt, K.M. (2009). Role of hydrophobic and electrostatic interactions in coiled coil stability and specificity. Biochemistry **48,** 10380-10388.

McGonigle, B., Bouhidel, K., and Irish, V.F. (1996). Nuclear localization of the Arabidopsis APETALA3 and PISTILLATA homeotic gene products depends on their simultaneous expression. Gene Dev **10,** 1812-1821.

Melzer, R., and Theißen, G. (2009). Reconstitution of 'floral quartets' in vitro involving class B and class E floral homeotic proteins. Nucleic Acids Res **37,** 2723-2736.

Melzer, R., Kaufmann, K., and Theissen, G. (2006). Missing links: DNA-binding and target gene specificity of floral homeotic proteins. Adv Bot Res **44,** 209-236.

Melzer, R., Verelst, W., and Theißen, G. (2009). The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro. Nucleic Acids Res **37,** 144-157.

Melzer, R., Wang, Y.Q., and Theißen, G. (2010). The naked and the dead: the ABCs of gymnosperm reproduction and the origin of the angiosperm flower. Semin Cell Dev Biol **21,** 118-128.

Melzer, R., Härter, A., Rümpler, F., Kim, S., Soltis, P.S., Soltis, D.E., and Theißen, G. (2014). DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. Ann Bot **114,** 1431-1443.

Mendes, M.A., Guerra, R.F., Berns, M.C., Manzo, C., Masiero, S., Finzi, L., Kater, M.M., and Colombo, L. (2013). MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in Arabidopsis. Plant Cell **25,** 2560-2572.

Messenguy, F., and Dubois, E. (2003). Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. Gene **316,** 1-21.

Meyerowitz, E.M., Smyth, D.R., and Bowman, J.L. (1989). Abnormal flowers and pattern-formation in floral development. Development **106,** 209-217.

Minato, N., Himeno, M., Hoshi, A., Maejima, K., Komatsu, K., Takebayashi, Y., Kasahara, H., Yusa, A., Yamaji, Y., Oshima, K., Kamiya, Y., and Namba, S. (2014). The phytoplasmal virulence factor TENGU causes plant sterility by downregulating of the jasmonic acid and auxin pathways. Sci Rep-Uk **4,** 7399–7405.

Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Research **41,** e121.

Mizukami, Y., and Ma, H. (1992). Ectopic expression of the floral homeotic gene Agamous in transgenic Arabidopsis plants alters floral organ identity. Cell **71,** 119-131.

Mizzotti, C., Mendes, M.A., Caporali, E., Schnittger, A., Kater, M.M., Battaglia, R., and Colombo, L. (2012). The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. Plant Journal **70,** 409-420.

Moitra, J., Szilak, L., Krylov, D., and Vinson, C. (1997). Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. Biochemistry **36,** 12567-12573.

Mondragon-Palomino, M., and Theissen, G. (2008). MADS about the evolution of orchid flowers. Trends in Plant Science **13,** 51-59.

Mondragon-Palomino, M., and Theissen, G. (2011). Conserved differential expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the flowers of Orchidaceae: refining the 'orchid code'. Plant Journal **66,** 1008-1019.

Mondragon-Palomino, M., Hiese, L., Harter, A., Koch, M.A., and Theissen, G. (2009). Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses. BMC Evol Biol **9,** 81.

Mooers, A.O., and Schluter, D. (1999). Reconstructing ancestor states with maximum likelihood: Support for one- and two-rate models. Syst Biol **48,** 623-633.

Moon, Y.H., Kang, H.G., Jung, J.Y., Jeon, J.S., Sung, S.K., and An, G. (1999). Determination of the motif responsible for interaction between the rice APETALA1/AGAMOUS-LIKE9 family proteins using a yeast two-hybrid system. Plant Physiol **120,** 1193-1204.

Moyle-Heyrman, G., Tims, H.S., and Widom, J. (2011). Structural constraints in collaborative competition of transcription factors against the nucleosome. Journal of Molecular Biology **412,** 634-646.

Muino, J.M., Smaczniak, C., Angenent, G.C., Kaufmann, K., and van Dijk, A.D.J. (2014). Structural determinants of DNA recognition by plant MADS-domain transcription factors. Nucleic Acids Res **42,** 2138-2146.

Münster, T., Wingen, L.U., Faigl, W., Werth, S., Saedler, H., and Theißen, G. (2001). Characterization of three GLOBOSA-like MADS-box genes from maize: evidence for ancient paralogy in one class of floral homeotic B-function genes of grasses. Gene **262,** 1-13.

