# And then there were many: MADS goes genomic☆

**Stefanie De Bodt[1], Jeroen Raes[1], Yves Van de Peer[1] and Günter Theißen[2]**

[1]Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium
[2]Friedrich-Schiller-University, Department of Genetics, Philosophenweg 12, D-07743 Jena, Germany

**During the past decade, MADS-box genes have become known as key regulators in both reproductive and vegetative plant development. Traditional genetics and functional genomics tools are now available to elucidate the expression and function of this complex gene family on a much larger scale. Moreover, comparative analysis of the MADS-box genes in diverse flowering and non-flowering plants, boosted by bioinformatics, contributes to our understanding of how this important gene family has expanded during the evolution of land plants. Therefore, the recent advances in comparative and functional genomics should enable researchers to identify the full range of MADS-box gene functions, which should help us significantly in developing a better understanding of plant development and evolution.**

Throughout plant evolution, MADS-box genes have been recruited as transcriptional regulators active in the development of diverse plant structures. Since the discovery of the first MADS-box genes more than a decade ago, biologists have made great progress in elucidating the roles of these genes in plant development. Expression studies and mutant analyses on MADS-box genes in diverse plant species such as *Arabidopsis*, *Antirrhinum majus* and maize (*Zea mays*), among others, revealed the crucial importance of MADS-box genes in the regulation of both reproductive (flower, seed, fruit) and vegetative (root, leaf) development [1]. Furthermore, MADS-box genes, used in the control of floral patterning, form the ideal genetic toolkit to study the diversification of flower architecture [2].

The MADS-box genes constitute a large gene family named after a few of its earliest members: *MCM1* (from yeast) [3], *AGAMOUS* (from *Arabidopsis*) [4], *DEFICIENS* (from *A. majus*) [5,6] and *SRF* (from *Homo sapiens*) [7]. The gene family can be divided into two main lineages, referred to as type I and type II, both of which are present in plants, animals and fungi. All members of the family possess the 180-nucleotide long (on average) MADS box [8], which encodes the domain of the transcription factors that is responsible for nuclear localization, DNA binding, dimerization and accessory factor binding [1,2,9].

In plants, type II MADS-domain proteins, referred to as MIKC-type proteins, possess three additional functional domains: a well-conserved K (keratin) domain, responsible for dimerization; a less-well-conserved I (intervening) domain, which constitutes a key regulatory determinant for the selective formation of DNA-binding dimers; and a variable C-terminal region, which is involved in transcriptional activation or in the formation of ternary or quaternary protein complexes [10–12] and contributes to functional specificity [13]. In contrast to type II genes, which have been the subject of extensive research, not much is known about the type I genes in plants. Except for the MADS box, the type I genes share no sequence similarity with type II genes. However, some type I genes share conserved C-terminal motifs with each other [14,15]. In addition, a third group of genes has recently been identified that possess only half of the MADS box or are overall highly divergent. These are referred to as MADS-like genes [14].

In this article, we present a survey of the recent progress that has been made in the field of MADS-box gene research, particularly the contribution of genomics, bioinformatics and protein–protein interaction studies to the understanding of the MADS-box gene family and the ways forward for plant developmental studies in the phylogenomics and phyloproteomics era.

## Genetics lays the foundations

The study of plant MADS-box genes was initially prompted by their importance in flower development. Gain- and loss-of-function phenotypes generated through T-DNA, transposon- or ethyl methane sulfonate-induced mutations in MADS-box genes have uncovered the function of many of these genes in diverse aspects of this process, ranging from the determination of flowering time (e.g. *FLOWERING LOCUS C*, *SUPPRESSOR OF OVER-EXPRESSION OF CONSTANS 1*) to the specification of floral meristem (e.g. *APETALA1*, *CAULIFLOWER*) and floral organ identity (e.g. *APETALA1*, *APETALA3*, *PISTILLATA*, *AGAMOUS*) [1]. As a result, developmental biologists, for example, have been able to clone almost all the genes providing the floral homeotic functions that, according to the ABC model, act in a combinatorial way to specify floral organ identity [16–18]. Later, more key players of the floral developmental pathway were

☆ Supplementary data associated with this article can be found at doi: 10.1016/j.tplants.2003.09.006

*Corresponding author:* Yves Van de Peer (yves.vandepeer@psb.ugent.be).

http://plants.trends.com 1360-1385/$ - see front matter © 2003 Elsevier Ltd. All rights reserved. doi:10.1016/j.tplants.2003.09.006

identified, leading to the extension of this model to the ABCDE and the protein-based quartet model [18].

Whereas research into floral developmental genes is progressing rapidly, the functional analysis of other MADS-box genes is lagging behind. Nevertheless, MADS-box genes have also been shown to function in the control of fruit development (*SHATTERPROOF1*, *SHATTERPROOF2*, *FRUITFULL*), seed development (e.g. *TRANSPARENT TESTA 16*) and root growth (e.g. *ARABIDOPSIS NITRATE-RESPONSIVE 1*) [19–24].

Unfortunately, when analysing large families such as the MADS-box gene family, one is confronted with several problems. First, owing to the high functional redundancy found in MADS-box genes, the construction of double or even multiple mutants is often inevitable to uncover the complete spectrum of gene functions by mutant phenotype. Because such studies are relatively time consuming, the prediction of functional redundancy by phylogeny reconstructions helps to minimize the effort [20,25–28]. In addition, the incomplete sampling of MADS-box genes in most organisms makes it difficult to assign the correct orthologous and paralogous relationships between genes, and restricts a comprehensive comparison of the gene functions [29]. Moreover, lineage-specific gene family expansion through gene duplication has led to extant plants having established orthologous relationships between clades of paralogous genes rather than between individual genes and could have led to differences in functional divergence of these duplicated genes in different plant lineages [30].

