Review TRENDS in Plant Science Vol.12 No.8

The floral genome: an evolutionary history of gene duplication and shifting patterns of gene expression

Douglas E. Soltis[1], Hong Ma[2], Michael W. Frohlich[3], Pamela S. Soltis[4], Victor A. Albert[5], David G. Oppenheimer[6], Naomi S. Altman[7], Claude dePamphilis[8] and Jim Leebens-Mack[9]

[1] Department of Botany and the Genetics Institute, University of Florida, Gainesville, FL 32611, USA
[2] Department of Biology, the Huck Institutes of the Life Sciences, and the Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, 201 Life Sciences Building, University Park, PA 16802, USA
[3] Department of Botany, Natural History Museum, London, SW7 5BD, UK
[4] Florida Museum of Natural History and the Genetics Institute, University of Florida, Gainesville, FL 32611, USA
[5] Natural History Museum, University of Oslo, PO Box 1172, Blindern, NO-0318 Oslo, Norway
[6] Department of Botany, UF Genetics Institute and Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, FL 32610, USA
[7] Department of Statistics and the Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA
[8] Department of Biology, the Huck Institutes of the Life Sciences, and the Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, 201 Life Sciences Building, University Park, PA 16802, USA
[9] Department of Plant Biology, University of Georgia, Athens, GA 30602, USA

Through multifaceted genome-scale research involving phylogenomics, targeted gene surveys, and gene expression analyses in diverse basal lineages of angiosperms, our studies provide insights into the most recent common ancestor of all extant flowering plants. MADS-box gene duplications have played an important role in the origin and diversification of angiosperms. Furthermore, early angiosperms possessed a diverse tool kit of floral genes and exhibited developmental 'flexibility', with broader patterns of expression of key floral organ identity genes than are found in eudicots. In particular, homologs of B-function MADS-box genes are more broadly expressed across the floral meristem in basal lineages. These results prompted formulation of the 'fading borders' model, which states that the gradual transitions in floral organ morphology observed in some basal angiosperms (e.g. *Amborella*) result from a gradient in the level of expression of floral organ identity genes across the developing floral meristem.

Origin of the flower – fossil and phylogenetic perspectives

Angiosperms represent one of the greatest terrestrial radiations. The oldest fossils date from the early Cretaceous [1], 130 million years ago (mya)–136 mya, followed by a rise to ecological dominance in many habitats before the end of the Cretaceous [1]. Recent molecular estimates have converged on 140 mya–190 mya, suggesting an even older origin in the early Cretaceous or late Jurassic [2,3].

The closest relatives of angiosperms remain a mystery. Modern gymnosperms form a clade; therefore, no single living gymnosperm lineage is the sister group of the angiosperms [4,5]. In addition, there are no known fossils representing unequivocal stem-group angiosperms (i.e. angiosperms that attach below the basal node of all extant angiosperms). When placed in the larger context of both living and fossil seed plants, angiosperms are nested within the pteridosperms ('seed ferns'), probably close to Caytoniales and Bennettitales [6–8] and perhaps close to Gigantopterids [9].

Molecular data have clarified much of the phylogeny of living angiosperms (e.g. Refs [3,8]), revealing *Amborella* (Amborellaceae), Nymphaeaceae (water lilies), and Austrobaileyales (star-anise and relatives) as sequential sister lineages to a clade including the magnoliids, the monocots and the eudicots. These basalmost lineages have emerged as pivotal for investigating the characteristics of the most recent common ancestor (MRCA) of all extant angiosperms. The Floral Genome Project (FGP) has been using genomic approaches to explore the evolution of floral development and to characterize the ancestral floral transcriptome, focusing on basal angiosperms [*Amborella*, *Nuphar* (spadderdock; Nymphaeaceae), representing the basalmost branches of the tree; *Persea* (avocado; Lauraceae) and *Liriodendron* (tulip poplar; Magnoliaceae) representing the magnoliid clade], a basal eudicot [*Eschscholzia* (California poppy; Papaveraceae)], and a basal monocot [*Acorus* (Acoraceae), sweet flag] (Figure 1a). Comparative analyses of ESTs, finished cDNA sequences, and gene expression patterns assessed through *in situ*, RT–PCR, and microarray experiments reveal both conservation and dynamism in the evolution of the floral genetic program.

Corresponding author: Soltis, D.E. (dsoltis@botany.ufl.edu).
Available online 19 July 2007.