Münster, T., Pahnke, J., DiRosa, A., Kim, J.T., Martin, W., Saedler, H., and Theißen, G. (1997). Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. P Natl Acad Sci USA **94,** 2415-2420.

Nagasawa, N., Miyoshi, M., Sano, Y., Satoh, H., Hirano, H., Sakai, H., and Nagato, Y. (2003). SUPERWOMAN1 and DROOPING LEAF genes control floral organ identity in rice. Development **130,** 705-718.

Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., Vonrhein, C., Moras, D., Romier, C., Bolognesi, M., and Mantovani, R. (2013). Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. Cell **152,** 132-143.

Nesi, N., Debeaujon, I., Jond, C., Stewart, A.J., Jenkins, G.I., Caboche, M., and Lepiniec, L. (2002). The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell **14,** 2463-2479.

Newman, J.R.S., Wolf, E., and Kim, P.S. (2000). A computationally directed screen identifying interacting coiled coils from Saccharomyces cerevisiae. P Natl Acad Sci USA **97,** 13203-13208.

Nilsson, I., Saaf, A., Whitley, P., Gafvelin, G., Waller, C., and von Heijne, G. (1998). Proline-induced disruption of a transmembrane alpha-helix in its natural environment. J Mol Biol **284,** 1165-1175.

Nougalli-Tonaco, I.A.N., Borst, J.W., de Vries, S.C., Angenent, G.C., and Immink, R.G.H. (2006). In vivo imaging of MADS-box transcription factor interactions. Journal of Experimental Botany **57,** 33-42.

Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hallman, J., Keech, O., Klasson, L., Koriabine, M., Kucukoglu, M., Kaller, M., Luthman, J., Lysholm, F., Niittyla, T., Olson, A., Rilakovic, N., Ritland, C., Rossello, J.A., Sena, J., Svensson, T., Talavera-Lopez, C., Theissen, G., Tuominen, H., Vanneste, K., Wu, Z.Q., Zhang, B., Zerbe, P., Arvestad, L., Bhalerao, R., Bohlmann, J., Bousquet, J., Garcia Gil, R., Hvidsten, T.R., de Jong, P., Mackay, J., Morgante, M., Ritland, K., Sundberg, B., Lee Thompson, S., Van de Peer, Y., Andersson, B., Nilsson, O., Ingvarsson, P.K., Lundeberg, J., and Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. Nature **497,** 579-584.

O'Maoileidigh, D.S., Graciet, E., and Wellmer, F. (2014). Gene networks controlling Arabidopsis thaliana flower development. New Phytologist **201,** 16-30.

O'Maoileidigh, D.S., Wuest, S.E., Rae, L., Raganelli, A., Ryan, P.T., Kwasniewska, K., Das, P., Lohan, A.J., Loftus, B., Graciet, E., and Wellmer, F. (2013). Control of reproductive floral organ identity specification in Arabidopsis by the C function regulator AGAMOUS. Plant Cell **25,** 2482-2503.

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R.C., Meldal, B., Melidoni, A.N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., van Roey, K., Cesareni, G., and Hermjakob, H. (2014). The MIntAct project-IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Research **42,** D358-D363.

Orlovskis, Z., and Hogenhout, S.A. (2016). A bacterial parasite effector mediates insect vector attraction in host plants independently of developmental changes. Front Plant Sci **7,** 885.

Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H.Y., Wei, W., Suzuki, S., Arashida, R., Nakata, D., Miyata, S., Ugaki, M., and Namba, S. (2004). Reductive evolution

suggested from the complete genome sequence of a plant-pathogenic phytoplasma. Nat Genet **36,** 27-29.

Pagel, M. (1999). The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst Biol **48,** 612-622.

Pajoro, A., Madrigal, P., Muino, J.M., Matus, J.T., Jin, J., Mecchia, M.A., Debernardi, J.M., Palatnik, J.F., Balazadeh, S., Arif, M., O'Maoileidigh, D.S., Wellmer, F., Krajewski, P., Riechmann, J.L., Angenent, G.C., and Kaufmann, K. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome Biol **15,** R41.

Parenicova, L., de Folter, S., Kieffer, M., Horner, D.S., Favalli, C., Busscher, J., Cook, H.E., Ingram, R.M., Kater, M.M., Davies, B., Angenent, G.C., and Colombo, L. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell **15,** 1538-1551.

Parry, D.A., Fraser, R.D., and Squire, J.M. (2008). Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. J Struct Biol **163,** 258-269.

