Plant Molecular Biology **42:** 115–149, 2000.

A short history of MADS-box genes in plants

Günter Theissen*, Annette Becker, Alexandra Di Rosa, Akira Kanno, Jan T. Kim, Thomas Münster, Kai-Uwe Winter and Heinz Saedler

*Max-Planck-Institut für Züchtungsforschung, Abteilung Molekulare Pflanzengenetik, Carl-von-Linné-Weg 10, 50829 Köln, Germany (*author for correspondence)*

Key words: angiosperm, development, evolution, fern, gymnosperm, MADS-box gene

Abstract

Evolutionary developmental genetics (evodevotics) is a novel scientific endeavor which assumes that changes in developmental control genes are a major aspect of evolutionary changes in morphology. Understanding the phylogeny of developmental control genes may thus help us to understand the evolution of plant and animal form. The principles of evodevotics are exemplified by outlining the role of MADS-box genes in the evolution of plant reproductive structures. In extant eudicotyledonous flowering plants, MADS-box genes act as homeotic selector genes determining floral organ identity and as floral meristem identity genes. By reviewing current knowledge about MADS-box genes in ferns, gymnosperms and different types of angiosperms, we demonstrate that the phylogeny of MADS-box genes was strongly correlated with the origin and evolution of plant reproductive structures such as ovules and flowers. It seems likely, therefore, that changes in MADS-box gene structure, expression and function have been a major cause for innovations in reproductive development during land plant evolution, such as seed, flower and fruit formation.

Introduction: on the origin of novel structures during evolution

We explain here what evolutionary developmental genetics (evodevotics) is, and how it may help us to understand the evolution of diversity and complexity in the living world. We present one of the most important corollaries of evodevotics, that changes in developmental control genes might be a major cause of evolutionary changes in morphology.

Higher organisms such as plants and animals impress us with their complexity and their diversity. Take plants as an example. Every tiny weed you can find on a little walk around the corner is by far more complex than anything we know from outside the living world, and the diversity of plants is breath-taking ranging, for example, from huge oak trees to microscopic green algae on their bark. Understanding the laws of nature that have generated that diversity and complexity is at the very heart of biology.

Initially, one can try to understand complex organisms from an engineer's point of view – an attitude which already has quite some explanatory power. For example, interpreting leaves as efficient sun-collectors explains why these are generally flat and oriented towards the sun. However, functional explanations have serious limitations in the living world. Why, for example, do the flowers of some plants have three organs (sepals, petals or tepals) in each whorl of their perianth (such as Liliaceae), while others have four (e.g. Brassicaceae) or five (e.g. Rosaceae), if any number of perianth organs is able to attract pollinators efficiently? Why do mammals usually walk on four limbs, insects on six and spiders on eight, if any even number of limbs allows efficient locomotion on land?

The difficulties with explanations that would satisfy engineers in the living world arise from the fact that all features of living organisms are a product both of necessity and chance during evolution [77]. Some

The MADS homepage: http://www.mpiz-koeln.mpg.de/mads/

aspects of living creatures merely trace back to chance events that became fixed during evolution and cannot be reduced to anything more meaningful. This is one of the reasons why living beings can be fully understood only from an evolutionary perspective which takes their unique 'history' into account. Unfortunately, it does not mean that evolutionary theory has already provided us with a complete understanding of the origin of complex and diverse structures in nature. On the contrary, understanding the mechanisms that generated complex organisms, such as oak trees and green algae, from bacteria-like ancestors is still one of the greatest intellectual challenges. The origin of novel structures or complete new body plans during evolution has been especially difficult to explain. Some of the problems arise from the fact that the 'classical', i.e. Darwinian evolutionary theory is a gradualistic one, which assumes that evolution proceeds in a countless number of very small steps, while, on the other hand, partial or intermediate structures might not have an adaptive value.

New ideas are needed to gain a better understanding of the origin of complexity and diversity in the living world or old ones have to be revitalized. One of the most promising concepts in that respect is evolutionary developmental biology, which has strong historical roots reaching back into the 19th century, but is now fashionable again under the term 'evo-devo' [37, 41, 124]. Evo-devo assumes that there is a close interrelationship between developmental and evolutionary processes. One of the reasons for this is an astonishing feature of higher organisms: that even the most complex organisms are generated by developmental processes that generally start with a single cell – the fertilized egg-cell (or zygote). Diversity and complexity thus do not only have evolutionary origins and causes, but also developmental ones [6]. In the case of multicellular organisms such as animals and plants, evolution of form is thus the evolution of developmental processes, and any phylogenetic innovation has to be compatible with the mode of development in a given organismic lineage. This is why development may put serious constraints on evolution, which could act both as negative forces preventing advantageous alterations as well as positive channels of preferred change [41].

From the close interdependence of development and evolution, one of the most important corollaries of evo-devo can be derived, namely that changes in developmental control genes might be a major cause of evolutionary changes in morphology [124]. Understanding the phylogeny of developmental control genes is therefore an important prerequisite for understanding the evolution of plant and animal form (note that we use 'developmental control gene' here as a convenient term for genes which significantly contribute to developmental processes; for a critical discussion of the term, see [125], and references therein). One can assume that the combination of evolutionary developmental biology with molecular genetics will provide deep insights into the mechanisms behind macroevolution. Since it is the genes that connect evolutionary and developmental processes, this novel combination of traditionally separated biological disciplines deserves a new name: evodevotics (for evolutionary developmental genetics). A very strong molecular genetic aspect clearly distinguishes evodevotics from its historical precursors.

In recent years, it has been discovered that the key developmental control genes are often members of a very limited number of multigene families which encode transcription factors. The paradigm for such gene families are the homeobox genes [35], which play a key role in the specification of the animal body plan in both development and evolution [56, 70, 114]. Many of the homeobox genes act as homeotic selector genes which are involved in differentiating different body regions from each other, probably by activating or repressing different sets of downstream genes ('target or realizator genes') in different parts of the body. Unfortunately, studying homeobox genes and animals alone will not allow us to detect all of the fundamental laws of macroevolution. All extant animals probably are relatively closely related members of a monophyletic group. Their body plans, though very diverse, were generated in a relatively short period of time about 540 million years ago (MYA) – hence that process has been termed the 'Cambrian explosion' [93]. In many cases, therefore, to distinguish necessities of macroevolutionary events from mere chance events that have been fixed in evolution, is impossible from studying only animals. For example, all animals specify their body plan in a very similar way, by using a well defined set of homeobox genes (*HOX* genes) which are organized in genomic clusters [101, 114]. However, the absence of *HOX* clusters in plants [73] tells us that the presence of such genes is not an absolute requirement for the evolution of complex multicellular body plans, a conclusion that could not have been drawn if only animal evodevotics would have been studied. Therefore, to understand better the general rules of the macroevolution of higher or-

ganisms, evolutionary lineages should be compared in which multicellular body plans originated independently from unicellular ancestors. It seems very likely that green plants have evolved multicellular development independently from that of animals and are thus a suitable system to compare with animal development [73].

MADS-box genes and the evodevotics of the flower

We argue that the flower is an ideal model for plant evodevotics, and that the phylogeny of MADS-box genes may have played an important role during the origin and evolution of flower development. We summarize what was known about the role of MADS-box genes in flower development of some genetic model systems – all being higher eudicotyledonous plants – before these genes were studied in a broader series of phylogenetically informative taxa in order to test their importance for flower evolution.