 1360-1385/$ – see front matter doi:10.1016/j.tplants.2007.06.012

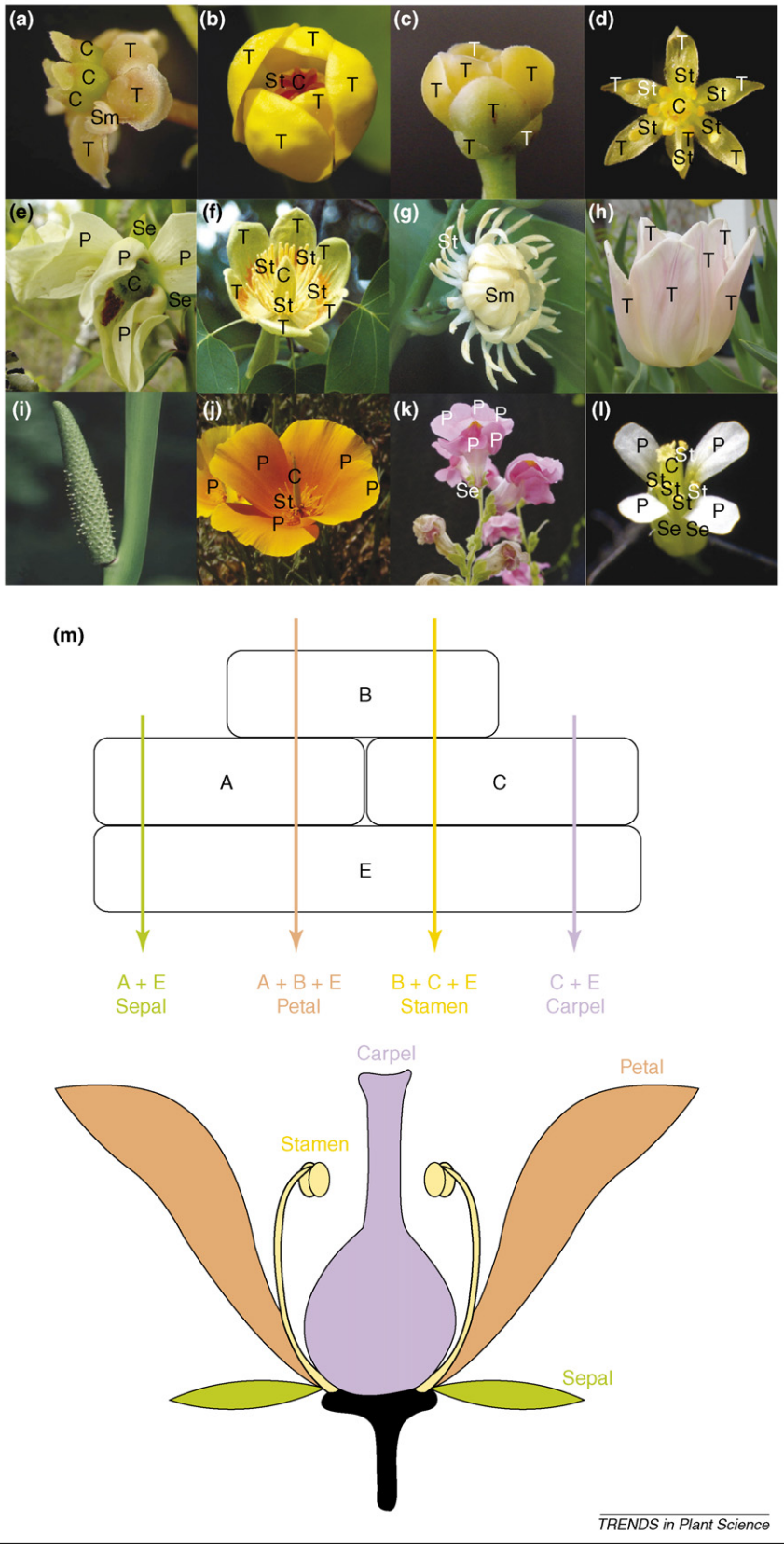

Figure 1. Floral diversity in basal angiosperms. **(a)** *Amborella trichopoda* (Amborellaceae) (photograph by S. Kim). **(b)** *Nuphar pumila* (Nympheaceae) (photograph by S. Kim). **(c)** *Illicium parviflorum* (Illiciaceae) (photograph by S. Kim). **(d)** *Persea americana* (Lauraceae) (photograph by M. Buzgo). **(e)** *Asimina longifolia* (Annonaceae) (photograph by S. Kim). **(f)** *Liriodendron tulipifera* (Magnoliaceae) (photograph by S. Kim). **(g)** *Eupomatia laurina* (Eupomatiaceae) (photograph by H. Teppner). **(h)** Cultivated *Tulipa* (Liliaceae) (photograph by S. Kim). **(i)** *Acorus americanus* (Acoraceae) (photograph by M. Buzgo). **(j)** *Eschscholzia californica* (Papaveraceae) (photograph by C. Chen). **(k)** *Antirrhinum majus* (Plantaginaceae (photograph by S. Kim). **(l)** *Arabidopsis thaliana* (Brassicaceae) (photograph by H. Ma). Abbreviations: C, carpel; P, petal; Se, sepal; Sm, staminode; St, stamen; T, tepal. **(m)** Extended ABCE model. In *Arabidopsis*, *APETALA1* (*AP1*) and *AP2* are the A-function genes, *AP3* and *PISTILLATA* (*PI*) are the B-function genes, and *AGAMOUS* (*AG*) is the C-function gene. In *Antirrhinum*, the homolog of *AP1* is *SQUAMOSA*, homologs of *AP2* are *LIPLESS1* and *LIPLESS2*, homologs of *AP3* and *PI* are *DEFICIENS* (*DEF*) and *GLOBOSA* (*GLO*), respectively, and the homolog of *AG* is *PLENA* (*PLE*). However, *SQUAMOSA* and *LIPLESS* genes do not perform a classic A function [17].

The ABC model and MADS-box genes – sequence conservation and shifting expression

Developmental genetic investigations of the eudicot models *Arabidopsis* and *Antirrhinum* (Figure 1a) have identified >80 genes crucial for normal floral development, including genes involved in flower initiation and organ development [10,11]; however, the number of genes active during flower development is certainly much larger (e.g. Ref. [12]), and mechanisms controlling some important attributes of the flower (such as the number of floral whorls and number of organs in a whorl) remain poorly understood. The best-known genes controlling floral organ identity are the A-, B- and C-function genes [13,14]. As proposed in the original ABC model, the A function alone specifies sepal identity, the A and B functions together control petal identity; the B and C functions together control stamen identity; and the C function alone specifies carpel identity (Figure 1b). The ABC model also specifies that A-function genes repress the expression of C-function genes.

Whereas the ABC model has served as a unifying paradigm for floral developmental research for more than a decade, only the B and C functions tend to be highly conserved, with the A function perhaps important only in some eudicots, such as *Arabidopsis* [14–18]. Except for the *Arabidopsis* A-function gene *APETALA2* (*AP2*) and homologs, all ABC genes are members of the MADS-box gene family and encode putative transcription factors [14]. Additional MADS-box genes control ovule identity in eudicots (the D function) [19] and confer the E function (*SEPALLATA1–SEPALLATA4*, *SEP1–SEP4* [20,21]), which is required for specification of all floral organs in *Arabidopsis*. We will not consider the D function separately here because the canonical D-function gene *SEEDSTICK* is closely related to the C-function gene *AGAMOUS* (*AG*), but D-function genes are not always orthologous [22,23]. However, the E function plays a major role in the formation of floral organs and is closely allied with ABC functions. Hence, the updated model of floral organ identity in well-studied angiosperms is the ABCE model, and we will use that designation (Figures 1b and 2).

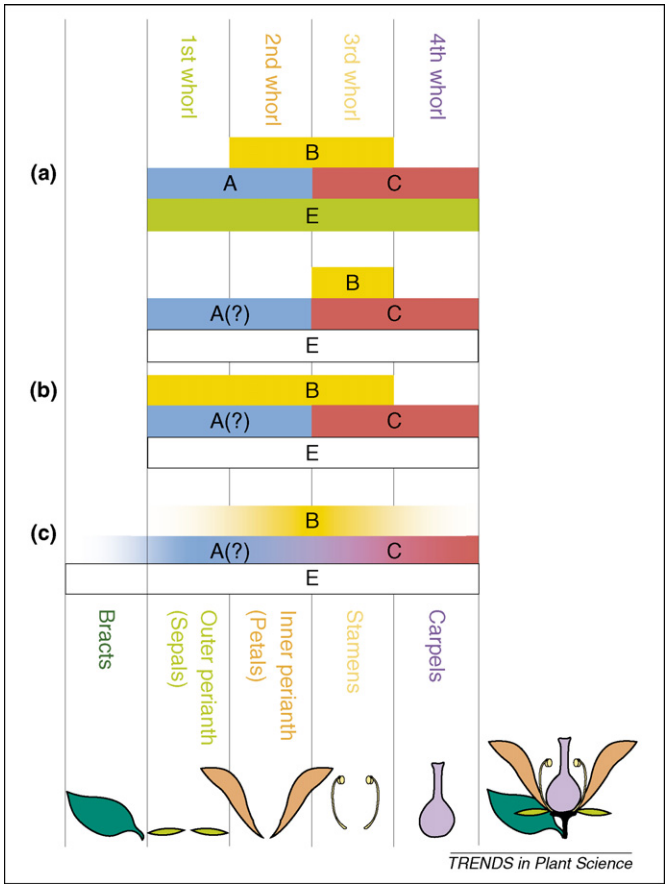

Figure 2. Models of genetic control of floral organ identity. **(a)** Classic ABCE model [13,19,20,70,71]. We are not considering the D function separately here because D-function genes are phylogenetically closely related to the C-function gene *AG* and its close homologs, and because genes controlling the D function are not always orthologous. The fundamental differences in floral morphology of some basal angiosperms – as well as many monocots and some basal eudicots – compared with the core eudicots suggested that there might be underlying differences in the regulation of floral organ identity genes in these lineages [8,30,47]. Some monocots have two outer cycles or whorls of colorful floral organs (tepals) that are not morphologically differentiated into clear sepals and petals; the classic ABC model might not apply to such flowers [72]. **(b)** The 'sliding boundary' model [73] (or 'modified ABC model' [72] or 'shifting boundary' model [30,74]) explains the presence of morphologically identical, petaloid, inner and outer whorls of parts (as in the monocots *Lilium* and *Tulipa* and some basal eudicots, including *Ranunculus*). The shifting or sliding boundary model permits the boundary of B function to slide across the developing flower from its restricted location in *Arabidopsis* to include the outer floral whorl. **(c)** The fading borders model [40,41] attempts to explain the gradual transition between floral parts in some basal angiosperms.