## Genomics reveals new roads ahead

Since the beginning of the 21st century, plant molecular biology has been flooded with a previously unseen amount of sequence data. The completion of the genome sequences of *Arabidopsis* and rice (*Oryza sativa*) now allows the investigation of the full complement of MADS-box genes in both eudicots and monocots [31–33]. The genome-wide structural annotation of the MADS-box gene family in these organisms has resulted in the discovery of more than 100 such genes in *Arabidopsis* (104 genes in Ref. [14], 107 in Ref. [15], 105 in Ref. [34]) and 71 such genes in rice [14] (TIGR annotation; S. De Bodt *et al.*, unpublished). A list of MADS-box genes in selected model species can be found as in the supplementary table in the HTML version of this article[1] on our website (http://www.psb.ugent.be/bioinformatics/ and). The true number of MADS-box genes in rice might be higher than 71 because annotation of the rice genome is far from complete. Structural annotation of the novel type I subfamily in the *Arabidopsis* and rice genomes has resulted in the discovery of 40 (+7 MIKC*, see below) and 37 MADS-box genes, respectively. Additionally, 20 highly diverged MADS-like genes have been identified in the *Arabidopsis* genome, for which no rice homologues have yet been found [14] [S. De Bodt *et al.*, unpublished; see Supplementary Material (http://www.psb.ugent.be/bioinformatics/ and supplementary Table)].[1] However, in the rice genome, there are several genes that possess remnants of the MADS box but have degenerated into pseudogenes through the insertion of stop codons. By contrast, all *Arabidopsis* MADS-like genes consist of complete open reading frames.

The genome-wide identification of MADS-box genes has led to new views of the evolution of the gene family. Through the ongoing *Arabidopsis* genome-sequencing project, a great amount of new data became available that was used to infer the phylogeny of the MADS-box gene family [8]. This phylogenetic analysis, comprising 45 MADS-box genes from *Arabidopsis* and representative genes from animal and fungal species, uncovered for the first time the existence of two MADS-box lineages (type I and type II) in plants, animals and fungi [8] (Figure 1). The authors suggested that the two lineages arose through an ancestral duplication that occurred in a common ancestor of plants, animals and fungi, and that the K domain (specific to plant type II genes) probably evolved in the plant lineage after its divergence from the animals and fungi. Structural analysis of all MADS-box genes has indicated two main differences between type I and type II genes: (i) the absence of the K box in type I genes; and (ii) most type I MADS-box genes are single exon genes, whereas type II genes consist of seven exons on average [8,14].

Phylogeny reconstructions based on a more extensive set of MADS-domain sequences indicated that seven *Arabidopsis* sequences originally assigned to a subtype of type I genes (termed class O genes) might actually represent deviant type II genes (hence now termed MIKC*-type genes) [14]. These genes constitute a novel subtype of plant MIKC-type (type II) genes, which have been marked with an asterisk to distinguish them from the 'classical' MIKC-type genes (which are hence also termed MIKC$^{c}$-type genes) [35]. First analyses have shown that MIKC*-type genes might be expressed mainly in pollen [34]. In line with this, a novel MADS-box gene closely related to the MIKC*-type genes from *Arabidopsis* was recently identified in *Nicotiana tabacum* through its differential expression in pollen [36] (see Supplementary Material, http://www.psb.ugent.be/bioinformatics/ and supplementary Table).[1]

### Type I genes

Plant type I genes (in the narrow sense, i.e. not including the MIKC*-type genes) have an evolutionary dynamic that is significantly different from that of both animal type I (SRF) and plant type II (MIKC) genes. For example, their rate of evolution is much higher than that of plant type II genes [14]. One possible explanation for this is that the functional constraint on type I genes is lower than on type II genes, and that type I genes are therefore of less functional importance to the plant. This could be why no mutant loss-of-function phenotype has ever been reported for a plant type I gene. By contrast, single or (for redundant genes) multiple mutant phenotypes are known for 18 type II MADS-box genes from *Arabidopsis*, and for many other plant type II genes [29]. The absence of mutant phenotypes for type I genes could be because of their functional redundancy with other genes, which have also been shown for plant type II genes [20,25–28]. Another

[1] See supplementary table in the HTML version of this article.