We believe that flowers and their phylogenetic precursors are an ideal model system to study the linkage between development, genes and evolution. Floral morphology is the predominant source of characters for angiosperm taxonomy and phylogeny reconstruction [26]. Accordingly, the evolution of floral form has been studied quite extensively, although important questions concerning the origin and diversification of flowers have remained unanswered [21]. For the same reasons, flower development has been studied at high resolution in quite a number of different species (e.g. [28]). The most important advantages of flower evodevotics, however, are provided by genetics. A number of flowering plant model species, such as *Arabidopsis thaliana* (mouse-ear cress), *Petunia hybrida* (petunia), *Nicotiana tabacum* (tobacco) and *Oryza sativa* (rice) can routinely be transformed with genes from other species, so that the conservation of gene function can be determined by transgenic technology. Moreover, flower development is one of the best understood morphogenetic processes of plants on the genetic level. An impressive number of studies in recent years has culminated in the insight that inflorescence and flower development in higher eudicotyledonous flowering plants are determined by a network of regulatory genes that is organized in a hierarchical fashion (Figure 1) ([131]; for reviews, see [88, 123–125]). Close to the top of that hierarchy are 'late- and early-flowering genes' that are triggered by environmental factors such as day length, light quality and temperature. These genes mediate the switch from vegetative to reproductive development, perhaps by activating meristem identity genes. Meristem identity genes 'control' the transition from vegetative to inflorescence and from inflorescence to floral meristems. Within floral meristems, cadastral genes set the boundaries of floral organ identity gene functions, thus defining the different floral whorls. Some intermediate genes possibly mediate between floral meristem and organ identity genes. Floral organ identity genes (homeotic selector genes; 'ABC genes') specify the organ identity within each whorl of the flower by activating 'realizator genes'. In a classical model, three classes of homeotic gene activities ('homeotic functions') have been proposed, called A, B and C (Figures 1 and 5) [20]. Within any one of the four flower whorls, expression of A alone specifies sepal formation. The combination AB specifies the development of petals, and the combination BC specifies stamen formation. Expression of the C function alone determines the development of carpels. The model also proposed that the A and C function genes negatively regulate each other (meaning that they also exert 'cadastral' functions) and that the B function is restricted to the second and third whorls independently of A and C functions. In wild-type flowers, the A function is expressed in the first and second floral whorl, the B function in the second and third whorl, and the C function in the third and fourth whorl. Therefore, sepals, petals, stamens and carpels are specified in whorls one, two, three and four, respectively (for recent reviews of the ABC model, see [103, 124, 132]). The ABC model was largely based on the analysis of *Arabidopsis* mutants, albeit *Antirrhinum* was also considered [20].

Although the ABC model is quite elegant, it fails to explain some complications. Mutations in B and C function genes, for example, have effects in addition to homeotic changes of organ identity. Loss-of-C-function mutants form flowers with an undetermined number of floral organs, indicating that C function genes not only specify organ identity, but are also necessary to confer floral determinacy. *Antirrhinum* loss-of-B-function mutants lack the fourth floral whorl, suggesting that the B function genes not only specify second and third whorl organ identity, but are also necessary for fourth whorl formation [128]. Aside from that, the B and C function mutants are usually clear-cut. On the contrary, there are notorious problems with the A function. The flowers of strong loss-of-A-function mutants of *Arabidopsis*, for example, often lack the second whorl, while weaker alleles do not

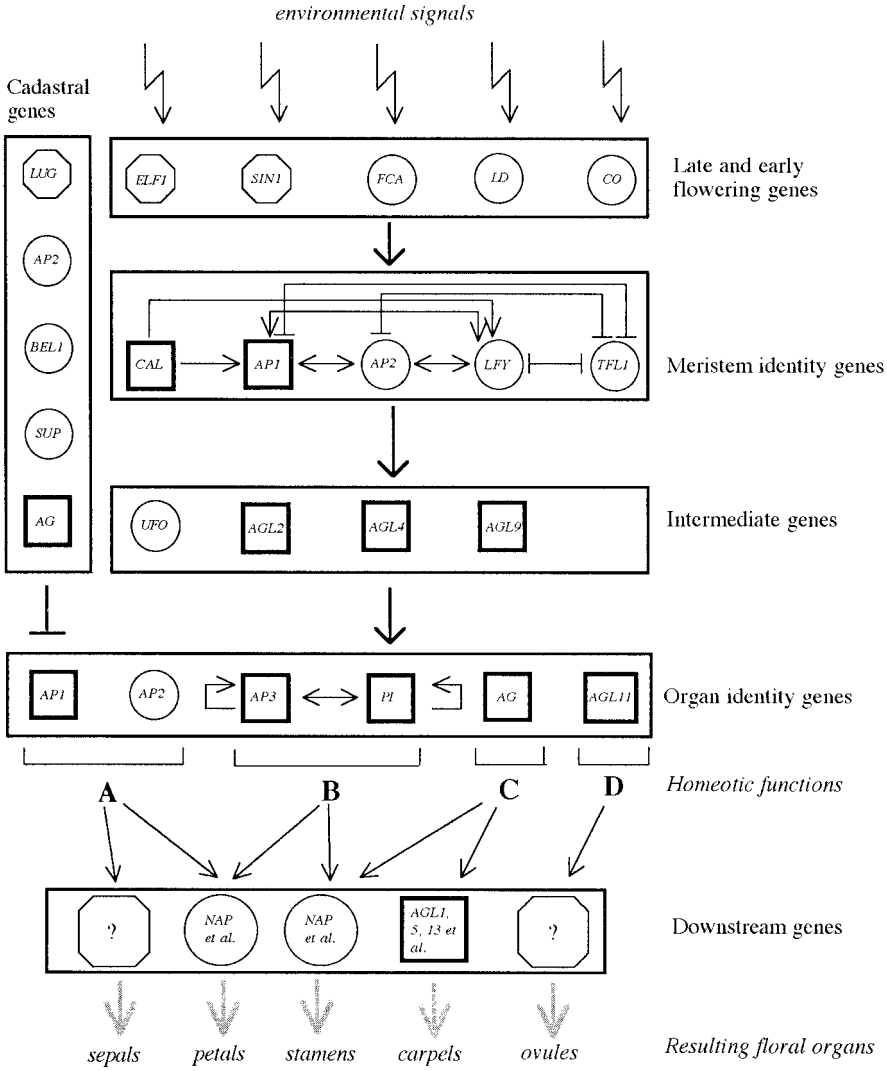


Figure 1. An extremely simplified and preliminary depiction of the genetic hierarchy that 'controls' flower development in *Arabidopsis thaliana*. Examples of the different types of genes within each hierarchy level are boxed. MADS-box genes are shown as open squares with thick lines, non-MADS-box genes as circles, and genes whose sequence has not been reported yet as octagons. The position of the genes was taken from the literature, as cited within this or other reviews [123–125]. Regulatory interactions between the different genes or blocks of genes are symbolized by arrows (activation), double arrows (synergistic interaction) or barred lines (inhibition, antagonistic interaction). Information about these interactions has been compiled from the review articles cited above. For a better overview, by far not all of the genes involved in flower development are shown (for review see [88]), and interactions (activation, repression) between the different hierarchy levels are depicted only globally (for some interactions between individual genes, see e.g. [124]). Absence of lines or arrows between genes means that an interaction has not been experimentally demonstrated yet, not that it does not exist. For the downstream genes, just one symbol is shown for every type of floral organ, though whole cascades of many direct target genes and further downstream genes are probably activated in each organ. The carpel-specific genes shown (*AGL*s) are only putative examples. Abbreviations used: AG, AGAMOUS; AGL1, 2, 4, 5, 9, 11, 13, AGAMOUS-LIKE GENE1, 2, 4, 5, 9, 11, 13; AP1, 2, 3, APETALA1, 2, 3; BEL1, BELL1; CAL, CAULIFLOWER; CO, CONSTANS; ELF1, EARLY FLOWERING1; LD, LUMINIDEPENDENS; LFY, LEAFY; LUG, LEUNIG; NAP, NAC-LIKE, ACTIVATED BY AP3/PI; PI, PISTILLATA; SIN1, SHORT INTEGUMENTS1; SUP, SUPERMAN; UFO, UNUSUAL FLORAL ORGANS; TFL, TERMINAL FLOWER.

have a full homeotic conversion of floral organs. Thus, 'ideal' mutants, in which the first- and second-whorl organs are homeotically transformed into carpels or stamens, respectively, actually do not exist. Mutants that are primarily caused by a loss of the A function are only known from *Arabidopsis*. *Antirrhinum* mutants with a similar phenotype are due to ectopic expression of a C function gene in whorls 1 and 2 of the flowers [9]. Searches for A function genes in petunia by a candidate gene approach inspired by results from *Arabidopsis* also remained negative [67], suggesting that the A function is phylogenetically less well conserved than the B and C functions. The confusion with the A function is a good example of problems that become less enigmatic when considered in an evolutionary perspective. It seems that some of the problems with defining the A function simply reflect the quite recent and multiple origin of the floral perianth (sepals and petals). Compared to the perianth organs, stamens and carpels (or their homologues from nonflowering plants), which are specified by B and C function genes, are evolutionarily more ancient and robust structures (see below).