MADS-box gene duplications and evolution

Homologs of most floral MADS-box genes have been identified from basalmost angiosperms and basal eudicots, in addition to core eudicots (Table 1), and show patterns of gene duplication and loss [21–24]. Phylogenetic analyses of MADS-box genes indicate many gene duplications occurred either before or close to the origin of angiosperms [21,23,25,26]. These gene duplications could have arisen simultaneously, via polyploidy, an important force throughout angiosperm history [8]. Genomic analyses have implicated polyploidization before or coincident with the origin of the angiosperms and the divergence of major core eudicot lineages [27]. Polyploidy has also been prevalent in many angiosperm clades (e.g. Ref. [28]), including several basal angiosperm lineages [29], and might have been responsible for the co-occurrence of duplicated MADS-box genes. Importantly, the patterns of duplications in the *APETALA1* (*AP1*), *APETALA3*/*PISTILLATA* (*AP3* and *PI*, respectively), *AG* and *SEP* subfamilies [21–23,25] are similar and roughly correspond to major events in angiosperm phylogeny (e.g. origin of eudicots), suggesting that MADS duplications might have played an important role in the origin and diversification of the angiosperms.

The duplication that resulted in the two B-function gene lineages (*AP3* and *PI*) occurred 230 mya–290 mya [25], ~100 million years before the oldest fossil flowers, shortly after the split between extant gymnosperms and angiosperms and on the 'stem' lineage of extant flowering plants. Although the B function controls in part the identity of petals, no such structures are recognizable in the fossil record during the 100-million-year period after the estimated time of the *AP3*/*PI* duplication, suggesting that the joint expression of *AP3* and *PI* did not immediately result in the formation of petals [25]. *AP3*- and *PI*-homologs also specify stamen identity in angiosperms (in conjunction with *AG* and *SEP* genes), so specification

Table 1. MADS-box gene sequences identified in Floral Genome Project expressed sequence tag sets and via screening using degenerate primers[a]

Subfamilies	Unpl	*AG*	*ST11*	*DEF/ GLO*	*DEF*	*GLO*	*SEP*	*AGL6*	*SQUA*	*TM3*	*TM8*	*MIKC**
Functional group		C and D		B			E		A			
Gymnosperms												
Welwitschia mirabilis	3	–	1	1	–	–	–	1	–	1	–	–
Zamia fischeri	2	–	–	1	–	–	–	3	–	–	–	–
Basalmost angiosperms												
Amborella trichopoda	1	2	1	–	2	1	2	1	1	–	1	–
Nuphar advena	1	1	–	–	2	1	2	1	1	–	–	–
Magnoliids												
Liriodendron tulipifera	–	–	1	–	1	–	1	2	–	–	–	1
Persea americana	–	2	1	–	2	1	2		1	1	1	–
Saruma henryi	–	1	–	–	1	–	2	1	–	–	–	–
Monocots												
Acorus americanus	–	–	1	–	–	1	2	1	–	1	–	–
Asparagus officinalis	–	1	–	–	–	2	2	2	–	1	–	–
Yucca filamentosa	–	–	1	–	–	2	–	–	–	–	–	–
Eudicots												
Eschscholzia californica	–	2	–	–	2	1	2	1	–	–	–	–
Cucumis sativus	–	–	–	–	–	–	–	1	–	–	–	–

Abbreviations: *AG, AGAMOUS*; *AGL, AGAMOUS-like*; *DEF, DEFICIENS*; *GLO, GLOBOSA*; *MIKC**, a type of MIKC gene with unusual I- and K-regions compared with classical MIKC-type genes; *SEP, SEPALLATA*; *SQUA, SQUAMOSA*; *ST11, Solanum tuberosum MADS11*; *TM3, TOMATO MADS3*; *TM8, TOMATO MADS8*; Unpl, identified as MADS-box gene family member but unplaced in known subfamilies (ambiguous subfamily identification because of short sequence or potential presence of new subfamilies); –, indicates no data.

[a]The numbers of genes identified for each clade and taxon are shown.

of microsporophylls (i.e. leaf-like organs bearing microsporangia [30]) might have been the original role of *AP3* and *PI* homologs. In extant gymnosperms, the (single-copy) sister gene of the *AP3* + *PI* clade appears to have that role [31]. Nonetheless, it is possible that co-expression of *AP3*- and *PI*-homologs mediated the evolutionary innovation of animal-attractive, petal-like organs well before the appearance of flowers. The presence of epidermal features on gymnosperm reproductive organs, such as papillose epidermal cells with cuticular ridges, which in angiosperms can contribute to petal color saturation [32] and to the release of fragrances [33], suggests that these fossil gymnosperms were insect-pollinated [34]. The fossil groups Bennettitales and Cheirolepidiaceae, both believed to be insect-pollinated, have papillae on the epidermises of their reproductive structures [35,36].

The C-function gene *AG* is a member of a subfamily that includes three other *Arabidopsis* genes, *SHATTERPROOF1*, *SHATTERPROOF2* (*SHP1*, *SHP2*), and *SEEDSTICK* (*STK*) [22,23]. The duplication events that gave rise to this group have resulted in paralogs that retain similar biochemical activities, with subfunctionalization and/or neofunctionalization perhaps primarily driven by shifts in expression pattern [23]. Expression patterns and gene function can be evolutionarily labile. For example, *AG* and the *Antirrhinum PLE* gene share C-function, but they are paralogs rather than orthologs [37]. As noted, the E function plays a major role in the formation of floral organs, and the corresponding *SEP* subfamily of MADS-box genes has also undergone multiple duplication events, including one before the origin of the extant angiosperms [21]. The *SEP1–SEP3* genes have retained largely overlapping E-function in *Arabidopsis* [20]. The apparent absence of any *SEP* homologs in extant gymnosperms and the function of *SEP* genes in promoting all floral organ identities, even when ectopically expressed in leaves, suggest that the function of *SEP* genes might have played a key role in the origin of the flower [21].