Pelaz, S., Tapia-Lopez, R., Alvarez-Buylla, E.R., and Yanofsky, M.F. (2001). Conversion of leaves into petals in Arabidopsis. Current Biology **11,** 182-184.

Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E., and Yanofsky, M.F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. Nature **405,** 200-203.

Perez, K., Yeam, I., Jahn, M.M., and Kang, B.C. (2006). Megaprimer-mediated domain swapping for construction of chimeric viruses. J Virol Methods **135,** 254-262.

Pinyopich, A., Ditta, G.S., Savidge, B., Liljegren, S.J., Baumann, E., Wisman, E., and Yanofsky, M.F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature **424,** 85-88.

Polach, K.J., and Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. Journal of Molecular Biology **258,** 800-812.

Potapov, V., Kaplan, J.B., and Keating, A.E. (2015). Data-driven prediction and design of bZIP coiled-coil interactions. Plos Comput Biol **11,** e1004046.

Puranik, S., Acajjaoui, S., Conn, S., Costa, L., Conn, V., Vial, A., Marcellin, R., Melzer, R., Brown, E., Hart, D., Theissen, G., Silva, C.S., Parcy, F., Dumas, R., Nanao, M., and Zubieta, C. (2014). Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in Arabidopsis. Plant Cell **26,** 3603-3615.

Purugganan, M.D., Rounsley, S.D., Schmidt, R.J., and Yanofsky, M.F. (1995). Molecular evolution of flower development - Diversification of the plant Mads-box regulatory gene family. Genetics **140,** 345-356.

Ree, R.H., and Donoghue, M.J. (1999). Inferring rates of change in flower symmetry in asterid angiosperms. Syst Biol **48,** 633-641.

Remenyi, A., Scholer, H.R., and Wilmanns, M. (2004). Combinatorial control of gene expression. Nat Struct Mol Biol **11,** 812-815.

Rhee, H.S., Bataille, A.R., Zhang, L.Y., and Pugh, B.F. (2014). Subnucleosomal structures and nucleosome asymmetry across a genome. Cell **159,** 1377-1388.

Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. Adv Protein Chem **34,** 167-339.

Riechmann, J.L., and Meyerowitz, E.M. (1997). MADS domain proteins in plant development. Biol Chem **378,** 1079-1101.

Riechmann, J.L., Wang, M.Q., and Meyerowitz, E.M. (1996a). DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Research **24,** 3134-3141.

Riechmann, J.L., Krizek, B.A., and Meyerowitz, E.M. (1996b). Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. P Natl Acad Sci USA **93,** 4793-4798.

Rijpkema, A.S., Zethof, J., Gerats, T., and Vandenbussche, M. (2009). The petunia AGL6 gene has a SEPALLATA-like function in floral patterning. Plant Journal **60,** 1-9.

Ronai, Z., Wang, Y., Khandurina, J., Budworth, P., Sasvari-Szekely, M., Wang, X., and Guttman, A. (2003). Transcription factor binding study by capillary zone electrophoretic mobility shift assay. Electrophoresis **24,** 96-100.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19,** 1572-1574.

Rost, B. (1999). Twilight zone of protein sequence alignments. Protein Engineering **12,** 85-94.

Rudall, P.J., Remizowa, M.V., Prenner, G., Prychid, C.J., Tuckett, R.E., and Sokoloff, D.D. (2009). Nonflowers near the base of extant angiosperms? Spatiotemporal arrangement of organs in reproductive units of Hydatellaceae and its bearing on the origin of the flower. Am J Bot **96,** 67-82.

Ruelens, P., de Maagd, R.A., Proost, S., Theissen, G., Geuten, K., and Kaufmann, K. (2013). FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. Nature communications **4,** 2280.

Ruelens, P., Zhang, Z.C., van Mourik, H., Maere, S., Kaufmann, K., and Geuten, K. (2017). The origin of floral organ identity quartets. Plant Cell **29,** 229-242.

Rümpler, F., Melzer, R., and Theißen, G. (2015a). Character-state reconstruction to infer ancestral protein-protein interaction patterns. Bio-protocol **5**.

Rümpler, F., Theißen, G., and Melzer, R. (2017). Sequence features of MADS-domain proteins that act as hubs in the protein-protein interaction network controlling flower development. bioRxiv, 125294.

Rümpler, F., Gramzow, L., Theißen, G., and Melzer, R. (2015b). Did convergent protein evolution enable phytoplasmas to generate 'zombie plants'? Trends in Plant Science **20,** 798-806.