Based on studies in petunia, the ABC model was recently extended by a D function [4]. When ectopically expressed, the D function genes *FBP7* and *FBP11* from petunia induce the formation of ectopic ovules on the perianth organs of transgenic flowers. They have, therefore, been defined as master control genes of ovules.

Arabidopsis genes providing the three homeotic activities A, B and C are known. The A function is contributed by both *APETALA1* (*AP1*) and *APETALA2* (*AP2*), the B function by *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), and the C function by *AGAMOUS* (*AG*). In *Antirrhinum*, the B function is provided by *DEFICIENS* (*DEF*) and *GLOBOSA* (*GLO*), the C function by *PLENA* (*PLE*). D function genes have been mutationally defined only in petunia so far, but sequence similarity suggests that the corresponding gene in *Arabidopsis* is *AGL11* (Figure 1) [4].

All these genes have been cloned. Except for *AP2*, all of them share a highly conserved, ca. 180 bp long DNA sequence, called the MADS-box. It encodes the DNA-binding domain of the respective MADS-domain transcription factors ([20, 111, 115, 123, 132, 137]; for recent reviews about MADS-box genes, see [103, 112, 123, 124]). MADS is an acronym for the four founder proteins MCM1 (from brewer's yeast, *Saccharomyces cerevisiae*), AGAMOUS (from *Arabidopsis*), DEFICIENS (from *Antirrhinum*), and SRF (a human protein), on which the definition of this gene family is based [111].

Within the hierarchical gene network contributing to flower development, MADS-box genes are not only dominant among the organ identity genes, but are well represented also at other levels, i.e. the levels of meristem identity genes, intermediate genes, cadastral genes, and possibly even downstream genes (Figure 1). In contrast to the *HOX* genes of animals, which are organized in genomic clusters, the MADS-box genes of plants are scattered throughout the entire plant genomes [31, 63].

MADS-domain proteins, like many other eukaryotic transcription factors, have a modular structural organization [112]. In the cases of almost all known seed plant MADS-domain proteins, it is very similar, including a MADS (M), intervening (I), keratin-like (K) and C-terminal (C) domain [66, 97, 123]. Genes encoding this type of protein hence have been termed MIKC-type MADS-box genes [85].

The MADS domain is by far the most highly conserved region of the proteins [97]. In most cases, it is found at the N-terminus of the putative proteins, although some plant proteins contain additional residues N-terminal to the MADS domain (NMIKC-type proteins). The MADS domain is the major determinant of DNA binding, but it also performs dimerization and accessory factor-binding functions [112]. Part of it folds into a novel structural motif for DNA interaction, an antiparallel coiled coil of α-helices that lies flat on the DNA minor groove [91]. In line with the conserved nature of their DNA-binding domain, MADS-domain proteins bind to similar DNA sites based on the consensus sequence $CC(A/T)_6GG$, which is called a CArG box (CC-A-rich-GG). CArG boxes are present in the promoter regions of many genes that are probably regulated by MADS-box genes [112, 127].

The I domain, directly downstream of the MADS domain, comprises ca. 30 amino acids, but is somewhat variable in length [66, 85]. It is only relatively weakly conserved among plant MADS-domain proteins [97]. In some *Arabidopsis* MADS-domain proteins, it was shown that the I domain constitutes a key molecular determinant for the selective formation of DNA-binding dimers [103]. The K domain, which is not present in any of the animal and fungal MADS-domain proteins known so far [123, 124], is characterized by a conserved, regular spacing of hydrophobic residues, which is proposed to allow for the formation of an amphipathic helix. It is assumed

that such an amphipathic helix interacts with that of another K domain-containing protein to promote dimerization [103, 112]. The most variable region, both in sequence and length, is the C domain at the C terminus of the MADS-domain proteins. The function of this domain is unknown, and it has been shown to be dispensable for DNA binding and protein dimerization in at least some floral homeotic MADS-domain proteins (see, for example, [139]). The C domain could be involved in transcriptional activation or the formation of multimeric transcription factor complexes.

According to the reasoning of evodevotics, understanding the origin and evolution of flower development depends on an understanding of the origin and evolution of the gene network governing flower development. Changes in gene number, expression and interaction thus all could have contributed to the evolution of flowers. Since MADS-box genes play such an important role in the network of flower development, understanding the phylogeny of MADS-box genes might strongly improve our understanding of flower evolution.

Phylogeny reconstructions disclosed that the MADS-box gene family is composed of several defined gene clades [26, 85, 97, 123, 124]. Most clade members share highly related functions and similar expression patterns. For example, the MADS-box genes providing the floral homeotic functions A, B and C each fall into separate clades, namely *SQUAMOSA*-like (A function), *DEFICIENS*- or *GLOBOSA*-like (B function), and *AGAMOUS*-like genes (C function) [26, 97, 123] (for the rules to name MADS-box gene clades used here, see [123]). The D function genes determining ovule identity [4] also belong to the clade of *AGAMOUS*-like genes [123]. Therefore, the establishment of the mentioned gene clades was probably an important event towards the establishment of the floral homeotic functions [123]. Thus the question arises as to when these gene clades arose during evolution and how some of their members were transformed into floral homeotic genes. To answer this, MADS-box genes have to be studied in phylogenetically informative taxa. Initially, plant MADS-box genes had been investigated only in a very limited taxonomic range, i.e. the genetic model plants, which are all higher eudicots. Meanwhile, however, the situation has changed considerably: while *Arabidopsis* and the like are still the favorites of hard-core developmental biologists, quite a number of scientists with evolutionary or agronomic interests have started to take plant diversity into account. In the following sections, we will outline what we have learned recently about MADS-box genes in non-flowering plants, basal angiosperms, and monocots. Then we briefly describe some new insights obtained from the eudicots. We will use these data to reconstruct the evolution of the MADS-box gene family and its relationship to floral evolution, i.e. we will tell a short natural history of MADS-box genes in plants. First, however, we will briefly speculate about the origin of plant MADS-box genes.

On the origin and major subdivisions of the MADS-box gene family

*We briefly describe what is known about MADS-box genes in animals and fungi, and report that homologues of MADS-box genes may even exist in bacteria. The MADS-box gene family proper of eukaryotes can be subdivided into three major clades. Representatives of two of these clades (*ARG80- *and* MEF2*-like genes) have only been found in animals and fungi so far, whereas members of the third group (MIKC-type genes) seem to be restricted to plants.*

The origin of the MADS-box gene family is unclear. Some bacterial proteins, such as members of the *UspA* family of stress response proteins known from *Escherichia coli* and *Haemophilus influenzae*, contain short sequence stretches that could be homologous to a part of the MADS domain [87]. However, sequence similarity between the bacterial proteins and the MADS domains is so low that special strategies of sequence database search were needed to detect it. Anyhow, it seems likely that a precursor of the MADS-type DNA-binding domain evolved before the separation of bacterial and eukaryotic lineages [87] about 2–3.5 billion years ago [69]. Interestingly, a coiled-coil structure is predicted in the downstream portion of UspA-like proteins [87]. Since the K domain of plant MADS-domain proteins is also assumed to adopt a coiled-coil structure [66, 112], even the K domain may have bacterial roots.