Broader expression in basal angiosperms

Expression patterns of MADS-box genes in eudicots and grasses are often consistent with gene functions as revealed by genetic studies, and support the ABCE model. Strong expression of eudicot *AP3* and *PI* homologs is typically limited to petals and stamens, where these genes are required for organ identity specification [14,15]. The expression of MADS-box genes in basal angiosperm flowers is generally consistent with the ABCE model [38]. However, homologs of the B-function genes *AP3* and *PI* are broadly expressed in tepals, stamens and carpels in many basal angiosperms, including representatives of the three basalmost lineages, *Amborella*, water lilies, and *Illicium* of Austrobaileyales, as well as in members of the magnoliid clade (e.g. *Magnolia*) [38]. Although the ancestral expression of C-function homologs is not inferred to be appreciably broader than that of extant core eudicots, C-function homologs are expressed in the perianth whorls (tepals) of the basal angiosperms *Illicium* [38] and *Persea* [39].

Broad and varying expression of the homologs of floral organ identity genes in basal angiosperms suggests more floral developmental variation among lineages as well as lability within individual species. This developmental 'flexibility' correlates well with the variability in floral form observed in basal angiosperms. However, this pattern contrasts with the restricted B-class expression and pronounced floral canalization observed in core eudicots. Flexibility in the expression pattern of floral regulators is

a fundamental component of the fading borders model (see below).

Fading borders: a model for early angiosperms?

The broader expression of B- (and to a lesser extent, C-) function homologs in basal angiosperms indicates that the ABCE model as developed for eudicots is not perfectly applicable to extant basal angiosperms and, by inference, the earliest angiosperms. The floral morphology of many basal angiosperms provides a crucial hint to what might be a more appropriate model for these plants. In some basal angiosperms (e.g. *Amborella* and *Illicium*), floral organs are spirally arranged with a transition from bracts to outer and inner tepals, from tepals to stamens, and finally to carpels (Figure 1a). These gradual intergradations of floral organs cannot be easily explained by the classic ABCE model and, together with floral developmental studies, prompted the development of the 'fading borders' model [40,41]. This model posits that the gradual transitions in floral organ morphology result from a gradient in the level of expression of floral organ identity genes across the developing floral meristem (Figure 2). Weak expression at the margin of the range of 'activity' of a gene overlaps with the expression of another regulator in adjacent cells. This pattern of overlapping expression would result in the formation of morphologically intermediate floral organs rather than organs that are clearly distinct [16,38]. Recent data from the expression of B-function genes in *Amborella* lend support to this model [38].

The existence of gradients of MADS-box gene expression across apical meristems of *Gerbera* (Asteraceae) has recently been described [42], albeit for the capitulum (the condensed inflorescence of composites) rather than for a single flower. These authors postulate that relatively simple threshold models comprising short-range activation, long-range inhibition, and non-linear regulatory feedback loops would be enough to establish a morphogenetic gradient that could account for the observed differential gene expression.

The A-function mystery

Whereas sequence and expression analyses suggest that B, C and E functions might be broadly conserved, the A function, specification of sepals and petals and repression of C-function genes, has not been widely observed. Indeed, although researchers often refer to the ABC model or recent variants (the ABCE or ABCDE model), the A function has so far been demonstrated only in *Arabidopsis* [17,18]. Brendan Davies *et al.* [17] argued that the A function as specified in the ABC model [13] is not even applicable to *Antirrhinum* or indeed the vast majority of flowering plants. Indeed, the original model for floral organ determination in *Antirrhinum* involved only B and C functions [43]. Although homologs of the A-function gene *AP2* are apparently expressed in some basal angiosperms (e.g. *Amborella*), expression of *AP1* homologs in *Nuphar* and *Magnolia* is higher in leaves and carpels than in perianth [38]. However, it is important to distinguish A function from expression of *AP1* and *AP2* homologs because the 'A function' might be conferred from genes other than homologs of the A-function genes of *Arabidopsis* [16], and different genes might be responsible for the specification of the perianth and the repression of C-function genes. Hence, at this point, whether an A function is important in basal angiosperms remains a major question.

Origin of the flower

Many theories have been proposed for the origin of the flower (reviewed in Refs [5,8]), but only recently have models been based on known gene regulatory networks [44–47]. The Mostly Male theory of flower origins was the first to be based on evidence from genes, in particular, homologs of the *Arabidopsis* gene *LEAFY* (*LFY*), as well as on mutant phenotypes and the morphologies of extant and fossil plants [5,34,44]. Alternatives to the Mostly Male theory include models that involve pleiotropy [45]. Other authors have proposed 'homeotic' models [46,47]. David Baum and Lena Hileman [47] proposed three steps in flower origin and evolution: the origin of bisexuality, floral determinacy with axis compression, and the origin of petals. Importantly, recent gene expression studies of the two *LFY* paralogs in gymnosperms [48] do not support the Mostly Male theory, although the morphological evidence remains. More comprehensive analyses of *LFY*-like gene expression from gymnosperms must be incorporated into any future model that includes modifications of such genes as central to the evolution of the flower.

The ancestral floral genetic program

Basal angiosperms already had a diverse array of floral regulators, comparable to that observed in derived eudicot models. The prevalence of the broader pattern of expression of B-, and to a lesser extent C-function genes in basal angiosperms suggests that broader expression characterized the earliest (ancestral) flowering plants, and phylogenetic reconstructions support this hypothesis [38]. Although there is evidence for weak expression of *AP3* and *PI* homologs in tissues other than stamens and petals in core eudicots during some stages of development, in basal angiosperms, *AP3* and *PI* homolog expression is both considerable and broad. Canalization of MADS-box gene expression in accord with the ABCE-model (as described for *Arabidopsis* and *Antirrhinum* [13,43]) originated later. That is, the more restricted pattern of floral gene expression in these eudicots is derived (Figure 3) [16,38].

The floral transcriptome

A rich tool kit

As data have emerged from major EST projects, it has become possible to make broad comparisons of some of the numerous genes and gene families that are involved in normal floral development [11,12]. Particularly useful have been ESTs generated for a suite of basal angiosperms (http://www.floralgenome.org) [49]. Importantly, many genes identified in rice and *Arabidopsis* have clear homologs in basal angiosperms [49]. The data indicate that early angiosperms already possessed a diverse assemblage of floral genes.

Focused studies of gene families involved in flowering in model organisms (e.g. *Arabidopsis*) further illustrate the rich tool kit of floral genes already present in early angiosperms, as well as contrasting modes of gene family

