SEPALLATA gene diversification: brave new whorls

Simon T. Malcomber and Elizabeth A. Kellogg

Department of Biology, University of Missouri – St. Louis, One University Boulevard, Saint Louis, MO 63121, USA

***SEPALLATA* (*SEP*) genes form an integral part of models that outline the molecular basis of floral organ determination and are hypothesized to act as co-factors with ABCD floral homeotic genes in specifying different floral whorls. The four *SEP* genes in *Arabidopsis* function redundantly, but the extent to which *SEP* genes in other flowering plants function similarly is unknown. Using a recent 113-gene *SEP* phylogeny as a framework, we find surprising heterogeneity among *SEP* gene C-terminal motifs, mRNA expression patterns, protein–protein interactions and inferred function. Although some *SEP* genes appear to function redundantly, others have novel roles in fruit maturation, floral organ specification and plant architecture, and have played a major role in floral evolution of diverse plants.**

SEPALLATA gene duplication and loss

Like other MADS box transcription factors, genes of the *SEPALLATA* (*SEP*) subfamily have been studied extensively in *Arabidopsis*. The four *SEP* genes in *Arabidopsis*, designated *Arabidopsis thaliana SEP1* (*AtSEP1*, previously known as *AGL2*), *AtSEP2* (*AGL4*), *AtSEP3* (*AGL9*) and *AtSEP4* (*AGL3*), function largely redundantly with single *sep* mutants having only subtle phenotypes [1,2]. By contrast, flowers of *sep1 sep2 sep3* (*sepallata*) triple mutants consist entirely of sepal-like organs [1] and those of *sep1 sep2 sep3 sep4* quadruple mutants are entirely comprised of leaf-like structures (Figure 1b) [2]. *Arabidopsis SEP* genes thus function in the development of each floral whorl and in meristem determinacy [2]. Each *SEP* protein interacts with one or more of the other MADS box transcription factors.

Perhaps because of the extensive redundancy and broadly similar function of *Arabidopsis SEP* genes, the tremendous diversity in the gene family has often been over-looked. The number of *SEP* genes, and their developmental and biochemical functions, vary across species and, thus, the extent of redundancy of any particular gene likewise varies. Here, we highlight the intricate and intriguing comparative biology of the *SEP*-subfamily and propose some mechanisms by which *SEP* function might have diversified.

Comparative study of the genes has been helped enormously by the recent phylogenetic analysis by Laura Zahn and colleagues [3] of all the available *SEP* genes, including several newly sequenced from basal angiosperms. Their work shows multiple duplications of *SEP* genes, the first occurring before the origin of extant angiosperms producing what we will call the *SEP3* and *LOFSEP* clades, containing *Arabidopsis AtSEP3*, and rice *LEAFY HULL STERILE1* (*LHS1*), *Oryza sativa* MADS5 (*OsMADS5*) and *OsMADS34* genes, petunia *FLORAL BINDING PROTEIN9* (*PhFBP9*) and *PhFBP23* genes, and *Arabidopsis AtSEP1*, *AtSEP2* and *AtSEP4* genes, respectively (Figure 2). Within the *SEP3* clade, an early diverging cluster of Asteraceae genes (*ASTERACEAE SEP3*) hints at a second duplication early in angiosperm evolution. Additional duplications occurred at the base of grasses (indicated by *GR*◉ in Figure 2) and more recently in other families (indicated by ◉ in Figure 2). Within the *LOFSEP* clade, duplications at or near the base of core eudicots produced the *SEP1/2, FBP 9/23* and *SEP4* clades (*EU*◉, Figure 2). A duplication within the *SEP1/2* clade before the origin of Brassicaceae produced the *AtSEP1* and *AtSEP2* genes. Additional duplications within the *SEP1/2* and *FBP9/23* clades occurred near the origin of Solanaceae (*SO*◉), and before the origin of *Malus* (◉); those in *Malus* are associated with polyploidy within Rosaceae. *Arabidopsis* lacks an *FBP9/23* gene, suggesting loss during Rosid diversification. The large *LOFSEP* clade of monocot sequences (M) shows duplications near the base of grasses, producing the *LHS1*, *OsMADS5* and *OsMADS34* subclades. Polyploidy in maize has produced additional duplicates.

Orthologs (genes related by descent at speciation) of *AtSEP1* and *AtSEP2* are restricted to Brassicaceae, whereas orthologs of *AtSEP4* occur only in core eudicots and are absent from non-core eudicots, monocots and basal angiosperms. Therefore, extrapolating the developmental role and biochemical function from *Arabidopsis* to other species needs to be tested.

Duplicate genes are thought to diverge in function over time, with long-term retention of truly redundant genes unlikely [4]. One copy will either decay to a pseudogene or will acquire a new function ('neofunctionalization'), or both genes will partition the ancestral function ('subfunctionalization') [5]. We see evidence of all phenomena in the *SEP* subfamily.

Evolution of *SEP* C-terminal motifs

Michiel Vandenbussche and colleagues [6] identified seven motifs in the C-terminus of the genes (numbered in Figure 2), some of which activate transcription [7]. The *SEP3* gene *Dendrobium DgMADS1* and the *SEP4* gene

Corresponding author: Malcomber, S.T. (malcombers@msx.umsl.edu).
Available online 15 August 2005