Since MADS-box genes have been found in extant plants, animals and fungi, it is quite safe to assume that the last common ancestor of these eukaryotic taxa, which existed about one billion years ago, had already at least one gene with a true MADS box [123]. The MADS-box gene family can be subdivided into three major clades, *ARG80*-like genes (also called the '*SRF* gene family'), *MEF2*-like genes and MIKC-type genes [45, 85, 123, 124]. While *ARG80*- and *MEF2*-like genes have been found only in animals and fungi

so far, MIKC-type genes seem to be restricted to plants (Figure 2). The presence of *ARG80*- and *MEF2*-like genes could represent a synapomorphy of animals and fungi, separating these taxa from plants (Figure 2). However, due to the limited sampling of MADS-box genes in any taxon one cannot exclude that, for example, *ARG80*- or *MEF2*-like genes are also present in plants. Moreover, although the hypothesis that animals and fungi are more closely related to each other than both are to plants is supported by quite a number of molecular data, there is also evidence for alternative relationships [130]. The picture drawn in Figure 2 is thus possibly not the last word on this subject.

MADS-box genes in animals and fungi are involved in a diverse range of biological activities (reviewed in [112, 123]). A common denominator of most MADS-domain proteins is that they control aspects of development or cell differentiation. Let us take the *ARG80*-like genes as an example, which include *ARG80* and *MCM1* from brewer's yeast and the *SRF* genes from animals. While *ARG80* is involved in regulating genes encoding arginine-metabolizing enzymes, *MCM1* is involved in a broader range of functions: in cooperation with different associated factors it represses or activates the transcription of many genes involved in diverse aspects of the yeast cell cycle and cell growth, metabolism (including that of arginine) and specialization. The role of *MCM1* in the determination of yeast cell type is especially well known [112, 123]. Recently, putative orthologues of *MCM1* have also been reported from distant fungal relatives of brewer's yeast, i.e. the fission yeast *Schizosaccharomyces pombe* (*MAP1* gene) and the smut fungus *Ustilago maydis* (*UMC1* gene) [61, 136]. The SRF (serum response factor) of vertebrates is involved in immediate-early gene and muscle-specific gene transcription. Its orthologue from *Drosophila* (*DSRF*) plays a role in tracheal development (reviewed in [112, 123]).

Members of the clade of *MEF2*-like genes are key components in muscle-specific gene regulation in animals [90], but probably also have functions in non-muscle cells. For more details about animal and fungal MADS-box genes we refer to other reviews on this topic and the original work cited therein [45, 112, 123].

Somewhere in the lineage that led to extant green plants, MADS-box genes appeared in which the MADS-box was followed by the I-, K- and C-regions, and the MIKC-type genes were born. The molecular mechanism that generated them is unknown. It seems that extant MIKC-type genes are more closely related to *MEF2*-like genes than to *ARG80*-like genes, implying that the last common ancestor of MIKC-type genes was more *MEF2*- than *ARG80*-like [123]. Molecular clock analyses and studies on MADS-box genes in ferns have helped recently to get better estimates about the time interval in the past when the first MIKC-type genes appeared, as described below.

MADS-box genes in ferns

We summarize data suggesting that the last common ancestor of ferns and seed plants about 400 MYA contained at least two different MIKC-type MADS-box genes that were homologues, but not orthologues, of floral homeotic genes. These genes probably had expression patterns and functions that were more general than those of the highly specialized floral homeotic genes from extant flowering plants.

After colonization of land, roughly about 500 MYA, land plants (today comprising liverworts, hornworts, mosses and vascular plants) evolved body structures of increasing complexity [57]. Extant vascular plants, for example, range from relatively simple clubmosses (lycopsids), horsetails (equisetopsids), whisk ferns (Psilotaceae) and ferns (filicopsids) to complex seed plants (spermatopsids), comprising gymnosperms and angiosperms (flowering plants *sensu stricto*) [57]. Although MADS-box gene cDNAs have already been isolated from a moss (MIKC-type; our unpublished data) and a clubmoss (see the citation in [3]), the most basal plants from which MADS-box gene sequences have been published so far are ferns [22, 46, 58, 85]. Among the land plants ferns are of considerable scientific interest because they are very likely the sister group of the seed plants. The two groups diverged about 400 MYA [36, 117]. Ferns have several characteristics that are primitive with respect to vascular plants as a whole [7]. For example, they produce naked sporangia at the abaxial sides of their leaves which lack accessory organs such as integuments. Ferns thus do not form ovules or seeds, and generally they also do not aggregate their sporophylls into flower-like structures. Most ferns are homosporous, i.e. their sporangia produce only one type of haploid reproductive spores, starting from diploid spore mother cells that undergo meiosis. In contrast to the megaspores of seed plants, the spores of ferns are shed, and the haploid gametophytes developing from them are entirely independent of the

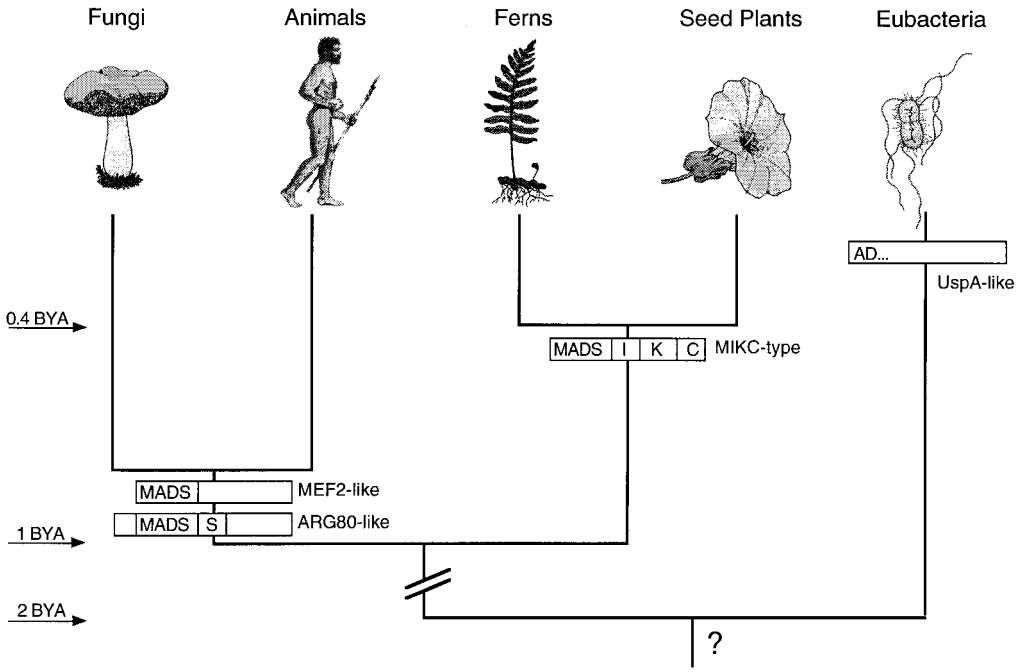


Figure 2. The major clades of MADS-box genes in the evolution of life. A phylogenetic tree of some major taxa of living organisms is shown. The ages (in BYA, billion years ago) given at some nodes of the tree are very rough estimates. Some aspects of the topology of the tree are controversial, for example, that fungi are more closely related to animals than to plants (see text). At some internal branches of the tree simplified domain structures of representative members of the three major clades of MADS-domain proteins are shown (ARG80- and MEF2-like proteins, MIKC-type proteins); in addition, a representative of a group of putative distant relatives of MADS-domain proteins, i.e. UspA-like proteins, is shown at the eubacterial branch. 'MADS', 'I', 'K' and 'C' denote the MADS-, I-, K- and C-domains, respectively. 'S' stands for the SAM domain, present in *S*RF, *A*RG80 and *M*CM1. 'AD' symbolizes the presence of a domain with sequence similarity to a part of the M*AD*S domain within the UspA-like proteins. The different gene types have been established during the time interval represented by the respective branches of the phylogenetic tree, at the latest. This could be concluded from the presence of respective clade members in extant taxa. For example, *MEF2*- and *ARG80*-like genes have been isolated from animals and fungi so far, but not from plants.

spore-producing plant (the sporophyte). On the gametophyte, sexual organs (archegonia and antheridia) are formed that produce egg and sperm cells, respectively. Fertilization results in a diploid zygote which develops into a new sporophytic generation.