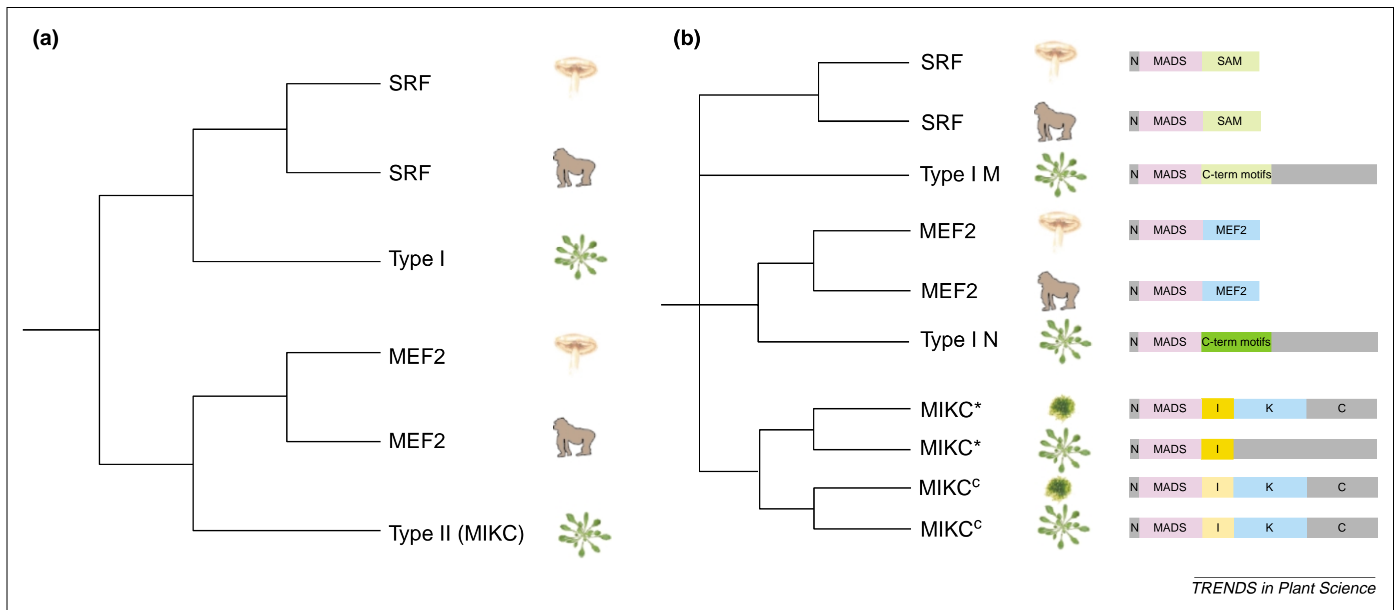



**Figure 1**. Evolution and structure of MADS-box genes of higher plants (indicated by an image of *Arabidopsis*), mosses (indicated by an image of a moss), animals (indicated by an image of a gorilla) and fungi (indicated by the image of a fungus), according to **(a)** Ref. [8] and **(b)** new data and alternative approaches. Abbreviation: C-term, C-terminal.

explanation is that plant type I genes have only subtle functions or work only under exceptional environmental conditions. In line with this, the expression level of most plant type I genes, if they are expressed at all, is much lower than that of type II genes. For example, for many type I genes, expression could only be detected by reverse transcription polymerase chain reaction (RT-PCR), whereas it was impossible to detect the expression of most type I genes through RNA gel blot analysis, macroarray or *in situ* hybridization [15,34]. In addition, there are some cases in which expression was detected using a macroarray approach, whereas expression was not detectable via RT-PCR (e.g. *AGL103*, *AGL34*) [15,34]. A meticulous analysis (e.g. including more tissues and conditions) of the expression patterns of these genes will be needed to resolve these issues.

Another reason why no type I gene mutant phenotype is known could be that type I genes are (evolving into) pseudogenes. Because many type I genes consist only of a single exon, one could assume that they represent processed pseudogenes. There is evidence that at least one type I MADS-box gene (At5g49490) is such a pseudogene because a poly A-tail is found downstream of the gene (S. De Bodt *et al.*, unpublished). Another possibility is that these genes arose through (retro)transposition. In this case, we expect to find repeat and known transposon-like sequences close to these genes. For some type I genes, short repeats can indeed be found 1kb up- and downstream, but for most type I genes, remnants of (retro)-transposition cannot be found. However, this does not rule out (retro)transposition events early in the history of type I genes. Moreover, type I genes are mainly located on chromosomes 1 and 5 [14,15,34], which can be assigned to the preferential local integration of several plant transposons. In addition, it is still possible that plant type I genes represent an absolutely novel and unprecedented class of transposable elements lacking any sequence hallmarks defined before. Transposons carrying a MADS box would not be unprecedented. In maize and its relatives, *En/Spm*-like transposable elements have been identified that have captured a MADS box and have distributed it throughout the maize genome [37,38]. However, these elements contain an *AGAMOUS*-like (hence type II) MADS box, and share no other domains with the type I genes.

Do type I genes have a function? The fact that type I genes in *Arabidopsis* and rice contain similar C-terminal motifs suggests sequence conservation caused by functional constraint, in spite of the high evolutionary rate of type I genes. However, function does not necessarily imply a function for the host plant. The alternative could be that type I sequences, rather than being conventional genes, represent transposable elements or some other kind of 'selfish' sequence elements.

Recently, it has been shown that the type I gene *PHERES1* (*AGL37*) is transiently expressed during embryo and endosperm development, and that upregulation of *PHERES1* in Polycomb-group gene mutants such as *medea* causes developmental defects including seed abortion. Moreover, *PHERES1* is obviously a direct target gene of some Polycomb-group proteins including MEDEA [39]. These findings raise the hope that a function can soon be assigned to at least one plant type I MADS-box gene.

However, no current hypothesis about plant type I genes fits all the data satisfactorily, and maybe no single hypothesis ever will if type I genes are a phylogenetically or functionally heterogeneous class of genes. To solve the frustrating conundrum of type I genes, comprehensive and careful analysis of type I gene mutants obtained by, for example, reverse genetic screens, will elucidate whether these sequence elements are functionally important to the plants. To circumvent putative problems with redundancy, the generation of double or even multiple gene knockouts (guided by phylogeny reconstructions) might prove

necessary. Although loss-of-function phenotypes for several genes will almost certainly exclude the transposon hypothesis (at least for the respective genes), the inability to identify phenotypes would be less conclusive, because lack of a recognizable phenotype does not necessarily mean that the gene has no function. However, in these cases, the defining characteristic of transposons (their ability to change their chromosomal position) might reveal the mobile character of these sequence elements. Transposition of mobile elements might be observed by Southern blot analysis or a technique called transposon display, as recently demonstrated for an active transposon family in rice [40].

### *Type II genes*

Whereas the function of type I genes largely remains a mystery, the functional importance of type II MADS-box genes has been shown through both the functional characterization of single MADS-box genes and moderate- to large-scale cDNA sequencing projects in diverse plants such as the eudicot angiosperm *Petunia hybrida* [41], the monocot *Z. mays* [42], the gymnosperms *Gnetum gnemon* [43,44], *Pinus radiata* [45–47], *Picea abies* [48] and *Ginkgo biloba* [49], the fern *Ceratopteris richardii* [50,51], and the moss *Physcomitrella patens* [35]. An overview of the current status of MADS-box gene sequences from *Arabidopsis*, *O. sativa*, *Z. mays* and *P. hybrida*, and their functions (where known) in these plants is given in the Supplementary Material (http://www.psb.ugent.be/bioinformatics/ and supplementary Table)[1].