 1360-1385/$ - see front matter doi:10.1016/j.tplants.2005.07.008

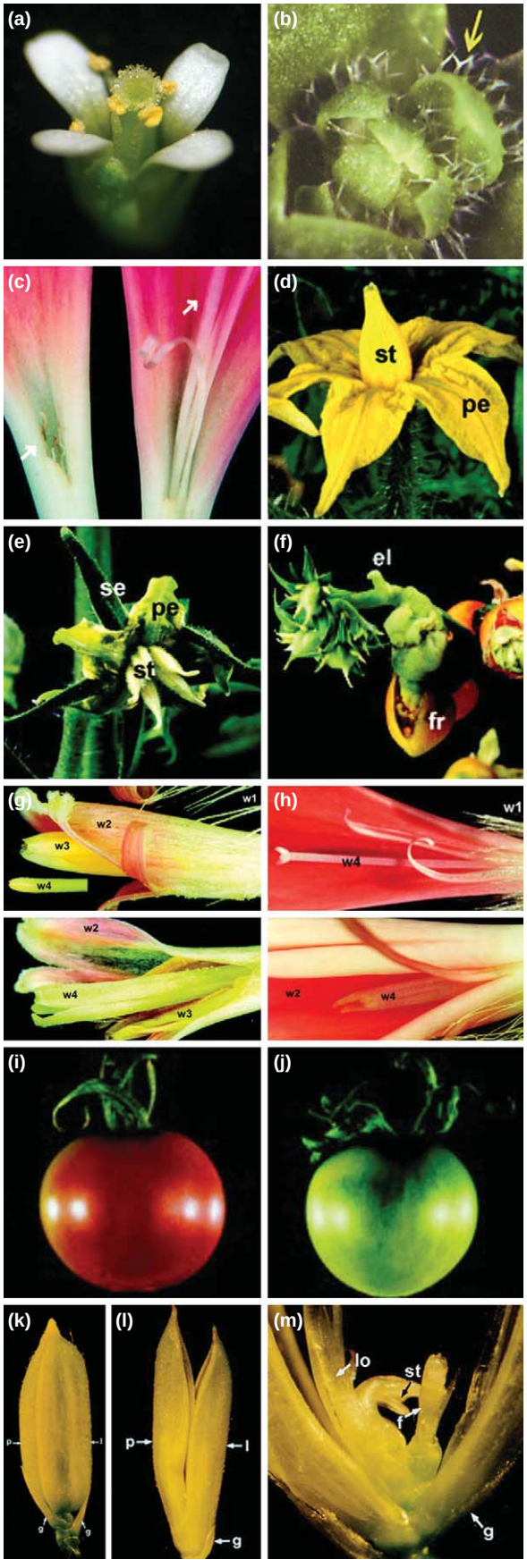

Figure 1. *SEP* gene mutant and down-regulated phenotypes. **(a,b)** *Arabidopsis thaliana*. (a) Wild-type flower. (b) Flower of *sep1 sep2 sep3 sep4* quadruple mutant: all floral whorls are converted into leaf-like structures with trichomes (arrow). Reproduced, with permission, from Ref. [2]. **(c)** *Gerbera hybrida*: on the left, normal ray florets with sterile staminodes (arrow); on the right, ray florets from a plant with *GhGRCD1* (*SEP3* clade) down-regulated, showing petal-like sterile staminodes (arrow). Reproduced, with permission, from Ref. [14]. **(d–f)** Tomato (*Lycopersicon esculentum*). Reproduced, with permission, from Ref. [26]. (d) Wild-type flower. (e) Flower of plant with *LeTM29* (*SEP1/2* clade) down-regulated, showing greenish petals and separated stamens. (f) Ectopic shoot developing from the fruit of a plant with *LeTM29* down-regulated. **(g,h)** *Gerbera hybrida*. Reproduced, with permission, from Ref. [27]. (g) Upper panel: disk flowers of wild-type flower. Lower panel: disk flowers of plant with *GhGRCD2* (*SEP1/2* clade) down-regulated. Green leaf-like organs replace pistils in whorl 4 (W4). (h) Upper panel: wild-type ray floret. Lower panel: ray flower of plant with *GhGRCD2* down-regulated, showing petal-like organs in W4. **(i,j)** Tomato (*Lycopersicon esculentum*). Reproduced, with permission, from Ref. [34]. (i) Wild-type fruit. (j) Unripe fruit of *rin* mutant (of the *LeMADSRIN* gene, *SEP4* clade). **(k–m)** Rice (*Oryza sativa*). Reproduced, with permission, from Ref. [40]. (k) Wild-type spikelet. (l) Spikelet of *lhs1* mutant (*LHS1* clade) with elongated palea and lemma. (m) *lhs1* spikelet with leafy lodicules and new flower emerging. Abbreviations: el, ectopic leaf; f, flower; fr, fruit; g, sterile lemmas; l, lemma; lo, lodicules; p, palea; pe, petal; se, sepal; st, stamens; w1, whorl 1; w2, whorl 2; w3, whorl 3; w4, whorl 4.

Capsicum CaPepMADS have truncated C-termini and lack any of the *SEP* C-terminal motifs. The internal *AGL2* motif is conserved in most *SEP* genes, but is appreciably shorter in rice *OsMADS34* (indicated by an arrow in Figure 2), and is disrupted by insertion of the amino acids LCR in *Eupomatia EbAGL2*, which might disrupt function.

C-terminal motifs are distinct between *SEP3* and *LOFSEP* proteins but are generally similar within these major clades. Exceptions are the novel motif in the *ASTERACEAE SEP3* clade (motif 8, Figure 2) and in each of the grass duplicates in the *LOFSEP* clade. Diversity of sequence is the first hint of diversity of function.

SEP gene expression, interactions and developmental role

SEP3 clade

Despite similar sequences, *SEP3* genes vary in expression pattern. All are expressed in inflorescences but expression in vegetative tissues has originated more than once. *SEP3* genes are expressed in the inner three floral whorls of all species except *Aranda AdOM1*, which lacks expression in stamens and carpels [8]. A broader expression domain that includes the outer whorl has originated multiple times [9–15] (Figure 3). Expression is detected in the fruit of all species where information is available.

Arabidopsis Atsep3 and petunia *Phfbp2* mutants, and tomato and birch plants with co-suppressed *LeTM5* and *BpMADS1* genes, respectively, show disruption in the organs in which the RNA is expressed [16–19]. However, down-regulation of *Gerbera GhGRCD1* affects only the sterile stamens (Figure 1c) [14], even though *GhGRCD1* RNA is present in all floral whorls. Thus, in *Gerbera*, either expression does not indicate function in sepals, petals or gynoecium, or a duplicate gene is partially redundant with *GhGRCD1*. Compared to other *SEP3* proteins, the C-terminus of *GhGRCD1* is divergent (Figure 2), which might also indicate modified function.

Over-expressing *SEP* genes often produces early flowering plants, as commonly occurs with over-expression of other MADS-box genes such as *AP1/FUL* [20]. Over-expression of *Arabidopsis*, tobacco and rice *SEP3* genes in *Arabidopsis* or tobacco does not affect floral