Ruokolainen, S., Ng, Y., Albert, V.A., Elomaa, P., and Teeri, T.H. (2010). Large scale interaction analysis predicts that the Gerbera hybrida floral E function is provided both by general and specialized proteins. Bmc Plant Biol **10,** 129.

Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004). The database of interacting proteins: 2004 update. Nucleic Acids Research **32,** D449-D451.

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Federhen, S., Feolo, M., Fingerman, I.M., Geer, L.Y., Helmberg, W., Kapustin, Y., Krasnov, S., Landsman, D., Lipman, D.J., Lu, Z.Y., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Miller, V., Karsch-Mizrachi, I., Ostell, J., Panchenko, A., Phan, L., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Wang, Y.L., Wilbur, W.J., Yaschenko, E., and Ye, J. (2012). Database resources of the National Center for Biotechnology Information. Nucleic Acids Research **40,** D13-D25.

Schneider, B., Torres, E., Martin, M.P., Schroder, M., Behnke, H.D., and Seemuller, E. (2005). 'Candidatus Phytoplasma pini', a novel taxon from Pinus silvestris and Pinus halepensis. Int J Syst Evol Micr **55,** 303-307.

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic-control of flower development by homeotic genes in Antirrhinum majus. Science **250,** 931-936.

Schwarz-Sommer, Z., Hue, I., Huijser, P., Flor, P.J., Hansen, R., Tetens, F., Lonnig, W.E., Saedler, H., and Sommer, H. (1992). Characterization of the Antirrhinum floral homeotic Mads-box gene deficiens - Evidence for DNA-binding and autoregulation of its persistent expression throughout flower development. Embo J **11,** 251-263.

Senear, D.F., and Brenowitz, M. (1991). Determination of binding constants for cooperative site-specific protein-DNA interactions using the gel mobility-shift assay. J Biol Chem **266,** 13661-13671.

Seok, H.Y., Park, H.Y., Park, J.I., Lee, Y.M., Lee, S.Y., An, G., and Moon, Y.H. (2010). Rice ternary MADS protein complexes containing class B MADS heterodimer. Biochem Biophys Res Commun **401,** 598-604.

Shchennikova, A.V., Shulga, O.A., Immink, R., Skryabin, K.G., and Angenent, G.C. (2004). Identification and characterization of four chrysanthemum MADS-box genes, belonging to the APETALA1/FRUITFULL and SEPALLATA3 subfamilies. Plant Physiol **134,** 1632-1641.

Silva, C.S., Puranik, S., Round, A., Brennich, M., Jourdain, A., Parcy, F., Hugouvieux, V., and Zubieta, C. (2016). Evolution of the plant reproduction master regulators LFY and the MADS transcription factors: The role of protein structure in the evolutionary development of the flower. Front Plant Sci **6,** 1193.

Slattery, M., Zhou, T.Y., Yang, L., Machado, A.C.D., Gordan, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. Trends in Biochemical Sciences **39,** 381-399.

Smaczniak, C., Immink, R.G., Angenent, G.C., and Kaufmann, K. (2012a). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. Development **139,** 3081-3098.

Smaczniak, C., Immink, R.G., Muino, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q.D., Liu, S., Westphal, A.H., Boeren, S., Parcy, F., Xu, L., Carles, C.C., Angenent, G.C., and Kaufmann, K. (2012b). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proc Natl Acad Sci USA **109,** 1560-1565.

Smirnova, E., Shurland, D.L., Newman-Smith, E.D., Pishvaee, B., and van der Bliek, A.M. (1999). A model for dynamin self-assembly based on binding between three different protein domains. Journal of Biological Chemistry **274,** 14942-14947.

Soboleva, T.A., Nekrasov, M., Ryan, D.P., and Tremethick, D.J. (2014). Histone variants at the transcription start-site. Trends in Genetics **30,** 199-209.

Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res **33,** W244-248.

Soltis, D.E., Chanderbali, A.S., Kim, S., Buzgo, M., and Soltis, P.S. (2007). The ABC model and its applicability to basal angiosperms. Annals of Botany **100,** 155-163.

Soltis, P.S., Soltis, D.E., Kim, S., Chanderbali, A., and Buzgo, M. (2006). Expression of floral regulators in basal angiosperms and the origin and evolution of ABC-function. Adv Bot Res **44,** 483-506.

Soltis, P.S., Brockington, S.F., Yoo, M.J., Piedrahita, A., Latvis, M., Moore, M.J., Chanderbali, A.S., and Soltis, D.E. (2009). Floral variation and floral genetics in basal angiosperms. Am J Bot **96,** 110-128.