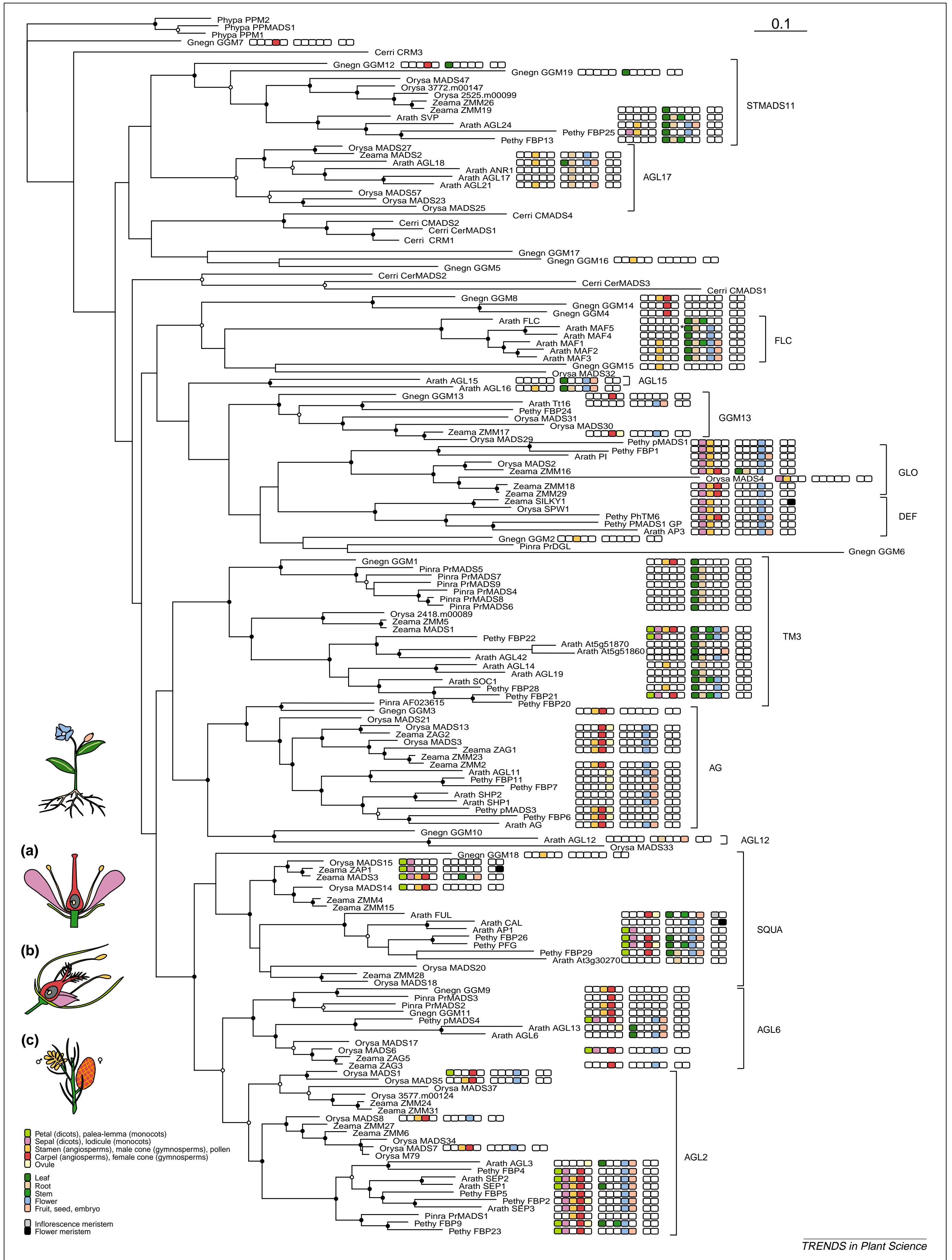
The cDNA sequencing efforts have allowed phylogenetic analyses of type II MADS-box genes. These genes can be subdivided into distinct clades, each clade comprising orthologues from different seed plants. However, MADS-box genes from ferns and mosses cannot be assigned to any of these clades as yet and probably possess a more ubiquitous expression and function than their counterparts in flowering plants [2,35,50,51]. Thus, the study of these genes enables us to correlate the appearance of new types (clades) of developmental control genes with the origin of novel morphological structures (such as ovules or seeds and flowers) in plants [2].

Two different approaches have been used to date the origin of the distinct clades of type II MADS-box genes and to correlate the evolution of MADS-box genes with the divergence of major plant lineages. One group studied gene sampling [29,44] and obtained estimates of 300 million to 400 million years for the origins of many type II gene clades, whereas others [52] used molecular-clock-based dating, leading to much older age estimates. The second study [52] implies that class B and class C floral homeotic gene lineages originated ~660 million and 570 million years ago, respectively – before the separation of the lineages that led to mosses, ferns and seed plants [52]. This suggests that representatives of these clades were either lost in extant mosses and ferns or are present but have simply not been identified. Another explanation is that type II genes in the lineage that led to extant ferns evolved faster than genes in the seed plant lineage, so that fern orthologues of seed plant genes cannot be recognized anymore. Alternatively, molecular clock estimates extrapolating from gymnosperm and angiosperm data might overestimate the ages of the clades because type II gene evolution in the lineage that led to extant seed plants could have been much faster 300 million to 400 million years ago (after the fern lineage split off) and slowed down 300 million years ago, after the angiosperm–gymnosperm split. If so, it would be interesting to find out which changes (e.g. in gene functions, modes of protein–protein interactions) can be correlated with these differences in evolutionary rate.

Figure 2 shows a phylogenetic tree of MIKC$^{c}$ genes from diverse plant lineages with their expression patterns in distinct tissues. This generally corroborates the view that members of the same gene subfamilies tend to have similar expression patterns [53], but it also demonstrates that this correlation is stricter for genes involved in flower formation than for genes mainly expressed elsewhere [29]. In some cases, lineage-specific expansions led to the occurrence of orthologous pairs of genes that possess a distinct pattern of divergence on expression level; both genes of one pair have kept the expression pattern of their ancestral gene, suggesting functional redundancy, whereas genes of the other pair have subdivided the expression pattern, resulting in genes with a more specific functional activity (e.g. expression of *AP1/CAL* from *Arabidopsis* and *PFG/FBP26* from *P. hybrida* in reproductive structures).