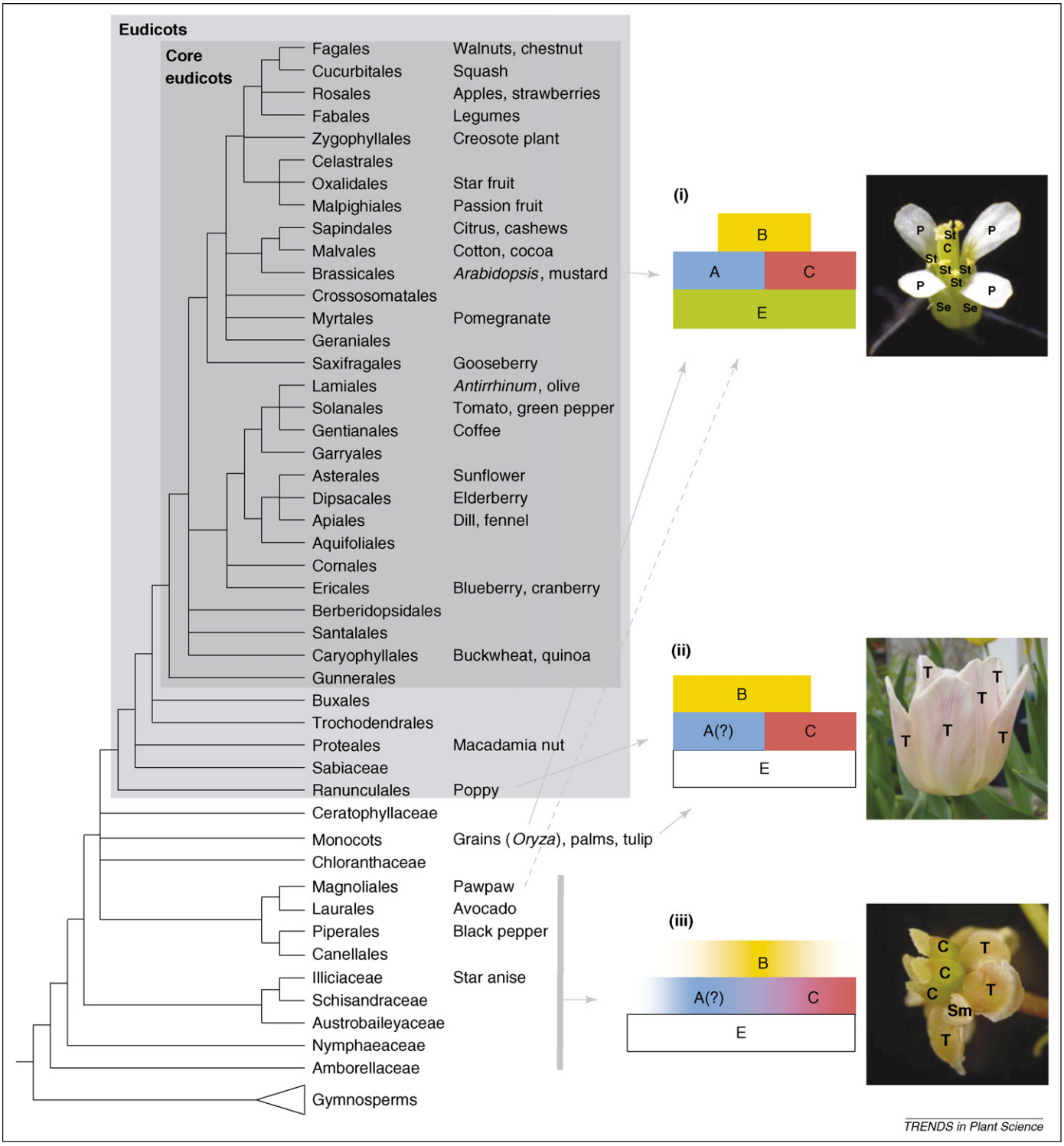

Figure 3. Evolution of expression patterns of floral MADS-box genes in angiosperms (modified from Ref. [38]). Summary tree for flowering plants with placement of model organisms and illustrations of floral diversity. Known or postulated expression patterns are shown on the right for organ identity genes: (i) ABC model developed for core eudicots [13] and some monocots; (ii) an example of shifting boundary model applied for some basal eudicots [75] and monocots [76]; (iii) fading borders model proposed for the basalmost angiosperms (Amborellaceae, Nymphaeaceae, and Austrobaileyaceae, Schisandraceae and Illiciaceae), as well as other basal angiosperms (e.g. some magnoliids [40,41]). Broken arrow indicates that a scheme similar to the classic ABC model might apply to at least one basal angiosperm (*Asimina*, pawpaw). '?' indicates uncertainty regarding A function. The eudicot clade is highlighted in gray, with core eudicots in dark gray. Basal angiosperms are a non-monophyletic group made up of all lineages outside of the eudicots; monocots are sometimes considered basal angiosperms based on their origin among other early lineages of flowering plants [8]. Likewise, basal eudicots are those lineages of eudicots outside of the core eudicots (highlighted in light gray). Abbreviations: C, carpel; P, petal; Se, sepal; Sm, staminode; St, stamen; T, tepal.

evolution. For example, *SHAGGY-like* kinase genes diversified into four well-marked clades early on during angiosperm evolution, yielding the four subgroups reported for *Arabidopsis* [50]. In contrast to *SHAGGY-like* kinase genes (as well as MADS-box genes), the 19 and 22 *SKP1* homologs that have been identified in *Arabidopsis* and rice, respectively, are products of lineage-specific duplications since the divergence of the monocots and eudicots [51].

Genome-wide analyses of *Arabidopsis* and rice *SKP1*-like homologs indicate that retroposition is an important mechanism for the expansion of the plant *SKP1* gene family [52], which includes members that are important for normal floral organ development [11].

MicroRNAs (miRNAs) are short, noncoding RNAs present in eukaryotes that regulate the expression of other genes [53]. One miRNA (*miR172* – known to regulate *AP2/ERF*-genes in *Arabidopsis*) causes early flowering; when overexpressed it disrupts the specification of floral organ identity in *Arabidopsis* [54]. Analyses of the *AP2*-like gene family showed that the binding site of *miR172* in angiosperms is restricted to members of one lineage and is not present in all *AP2*-like genes. Furthermore, Sangtae Kim *et al.* [55] speculated that this mechanism of gene regulation by *miR172* is an ancient one, having originated before the divergence of extant gymnosperms and angiosperms ~290 mya–310 mya. Recent experiments involving transgenic *Nicotiana benthamiana* suggest similarities in the regulation of *AP2* homologs by *miR172* in rosid and asterid eudicots [56].

Other floral genes

ABCE-function transcription factors influence floral form through regulation of target genes, including the antagonistic interactions specified in the ABC model [13], but much remains to be learned about ABCE gene targets in model systems and about how the targets of organ identity genes vary across species. Microarray analyses of gene expression in *Arabidopsis* mutants with compromised A, B and C functions have implicated possible targets [57,58], and a similar approach has been used to investigate upstream regulators of flower development [59]. Homologs have been identified in basal angiosperms (through EST sequencing and limited targeted gene surveys) for many of the genes found to be up- or down-regulated during floral development in these *Arabidopsis* mutants [49], and experiments are underway to determine whether gene expression patterns are consistent with conserved function throughout angiosperm history.