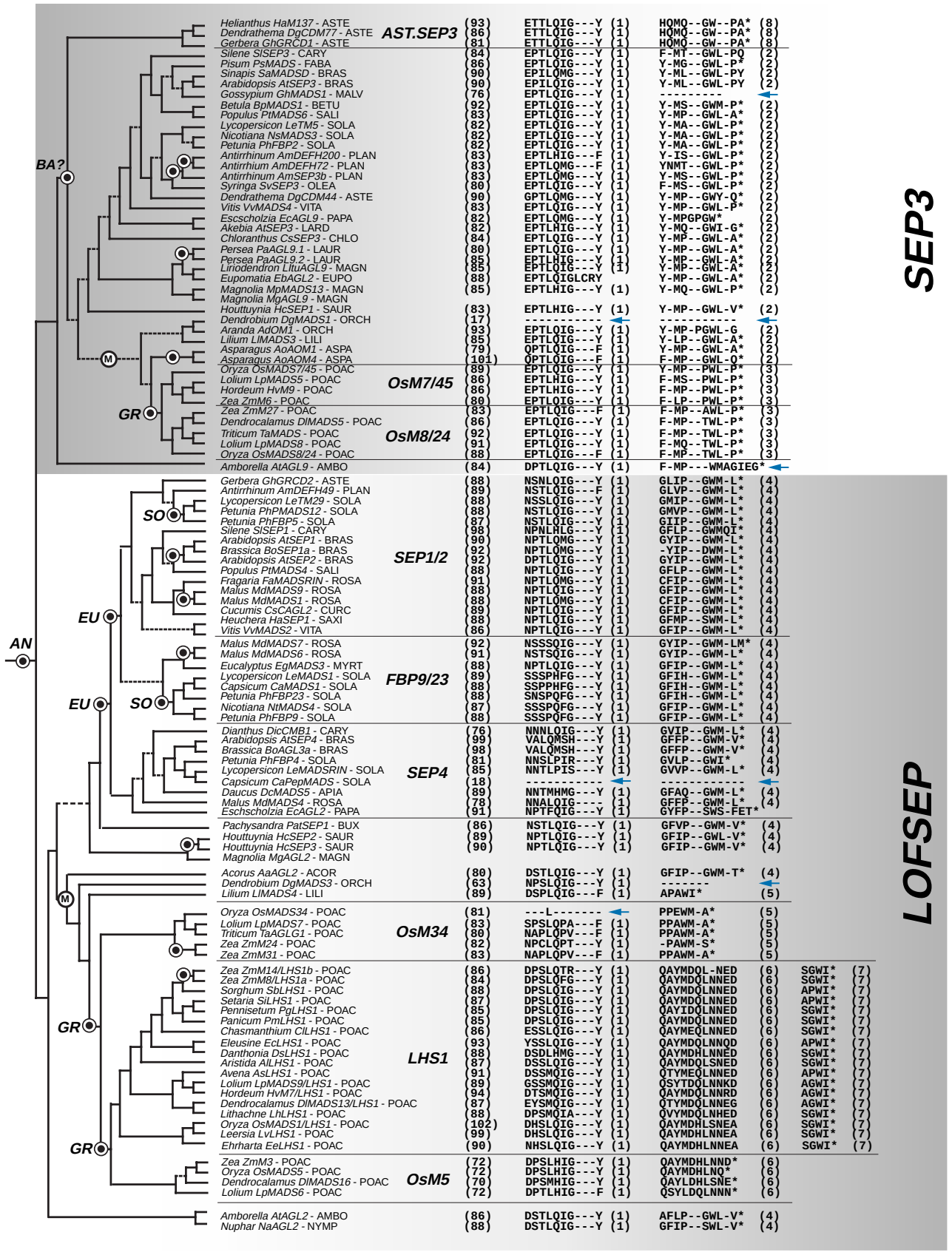

Figure 2. Phylogeny of 113 *SEP* genes. Tree topology adapted, with permission, from Ref. [3]. Branches made up of solid lines are supported by >70% maximum parsimony bootstrap (MPB); branches indicated by broken lines indicate <70% MPB. The ◉ symbol indicates an inferred duplication event. Abbreviations: AN, angiosperms; AST, Asteraceae; BA, basal angiosperms; EU, core eudicots; GR, grasses; M, monocots; OsM, OsMADS; SO, Solanaceae. The number in parentheses represents the length of the C-terminus in amino acids. Numbers (1–8) indicate the C-terminal motifs (1–7 were designated by Vandenbussche *et al.* [6]): (1) = internal *AGL2* motif; (2) = *AGL9/SEP3* motif; (3) = *ZmM7* motif; (4) = *AGL2/SEP1* terminal motif; (5) = *OsMADS34* motif; (6) = *ZmM3/OsMADS5* motif; (7) = *OsMADS1/LHS1* motif; (8) = *ASTERACEAE SEP3* motif. An ← symbol indicates that a C-terminal motif is absent or not discernable.