The characterization of MADS-box genes in ferns has focused so far on *Ceratopteris* because it has some features that qualify it as a plant model system. *Ceratopteris richardii*, for example, has a short sexual life cycle of less than 120 days. Moreover, it behaves like a diploid species and is well suited for genetic and developmental analyses [14].

cDNAS representing more than 15 different genomic loci containing a MADS box have already been isolated by three different research groups [22, 46, 58, 85]. Unfortunately, these groups used three different systems of gene nomenclature, which resulted in up to three different names for the same gene. In the following section, we always use the gene name that has been published first, but we will also mention synonymous names where these exist to facilitate comparison between the different studies.

In one study, cDNAs of 12 different genomic loci, designated *CRM1–CRM12* (for *Ceratopteris* MADS1–12) were isolated from *Ceratopteris richardii*, *Ceratopteris pteroides*, or both [22, 85]. *CRM8*, however, had been published earlier under the name of *CERMADS5* [58], so we adopt that name here. *CERMADS5* was later also called *CMADS2* [46]. Southern blot analysis indicated that *CRM1–CRM10* represent single-copy loci in the genome of *Ceratopteris richardii* [22, 85]. cDNAs of three additional genes, termed *CMADS1*, *CMADS4* and *CERMADS3*, have been isolated in two other studies [46, 58]. Most cDNAs of *Ceratopteris* MADS-box genes isolated so far show high sequence similarity to typical seed plant MADS-box genes with respect to MADS-domain sequence and overall domain structure, i.e. they can be classified as MIKC-type genes. There is no indication that domain shuffling occurred

within the genealogy of these genes. The similarity between fern and seed plant MADS-box genes clearly indicates that these genes share a common ancestor from which they were derived by gene duplications, sequence diversification and fixation [85]. The fern genes identified are thus clearly homologues of the MADS-type floral homeotic genes known from angiosperms.

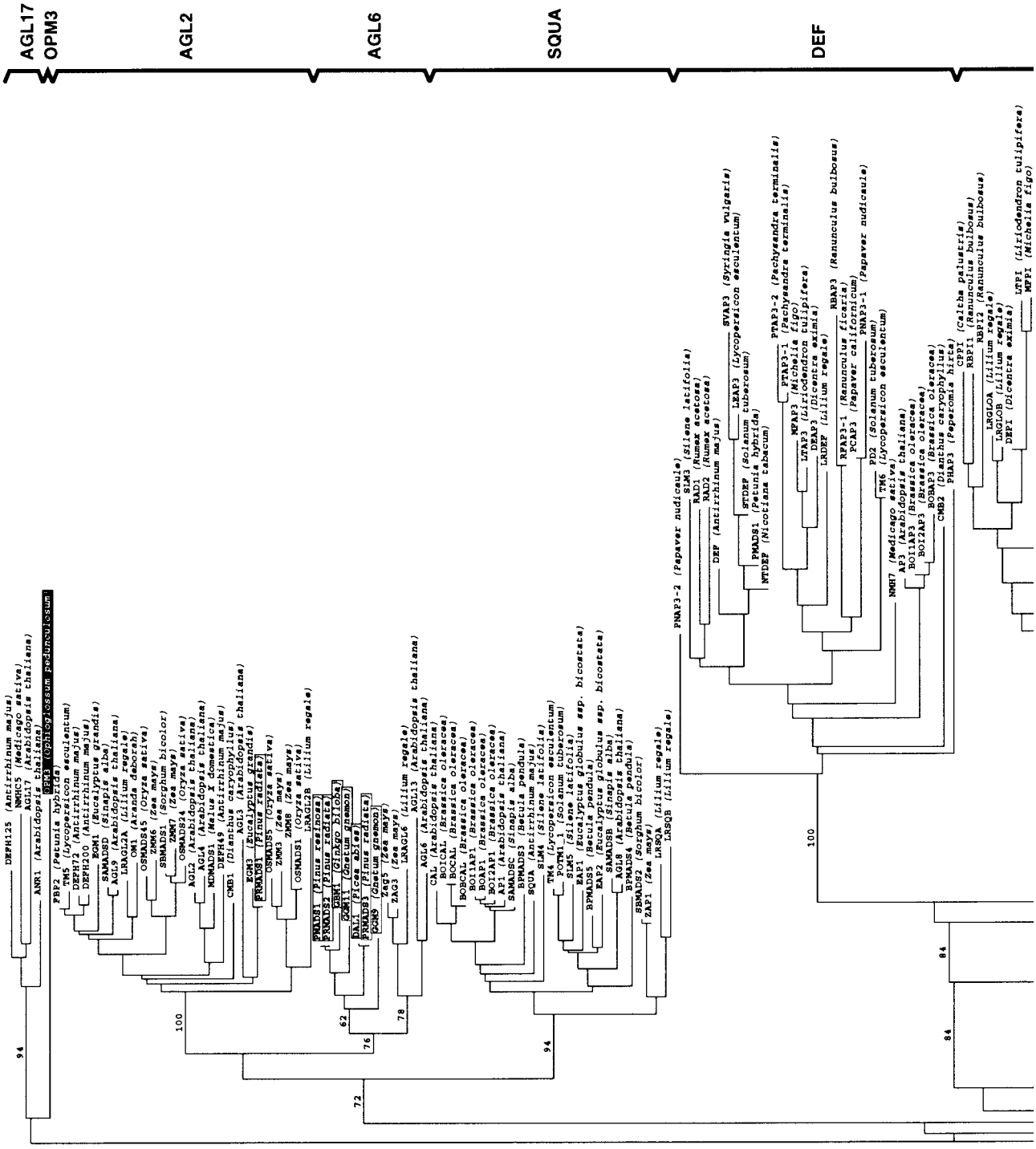
To determine the evolutionary relationships between the fern genes and the other known MADS-box genes, phylogeny reconstructions were carried out. They disclosed that the genes from *Ceratopteris* constitute three different gene clades, termed *CRM1*-, *CRM3*- and *CRM6*-like genes, which are interspersed among seed plant gene clades [46, 85]. The *CRM6*-like genes can be further subdivided into *CRM6*-like genes *sensu stricto* and *CRM7*-like genes (Figure 3). In some phylogenetic trees, monophyly of the *CRM6*-like genes is not well supported (Figure 3), in some others the *CRM6*-like genes *sensu stricto* even appear separated from the *CRM7*-like genes [58]. In a few other gene trees, however, even the *CRM1*- and *CRM6*-like genes form sister clades [85]. A conservative interpretation of all available data thus leads to the conclusion that at least two different MIKC-type MADS-box genes existed already in the last common ancestor of ferns and seed plants [85]. It seems more likely, however, that at least three or four different MIKC-type genes were already present in this species. On the other hand, it is obvious from the analyses carried out so far that many of the gene duplications which led to the large number of present-day MIKC-type genes occurred independently in the lineages that led to extant ferns and seed plants [85]. Although the MADS-box genes from *Ceratopteris* can be considered being homologous to the MIKC-type genes from other plants, including the floral homeotic genes, they are clearly not orthologues of specific floral homeotic genes. It seems likely, therefore, that the last common ancestor of ferns and seed plants contained only a relatively small number of MIKC-type genes compared to the large number of genes present in extant seed plants and ferns [85]. Alternative scenarios are conceivable, but appear far less parsimonious.

Molecular clock estimates suggest that MIKC-type genes started to diverge about 450–500 MYA, i.e. before the separation of the ferns and the seed plants [96]. The presence of at least two different MIKC-type genes in the last common ancestor of ferns and seed plants about 400 MYA is in good agreement with this estimation. Accordingly, it seems likely that the last common ancestor of MIKC-type genes existed during the Ordovician, when plants probably started to colonize the land [96]. Therefore, MIKC-type MADS-box genes probably had already been established in plants more basal than ferns. Cloning of a MIKC-type cDNA from the moss *Physcomitrella patens* supports this hypothesis (our unpublished data).