Gene discovery

A genomics approach has offered the promise of discovery of new floral developmental regulators [60,61] through comparisons of conserved, florally expressed genes in selected exemplar species. By focusing on those *Arabidopsis* and rice single-copy genes that were also found in floral EST libraries from basal angiosperms, the *Arabidopsis* T-DNA knockout resource could be used for functional analysis of the FGP-isolated cDNA in question. One gene discovered by this method was identified as a new floral developmental regulator based on the striking alteration in floral organ number shown by plants homozygous for the T-DNA knockout allele of its *Arabidopsis* homolog (P. Zheng and D.G. Oppenheimer, unpublished). This new factor was identified as a subunit of the Mediator complex that 'mediates' interactions between transcriptional regulators and RNA polymerase II during transcriptional activation. Like the transcriptional initiation complex, Mediator was generally believed to have a role in the control of all genes transcribed by RNA polymerase II. Loss of a Mediator subunit, therefore, would be predicted to result in overall transcriptional activation problems, resulting in a highly pleiotropic phenotype. The appearance of specific floral developmental defects in this mutant was unexpected, particularly given two decades of *Arabidopsis* forward genetic screens, and demonstrates that broad comparisons across angiosperms can highlight novel genes with strong mutant phenotypes, even in well-studied models. Other shared single-copy genes are being similarly investigated (e.g. Ref. [62]).

Microarray experiments

To determine the tissues or cells in which the floral genes of basal angiosperms are expressed, species-specific microarray and other approaches have been used to profile gene expression patterns. Custom microarrays containing *in-situ* synthesized 60-mer gene probes have been developed based on the ESTs generated for FGP species, and microarray experiments have so far been performed for the magnoliid *Persea* (A. Chanderbali *et al.*, unpublished) and the basal eudicot *Eschscholzia* (L. Zahn *et al.*, unpublished). RT–PCR, *in situ* hybridization, and microarray investigations coupled with developmental data suggested that the 'petals' of *Persea* (Figure 1a) and other Lauraceae are of staminal origin (A. Chanderbali *et al.*, unpublished) [39]. Furthermore, dozens of floral genes were most highly expressed in young fruit, including *MYB*, *YABBY* and *WRKY* transcription factors. Homologs of the MADS-box transcription factors *AG*, *AP3* and *SEP3* were also highly expressed in young fruit, complementing genetic evidence from tomato [63] that floral developmental regulators also play a role in fleshy fruit development.

Eschscholzia (Ranunculales; Figure 1a) represents the basalmost lineage of eudicots and is well suited as a comparative model for eudicot radiation and diversification [23]. Of >6000 unigenes examined, ~83% were differentially expressed, suggesting that most of the genes might have differential function in one or more of the tissues analyzed. Several genes are differentially expressed between one or more of the floral organs and leaves, including some putative homologs of known floral genes and new candidate regulator genes. The floral ESTs even include genes that are preferentially expressed in the leaf, such as genes encoding chloroplast and ribosomal proteins. Expression data were obtained for a large number of putative regulatory genes. In addition to MADS-box genes, genes encoding other transcription factors, such as HD-Zip, bZip, bHLH and MYB proteins, exhibit differential expression among organs. Future comparative analyses will assess the evolution of expression pattern in these gene families by placing these data for *Eschscholzia* onto gene trees.

Floral genes, floral structures and homology

In contrast to most other basal angiosperms, which have flowers with an undifferentiated perianth of morphologically similar tepals, *Asimina* (pawpaw) and other Annonaceae instead possess a bipartite perianth of distinct sepals and petals (Figure 1a) comparable to that observed in eudicots. This perianth differentiation is accompanied by a restriction of expression of B-function homologs to the

petals and stamens, rather than broad expression across perianth organs; meanwhile, C-class genes are expressed in stamens and carpels, but not in the perianth whorls [38]. This pattern of expression is identical to that of *Arabidopsis* and *Antirrhinum* and indicates that *Asimina* sepals and petals appear comparable to those of eudicots in terms of developmental genetic criteria. However, these sepals and petals represent an independent evolutionary origin from that of eudicots, indicating that on at least two occasions, modifications to the ancestral expression pattern of floral regulatory genes resulted in the 'classic' ABC model, with discrete zones of gene activity, and the formation of a bipartite perianth. There have also been multiple transitions to bipartite perianth structure in the monocot clade (e.g. within the Poaceae [64]).

Future prospects

Detailed functional genetic studies of a few key model eudicots, chiefly *Arabidopsis* and *Antirrhinum*, have provided enormous insights into the genetic control of flower development. However, with only a few exceptions, basal angiosperms are not yet amenable to true functional studies; most are woody, with long generation times and/or large size, making most species impractical as genetic models. However, herbaceous basal angiosperms that are transformable and have rapid life cycles, such as *Aristolochia* [Dutchman's pipe (Aristolochiaceae), B. Bliss, H. Ma, S. Maximova and C. dePamphilis, unpublished], might have great potential as new functional models. In addition, *Persea americana* (avocado) has promise as a functional genetics model for woody magnoliids. It can be transformed, and by grafting transformants onto older shoots, flowering and phenotype analysis can be achieved in less than two years [65]. Virus-induced gene silencing (VIGS) also allows functional analysis, by permitting targeted gene down-regulation in diverse plants, such as the basal eudicots *Papaver* [66] and *Aquilegia* [67].

Greatly improved high-throughput methods enable the complete transcriptome of a flower to be fully sampled within a reasonable timeframe and cost, permitting the discovery of all active genes and facilitating inferences based on what is, and what is not, expressed. Phylogenetically based comparisons among the promoters of functionally equivalent genes will identify islands of conserved DNA sequence that are likely to correspond to conserved binding sites for transcription factors. DNA sequences within these conserved islands will suggest which classes of transcription factors regulate particular genes. This 'phylogenetic footprinting' enables direct comparison of regulatory inputs to gene expression. Other methods, such as surface plasmon resonance, are being applied to measure quantitatively the protein–DNA as well as protein–protein interactions. These studies are essential for modeling transcription factor networks. Parallel technological advances should facilitate study of other components of gene regulatory systems, such as microRNAs and chromatin structure. Such approaches will greatly strengthen both forward and reverse genetic studies of diverse plants.

Evolutionary history leaves its mark in all aspects of life, including the genetic systems that control development. An evolutionary perspective has already enhanced understanding of well-studied models, as shown by the pioneering phylogenetic footprinting study of the *AG* promoter in Brassicaceae [68]. Organisms often have redundant genetic control systems to mediate important aspects of development that result in developmental homeostasis. The relative importance of these systems is likely to differ in different plants. Systems of minor importance in standard models can be of central importance in other plants; the discovery of such systems in non-model organisms might lead to their identification in standard models. An excellent example is the discovery of plasma membrane receptors for steroid hormones in *Arabidopsis* before the discovery of plasma membrane steroid hormone receptors in animals.

Much current research focuses on commonalities in developmental genetic systems across angiosperms. As shared systems become better understood, efforts will shift to developmental systems that undergo changes, resulting in evolutionary novelty. The most important will be developmental genetic changes that underlie instances of parallel and convergent evolution. Parallel evolution is the analog of the replicated experiments fundamental to the physical sciences. Such studies will show the range of evolutionary mechanisms that can bring about specific morphological and physiological changes. This framework will also provide genetic engineers with multiple examples of how to manipulate the morphology and physiology of plants for practical benefit. Parallel evolution in features of biological importance is far more common in plants than in animals; hence, angiosperms are the organisms of choice for such studies [69].