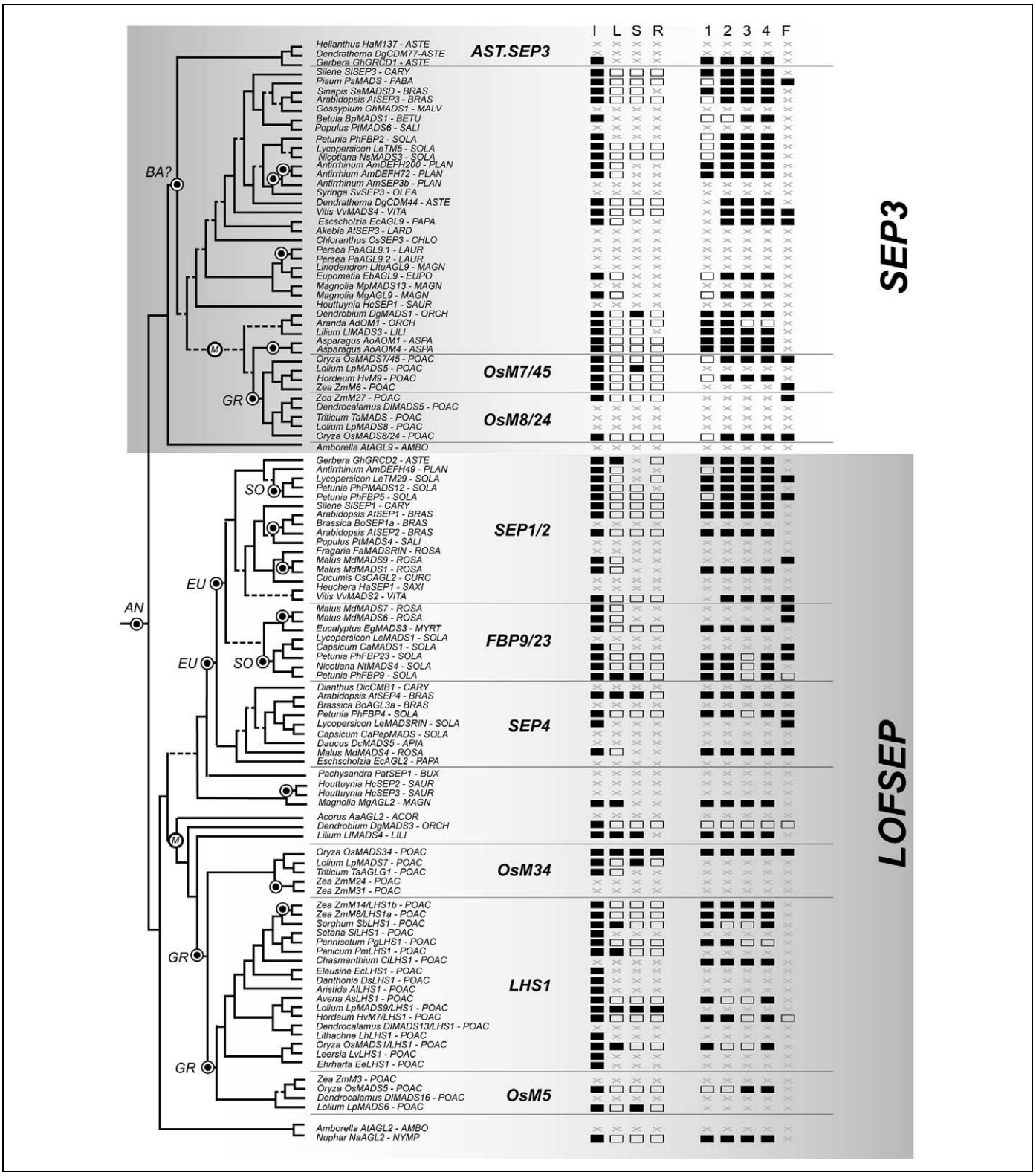

Figure 3. *SEP* phylogeny and mRNA expression profiles. Tree topology adapted, with permission, from Ref. [3]. Branches made up of solid lines are supported by >70% maximum parsimony bootstrap (MPB); branches indicated by broken lines indicate <70% MPB. The ◉ symbol indicates an inferred duplication event. Abbreviations: AN, angiosperms; AST, Asteraceae; BA, basal angiosperms; EU, core eudicots; GR, grasses; M, monocots; OsM, OsMADS; SO, Solanaceae. Abbreviations used in the panels indicating expression: I, inflorescence; L, leaf; S, stem or culm; R, roots; 1, whorl 1; 2, whorl 2; 3, whorl 3; 4, whorl 4; F, fruit. Filled rectangles indicate gene expression; unfilled rectangles indicate no gene expression; X indicates gene expression unknown. RNA expression: *Antirrhinum AmDEFH49, AmDEFH72, AmDEFH200* [12]; *Aristida AlLHS1* [38]; *Arabidopsis AtSEP1* [48], *AtSEP2* [49], *AtSEP3* [50], *AtSEP4* [32]; *Asparagus AoAOM1* [10], *AoAOM4* [11]; *Avena AsLHS1* [38]; *Betula BpMADS* [19]; *Capsicum CaMADS1* [31]; *Chasmanthium ClLHS1* [38]; *Danthonia DsLHS1* [38]; *Dendrathema DgCDM44* [51]; *Dendrobium DgMADS1, DgMADS3* [9]; *Ehrharta EeLHS1* [38]; *Eleusine EcLHS1* [38]; *Eschscholzia EcAGL9* [3]; *Eucalyptus EgMADS3* [52]; *Eupomatia EbAGL9* [25]; *Gerbera GhGRCD1* [14], *GhGRCD2* [27]; *Hordeum HvM7/LHS1, HvM9* [53]; *Leersia LvLHS1* [38]; *Lilium LIMADS3, LIMADS4* [15]; *Lithachne LhLHS1* [38]; *Lolium LpMADS5, LpMADS6, LpMADS7, LpMADS9/LHS1* [36]; *Lycopersicon LeMADSRIN* [34], *LeTM5* [18], *LeTM29* [26]; *Magnolia MgAGL2, MgAGL9* [25]; *Malus MdMADS1* [54], *MdMADS4* [33], *MdMADS6, MdMADS7, MdMADS9* [55]; *Nicotiana NsMADS3* [21], *NtMADS4* [30]; *Nuphar NaAGL2* [3]; *Oryza OsMADS1/LHS1* [42], *OsMADS5* [37], *OsMADS7/45, OsMADS8/24* [56], *OsMADS34* [35]; *Panicum PmLHS1* [38]; *Pennisetum PgLHS1* [38]; *Petunia PhFBP2, PhFBP4, PhFBP5, PhFBP9, PhFBP23* [29], *PhpMADS12* [20]; *Pisum PsMADS*, [57]; *Setaria SiLHS1* [38]; *Silene SlSEP1, SlSEP3* [58]; *Sinapis SaMADSD* [13]; *Sorghum SbLHS1* [38]; *Triticum TaAGLG1* [59]; *Vitis VvMADS2, VvMADS4* [60]; *Zea ZmM6, ZmM27* [61], *ZmM8/LHS1a, ZmM14/LHS1b* [39].

morphology [16,21,22]. However, overexpression of *Lilium LlMADS3* in *Arabidopsis* creates indeterminate flowers [15], indicating compromised C-class function [23]. Unfortunately, mutants are generally not available for species with expanded expression domains.

All *SEP3* proteins interact with *AP1/FUL* and C/D-class genes. This might explain some mutant phenotypes, which can be similar to those of *AP1/FUL* or *C/D*-class mutants. If complexes of proteins regulate the transition to flowering and the formation of determinate floral meristems, then disruption of any member of the complex will have the same effect. *AtSEP3* and *PhFBP2* both interact with *SOC1-like* proteins. *AtSEP3* and rice *OsMADS7* and *OsMADS8* all interact with *AGL6-like* proteins (Figure 4). *SEP* proteins are reported to interact only with B-class heterodimers; no direct interactions are reported between *SEP* and individual B-class proteins (Figure 4).

LOFSEP clade

Genes from *Amborella* and *Nuphar* are sister to the entire clade, and genes from *Houttuynia* and *Magnolia* are sister to the eudicot sequences, consistent with their phylogenetic position [24]. The *Nuphar* and *Magnolia* genes are expressed in all floral organs [3,25]. No functional data are available.

SEP1/2 subclade

All *SEP1/2* genes are expressed in inflorescences; *Gerbera GhGRCD2* is also expressed in vegetative tissues, presumably a derived expansion of the expression domain. Expression in sepals is common but not universal; all genes are expressed in the second, third and fourth whorls, and in the fruit. *Arabidopsis Atsep1* and *Atsep2* single mutants only have subtle changes in phenotype [1]; petunia *Phfbp5* insertion mutants are indistinguishable from wild type [17]. Co-suppression of the tomato *LeTM29* produces plants with altered petals, stamens, pistils and fruit; the fruit eventually develops ectopic shoots with leaves and secondary flowers (Figure 1e,f) [26]. *Gerbera* plants with co-suppressed *GhGRCD2* have indeterminate flowers, petal-like carpels and larger inflorescences with more flowers (Figure 1g,h) [27]. The *GhGRCD2* floral phenotype suggests reduced C-class gene activity, and the protein interacts with *Gerbera AG* homologs [27]. However, the inflorescence phenotype is novel, reflecting either novel protein function or co-suppression of related *SEP* genes.

The *SEP1/2* and *PhFBP5/pMADS12* duplications occurred independently in Brassicaceae and Solanaceae, respectively. Consistent with theory, the duplicate proteins have non-identical interaction patterns. *AtSEP1* and *PhFBP5* proteins interact with *AP1/FUL* proteins but neither duplicate (*AtSEP2* or *PhpMADS12*) does [28,29] (Figure 4). Similarly, *PhFBP5* interacts with B-class proteins but its duplicate does not [20]. In spite of the apparent developmental redundancy of the pairs, biochemical function has diverged.