The extensive analysis of the *Arabidopsis* MIKC-type genes has allowed the transfer of knowledge about functions to orthologous genes from other plants through the principle of 'phylogenomics' [54]. In particular, high functional redundancy [found through the analyses of *Arabidopsis* MADS-box genes (e.g. *SEPALLATA* and *SHATTERPROOF* genes)] can be anticipated in similar studies in other organisms [25–28]. Conversely, the analysis of MADS-box genes in species other than *Arabidopsis* has provided us with greater insights into *Arabidopsis* genes. For example, studies in the gymnosperm *G. gnemon* led to the discovery of a novel MADS-box gene subfamily with a sister-group relationship to the class B genes, having members in *G. gnenom* (*GGM13*) but also *Arabidopsis* [*Arabidopsis B-sister* (*ABS*)], *Z. mays* (*ZMM17*) and other plants [55]. An independent and parallel functional characterization of the *ABS* gene [or *TRANSPARENT TESTA16* (*TT16*)] demonstrated its involvement in endothelial cell specification and in the

**Figure 2**. Phylogenetic tree and expression patterns (where known) of MIKC genes from **(a)** the eudicots *Arabidopsis* (Arath) and *Petunia hybrida* (Pethy), **(b)** the monocots *Oryza sativa* (Orysa) and *Zea mays* (Zeama), **(c)** the gymnosperms *Pinus radiata* (Pinra) and *Gnetum gnemon* (Gnegn), the fern *Ceratopteris richardii* (Cerri), and the moss *Physcomitrella patens* (Phypa). The phylogenetic tree is constructed using MrBayes ($10^6$ generations, four chains). Nodes supported by posterior probabilities higher than 70 are denoted by a black dot, posterior probabilities between 50 and 70 by an open circle. The scale indicates 0.1 substitutions per site. Expression patterns (tissue specific) are extracted from literature on specific genes, and from genome-wide analyses [15,34]. In case of conflict, preference was given to indicate a gene as being expressed when more-sensitive approaches (e.g. RT-PCR) gave a positive result and others (e.g. macroarrays and northern blots) did not. The expression of genes that could only be detected through macroarray analysis [34] and not through other methods is marked with an asterisk (*).



genetic control of seed coat pigmentation [24]. In addition, MADS cDNA sequencing in the moss *P. patens* led to the identification of an additional class of MIKC genes that possess a divergent I box and are referred to as MIKC*-type genes, as mentioned above [35]. So, two interesting classes of MADS-box genes have been identified first in lower plants (i.e. a moss and a gymnosperm) rather than in the model plant *Arabidopsis*. Moreover, comparative analysis of the MADS-box gene family in angiosperms, and of the *APETALA/PISTILLATA* clade of genes in particular, uncovered distinct C-terminal motifs that can be correlated with their functional specificity. At least for some genes, it has recently been shown that these specific motifs probably arose through one (or more) nucleotide insertions or deletions, causing translational frame shifts, and subsequent sequence conservation [56]. What is remarkable about this finding is that frame-shift mutations in C-terminal regions of duplicate genes are selected for and hence the gene has been retained together with the unchanged gene duplicate. 3′-Terminal frame-shift mutations might therefore represent an important novel mechanism in the functional diversification of transcription factor gene families [56].

### Classification

The results of recent genome-wide studies, like those described above, urge an unambiguous definition and nomenclature for the different classes of MADS-box genes, preferably based on careful, evolutionary analyses (Table 1). Unfortunately, the phylogenetic analyses of the whole gene family in *Arabidopsis* and rice result in poorly resolved trees, mainly owing to the combination of a limited number of phylogenetically informative positions in the short MADS-domain (60 amino acids) and the large number of genes [14]. Therefore, type I MADS-box genes were first classified based on structural characteristics rather than on poorly resolved phylogenetic trees, resulting in the class M and N type I genes, which can be distinguished through the presence of conserved, C-terminal motifs [14]. Detailed phylogenetic analyses of these classes in both *Arabidopsis* and rice showed extensive expansion of the number of these genes after the divergence of monocots and eudicots [14]. To reconstruct the evolution of other complex gene families, alternative approaches have been used, such as limiting the number of genes and choosing only genes from a few representative species [57]. Another solution is to replace well-supported clades of genes by their ancestral sequence. The two latter approaches, applied to MADS-box genes, give a topology (Figure 1b) that contrasts with previous results [8] (Figure 1a). Although these approaches cannot unequivocally resolve the deeper branching order between subclades of MADS-box genes, they suggest a polyphyletic origin for different groups of type I genes. However, it remains difficult to elucidate the evolutionary relationships between these different groups of type I genes and their animal and fungal counterparts. More extensive sampling of MADS-box genes from diverse species, including basal plants, should contribute to the reconstruction of the evolutionary history of the MADS-box gene family. In the future, more large-scale sequencing projects (such as the floral genome project [58]) combined with high-throughput functional characterization approaches will undoubtedly enable more comprehensive comparative analyses (both functional and evolutionary) and should consequently enable us to gain deeper insights into the role of different classes of MADS-box genes (types I and II) in the evolution of the gene family and in plant development.

## Functional genomics provides the tools

The availability of complete genome sequences and large sets of expressed sequence tags has triggered the development of high-throughput methods to analyse these raw data functionally. Oligonucleotide and cDNA microarrays now allow the genome-wide analysis of spatial and temporal expression patterns [59,60]. To gain insights into the expression of regulatory genes such as MADS-box genes, specific arrays are being designed to profile these genes [61]. In addition, the effect of MADS-box gene perturbation can be analysed using microarrays, which allows the identification of the downstream genes in the developmental pathway. For example, the global identification of target genes regulated by class B floral homeotic genes *APETALA3* and *PISTILLATA* was conducted through the use of cDNA microarrays [62]. Similar analyses are being conducted, although on a smaller scale, in other plants.