With funding from the NSF Plant Genome Comparative Sequencing Program, a physical map will be developed over the next few years for *Amborella* as part of the Ancestral Angiosperm Genome Project (http://AncestralAngiosperm.org). Even coarse knowledge of gene order over large blocks of the *Amborella* genome should provide the data necessary to discern whether MADS-box gene duplications in ancient angiosperms were associated with a genome-wide duplication [29]. Furthermore, with the publication of the *Populus* genome and active genome sequencing projects for other well-placed eudicots (e.g. *Carica*, *Medicago*, *Solanum*, *Mimulus* and *Aquilegia*), great opportunities exist for genome-enabled comparative research on floral development and many other aspects of angiosperm evolution.

Acknowledgements

The Floral Genome Project was supported by a grant from the National Science Foundation (PGR-0115684). The success of this project was due in large part to the dedication of postdoctoral researchers and students: Sam Brockington, Matyas Buzgo, Liying Cui, Andre Chanderbali, Sangtae Kim, Jin Koh, Kerr Wall, Mi-Joong Yoo, Wei Hu, Jenny Arrington and Hongzhi Kong. We also thank the technicians: Lena Landherr, Yi Hu and Donglan Tian. We thank four anonymous reviewers for their helpful comments and Sangtae Kim for his help with several figures. We also acknowledge support from grant 154145 from the Research Council of Norway (V.A.A.) and DBI-0638595 for research related to the Floral Genome Project.

References

1 Friis, E.M. *et al.* (2006) Cretaceous angiosperm flowers: innovation and evolution in plant reproduction. *Paleogeogr., Paleoclimatol. Paleoecol.* 232, 251–293

2 Bell, C.D. *et al.* (2005) The age of the angiosperms: a molecular timescale without a clock. *Evolution Int. J. Org. Evolution* 59, 1245–1258
3 Leebens-Mack, J. *et al.* (2005) Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Mol. Biol. Evol.* 22, 1948–1963
4 Burleigh, J.G. and Mathews, S. (2004) Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. *Am. J. Bot.* 91, 1599–1613
5 Frohlich, M.W. (2006) Recent developments regarding the evolutionary origin of flowers. In *Developmental Genetics of the Flower (Advances in Botanical Research)* (Vol. 44) (Soltis, D.E. *et al.*, eds), pp. 64–116, Elsevier
6 Doyle, J.A. (2006) Seed ferns and the origin of angiosperms. *J. Torrey Bot. Soc.* 133, 169–209
7 Hilton, J. and Bateman, R.M. (2006) Pteridosperms are the backbone of seed-plant phylogeny. *J. Torrey Bot. Soc.* 133, 119–168
8 Soltis, D.E. *et al.* (2005) *Phylogeny and Evolution of Angiosperms,* Sinauer Associates
9 Taylor, D.W. *et al.* (2006) Biochemical evidence for the presence of the angiosperm molecular fossil oleanane in Paleozoic and Mesozoic non-angiospermous fossils. *Paleobiology* 32, 179–190
10 Zhao, D. *et al.* (2001) Genetic control of reproductive meristems. In *Meristematic Tissues in Plant Growth and Development* (McManus, M.T. and Veit, B., eds), pp. 89–142, Sheffield Academic Press
11 Zahn, L.M. *et al.* (2006) Beyond the ABC model: regulation of floral homeotic genes. In *Developmental Genetics of the Flower (Advances in Botanical Research)* (Vol. 44) (Soltis, D.E. *et al.*, eds), pp. 164–196, Elsevier
12 Zhang, X. *et al.* (2005) Genome-wide expression profiling and identification of gene activities during early flower development in *Arabidopsis*. *Plant Mol. Biol.* 58, 401–419
13 Coen, E.S. and Meyerowitz, E.M. (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37
14 Ma, H. (1994) The unfolding drama of flower development: recent results from genetic and molecular analyses. *Genes Dev.* 8, 745–756
15 Ma, H. and dePamphilis, C.W. (2000) The ABCs of floral evolution. *Cell* 101, 5–8
16 Soltis, P.S. *et al.* (2006) Expression of floral regulators in basal angiosperms and the origin and evolution of the ABC model. In *Developmental Genetics of the Flower (Advances in Botanical Research)* (Vol. 44) (Soltis, D.E. *et al.*, eds), pp. 484–502, Elsevier
17 Davies, B. *et al.* (2006) Flower development: the *Antirrhinum* perspective. In *Developmental Genetics of the Flower (Advances in Botanical Research)* (Vol. 44) (Soltis, D.E. *et al.*, eds), pp. 280–3315, Elsevier
18 Litt, A. (2007) An evaluation of A-function: evidence from the *APETALA*1and *APETALA2* gene lineages. *Int. J. Plant Sci.* 168, 73–91
19 Colombo, L. *et al.* (1995) The petunia MADS box gene *FBP11* determines ovule identity. *Plant Cell* 7, 1859–1868
20 Pelaz, S. *et al.* (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* 405, 200–203
21 Zahn, L.M. *et al.* (2005) The evolution of the *SEPALLATA* subfamily of MADS-box genes: a pre-angiosperm origin with multiple duplications throughout angiosperm history. *Genetics* 169, 2209–2223
22 Kramer, E.M. *et al.* (2004) Patterns of gene duplication and functional evolution during the diversification of the *AGAMOUS* subfamily of MADS box genes in angiosperms. *Genetics* 166, 1011–1023
23 Zahn, L.M. *et al.* (2006) Conservation and divergence in the *AGAMOUS* subfamily of MADS-box genes: evidence for independent sub- and neofunctionalization events. *Evol. Dev.* 8, 30–45
24 Kramer, E.M. and Zimmer, E.A. (2006) Gene duplication and floral developmental genetics of basal eudicots. In *Developmental Genetics of the Flower (Advances in Botanical Research)* (Vol. 44) (Soltis, D.E. *et al.*, eds), pp. 354–376, Elsevier
25 Kim, S. *et al.* (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
26 Irish, V. (2006) Duplication, diversification, and comparative genetics of angiosperm MADS-box genes. In *Developmental Genetics of the Flower (Advances in Botanical Research)* (Vol. 44) (Soltis, D.E. *et al.*, eds), pp. 130–151, Elsevier
27 Bowers, J.E. *et al.* (2003) Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422, 433–438
28 Blanc, G. and Wolfe, K.H. (2004) Widespread paleopolyploidy in model plant species inferred from age distribution of duplicate genes. *Plant Cell* 16, 1667–1678
29 Cui, L. *et al.* (2006) Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16, 738–749
30 Albert, V.A. *et al.* (1998) Ontogenetic systematics, molecular developmental genetics, and the angiosperm petal. In *Molecular Systematics of Plants, II* (Soltis, D.E. *et al.*, eds), pp. 349–374, Kluwer
31 Kaufmann, K. *et al.* (2005) MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347, 183–198
32 Kay, Q.O.N. *et al.* (1981) Pigment distribution, light reflection and cell structure in petals. *Bot. J. Linn. Soc.* 83, 57–84
33 Vogel, S. (1990) *The Role of Scent Glands in Pollination,* Smithsonian Institution
34 Frohlich, M.W. (2001) A detailed scenario and possible tests of the Mostly Male theory of flower evolutionary origins. In *Beyond Heterochrony: The Evolution of Development* (Zelditch, M.L., ed.), pp. 59–104, Wiley-Liss
35 Axsmith, B.J. *et al.* (2004) Conifer pollen cones from the Cretaceous of Arkansas: implications for diversity and reproduction in the Cheirolepidiaceae. *J. Paleontol.* 78, 402–409
36 Watson, J. and Sincock, C.A. (1992) *Bennettitales of the English Wealden,* The Palaeontogaphical Society
37 Causier, B. *et al.* (2005) Evolution in action: following function in duplicated floral homeotic genes. *Curr. Biol.* 15, 1508–1512
38 Kim, S. *et al.* (2005) Expression of floral MADS-box genes in basal angiosperms: implications on evolution of floral regulators and the perianth. *Plant J.* 43, 724–744
39 Chanderbali, A. *et al.* (2006) Genetic footprints of stamen ancestors guide perianth evolution in *Persea* (Lauraceae). *Int. J. Plant Sci.* 167, 1075–1089
40 Buzgo, M. *et al.* (2004) Floral developmental morphology of *Amborella trichopoda* (Amborellaceae). *Int. J. Plant Sci.* 165, 925–947
41 Buzgo, M. *et al.* (2005) The making of the flower. *Biologist* 52, 149–154
42 Laitinen, R.A. *et al.* (2006) Patterns of MADS-box gene expression mark flower-type development in *Gerbera hybrida* (Asteraceae). *BMC Plant Biol.* 6, 11
43 Schwarz-Sommer, Z. *et al.* (1990) Genetic control of flower development in *Antirrhinum majus*. *Science* 250, 931–936
44 Frohlich, M.W. and Parker, D.S. (2000) The Mostly Male theory of flower evolutionary origins: from genes to fossils. *Syst. Bot.* 25, 155–170
45 Albert, V.A. *et al.* (2002) Pleiotropy, redundancy and the evolution of flowers. *Trends Plant Sci.* 7, 297–301
46 Theissen, G. and Becker, A. (2004) Gymnosperm orthologues of class B floral homeotic genes and their impact on understanding flower origin. *Crit. Rev. Plant Sci.* 23, 129–148
47 Baum, D.A. and Hileman, L.C. (2006) A developmental genetic model for the origin of the flower. In *Flowering and its Manipulation* (Ainsworth, C., ed.), pp. 3–27, Blackwell Publishers
48 Vázquez-Loboa, A. *et al.* Characterization of the expression patterns of LEAFY/FLORICAULA and NEEDLY orthologues in female and male cones of the conifer genera *Picea*, *Podocarpus* and *Taxus*: implications for current evo-devo hypotheses for gymnosperms. *Evol. Dev.* (in press)
49 Albert, V.A. *et al.* (2005) Floral gene resources from basal angiosperms for comparative genomics research. *BMC Plant Biol.* 5, 5
50 Yoo, M-J. *et al.* (2006) Diversification of glycogen synthase kinase 3/SHAGGY-like kinase in plants. *BMC Plant Biol.* 6, 3
51 Kong, H. *et al.* (2004) Highly heterogeneous rates of evolution in the SKP1 gene family in plants and animals: functional and evolutionary implications. *Mol. Biol. Evol.* 21, 117–128
52 Kong, H. *et al.* (2007) Patterns of gene duplication in the plant *SKP1* gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. *Plant J.* 50, 873–885
53 Kidner, C. and Martienssen, R. (2004) Spatially restricted microRNA directs leaf polarity through *ARGONAUTE1*. *Nature* 428, 81–84
54 Park, W. *et al.* (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* 12, 1484–1495