The *SEP1/2* protein interactions also differ between species (Figure 4). *PhFBP5* interacts with another *SEP* protein, whereas no such interaction is reported in *Arabidopsis* [20,28,29]. Thus, despite similar developmental roles, the biochemical function of the *SEP1/2* proteins appears to be diverse among eudicots.

FBP9/23 subclade

Arabidopsis has lost this group of genes, so less is known about them. All are restricted to inflorescences except *PhFBP9*, which is also expressed in leaves and stems [29]. All genes are detected in sepals, petals and gynoecia, but only *Eucalyptus EgMADS3* is expressed in stamens. All genes except *PhFBP9* are expressed in fruit.

All available functional data are from Solanaceae. *Phfbp9* insertion mutants have altered plant architecture during the reproductive phase [17]. Ectopically expressing *Capsicum CaMADS1* and *Nicotiana NsMADS4* produces short, early flowering tobacco plants but no modifications in floral form [30,31]. As with other duplicate pairs, *PhFBP9* and *PhFBP23* have non-identical interactions. Both interact with *PhUNS* and *PhFBP28*, but *PhFBP9* also interacts with *PhFBP21* and *PhFBP22* [29]. Both also interact with C and D class proteins, but only *PhFBP23* interacts with *AP1/FUL* proteins [29]. Thus, members of the *FBP9/23* subclade appear to function redundantly with other *SEP* genes during floral development but might have other roles during the transition to flowering. This requires testing in families other than Solanaceae.

SEP4 subclade

Expression of *SEP4* genes varies among *Arabidopsis AtSEP4* (all above ground organs) [32], *Malus MdMADS4* (all floral whorls and fruit) [33], petunia *PhFBP4* (sepals, petals, pistil and fruit) [29] and *Lycopersicon LeMADS-RIN* (fruit) [34]. These diverse expression patterns suggest functional divergence. *Atsep4* mutants closely resemble wild-type plants [2], indicating largely redundant developmental function. Both *AtSEP4* and *PhFBP4* interact with *SOC1-like* proteins [20,28] but only *AtSEP4* interacts with *AP1/FUL* proteins (Figure 4). *PhFBP4* interacts with itself but *AtSEP4* does not [28,29].

Unlike other *SEP4* mutants, *rin* mutants of tomato (of *LeMADSRIN*) fail to produce ripe fruit (Figure 1j) [34]. A frameshift in the C-terminus, shortly after the internal *AGL2* motif, causes this mutation, disrupting the external *AGL2* motif and extending the gene product by 148 amino acids. Because the protein is so highly modified, it is not clear whether *LeMADSRIN* has a novel non-redundant role during fruit maturation or whether the *rin* mutation is a gain-of-function caused by the newly extended C-terminus. *SEP4* function with regard to developmental role and biochemical function appears to be highly variable across species.

Monocot LOFSEP subclade

The monocot *LOFSEP* clade comprises three subclades of grass genes, plus sequences from *Acorus, Dendrobium* and *Lilium*. Expression patterns within the clade are heterogeneous (Figure 2) [9,15,35–38]. Gene expression has been investigated extensively in the grass-specific *LHS1* clade. In all grasses investigated, *LHS1* is expressed in the palea and lemma (first whorl organs), but expression within the other organs varies widely from species to species. *LHS1*