Sommer, H., Beltran, J.P., Huijser, P., Pape, H., Lonnig, W.E., Saedler, H., and Schwarzsommer, Z. (1990). Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus - the protein shows homology to transcription factors. Embo J **9,** 605-613.

Stebbins, C.E., and Galan, J.E. (2001). Structural mimicry in bacterial virulence. Nature **412,** 701-705.

Sugawara, K., Honma, Y., Komatsu, K., Himeno, M., Oshima, K., and Namba, S. (2013). The alteration of plant morphology by small peptides released from the proteolytic processing of the bacterial peptide TENGU. Plant Physiol **162,** 2005-2014.

Sugio, A., and Hogenhout, S.A. (2012). The genome biology of phytoplasma: modulators of plants and insects. Curr Opin Microbiol **15,** 247-254.

Sugio, A., MacLean, A.M., and Hogenhout, S.A. (2014). The small phytoplasma virulence effector SAP11 contains distinct domains required for nuclear targeting and CIN-TCP binding and destabilization. New Phytologist **202,** 838-848.

Sugio, A., Kingdom, H.N., MacLean, A.M., Grieve, V.M., and Hogenhout, S.A. (2011). Phytoplasma protein effector SAP11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. P Natl Acad Sci USA **108,** E1254-E1263.

Sundström, J., and Engström, P. (2002). Conifer reproductive development involves B-type MADS-box genes with distinct and different activities in male organ primordia. Plant J **31,** 161-169.

Surkont, J., and Pereira-Leal, J.B. (2015). Evolutionary patterns in coiled-coils. Genome biology and evolution **7,** 545-556.

Takei, T., Okonogi, A., Tateno, K., Kimura, A., Kojima, S., Yazaki, K., and Miura, K. (2006). The effects of the side chains of hydrophobic aliphatic amino acid residues in an amphipathic polypeptide on the formation of a helix and its association. J Biochem **139,** 271-278.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution **30,** 2725-2729.

Tanabe, Y., Hasebe, M., Sekimoto, H., Nishiyama, T., Kitani, M., Henschel, K., Munster, T., Theißen, G., Nozaki, H., and Ito, M. (2005). Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. P Natl Acad Sci USA **102,** 2436-2441.

Theißen, G. (2001). Development of floral organ identity: stories from the MADS house. Curr Opin Plant Biol **4,** 75-85.

Theißen, G., and Saedler, H. (2001). Plant biology. Floral quartets. Nature **409,** 469-471.

Theißen, G., and Melzer, R. (2006). Combinatorial control of floral organ identity by MADS-domain transcription factors. In Annual Review of Plant Biology, K. Grasser, ed (New York: Blackwell Publishing), pp. 253-265.

Theißen, G., and Melzer, R. (2007). Molecular mechanisms underlying origin and diversification of the angiosperm flower. Annals of Botany **100,** 603-619.

Theißen, G., and Gramzow, L. (2016). Structure and evolution of plant MADS-domain transcription factors. In Plant Transcription Factors: Evolutionary, Structural and Functional Aspects, D.H. Gonzalez, ed (Philadelphia: Elsiver), pp. 127-138.

Theißen, G., Kim, J.T., and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. J Mol Evol **43,** 484-516.

Theißen, G., Melzer, R., and Rümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. Development **143,** 3259-3271.

Theißen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J.T., Munster, T., Winter, K.U., and Saedler, H. (2000). A short history of MADS-box genes in plants. Plant Mol Biol **42,** 115-149.

Theissen, G., and Saedler, H. (1995). MADS-box genes in plant ontogeny and phylogeny: Haeckel's 'biogenetic law' revisited. Curr Opin Genet Dev **5,** 628-639.

Thompson, B.E., Bartling, L., Whipple, C., Hall, D.H., Sakai, H., Schmidt, R., and Hake, S. (2009). bearded-ear encodes a MADS-box transcription factor critical for maize floral development. Plant Cell **21,** 2578-2590.

Todeschini, A.L., Georges, A., and Veitia, R.A. (2014). Transcription factors: specific DNA binding and specific gene regulation. Trends in Genetics **30,** 211-219.

Toruno, T.Y., Music, M.S., Simi, S., Nicolaisen, M., and Hogenhout, S.A. (2010). Phytoplasma PMU1 exists as linear chromosomal and circular extrachromosomal elements and has enhanced expression in insect vectors compared with plant hosts. Mol Microbiol **77,** 1406-1415.