The presence of a short peptide motif at the C-terminal end of the respective proteins suggests a close relationship between the *CRM3*-like genes (comprising *CRM3*, also called *CMADS6* [46], and *CRM9* up to now), and the *DEF*/*GLO*-like genes [60, and our unpublished results]. Based on the presence of a N-terminal extension in the derived proteins, a close relationship between *CRM6*/*7*-like genes, including *CRM6* (also called *CERMADS2*), *CERMADS3* and *CMADS1*, and the members of the *AG* clade has also been postulated [46]. However, we consider the respective evidences as weak, since they are based on the presence of small peptide sequences of limited sequence similarities. They thus do not necessarily define synapomorphies, but also could represent homoplasies (i.e. the recurrences of similarities in evolution). Using phylogeny reconstructions, clear sister group relationships between fern and seed plant MADS-box gene clades have not been established yet.

Since orthologues of floral homeotic genes have not been isolated so far from *Ceratopteris*, the question arises whether such genes actually exist in this taxon. MADS-box gene cDNA cloning has involved three independent research groups and different cloning techniques. Diverse phases of the fern life cycle and different plant tissues were used as mRNA sources. Moreover, probes and primers for cloning experiments were derived, at least in some cases, from *Arabidopsis* and *Antirrhinum* floral homeotic genes. However, the three research groups found only *CRM1*-, *CRM3*-, *CRM6*- and *CRM7*-like genes in quite a redundant fashion [22, 46, 58, 85]. Although the possibility remains that orthologues of floral homeotic genes are present in *Ceratopteris*, this appears less and less likely.

We wanted to verify that the apparent absence of floral homeotic gene orthologues is not merely a specific feature of *Ceratopteris* or its close relatives. Therefore, we applied cDNA cloning also to *Ophioglossum*, another fern which is only very distantly related to *Ceratopteris*. While *Ceratopteris* is a highly derived leptosporangiate fern, the Ophioglossales are eusporangiate ferns which branch off near the base of the fern tree [95]. cDNAs representing


AGL17
OPM3
AGL2
AGL6
SQUA
DEF
DEFH125 (Antirrhinum majus)
NMHC5 (Medicago sativa)
AGL17 (Arabidopsis thaliana)
ANR1 (Arabidopsis thaliana)
FBP2 (Petunia hybrida)
TM5 (Lycopersicon esculentum)
DEFH72 (Antirrhinum majus)
DEFH200 (Antirrhinum majus)
EGM1 (Eucalyptus grandis)
SAMADS8 (Sinapis alba)
AGL3 (Arabidopsis thaliana)
LRA1 (Lilium regale)
OM1 (Arabidopsis thaliana)
OSMADS45 (Oryza sativa)
ZMM6 (Zea mays)
SBMADS1 (Sorghum bicolor)
ZMM7 (Zea mays)
OSMADS14 (Oryza sativa)
AGL2 (Arabidopsis thaliana)
AGL4 (Arabidopsis thaliana)
MDMADS1 (Malus X domestica)
DEFH49 (Antirrhinum majus)
CMB1 (Dianthus caryophyllus)
AGL3 (Arabidopsis thaliana)
EGM3 (Eucalyptus grandis)
PRMADS1 (Pinus radiata)
OSMADS5 (Oryza sativa)
ZMM3 (Zea mays)
ZMM8 (Zea mays)
OSMADS1 (Oryza sativa)
LRAGL2B (Lilium regale)
PMADS1 (Pinus resinosa)
PRMADS2 (Pinus radiata)
DGM1 (Ginkgo biloba)
DGM11 (Gnetum gnemon)
DAL1 (Picea abies)
PRMADS3 (Pinus radiata)
DGM5 (Gnetum gnemon)
ZAG5 (Zea mays)
ZAG3 (Zea mays)
LRAGL6 (Lilium regale)
AGL13 (Arabidopsis thaliana)
AGL6 (Arabidopsis thaliana)
CAL (Arabidopsis thaliana)
BOICAL (Brassica oleracea)
BOCAL (Brassica oleracea)
BOBCAL1 (Brassica oleracea)
BOA1 (Brassica oleracea)
BOAP1 (Brassica oleracea)
BOI2AP1 (Brassica oleracea)
AP1 (Arabidopsis thaliana)
SAMADS9C (Sinapis alba)
BPMADS3 (Betula pendula)
SQUA (Antirrhinum majus)
SLM4 (Silene latifolia)
TM4 (Lycopersicon esculentum)
POTM1.1 (Solanum tuberosum)
SLM5 (Silene latifolia)
EAP1 (Eucalyptus globulus ssp. bicostata)
BPMADS5 (Betula pendula)
EAP2 (Eucalyptus globulus ssp. bicostata)
SAMADS8 (Sinapis alba)
AGL9 (Arabidopsis thaliana)
BPMADS4 (Betula pendula)
SBMADS2 (Sorghum bicolor)
ZAP1 (Zea mays)
LRSQA (Lilium regale)
LRSQB (Lilium regale)
PNAP3-2 (Papaver nudicaule)
SLM3 (Silene latifolia)
RAD1 (Rumex acetosa)
RAD2 (Rumex acetosa)
DEF (Antirrhinum majus)
SVAP3 (Syringia vulgaris)
LSEAP3 (Lycopersicon esculentum)
STDEF (Solanum tuberosum)
PMADS1 (Petunia hybrida)
NTDEF (Nicotiana tabacum)
PTAP3-2 (Pachysandra terminalis)
PTAP3-1 (Pachysandra terminalis)
NFAP3 (Michelia figo)
LTAP3 (Liriodendron tulipifera)
DEAP3 (Dicentra eximia)
LADEF (Lilium regale)
RBAP3 (Ranunculus bulbosus)
RPAP3-1 (Ranunculus ficaria)
PCAP3 (Papaver californicum)
PNAP3-1 (Papaver nudicaule)
PD2 (Solanum tuberosum)
TM6 (Lycopersicon esculentum)
NMH7 (Medicago sativa)
AP3 (Arabidopsis thaliana)
BOI1AP3 (Brassica oleracea)
BOI2AP3 (Brassica oleracea)
BOBAP3 (Brassica oleracea)
CMB2 (Dianthus caryophyllus)
PHAP3 (Peperomia hirta)
CPPI (Caltha palustris)
RBPI1 (Ranunculus bulbosus)
RBPI2 (Ranunculus bulbosus)
LRGLOA (Lilium regale)
LRGLOB (Lilium regale)
DEPI (Dicentra eximia)
LTPI (Liriodendron tulipifera)
MPPI (Michelia figo)
94
100
62
76
78
72
94
100
84
84