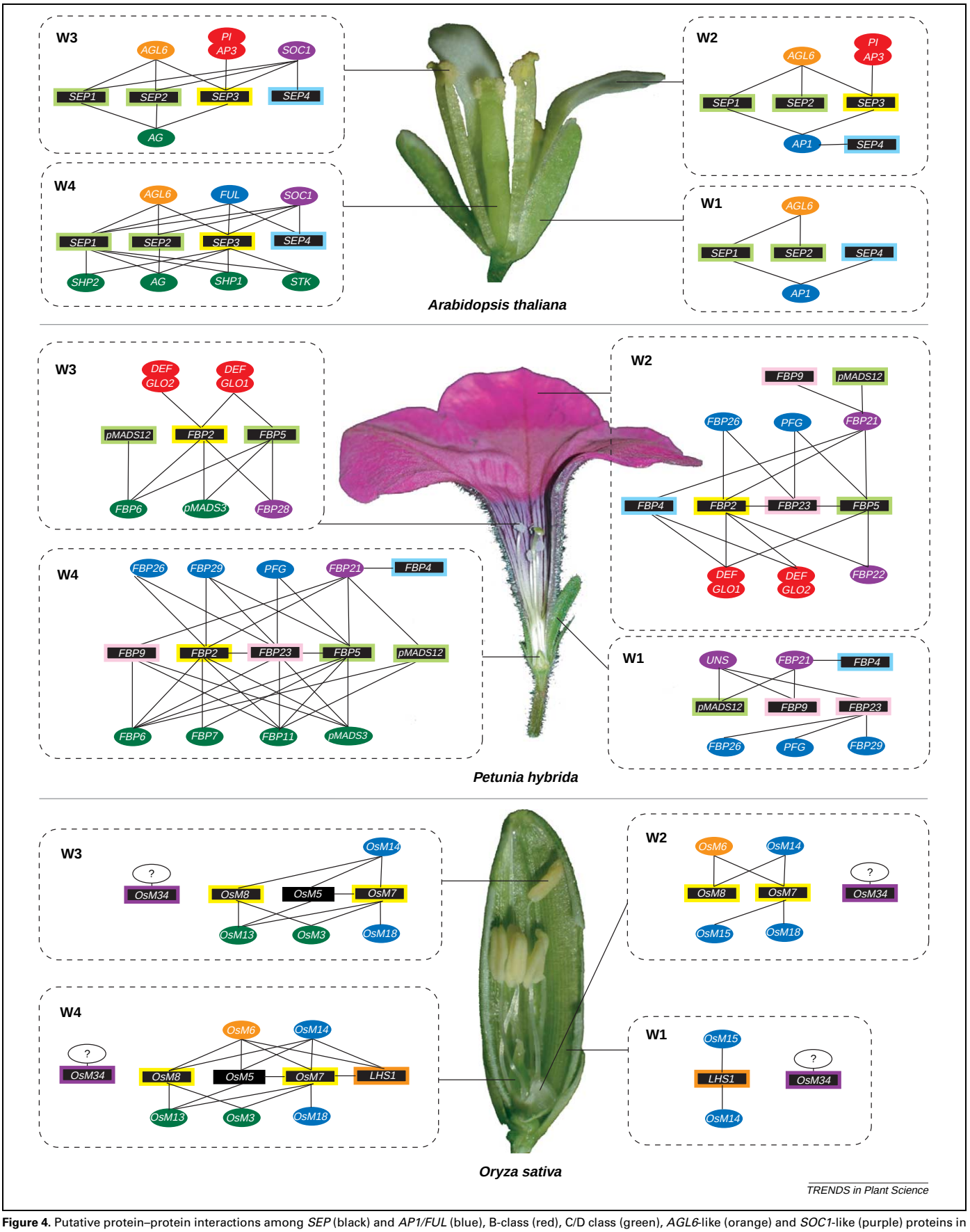

Figure 4. Putative protein–protein interactions among *SEP* (black) and *AP1/FUL* (blue), B-class (red), C/D class (green), *AGL6*-like (orange) and *SOC1*-like (purple) proteins in different floral whorls of *Arabidopsis thaliana*, petunia (*Petunia hybrida*) and rice (*Oryza sativa*) based on yeast two- and three-hybrid assays and RNA expression. A question mark indicates that the interacting protein is unknown. The different proteins are indicated by different colored rectangular borders: yellow = *SEP3*; green = *SEP1/2*; blue = *SEP4*; purple = *OsMADS34*; orange = *LHS1*; black = *OsMADS5*. Interactions among non-*SEP* proteins are omitted for clarity. Abbreviations: W1 = whorl 1, sepals (*Arabidopsis* and petunia) or palea and lemma (rice); W2 = whorl 2, petals (*Arabidopsis* and petunia) or lodicules (rice); W3 = whorl 3, stamens; W4 = whorl 4, pistil.

expression also varies with floret maturation in the spikelet: *LHS1* is expressed only in the terminal floret of spikelets that develop from top to bottom (basipetally) but is expressed in multiple florets of species that develop from bottom to top (acropetally) [38]. The former pattern supports the hypothesis that *LHS1* orthologs specify the terminal flower of the spikelet (the 'selector gene' hypothesis) [39]. In species with acropetal maturation of florets, the gene might have different or additional developmental roles.

The only available mutant for any monocot *LOFSEP* gene is the semi-dominant negative mutation in the rice *LHS1* gene. It has leafy palea, lemma and lodicules, fewer stamens and, occasionally, an extra pistil or floret (Figure 1l,m) [40]. Ectopic expression of this gene in rice produces plants with short panicles and irregularly positioned branches; the sterile lemmas of the two lower flowers are similar to the palea and lemma of the fertile upper flower [41]. Over-expression in tobacco promotes early flowering but does not affect floral morphology, suggesting that the rice protein is not recognized by the tobacco floral machinery [42].

Over-expression of rice *OsMADS5* promotes early flowering in rice but does not affect floral morphology [43]. Ectopic expression of *Lilium LlMADS4* in *Arabidopsis* has no phenotypic effect [15], presumably reflecting the considerable divergence between monocot and eudicot *LOFSEP* genes.

The duplicate rice proteins *LHS1* and *OsMADS5* both interact with the *AP1/FUL* protein *OsMADS14*, the *SEP3* protein *OsMADS7/45* and the *AGL6* protein *OsMADS6* (Figure 4). *LHS1* also interacts with the *AP1/FUL* protein *OsMADS15*, and *OsMADS5* interacts with the C-class protein *OsMADS13* and the *AGL6* protein *OsMADS17*, indicating possible divergence in biochemical function following the duplication [5]. No interaction information is available for any *OsMADS34* proteins. The function of the other monocot *LOFSEP* genes is completely unknown but their diverse expression patterns similarly point to multiple developmental roles.

SEP gene diversity

Although *SEP* genes are generally similar in specifying meristem and floral organ identity, they exhibit considerable heterogeneity in C-terminal motifs, mRNA expression patterns, protein–protein interactions and mutant phenotypes. Far from being a set of redundant genes with largely conserved function, *SEP* genes range from developmentally redundant, as in *Arabidopsis*, to non-redundant, with unique roles in fruit maturation (tomato *SEP4* gene *LeMADSRIN*), floral organ specification (rice *LHS1* gene *OsLHS1*), staminode identity (*Gerbera ASTERACEAE SEP3* gene *GhGRCD1*) and plant architecture (petunia *FBP9/23* gene *PhFBP9*).

MIKC-type MADS box proteins must dimerize to bind DNA [44]; the dimers are thought to form tetramers or even higher order complexes [45,46] but the precise structure of these multimers is unknown. *SEP* proteins appear to contribute to such multimeric complexes [45,46]. The available data suggest that the function of a *SEP* protein in a transcription factor complex is influenced by: (i) the organ in which it is expressed, (ii) the number and identity of other co-expressed *SEP* genes, (iii) the particular interactions of each *SEP* protein with other *SEP*s and other MADS box proteins, and (iv) the strength with which the complex can activate transcription. C-terminal motifs, which have diversified among *SEP* proteins, have been proposed as central controls of transcriptional activation, partner specificity in higher order complexes, DNA binding specificity, subcellular localization and/or the ability to attract interacting partners [6]. We recognize eight C-terminal motifs in *SEP* proteins; novel motifs correlate with duplications near the base of grasses in the *LOFSEP* clade and within basal angiosperms in the *SEP3* clade (Figure 2). Some of these motifs are required for transcription activation [7]. The proteins *Dendrobium DgMADS1* (*SEP3*) and *Capsicum CaPepMADS* (*SEP4*) lack all key motifs and might, therefore, be unable to activate transcription, although they could still help stabilize the transcription factor complex – this is testable. Likewise, members of *OsMADS34* clade, with their shortened C-terminus, might be unable to activate transcription. This would explain why the co-expressed *LHS1* proteins are non-redundant; transcriptional activation might be conferred by *LHS1* alone, even if *OsMADS34* is required for multimer formation. The dramatically longer C-terminus in the mutant tomato *SEP4* gene *LeMADSRIN* indicates its developmental importance but its precise role remains unknown. Evolution of a novel C-terminal motif correlates with the origin of a non-redundant role in staminodial development in the *ASTERACEAE SEP3* gene *Gerbera GhGRCD1*. *GhGRCD1* and the *SEP1/2* gene *GhGRCD2* both interact with the C-class proteins [14] but down-regulating *GhGRCD2* does not affect staminode morphology [27]. Experimental modification of the *GhGRCD1* C-terminal motif could test the influence of these C-class gene interactions and the effect of the protein on staminodes.

Plants also control *SEP* protein interactions by modifying expression patterns. Proteins that interact in a yeast two-hybrid system will only interact in the plant if they are expressed in the same cells. *SEP* RNA expression patterns show frequent gains and losses throughout the phylogeny (Figure 3). Loss of expression of any given *SEP* gene in an organ is generally complemented by retention of expression of a paralogous gene (related by descent at time of duplication). For example, despite frequent loss or gain of gene expression in the first (outer) floral whorl, every species for which data are available has at least one *SEP* gene expressed in that whorl. Thus, at the level of expression, the genes have apparently sub-functionalized, as predicted by theory. Gain of expression is also of considerable evolutionary interest. *PhFBP9* is the only petunia *SEP* gene detected in stems; novel protein

References for RNA expression: *Arabidopsis* [32,45,48–50,52,62–66]; petunia [20,29]; rice [35,37,42,56,67]. References for protein–protein interactions: *Arabidopsis* [28,45,68]; petunia [20,29], rice [7,67,69].

interactions could specify plant architecture during reproduction.

