The Company of **Biologists**

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

[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

(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.

DEVELOPMENT

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
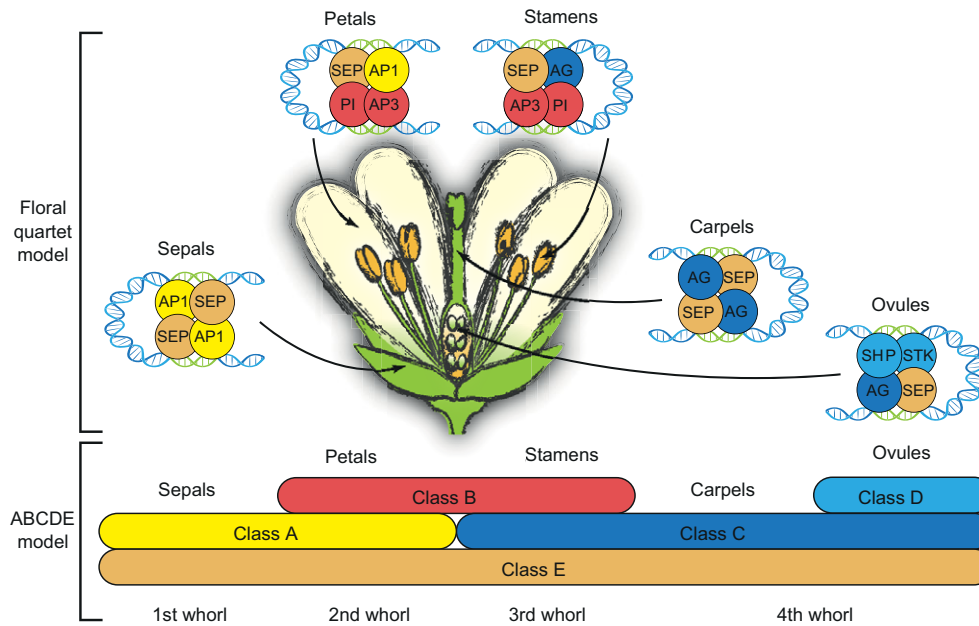
Fig. 1. The floral quartet model and the underlying ABCDE model of organ identity determination in *Arabidopsis thaliana*. The top part of the figure depicts a version of the floral quartet model, which maintains that the five floral organ identities (sepals, petals, stamens, carpels and ovules) are 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

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

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
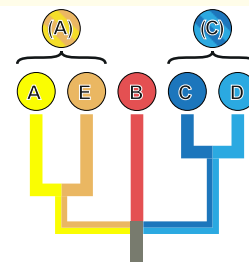
The highly simplified phylogenetic tree depicts the relationships between floral homeotic genes, proteins and functions as defined in the ABCDE model (Fig. 1; Gramzow and Theißen, 2010). While the deep branching of the tree is still largely unknown (indicated by the basal trifurcation) there is strong 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.
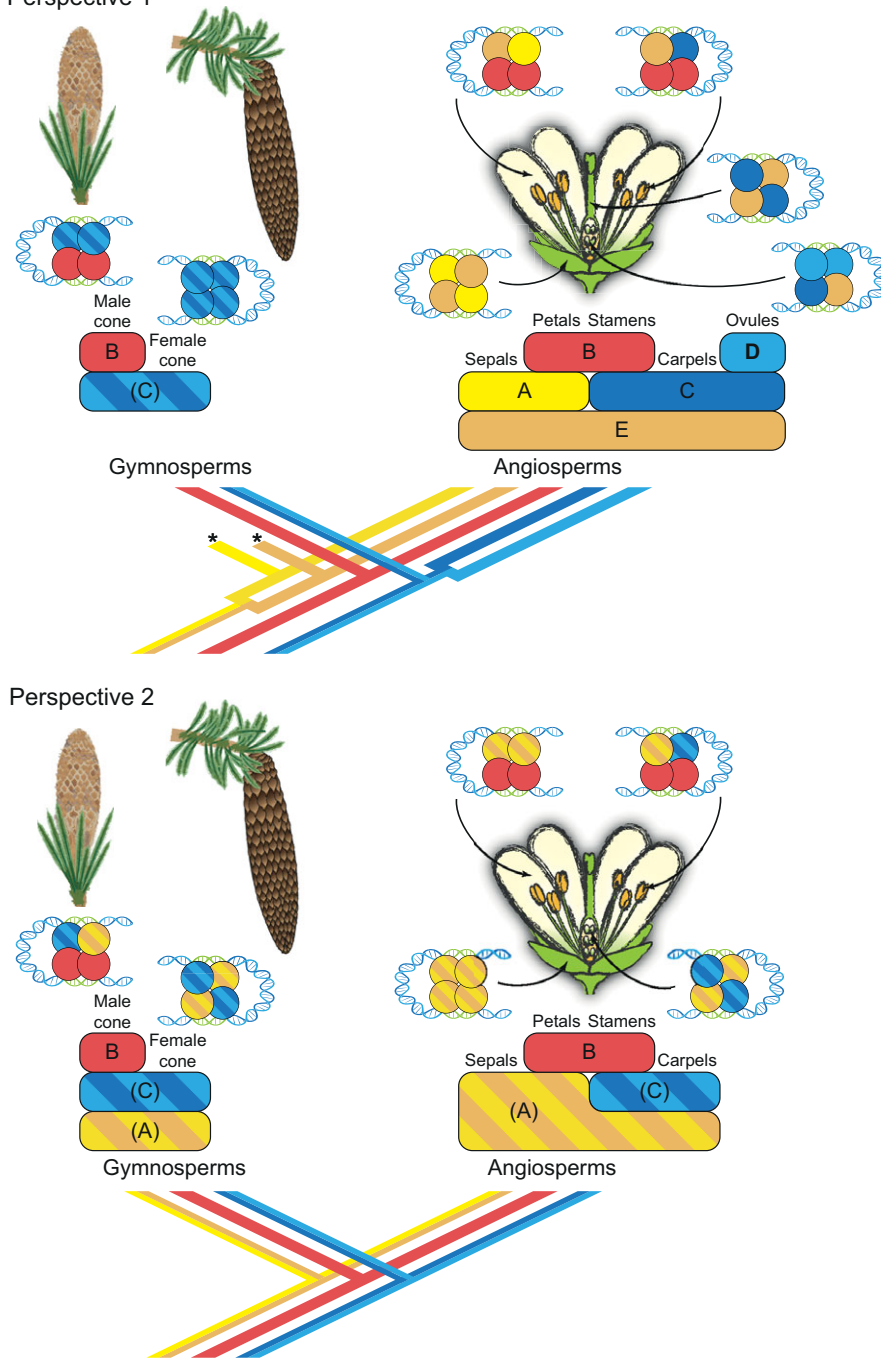
Fig. 2. Perspectives on floral homeotic functions in gymnosperms and angiosperms. Two perspectives (1 and 2) of how floral homeotic genes might function in gymnosperms and angiosperms are shown. The trees at the bottom of each panel illustrate proposed phylogenetic relationships between the five floral 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