UniProt Consortium. (2015). UniProt: a hub for protein information. Nucleic Acids Research **43,** D204-D212.

van Dijk, A.D.J., Morabito, G., Fiers, M., van Ham, R.C.H.J., Angenent, G.C., and Immink, R.G.H. (2010). Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein-protein interaction. Plos Computational Biology **6,** e1001017.

Vandenbussche, M. (2004). The duplicated B-class heterodimer model: Whorl-specific effects and complex genetic interactions in Petunia hybrida flower development. The Plant Cell Online **16,** 741-754.

Wang, Y.-Q., Melzer, R., and Theißen, G. (2010). Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm Gnetum gnemon provide a clue to the evolutionary origin of 'floral quartets'. The Plant Journal **64,** 177-190.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics **25,** 1189-1191.

Wellmer, F., Graciet, E., and Riechmann, J.L. (2014). Specification of floral organs in Arabidopsis. Journal of Experimental Botany **65,** 1-9.

Wernersson, R., and Pedersen, A.G. (2003). RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Research **31,** 3537-3539.

Whipple, C.J. (2004). Conservation of B-class floral homeotic gene function between maize and Arabidopsis. Development **131,** 6083-6091.

Whipple, C.J., and Schmidt, R.J. (2006). Genetics of grass flower development. Adv Bot Res **44,** 385-424.

Whitty, A. (2008). Cooperativity and biological complexity. Nat Chem Biol **4,** 435-439.

Win, J., Chaparro-Garcia, A., Belhaj, K., Saunders, D.G., Yoshida, K., Dong, S., Schornack, S., Zipfel, C., Robatzek, S., Hogenhout, S.A., and Kamoun, S. (2012). Effector biology of plant-associated organisms: concepts and perspectives. Cold Spring Harb Symp Quant Biol **77,** 235-247.

Winter, K.U., Saedler, H., and Theissen, G. (2002a). On the origin of class B floral homeotic genes: functional substitution and dominant inhibition in Arabidopsis by expression of an orthologue from the gymnosperm Gnetum. Plant J **31,** 457-475.

Winter, K.U., Becker, A., Munster, T., Kim, J.T., Saedler, H., and Theissen, G. (1999). MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. P Natl Acad Sci USA **96,** 7342-7347.

Winter, K.U., Weiser, C., Kaufmann, K., Bohne, A., Kirchner, C., Kanno, A., Saedler, H., and Theißen, G. (2002b). Evolution of class B floral homeotic proteins: obligate heterodimerization originated from homodimerization. Mol Biol Evol **19,** 587-596.

Wuest, S.E., O'Maoileidigh, D.S., Rae, L., Kwasniewska, K., Raganelli, A., Hanczaryk, K., Lohan, A.J., Loftus, B., Graciet, E., and Wellmer, F. (2012). Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. P Natl Acad Sci USA **109,** 13452-13457.

Xu, D., and Zhang, Y. (2012). Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins-Structure Function and Bioinformatics **80,** 1715-1735.

Yan, W.H., Chen, D.J., and Kaufmann, K. (2016). Molecular mechanisms of floral organ specification by MADS domain proteins. Current Opinion in Plant Biology **29,** 154-162.

Yang, C.Y., Huang, Y.H., Lin, C.P., Lin, Y.Y., Hsu, H.C., Wang, C.N., Liu, L.Y.D., Shen, B.N., and Lin, S.S. (2015). MicroRNA396-targeted SHORT VEGETATIVE PHASE is required to repress flowering and is related to the development of abnormal flower symptoms by the phyllody Symptoms1 effector. Plant Physiol **168,** 1702-1716.

Yang, Y.Z., and Jack, T. (2004). Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins. Plant Mol Biol **55,** 45-59.

Yang, Y.Z., Xiang, H.J., and Jack, T. (2003a). pistillata-5, an Arabidopsis B class mutant with strong defects in petal but not in stamen development. Plant Journal **33,** 177-188.

Yang, Y.Z., Fanning, L., and Jack, T. (2003b). The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. Plant J **33,** 47-59.

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., and Meyerowitz, E.M. (1990). The protein encoded by the Arabidopsis homeotic gene Agamous resembles transcription factors. Nature **346,** 35-39.

Yao, S.G., Ohmori, S., Kimizu, M., and Yoshida, H. (2008). Unequal genetic redundancy of rice PISTILLATA orthologs, OsMADS2 and OsMADS4, in lodicule and stamen development. Plant Cell Physiol **49,** 853-857.

Zahn, L.M., Leebens-Mack, J., dePamphilis, C.W., Ma, H., and Theissen, G. (2005a). To B or not to B a flower: The role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. J Hered **96,** 225-240.