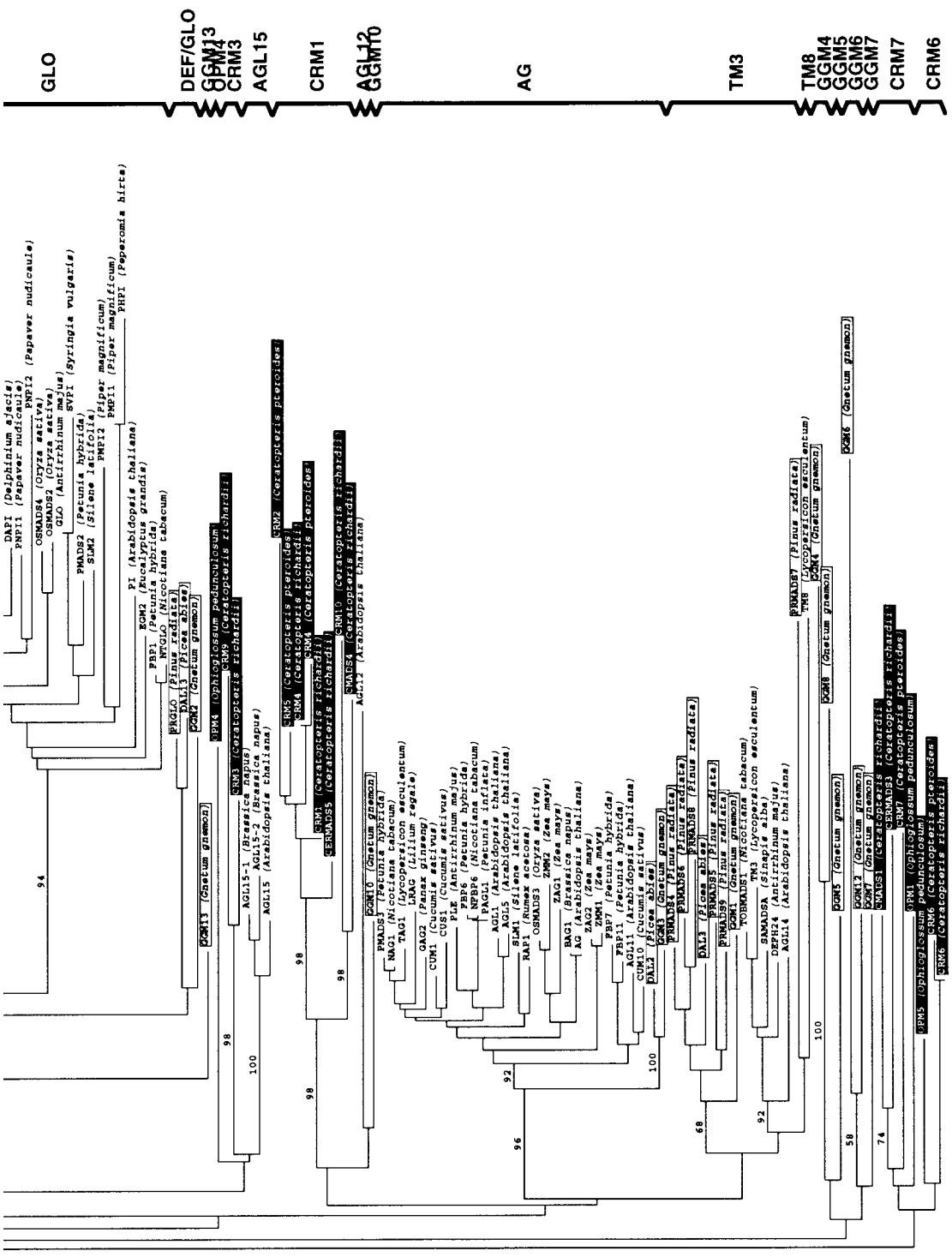


Figure 3. Evolution of MIKC-type MADS-box genes. The phylogenetic tree shows the relationships between the published MIKC-type MADS-domain proteins. Names of species from which the respective genes were isolated are given in parenthesis behind the protein names. Genes of gymnosperms are indicated by open boxes, those of ferns by inverted boxes. All other genes have been isolated from angiosperms. The names of the *Lilium regale* genes are preliminary. The numbers next to some nodes give bootstrap percentages, shown only for relevant nodes (e.g. those defining gene subfamilies), and only if above 50%. Subfamilies are labelled by brackets at the right margin. They generally represent monophyletic gene clades. The tree was generated by the neighbor-joining algorithm as described elsewhere [85]. An updated and colored version of the tree and the sequence alignment on which it is based, as well as information about the different genes (including accession numbers) are provided on the 'MADS homepage' within the worldwide web (URL: http://www.mpiz-koeln.mpg.de/mads/).

four different genes could be isolated so far, termed *OPM1* and *OPM3–OPM5* (*Ophioglossum pedunculosum* MADS1, 3–5) ([85], and our unpublished data). However, it turned out again that these genes are not members of any of the gene clades known from seed plants. While *OPM3* and *OPM4* do not fit into any clade defined so far, the other *OPM* genes seem to be *CRM6*- or *CRM7*-like genes, respectively (Figure 3) [85]. Although bootstrap support for this kind of grouping is often not very high, it suggests that *CRM6*- and *CRM7*-like genes were established at an early time point in fern evolution (Figure 4). Taken together, there is no evidence so far that orthologues of floral homeotic genes (such as *SQUA*-, *DEF*-, *GLO*- or *AG*-like genes) are present in extant ferns. The fact that no genes from ferns have been isolated that are within the gene clades known from seed plants, might be correlated to the absence of seed or flowering plant specific structures, such as ovules, carpels, stamens or floral perianth organs.

Molecular clock estimates suggested that the last common ancestor of the clade comprising *AGL2*-, *AGL6*- and *SQUA*-like genes (also termed 'AP1/AGL9 clade') existed about 370 MYA [96], i.e. after the lineage that led to extant seed plants had already separated from the lineage that led to present-day ferns. This estimation is in agreement with the fact that no distinctive *AGL2*-, *AGL6*- or *SQUA*-like genes have been found in ferns so far, while members of the *AGL2* and *AGL6* clades could be cloned from both gymnosperms and angiosperms, two seed plant lineages which separated about 300 MYA (see below).

Unfortunately, phylogeny reconstructions did not give specific clues to fern MADS-box gene function, and mutants or transgenic plants in which the expression of these genes is changed are also not yet available. Accordingly, the expression of several *Ceratopteris* genes was determined by northern and *in situ* hybridizations to get some idea about their function. It turned out that most genes are well expressed in the gametophytic as well as the sporophytic phase of the fern life cycle [22, 46, 85]. Exceptions are *CRM9* and *CMADS1*, which are much more strongly expressed in the sporophyte than in the gametophyte [22, 46]. Exclusive expression in hermaphroditic gametophytes was reported for *CRM3* in one study (termed *CMADS6* there) [46], but this result is controversial, because in other studies *CRM3* expression was also observed in sporophytes and male gametophytes [22, 85]. Preliminary data indicate that in male gametophytes expression of *CRM3* is in spermatides that develop within antheridia. In hermaphroditic gametophytes, *CRM3* expression was detected in meristematic cells [22]. Expression analysis in the sporophyte revealed that quite a number of genes are expressed in many tissues [22, 46]. An exception is *CMADS4*, which is predominantly expressed in roots [46]. Expression of *CRM3* and *CRM9*, for example, was found in the shoot axis as well as in fronds of juvenile plants. In cross-sections of fertile fronds, expression of *CRM3*, *CRM6* and *CRM9* was observed, with *CRM6* expression being relatively strong in sporangia [22]. *CMADS1* expression was observed in the shoot apical meristem, leaf primordia and the procambium [46]. As the leaves increase in cell number, *CMADS1* signals become stronger in all cells at the top of the leaf. As tissue systems differentiate, *CMADS1* expression gradually becomes restricted to three leaf parts: procambium, sporangium initials, and the regions that will give rise to the lamina, or pinnae. Signals are also observed in differentiated vascular bundles of the petiole, and in the root apical meristems and their associated provascular cell files. *CMADS1* expression can also be observed in developing sporangia, but not in the sporangia containing mature spores [46]. The expression patterns of *CRM1* (also called *CerMADS4* or *CMADS3* [46, 58]) and *CerMADS5* (for synonymous names, see above) are very similar to those of *CMADS1*, albeit weaker [46].