55 Kim, S. *et al.* (2005) Phylogeny and domain evolution in the *APETALA2*-like gene family. *Mol. Biol. Evol.* 23, 107–120
56 Mlotshwa, S. *et al.* (2006) Floral patterning defects induced by *Arabidopsis APETALA2* and microRNA172 expression in *Nicotiana benthamiana*. *Plant Mol. Biol.* 61, 781–793
57 Gomez-Mena, C. *et al.* (2005) Transcriptional program controlled by the floral homeotic gene AGAMOUS during early organogenesis. *Development* 132, 429–438
58 Wellmer, F. *et al.* (2004) Genome-wide analysis of spatial gene expression in *Arabidopsis* flowers. *Plant Cell* 16, 1314–1326
59 Schmidt, M. *et al.* (2003) Dissection of floral induction pathways using global expression analysis. *Development* 130, 6001–6012
60 Soltis, D.E. *et al.* (2002) Missing links: the genetic architecture of the flower and floral diversification. *Trends Plant Sci.* 7, 22–30
61 Baum, D.A. *et al.* (2002) Response: Missing links: the genetic architecture of the flower and floral diversification. *Trends Plant Sci.* 71, 31–32
62 Leebens-Mack, J. *et al.* (2006) A genomics approach to the study of ancient polyploidy and floral developmental genetics. In *Developmental Genetics of the Flower (Advances in Botanical Research)* (Vol. 44) (Soltis, D.E. *et al.*, eds), pp. 528–542, Elsevier
63 Vrebalov, J. *et al.* (2002) A MADS-box gene necessary for fruit ripening at the tomato *Ripening-Inhibitor* (*Rin*) locus. *Science* 296, 343–346
64 Whipple, C.J. *et al.* (2007) Conservation of B class gene expression in the second whorl of a basal grass and outgroups links the origin of lodicules and petals. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1081–1086
65 Raharjo, SH-T. and Litz, R.E. (2005) Micrografting and ex vitro grafting for somatic embryo rescue and plant recovery in avocado (*Persea americana*). *Plant Cell Tiss. Org. Cult.* 82, 1–9
66 Hileman, L.C. *et al.* (2005) Virus-induced gene silencing is an effective tool for assaying gene function in the basal eudicot species *Papaver somniferum* (opium poppy). *Plant J.* 44, 334–341
67 Kramer, E.M. *et al.* (2007) Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia*. *Plant Cell* 19, 750–766
68 Hong, R.L. *et al.* (2003) Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. *Plant Cell* 15, 1296–1309
69 Frohlich, M.W. (2006) Recommendations and goals for evo-devo research: scenarios, genetic constraint, and developmental homeostasis. In *Monocots: Comparative Biology and Evolution* (Columbus, J.T. *et al.*, eds), pp. 172–187, Rancho Santa Ana Botanic Garden
70 Ditta, G. *et al.* (2004) The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr. Biol.* 14, 1935–1940
71 Theissen, G. (2001) Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* 4, 75–85
72 Van Tunen, A.J. *et al.* (1993) Floral organogenesis in *Tulipa*. *Flowers Newslett.* 16, 33–38
73 Kramer *et al.* (2003) Complex patterns of gene duplication in the *APETALA3* and *PISTILLATA* lineages of the Ranunculaceae. *Int. J. Plant Sci.* 164, 1–11
74 Bowman, J.L. (1997) Evolutionary conservation of angiosperm flower development at the molecular and genetic levels. *J. Biosci.* 22, 515–527
75 Kramer, E.M. and Irish, V.F. (1999) Evolution of genetic mechanisms controlling petal development. *Nature* 399, 144–148
76 Kanno, A. *et al.* (2003) Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). *Plant Mol. Biol.* 52, 831–841