Protein interactions can be predicted by combining patterns of RNA expression and yeast two- and three-hybrid interaction data (Figure 4) with similar protein interaction partners at the same developmental stage, implying redundant biochemical function. *AtSEP1* and *AtSEP3*, and *PhFBP2* and *PhFBP5* have largely common sets of interaction partners in whorls 2–4 (W2–4), consistent with their overlapping developmental roles [20,28,29]. By contrast, *AtSEP2* and *AtSEP4* interact with only a subset of *AtSEP1* and *AtSEP3* interacting proteins, suggesting more restricted biochemical functions (Figure 4). In rice, only two *SEP* genes are expressed in W1 versus three to five in W2–W4 (Figure 4). Although differences in activation are also possible, the unique role of rice *LHS1* in W1 might be caused by protein–protein interactions different from those of the co-expressed *SEP* gene *OsMADS34*. A similar situation might occur with other *SEP* genes, such as the tomato *LeMADSRIN*, which is involved in the early stages of fruit ripening.

Future prospects

Although RNA expression profiles exist for diverse species, much remains to be learned about *SEP* genes. Dissecting individual and combined roles will require functional analyses, particularly in non-model species, and evidence of *SEP* protein localization and interactions. Yeast two-hybrid assays provide only a first indication of protein–protein interaction, with three- and four-hybrid assays in yeast, in protoplasts [47] and *in planta* necessary to examine how protein complexes contribute to organ identity in different species. This aspect of *SEP* protein biology is only beginning to be explored and is likely to yield more intriguing insights.

In summary, the complex pattern of gene duplication (and presumed loss) has facilitated diversification in expression pattern, biochemical function and, in some cases, developmental role. Far from being a set of redundant genes with identical functions, diversity in the *SEP* gene family mirrors the morphological diversity of the flowering plants.

Acknowledgements

We thank Andrew Doust, Claire Hemingway, Jill Preston and two anonymous reviewers for comments on an earlier draft. Marty Yanofsky and Gary Ditta kindly provided information on *Arabidopsis SEP4*. This work was supported by National Science Foundation grant DBI-0110189 (to E.A.K).

References

1 Pelaz, S. *et al*. (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* 405, 200–203
2 Ditta, G. *et al*. (2004) The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr. Biol.* 14, 1935–1940
3 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, 2225–2239
4 Ohno, S. (1970) *Evolution by Gene Duplication*, Allen & Unwin; Springer-Verlag
5 Force, A. *et al*. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545
6 Vandenbussche, M. *et al*. (2003) Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucleic Acids Res.* 31, 4401–4409
7 Lim, J. *et al*. (2000) Two rice MADS domain proteins interact with *OsMADS1*. *Plant Mol. Biol.* 44, 513–527
8 Lu, Z.X. *et al*. (1993) Nucleotide sequence of a flower-specific MADS box cDNA clone from orchid. *Plant Mol. Biol.* 23, 901–904
9 Yu, H. and Goh, C.J. (2000) Identification and characterization of three orchid MADS-box genes of the *AP1/AGL9* subfamily during floral transition. *Plant Physiol.* 123, 1325–1336
10 Caporali, E. *et al*. (2000) The MADS box gene *AOM1* is expressed in reproductive meristems and flowers of the dioecious species *Asparagus officinalis*. *Sex. Plant Reprod.* 13, 151–156
11 Losa, A. *et al*. (2004) *AOM3* and *AOM4*: two MADS box genes expressed in reproductive structures of *Asparagus officinalis*. *Sex. Plant Reprod.* 16, 215–221
12 Davies, B. *et al*. (1996) Multiple interactions amongst floral homeotic MADS box proteins. *EMBO J.* 15, 4330–4343
13 Bonhomme, F. *et al*. (1997) Characterization of *SaMADS D* from *Sinapis alba* suggests a dual function of the gene: in inflorescence development and floral organogenesis. *Plant Mol. Biol.* 34, 573–582
14 Kotilainen, M. *et al*. (2000) *GRCD1*, an *AGL2*-like MADS box gene, participates in the C function during stamen development in *Gerbera hybrida*. *Plant Cell* 12, 1893–1902
15 Tzeng, T.Y. *et al*. (2003) Two lily *SEPALLATA*-like genes cause different effects on floral formation and floral transition in *Arabidopsis*. *Plant Physiol.* 133, 1091–1101
16 Pelaz, S. *et al*. (2001) *APETALA1* and *SEPALLATA3* interact to promote flower development. *Plant J.* 26, 385–394
17 Vandenbussche, M. *et al*. (2003) Toward the analysis of the petunia MADS box gene family by reverse and forward transposon insertion mutagenesis approaches: B, C, and D floral organ identity functions require *SEPALLATA*-like MADS box genes in petunia. *Plant Cell* 15, 2680–2693
18 Pnueli, L. *et al*. (1994) The *TM5* MADS box gene mediates organ differentiation in the three inner whorls of tomato flowers. *Plant Cell* 6, 175–186
19 Lemmetyinen, J. *et al*. (2004) Functional characterization of *SEPALLATA3* and *AGAMOUS* orthologues in silver birch. *Physiol. Plant.* 121, 149–162
20 Ferrario, S. *et al*. (2003) The MADS box gene *FBP2* is required for *SEPALLATA* function in petunia. *Plant Cell* 15, 914–925
21 Jang, S. *et al*. (1999) Ectopic expression of tobacco MADS genes modulates flowering time and plant architecture. *Mol. Cells* 9, 576–586
22 Kang, H.G. *et al*. (1997) Characterization of two rice MADS box genes that control flowering time. *Mol. Cells* 7, 559–566
23 Mizukami, Y. and Ma, H. (1995) Separation of *AG* function in floral meristem determinacy from that in reproductive organ identity by expressing antisense *AG* RNA. *Plant Mol. Biol.* 28, 767–784
24 Angiosperm Phylogeny Group. (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botan. J. Linn. Soc.* 141, 399–436
25 Kim, S. *et al*. (2005) Sequence and expression studies of A-, B-, and E-class MADS-box homologues in *Eupomatia* (Eupomatiaceae): Support for the bracteate origin of the calyptra. *Int. J. Plant Sci.* 166, 185–198
26 Ampomah-Dwamena, C. *et al*. (2002) Down-regulation of *TM29*, a tomato *SEPALLATA* homolog, causes parthenocarpic fruit development and floral reversion. *Plant Physiol.* 130, 605–617
27 Uimari, A. *et al*. (2004) Integration of reproductive meristem fates by a *SEPALLATA*-like MADS-box gene. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15817–15822
28 de Folter, S. *et al*. (2005) Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. *Plant Cell* 17, 1424–1433
29 Immink, R.G.H. *et al*. (2003) Analysis of the petunia MADS-box transcription factor family. *Mol. Genet. Genomics* 268, 598–606
30 Jang, S. *et al*. (2002) Characterization of tobacco MADS-box genes involved in floral initiation. *Plant Cell Physiol.* 43, 230–238
31 Sung, S.K. *et al*. (2001) Characterization of MADS box genes from hot pepper. *Mol. Cells* 11, 352–359
32 Huang, H. *et al*. (1995) The *Arabidopsis* MADS-box gene *AGL3* is widely expressed and encodes a sequence-specific DNA-binding protein. *Plant Mol. Biol.* 28, 549–567