**Table 1. Classification of the MADS-box genes in *Arabidopsis***

| Alvarez-Buylla *et al.* [8] | De Bodt *et al.* [14] | Parenicova *et al.* [15] | Kofuji *et al.* [34] |
|---|---|---|---|
| Type I (SRF-like) | Type I M | Mα[a,d,e] | M[a,f,g] |
| Type I (SRF-like) | Type I N | Mγ[c,e] | M |
| Type I (SRF-like) | MIKC*[g,i] | Mδ[g] | MIKC* |
| Type I (SRF-like) | Type I O[e,f,h] | – | M |
| Type I (SRF-like) | MADS-like | Mβ[b,e] | M |
| Type II (MEF2-like) | Type II | MIKC[f] | MIKC[ci] |

[a]At1g29960 and At1g54760 are assigned to class Mα by Parenicova *et al.* [15], to class M by Kofuji *et al.* [34] and are not identified as MADS-box genes by De Bodt *et al.* [14].
[b]At4g02240 and At5g37420 are assigned to class Mβ by Parenicova *et al.* [15] and are not identified as MADS-box genes by De Bodt *et al.* [14] and Kofuji *et al.* [34].
[c]At2g15660 is assigned to class Mγ according to Parenicova *et al.* [15] and is not identified as a MADS-box gene by De Bodt *et al.* [14] and Kofuji *et al.* [34].
[d]At1g46408 was identified for the first time in Parenicova *et al.* [15] and belongs to class Mα.
[e]At1g72350 and At1g17310 are assigned to class Mα, At5g06500 and At1g22590 to class Mγ and At5g26950, At5g58890 and At5g55690 to class Mβ by Parenicova *et al.* [15], and belong to type I O according to De Bodt *et al.* [14].
[f]At1g31140 is assigned to class type I O by De Bodt *et al.* [14], to class M by Kofuji *et al.* [34], but to class MIKC by Parenicova *et al.* [15].
[g]Originally considered Type I O, but then identified as MIKC* by De Bodt *et al.* [14]. At2g26320 is assigned to MIKC* by De Bodt *et al.* [14], to class Mδ by Parenicova *et al.* [15], and to class M by Kofuji *et al.* [34].
[h]Type I O genes *sensu stricto* are class O genes according to De Bodt *et al.* [14], except the MIKC* genes mentioned in the same paper.
[i]Term MIKC[c] and MIKC* introduced by Henschel *et al.* [35].

For example, the effects of tomato *ripening-inhibitor* (*rin*) and *non-ripening* (*nor*) mutants on gene expression are being investigated using various genomics tools [63].

Large-scale interaction studies such as yeast two- and three-hybrid screens and fluorescence-resonance-energy transfer (FRET) analyses provide insights into protein–protein and RNA–protein interactions [64–66]. FRET analysis has been shown to be effective in the identification of dimeric complexes of MADS-domain proteins involved in flower development *in planta* [66, 69]. Moreover, yeast one-hybrid experiments are used to detect protein–DNA interactions and to isolate new proteins that bind to a specific target (regulatory) element [67]. These experiments can be conducted on a large scale when an extensive collection of promoters and their *cis*-acting regulatory elements is available for plants. Chromatin immunoprecipitation recently allowed the identification of several targets of AGL15 [68,69]. The development of microarrays containing regulatory regions for all *Arabidopsis* genes should speed up the detection of candidate target genes using this approach, as has been demonstrated in yeast [70] and *H. sapiens* [71]. In parallel with these experimental studies, *in silico* analyses of promoters can be exerted using clusters of co-regulated genes or through a comparative approach using homologous genes in different organisms [72–74].

As such, a complete survey of all genes, including the largely unexplored type I MADS-box and MADS-like genes, can be compiled in an efficient way, giving a glimpse of the processes in which these genes are active and that can be used to select interesting genes for more in-depth analyses on both the RNA and protein level.

## Conclusion and outlook

These are exciting times in the MADS world. The availability of complete genomes and the rise of novel sophisticated technologies open up many possibilities for plant research. Thanks to the combination of comparative developmental biology and genomics, exciting new insights are being revealed in the evolution of development and the underlying regulatory mechanisms. To be most profitable, efforts should focus on plant species of evolutionary importance, for which genetic and genomic tools exist or can be developed [75]. However, the choice of adequate model systems is not self-evident owing to the large genome size and the long generation time of many phylogenetically interesting plants (e.g. gymnosperms). By contrast, the moss *P. patens* is an example of a particularly interesting and useful plant model organism, not only because it has a small genome and is easy to grow, but also because it is the only land plant that is amenable to efficient gene targeting via homologous recombination [76,77].

We believe that only an integrative approach, combining classical genetics, functional genomics, bioinformatics and comparative genomics, will be able to unravel the evolution and functional divergence of large transcription factor families such as the MADS-box gene family. Probably, future research will even go beyond this comprehensive 'phylogenomics' approach because there is much evidence that the specificity of MADS-box gene action is conferred by combinatorial protein–protein interactions [78]. Examples include the quartet model [18] and some others, termed 'the second model' and 'a third model' [79], describing the specification of floral organ identity. We predict, therefore, that future studies will focus more and more on trying to understand MADS-domain protein–protein interactions. Using techniques such as X-ray crystallography, nuclear magnetic resonance, FRET, gel retardation assays and the yeast two-hybrid system in a phylogenetic context, 'phyloproteomics' of MADS-domain transcription factors might be on the horizon.


## Acknowledgements

S.D.B. is indebted to the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT) for a predoctoral fellowship. We thank three anonymous reviewers for helpful comments on an earlier version of the manuscript.