Zahn, L.M., King, H.Z., Leebens-Mack, J.H., Kim, S., Soltis, P.S., Landherr, L.L., Soltis, D.E., dePamphilis, C.W., and Ma, H. (2005b). The evolution of the SEPALLATA subfamily of MADS-Box genes: A preangiosperm origin with multiple duplications throughout angiosperm history. Genetics **169,** 2209-2223.

Zhang, Z., Coenen, H., Ruelens, P., Hazarika, R.R., Al Hindi, T., K. Oguis, G., van Noort, V., and Geuten, K. (2016). Resurrected protein interaction networks reveal the innovation potential of ancient whole genome duplication. bioRxiv**,** 074989.

Zhu, B.Y., Zhou, N.E., Kay, C.M., and Hodges, R.S. (1993). Packing and hydrophobicity effects on protein folding and stability - effects of beta-branched amino-acids, valine and isoleucine, on the formation and stability of 2-stranded alpha-helical coiled coils leucine zippers. Protein Sci **2,** 383-394.

Zobell, O., Faigl, W., Saedler, H., and Münster, T. (2010). MIKC* MADS-Box Proteins: Conserved Regulators of the Gametophytic Generation of Land Plants. Molecular Biology and Evolution **27,** 1201-1211.

6 Acknowledgements

Mein Dank gilt zuallererst meinem Doktorvater Günter, der mich in den letzten Jahren nicht nur auf fachlicher sondern auch auf menschlicher Ebene sehr unterstützt hat. Deine, im positiven Sinne, antiautoritäre Art der Betreuung hat mir immer genug Freiraum gelassen, um meinen eigenen Ideen nachzugehen. Zugleich hast Du aber immer dafür gesorgt, dass ich nicht den Fokus verliere. Ich danke Dir außerdem für die unermüdliche Bereitschaft, die einzelnen Abschnitte dieser Dissertation Korrektur zu lesen und das mitunter auch mitten in der Nacht. Neben der fachlichen Unterstützung möchte ich Dir auch für die gemeinschaftlichen sportlichen Höchstleistungen danken, die wir in den letzten Jahren vielfach erbracht haben.

Der Deutschen Forschungsgemeinschaft (DFG) danke ich für die Finanzierung des Großteils meiner Promotionsphase.

Prof. Richard Immink möchte ich für seine Bereitschaft danken, meine Dissertation als externer Gutachter zu bewerten.

Darüber hinaus möchte ich mich bei Rainer, Susanne, Christian, Andrea, Teresa, Lydia, Dajana, Nina, Lisa, Katharina, Benni, Khushboo, Sandra, Corni, Sabine, Ulrike, Heidi und allen anderen aktuellen und ehemaligen Mitarbeitern am Lehrstuhl für Genetik für das sehr angenehme Arbeitsklima und die hervorragende Zusammenarbeit bedanken. Insbesondere Rainer danke ich zusätzlich für die ausgezeichnete inoffizielle Zweitbetreuung meiner Promotion, die super Zusammenarbeit und dafür, dass ich Dich ständig mit Fragen löchern durfte, selbst nachdem Du nach Irland ausgewandert bist.

Bei meinen Freunden aus Erfurt und dem Rest der Welt, insbesondere bei Kai, Jule, Smart, Dominique, Stefan, Walli, Johannes, Felix, Eva, Renè, Christin, Lars, Anja, Michi, Yvonne, Chrissi, Robin, Carsten, Sascha, Martin, Katja, Valle, Robert, Faff, Dannern und Bennern bedanke ich mich für ihr Verständnis dafür, dass ich mich insbesondere in der letzten Phase meiner Promotion sehr rar gemacht habe. Wir holen das alles nach!

Außerdem möchte ich mich bei meinen lieben Eltern Liane und Michael, meinem Bruder Hannes und seiner Frau Carolin bedanken. Ihr habt mir während der Promotion immer Mut zugesprochen und mir das Gefühl gegeben, dass sie der Mühe wert ist. Bei meiner kleinen Nichte Lotta möchte ich mich dafür bedanken, dass sie mich, trotz meiner viel zu seltenen Besuche, immer wieder freudig empfängt. Jetzt, nach der Abgabe dieser Arbeit, werde ich mich häufiger sehen lassen, versprochen!

Ich möchte auch meiner Oma Elfriede dafür danken, dass wir uns immer über die neusten Erkenntnisse in der Spitzenforschung austauschen konnten und dafür, dass sie mich in den letzten Jahren immer wieder gefragt hat, ob ich nicht langsam mit studieren fertig werden will und endlich anfange zu arbeiten. Deine Beharrlichkeit hat sicherlich dafür gesorgt, dass ich mir nicht noch mehr Zeit gelassen habe.