33 Sung, S.K. *et al*. (2000) Developmentally regulated expression of two MADS-box genes, MdMADS3 and MdMADS4, in the morphogenesis of flower buds and fruits in apple. *Planta* 210, 519–528
34 Vrebalov, J. *et al*. (2002) A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (*rin*) locus. *Science* 296, 343–346
35 Pelucchi, N. *et al*. (2002) Comparative analysis of rice MADS-box genes expressed during flower development. *Sex. Plant Reprod.* 15, 113–122
36 Petersen, K. *et al*. (2004) MADS-box genes from perennial ryegrass differentially expressed during transition from vegetative to reproductive growth. *J. Plant Physiol.* 161, 439–447
37 Kang, H.G. and An, G.H. (1997) Isolation and characterization of a rice MADS box gene belonging to the *AGL2* gene family. *Mol. Cells* 7, 45–51
38 Malcomber, S.T. and Kellogg, E.A. (2004) Heterogeneous expression patterns and separate roles of the *SEPALLATA* gene *LEAFY HULL STERILE1* in grasses. *Plant Cell* 16, 1692–1706
39 Cacharrón, J. *et al*. (1999) Expression of MADS box genes *ZMM8* and *ZMM14* during inflorescence development of *Zea mays* discriminates between the upper and the lower floret of each spikelet. *Dev. Genes Evol.* 209, 411–420
40 Jeon, J.S. *et al*. (2000) *leafy hull sterile1* is a homeotic mutation in a rice MADS box gene affecting rice flower development. *Plant Cell* 12, 871–884
41 Prasad, K. *et al*. (2001) Ectopic expression of rice *OsMADS1* reveals a role in specifying the lemma and palea, grass floral organs analogous to sepals. *Dev. Genes Evol.* 211, 281–290
42 Chung, Y.Y. *et al*. (1994) Early flowering and reduced apical dominance result from ectopic expression of a rice MADS box gene. *Plant Mol. Biol.* 26, 657–665
43 Jeon, J.S. *et al*. (2000) Production of transgenic rice plants showing reduced heading date and plant height by ectopic expression of rice MADS-box genes. *Mol. Breed.* 6, 581–592
44 Shore, P. and Sharrocks, A.D. (1995) The MADS-box family of transcription factors. *Eur. J. Biochem.* 229, 1–13
45 Honma, T. and Goto, K. (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525–529
46 Theissen, G. and Saedler, H. (2001) Plant biology. Floral quartets. *Nature* 409, 469–471
47 Immink, R.G.H. *et al*. (2002) Analysis of MADS box protein-protein interactions in living plant cells. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2416–2421
48 Flanagan, C.A. and Ma, H. (1994) Spatially and temporally regulated expression of the MADS-box gene *AGL2* in wild-type and mutant *Arabidopsis* flowers. *Plant Mol. Biol.* 26, 581–595
49 Savidge, B. *et al*. (1995) Temporal relationship between the transcription of two *Arabidopsis* MADS box genes and the floral organ identity genes. *Plant Cell* 7, 721–733
50 Mandel, M.A. and Yanofsky, M.F. (1998) The *Arabidopsis AGL9* MADS box gene is expressed in young flower primordia. *Sex. Plant Reprod.* 11, 22–28
51 Shchennikova, A.V. *et al*. (2004) Identification and characterization of four chrysanthemum MADS-box genes, belonging to the *APETALA1/FRUITFULL* and *SEPALLATA3* subfamilies. *Plant Physiol.* 134, 1632–1641
52 Southerton, S.G. *et al*. (1998) Eucalypt MADS-box genes expressed in developing flowers. *Plant Physiol.* 118, 365–372
53 Schmitz, J. *et al*. (2000) Cloning, mapping and expression analysis of barley MADS-box genes. *Plant Mol. Biol.* 42, 899–913
54 Sung, S.K. and An, G. (1997) Molecular cloning and characterization of a MADS-box cDNA clone of the Fuji apple. *Plant Cell Physiol.* 38, 484–489
55 Yao, J.L. *et al*. (1999) Seven MADS-box genes in apple are expressed in different parts of the fruit. *J. Am. Soc. Hortic. Sci.* 124, 8–13
56 Greco, R. *et al*. (1997) MADS box genes expressed in developing inflorescences of rice and sorghum. *Mol. Gen. Genet.* 253, 615–623
57 Buchner, P. and Boutin, J.P. (1998) A MADS box transcription factor of the *AP1/AGL9* subfamily is also expressed in the seed coat of pea (*Pisum sativum*) during development. *Plant Mol. Biol.* 38, 1253–1255
58 Matsunaga, S. *et al*. (2004) Characterization of two *SEPALLATA* MADS-box genes from the dioecious plant *Silene latifolia*. *Sex. Plant Reprod.* 17, 189–193
59 Yan, L. *et al*. (2003) Positional cloning of the wheat vernalization gene *VRN1*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6263–6268
60 Boss, P.K. *et al*. (2002) Cloning and characterisation of grapevine (*Vitis vinifera* L.) MADS-box genes expressed during inflorescence and berry development. *Plant Sci.* 162, 887–895
61 Lid, S.E. *et al*. (2004) Knock-out mutants of two members of the *AGL2* subfamily of MADS-box genes expressed during maize kernel development. *Plant Sci.* 167, 575–582
62 Parenicova, L. *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
63 Samach, A. *et al*. (2000) Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science* 288, 1613–1616
64 Pinyopich, A. *et al*. (2003) Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424, 85–88
65 Mandel, M.A. *et al*. (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* 360, 273–277
66 Mandel, M.A. and Yanofsky, M.F. (1995) The *Arabidopsis AGL8* MADS box gene is expressed in inflorescence meristems and is negatively regulated by *APETALA1*. *Plant Cell* 7, 1763–1771
67 Moon, Y.H. *et al*. (1999) Determination of the motif responsible for interaction between the rice *APETALA1/AGAMOUS-LIKE9* family proteins using a yeast two-hybrid system. *Plant Physiol.* 120, 1193–1204
68 Favaro, R. *et al*. (2003) MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell* 15, 2603–2611
69 Cooper, B. *et al*. (2003) A network of rice genes associated with stress response and seed development. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4945–4950