## References


1 Ng, M. and Yanofsky, M.F. (2001) Function and evolution of the plant MADS-box gene family. *Nat. Rev. Genet.* 2, 186–195

2 Theißen, G. *et al.* (2000) A short history of MADS-box genes in plants. *Plant Mol. Biol.* 42, 115–149

3 Passmore, S. *et al.* (1988) *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MAT α cells. *J. Mol. Biol.* 204, 593–606

4 Yanofsky, M.F. *et al.* (1990) The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* 346, 35–39

5 Sommer, H. *et al.* (1990) *DEFICIENS*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J.* 9, 605–613

6 Schwarz-Sommer, Z. *et al.* (1992) Characterization of the *Antirrhinum* floral homeotic MADS-box gene *DEFICIENS*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *EMBO J.* 11, 251–263

7 Norman, C. *et al.* (1988) Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-Fos serum response element. *Cell* 55, 989–1003

8 Alvarez-Buylla, E.R. *et al.* (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5328–5333

9 Immink, R.G. *et al.* (2002) Analysis of MADS box protein–protein interactions in living plant cells. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2416–2421

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

11 Egea-Cortines, M. *et al.* (1999) Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *EMBO J.* 18, 5370–5379

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

13 Lamb, R.S. and Irish, V.F. (2003) Functional divergence within the *APETALA3/PISTILLATA* floral homeotic gene lineages. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6558–6563

14 De Bodt, S. *et al.* (2003) Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. *J. Mol. Evol.* 56, 573–586

15 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

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

17 Weigel, D. and Meyerowitz, E.M. (1994) The ABCs of floral homeotic genes. *Cell* 78, 203–209

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

19 Rounsley, S.D. *et al.* (1995) Diverse roles for MADS box genes in *Arabidopsis* development. *Plant Cell* 7, 1259–1269

20 Liljegren, S.J. *et al.* (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404, 766–770
21 Zhang, H. and Forde, B.G. (2000) Regulation of *Arabidopsis* root development by nitrate availability. *J. Exp. Bot.* 51, 51–59
22 Burgeff, C. *et al.* (2002) MADS-box gene expression in lateral primordia, meristems and differentiated tissues of *Arabidopsis thaliana* roots. *Planta* 214, 365–372
23 Ferrandiz, C. *et al.* (2000) Negative regulation of the *SHATTERPROOF* genes by *FRUITFULL* during *Arabidopsis* fruit development. *Science* 289, 436–438
24 Nesi, N. *et al.* (2002) The *TRANSPARENT TESTA16* locus encodes the *Arabidopsis* BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* 14, 2463–2479
25 Riechmann, J.L. *et al.* (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290, 2105–2110
26 Pelaz, S. *et al.* (2000) B and C floral organ identity function require *SEPALLATA* MADS-box genes. *Nature* 405, 200–203
27 Smyth, D. (2000) A reverse trend – MADS functions revealed. *Trends Plant Sci.* 5, 315–317
28 Pinyopich, A. *et al.* (2003) Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424, 85–88
29 Becker, A. and Theißen, G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylog. Evol.* (in press)
30 Theißen, G. and Becker, A. The ABCs of flower development in *Arabidopsis* and rice. *Prog. Bot.* (in press)
31 Arabidopsis Genome Initiative, (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
32 Goff, S.A. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–100
33 Yu, J. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296, 79–92
34 Kofuji, R. *et al.* (2003) Evolution and divergence of MADS-box gene family based on genome wide expression analyses. *Mol. Biol. Evol.* 10.1093/molbev/msg216 (http://mbe.oupjournals.org/cgi/reprint/msg216v1)
35 Henschel, K. *et al.* (2002) Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. *Mol. Biol. Evol.* 19, 801–814
36 Steiner, C. *et al.* (2003) Two novel genes are differentially expressed during early germination of the male gametophyte of *Nicotiana tabacum*. *Biochim. Biophys. Acta* 1625, 123–133
37 Fischer, A. *et al.* (1995) Chromosomal mapping of the MADS-box multigene family in *Zea mays* reveals dispersed distribution of allelic genes as well as transposed copies. *Nucleic Acids Res.* 23, 1901–1911
38 Montag, K. *et al.* (1996) The *ZEM2* family of maize MADS box genes possess features of transposable elements. *Maydica* 41, 241–254
39 Köhler, C. *et al.* (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. *Genes Dev.* 17, 1540–1553
40 Jiang, N. *et al.* (2003) An active DNA transposon family in rice. *Nature* 421, 163–167
41 Immink, R.G. *et al.* (2003) Analysis of the petunia MADS-box transcription factor family. *Mol. Genet. Genomics* 268, 598–606
42 Münster, T. *et al.* (2002) Maize MADS-box genes galore. *Maydica* 47, 287–301
43 Winter, K.U. *et al.* (1999) MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 96, 7342–7347
44 Becker, A. *et al.* (2000) MADS-box gene diversity in seed plants 300 million years ago. *Mol. Biol. Evol.* 17, 1425–1434
45 Mouradov, A. *et al.* (1999) A *DEF/GLO*-like MADS-box gene from a gymnosperm: *Pinus radiata* contains an ortholog of angiosperm B class floral homeotic genes. *Dev. Genet.* 25, 245–252
46 Mouradov, A. *et al.* (1998) Family of MADS-box genes expressed early in male and female reproductive structures of Monterey pine. *Plant Physiol.* 117, 55–62
47 Walden, A.R. *et al.* (1998) A large family of TM3 MADS-box cDNAs in *Pinus radiata* includes two members with deletions of the conserved K domain. *Plant Sci.* 138, 167–176
48 Rutledge, R. *et al.* (1998) Characterization of an AGAMOUS homologue from the conifer black spruce (*Picea mariana*) that produces floral homeotic conversions when expressed in *Arabidopsis*. *Plant J.* 15, 625–634
49 Jager, M. *et al.* (2003) MADS-box genes in *Ginkgo biloba* and the evolution of the *AGAMOUS* family. *Mol. Biol. Evol.* 20, 842–854
50 Münster, T. *et al.* (1997) Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proc. Natl. Acad. Sci. U. S. A.* 94, 2415–2420
51 Hasebe, M. *et al.* (1998) Characterization of MADS homeotic genes in the fern *Ceratopteris richardii*. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6222–6227
52 Nam, J. *et al.* (2003) Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol. Biol. Evol.* 20, 1435–1447
53 Theißen, G. *et al.* (1996) Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* 43, 484–516
54 Eisen, J.A. and Wu, M. (2002) Phylogenetic analysis and gene functional predictions: phylogenomics in action. *Theor. Popul. Biol.* 61, 481–487
55 Becker, A. *et al.* (2002) A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Mol. Genet. Genomics* 266, 942–950
56 Vandenbussche, M. *et al.* (2003) Structural diversification and neo-functionalisation during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucleic Acids Res.* 31, 4401–4409
57 Bharathan, G. *et al.* (1997) Did homeodomain proteins duplicate before the origin of angiosperms, fungi, and metazoa? *Proc. Natl. Acad. Sci. U. S. A.* 94, 13749–13753
58 Soltis, D.E. *et al.* (2002) Missing links: the genetic architecture of flowers and floral diversification. *Trends Plant Sci* 7, 22–34.(erratum: *Trends Plant. Sci.* 7, 95)
59 Lockhart, D.J. and Winzeler, E.A. (2000) Genomics, gene expression and DNA arrays. *Nature* 405, 827–836
60 Schulze, A. and Downward, J. (2001) Navigating gene expression using microarrays – a technology review. *Nat. Cell Biol.* 3, E190–E195
61 Paz-Ares, J. (2002) REGIA, an EU project on functional genomics of transcription factors from *Arabidopsis thaliana*. *Comp. Funct. Genomics* 3, 102–108
62 Zik, M. and Irish, V.F. (2003) Global identification of target genes regulated by *APETALA3* and *PISTILLATA* floral homeotic gene action. *Plant Cell* 15, 207–222
63 Moore, S. *et al.* (2002) Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato. *J. Exp. Bot.* 53, 2023–2030
64 Fields, S. and Song, O. (1989) A novel genetic system to detect protein–protein interactions. *Nature* 340, 245–246
65 Sengupta, D.J. *et al.* (1996) A three-hybrid system to detect RNA–protein interactions *in vivo*. *Proc. Natl. Acad. Sci. U. S. A.* 93, 8496–8501
66 Immink, R.G. and Angenent, G.C. (2002) Transcription factors do it together: the hows and whys of studying protein–protein interactions. *Trends Plant Sci.* 7, 531–534
67 Luo, Y. *et al.* (1996) Cloning and analysis of DNA-binding proteins by yeast one-hybrid and one–two-hybrid systems. *Biotechniques* 20, 564–568
68 Fernandez, D.E. *et al.* (2000) The embryo MADS domain factor AGL15 acts postembryonically. Inhibition of perianth senescence and abscission via constitutive expression. *Plant Cell* 12, 183–198
69 Wang, H. *et al.* (2002) A chromatin immunoprecipitation (ChIP) approach to isolate genes regulated by AGL15, a MADS domain protein that preferentially accumulates in embryos. *Plant J.* 32, 831–843
70 Ren, B. *et al.* (2000) Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309
71 Weinmann, A.S. and Farnham, P.J. (2002) Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* 26, 37–47
72 Koch, M.A. *et al.* (2001) Comparative genomics and regulatory