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
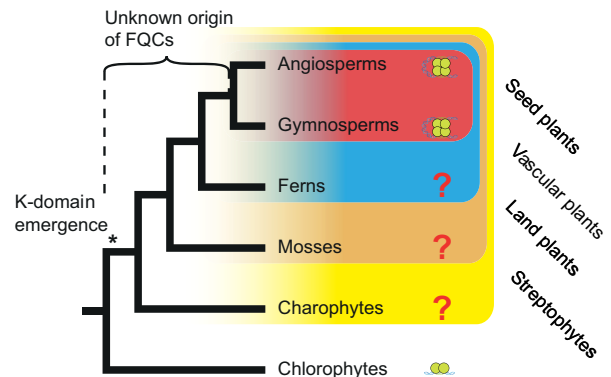

Fig. 3. The emergence of FQC formation during green plant evolution. The tree illustrates a simplified green plant phylogeny. Within the green plant phylogeny, seed plants (angiosperms and gymnosperms) together with ferns represent the clade of vascular plants; vascular plants and mosses represent the 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.

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

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.

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.

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

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.
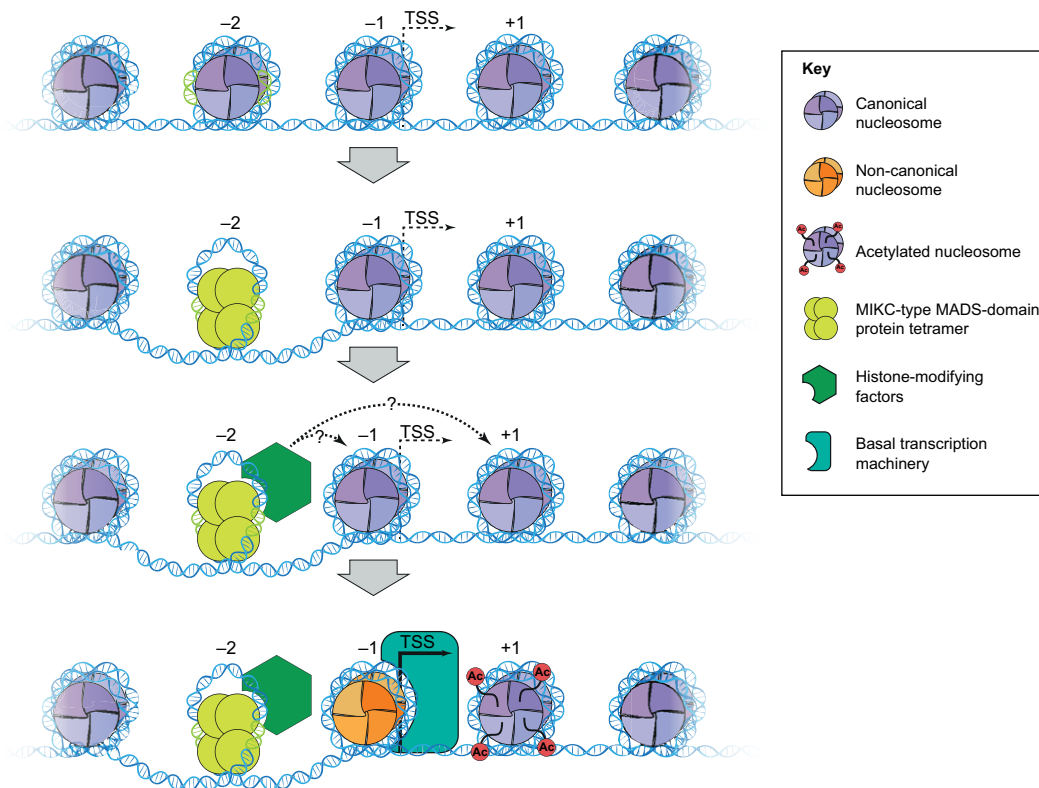
Our model is based on similarities between FQCs on the one hand, and (half-) nucleosomes and the transcription factor NF-Y, which mimics H2A/H2B-DNA nucleosome assembly (Nardini et al., 2013), on the other hand. Both FQCs and half-nucleosomes are composed of tetramers of similar proteins. Moreover, DNA 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.

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.

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.

DEVELOPMENT

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 MIKCᶜ-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.

DEVELOPMENT

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

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.

DEVELOPMENT