Zu guter Letzt möchte ich meiner süßen Jessi danken, dafür, dass sie sich vor inzwischen fast drei Jahren dafür entschieden hat, ihre Masterarbeit am Lehrstuhl für Genetik zu machen, sodass wir uns dort begegnen und uns kennenlernen konnten. Jessi, Du hast mir bei der Anfertigung dieser Arbeit und auch bei allem anderen unglaublich geholfen. Mit Dir macht einfach alles Spaß! Selbst die zermürbenden letzten zwei Tage vor der Abgabe, die eigentlich ein sehr langer Tag waren, haben sich mit Dir angefühlt, als würden wir eine Party feiern und einfach durchmachen. Das lag sicherlich nicht zuletzt daran, dass zu viel Koffein bei Dir wirkt wie Alkohol bei normalen Menschen.

7 Declaration of authorship (Ehrenwörtliche Erklärung)

Hiermit erkläre ich ehrenwörtlich, dass mir die geltende Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena bekannt ist und ich die vorliegende Arbeit selbst angefertigt habe. Alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen habe ich in meiner Arbeit angegeben. Bei der Auswertung des Materials sowie beim Verfassen dieser Dissertation haben mich die in der Danksagung dieser Arbeit genannten Personen unterstützt.

Ferner erkläre ich ehrenwörtlich, für die Anfertigung der Arbeit keinen Promotionsberater in Anspruch genommen zu haben und dass Dritte weder mittelbar noch unmittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die vorliegende Dissertation habe ich bisher nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung vorgelegt. Auch habe ich weder diese Dissertation noch eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung bei einer anderen Hochschule als Dissertation eingereicht.

Florian Rümpler

Jena, 29.06.2017

9 Publications and conference contributions

Publications

Melzer, R., Härter, A., Rümpler, F., Kim, S., Soltis, P. S., Soltis, D. E., Theißen, G. (2014). DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. *Ann. Bot.* 114, 1431-1443.

Rümpler, F., Theißen, G., Melzer, R. (2015). Character-state reconstruction to infer ancestral protein-protein interaction patterns. *Bio-protocol* 5, published online.

Rümpler, F., Gramzow, L., Theißen, G., Melzer, R. (2015). Did convergent protein evolution enable phytoplasmas to generate 'zombie plants'? *Trends Plant Sci.* 20, 798-806.

Theißen, G., Melzer, R., Rümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development* 143, 3259-3271.

Lenser, T., Graeber, K., Cevik, Ö.S., Adigüzel, N., Dönmez, A.A., Grosche, C., Kettermann, M., Mayland-Quellhorst, S., Mérai, Z., Mohammadin, S., Nguyen, T.P., Rümpler, F., Schulze, C., Sperber, K., Steinbrecher, T., Wiegand, N., Strnad, M., Scheid, O.M., Rensing, S.A., Schranz, M.E., Theißen, G., Mummenhoff, K., Leubner-Metzger, G. (2016). Developmental control and plasticity of fruit and seed dimorphism in *Aethionema arabicum*. *Plant Physiol.* 173, 1691-1707.

Rümpler, F., Theißen, G., Melzer, R. (2017). Sequence features of MADS-domain proteins that act as hubs in the protein-protein interaction network controlling flower development. *bioRxiv*,125294.

Conference contributions

Rümpler, F., Melzer, R., Theißen, G. (June 2012). Defining regions of SEP3 important for cooperative DNA-binding. Poster presented at Plant Science Student Conference (PSSC) 2012, Gatersleben, Germany.

Rümpler, F., Melzer, R., Theißen, G. (September 2012). Evidences that the K-domain of the floral homeotic protein SEP3 does not fold into a 'standard' coiled-coil structure, but a leucine zipper. Poster presented at Gesellschaft für Entwicklungsbiologie (GfE) School 2012, Günzburg, Germany.

Rümpler, F., Theißen, G., Melzer, R. (June 2013). What makes the glue sticky? - Defining regions of SEP3 important for cooperative DNA-binding. Poster presented at Workshop on Molecular Mechanisms Controlling Flower Development 2013, Presqu'lie de Giens, France.

Rümpler, F., Gramzow, L., Theißen, G., Melzer, R. (July 2015). Convergent evolution enabled phytoplasmas to generate 'zombie plants'. Poster presented at Society for Molecular Biology and Evolution (SMBE) Meeting 2015, Vienna, Austria.