evolution: conservation and function of the *CHS* and *APETALA3* promoters. *Mol. Biol. Evol.* 18, 1882–1891
73 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
74 Rombauts, S. *et al.* (2003) Computational approaches to identify promoters and *cis*-regulatory elements in plant genomes. *Plant Physiol.* 132, 1162–1176
75 Pryer, K.M. *et al.* (2002) Deciding among green plants for whole genome studies. *Trends Plant Sci.* 7, 550–554
76 Rensing, S.A. *et al.* (2002) Moss transcriptome and beyond. *Trends Plant Sci.* 7, 535–538
77 Schaefer, D. (2001) Gene targeting in *Physcomitrella patens*. *Curr. Opin. Plant Biol.* 4, 143–150
78 Messenguy, F. and Dubois, E. Role of MADS-box proteins and their cofactors as promoter architects in combinatorial control of gene expression and cell development. *Gene* (in press)
79 Jack, T. (2001) Relearning our ABCs: new twists on an old model. *Trends Plant Sci.* 6, 310–316

***Endeavour***

the quarterly magazine for the history and philosophy of science

You can access *Endeavour* online either through your *BioMedNet* Reviews subscription or via *ScienceDirect*, where you'll find a collection of beautifully illustrated articles on the history of science, book reviews and editorial comment.

featuring

**The pathway to the cell and its organelles: one hundred years of the Golgi apparatus** by M. Bentivoglio and P. Mazzarello
**Joseph Fourier, the 'greenhouse effect' and the quest for a universal theory of terrestrial temperatures** by J.R. Fleming
**The hunt for red elixir: an early collaboration between fellows of the Royal Society** by D.R. Dickson
**Art as science: scientific illustration 1490–1670 in drawing, woodcut and copper plate** by C.M. Pyle
**The history of reductionism versus holistic approaches to scientific research** by H. Andersen
**Reading and writing the Book of Nature: Jan Swammerdam (1637–1680)** by M. Cobb
**Coming to terms with ambiguity in science: wave–particle duality** by B.K. Stepansky
**The role of museums in history of science, technology and medicine** by L. Taub
**The 'internal clocks' of circadian and interval timing** by S. Hinton and W.H. Meck
**The troubled past and uncertain future of group selectionism** by T. Shanahan
**A botanist for a continent: Ferdinand Von Mueller (1825–1896)** by R.W. Home
**Rudolf Virchow and the scientific approach to medicine** by L. Benaroyo
**Darwinism and atheism: different sides of the same coin?** by M. Ruse
**Alfred Russel Wallace and the flat earth controversy** by C. Garwood
**John Dalton: the world's first stereochemist** by Dennis H. Rouvray
**Forensic chemistry in 19th-century Britain** by N.G. Coley
**Owen and Huxley: unfinished business** by C.U.M. Smith
**Characteristics of scientific revolutions** by H. Andersen

and much, much more . . .

Locate *Endeavour* in the *BioMedNet Reviews* collection.
Log on to http://reviews.bmn.com, hit the 'Browse Journals' tab and scroll down to *Endeavour*