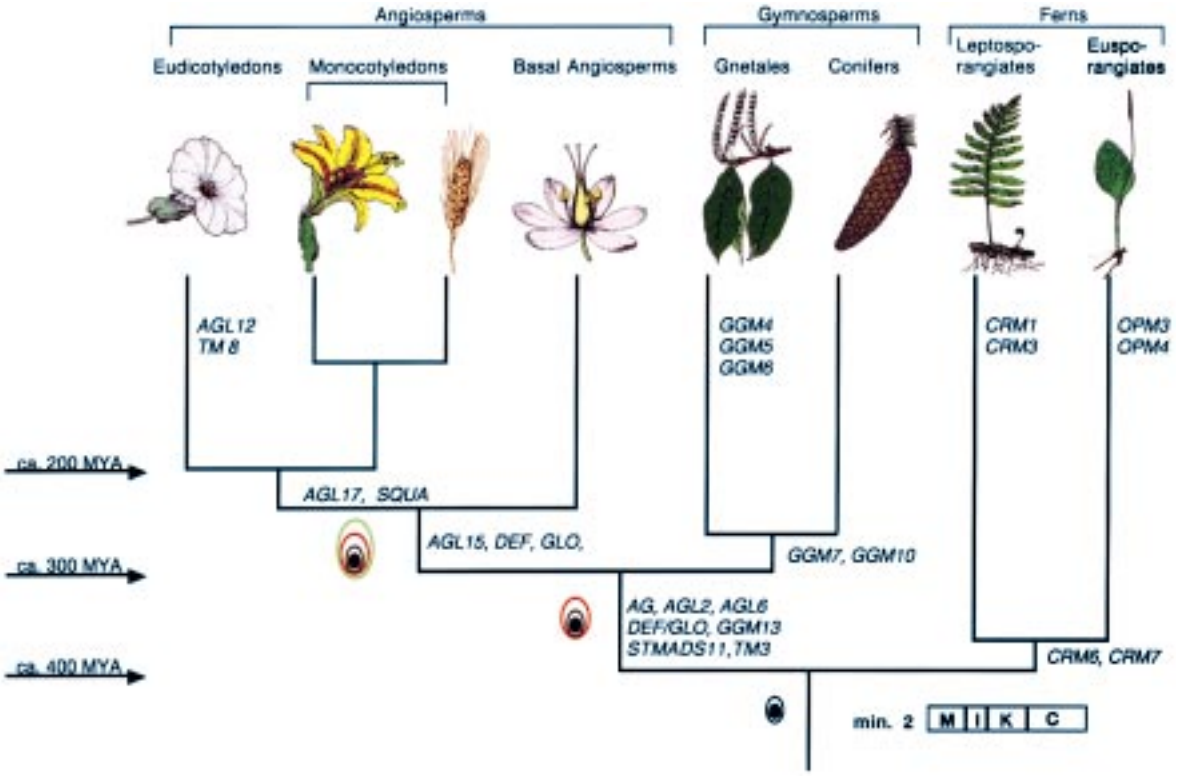
The expression of most fern genes in both major phases of the life cycle is in remarkable contrast to the situation in seed plants, where expression of a MIKC-type gene in the gametophytic phase has been demonstrated to date only in a single case, the *AGL17*-like gene *DEFH125* from *Antirrhinum* [140], although many MADS-box genes are expressed in stamens, carpels or ovules. Expression in both sporophytes and gametophytes suggests a more ubiquitous function of the fern genes in the control of development or cell differentiation than the temporally and spatially quite restricted functions of the homeotic genes determining floral organ identity of angiosperms.

MADS-box genes with a relatively ubiquitous expression in the sporophytic phase do also exist in seed plants. Examples are most members of the clade of *TM3*-like genes and *AGL3*, an *AGL2*-like gene (reviewed in [123]). As indicated above, a close relationship between the fern gene clades and the *TM3*- or *AGL2*-like genes from seed plants cannot be demonstrated. However, the organs in which the organ identity genes of seed plants are specifically expressed were very likely not present in the last common ancestor of ferns and seed plants. It seems

plausible, therefore, that the rather ubiquitous expression of most fern MADS-box genes and of some MADS genes from seed plants represents the ancestral state of MIKC-type gene expression. The highly organ-specific expression of the floral homeotic genes of angiosperms is thus very likely a derived condition that was achieved during the processes in which some MIKC-type genes were recruited as floral homeotic genes. If so, spatiotemporal restriction of gene expression was an important aspect during the co-option of MADS-box genes as homeotic selector genes of specialized plant organs. This gene recruitment must have occurred in the lineage that led to seed plants after the lineage that led to extant ferns had already branched off. It has been speculated that the restriction of MADS-box gene expression may have been caused by the evolution of other genes that regulate the MADS-box genes, such as relatives of *LEAFY* or *CURLY LEAF* [46]. However, these changes in expression patterns could also have been caused by mutations in *cis*-regulatory elements controlling MADS-box gene expression [124]. In some precedent cases, concerning anthocyanin biosynthesis and growth form in maize, the molecular basis of evolutionary changes in gene expression in plants has been clarified recently. In these cases it turned out that *cis*-regulatory elements, not *trans*-acting factors, were responsible for changes in gene expression (examples cited in [6]). It has even been argued that modifications in the *cis*-regulatory regions of transcriptional regulators represent a predominant mode for the evolution of novel plant forms [23]. Besides *trans*-acting factors, evolutionary changes in MADS-box gene promoters should therefore be seriously considered as a possible cause for the changes in MADS-box gene expression during evolution.

Besides the rather ubiquitous spatiotemporal expression of most genes, several *Ceratopteris* MADS-box genes also display some other features that are atypical of seed plant MADS-box genes. For example, there is evidence that the primary transcripts of a relatively large fraction of genes, including *CRM1*, *CRM4* (also called *CerMADS1*), *CRM6* (also called *CerMADS2*) and *CRM9* are alternatively spliced [22, 58]. For comparison, although more than 150 different MIKC-type genes have been reported so far from seed plants (Figure 3), alternative splicing has been reported only in a single case [62]. However, alternative splicing is typical of *MEF2*-like MADS-box genes from animals (for reviews, see [90, 123]) and has also been documented in cases of some transposon-like elements containing a MADS box which have been isolated from the flowering plant, maize (see below) [31, 78, 79]. Alternative splicing, therefore, may represent an ancient mechanism to increase the diversity of protein products from individual MADS-box genes that has been reduced in seed plants. One should not be too surprised, however, if alternative splicing plays a more important role in seed plants than currently thought.

Another unusual feature of some fern MADS-box genes concerns their structure. While the majority of fern cDNAs have the potential to encode perfect (N)MIKC-type proteins, cDNAs of several other loci, including *CRM11*, *CRM12* and *CMADS5*, also show high sequence similarity to MADS-box gene cDNAs, but do not contain continuous open reading frames, due to the presence of in-frame stop codons or nucleotide insertions or deletions [22, 46]. Whether the respective genomic loci have a function is unclear. In principle, they could encode truncated proteins that work as transcriptional modulators. They even may encode full-length proteins generated by programmed frameshifting (ribosome hopping). Also a function apart from the protein level is conceivable. Alternatively, these loci may simply represent nonfunctional pseudogenes that got into the vicinity of promoters and are therefore transcribed. In mammalian genomes, nonfunctional pseudogenes are often created through reverse transcription of mRNA and integration of the copy DNA into the genome. A similar mechanism might work in ferns. This hypothesis is supported by the fact that, in contrast to *CRM1–10*, Southern hybridizations revealed several *CRM12*-like loci in the *Ceratopteris* genome even under high-stringency hybridization conditions [22]. Analysis of genomic loci such as *CRM11* and *CRM12* might give further clues to their origin. It is interesting to note that our observations are not unprecedented. It has been reported that the majority of genomic clones of homologues to the chlorophyll *a/b*-binding (CAB) protein that have been isolated from the homosporous fern *Polystichum munitum* are defective. A major cause is, again, the presence of in-frame stop codons and nucleotide insertions or deletions [94]. Whether the probably defective *CAB* genes are transcribed has not been reported. One of the explanations for the *CAB* gene defects is gene silencing upon polyploidization [94]. However, since we found multiple copies for only a minority of *Ceratopteris* MADS-box genes, this does not seem to be a likely explanation for the structurally aberrant MADS-box gene cDNAs reported here.


Angiosperms
Gymnosperms
Ferns
Eudicotyledons
Monocotyledons
Basal Angiosperms
Gnetales
Conifers
Leptospo-rangiates
Euspo-rangiates
AGL12
TM 8
GGM4
GGM5
GGM8
CRM1
CRM3
OPM3
OPM4
ca. 200 MYA
AGL17, SQUA
ca. 300 MYA
AGL15, DEF, GLO,
GGM7, GGM10
AG, AGL2, AGL6
DEF/GLO, GGM13
STMADS11,TM3
ca. 400 MYA
CRM6, CRM7
min. 2
M
I
K
C

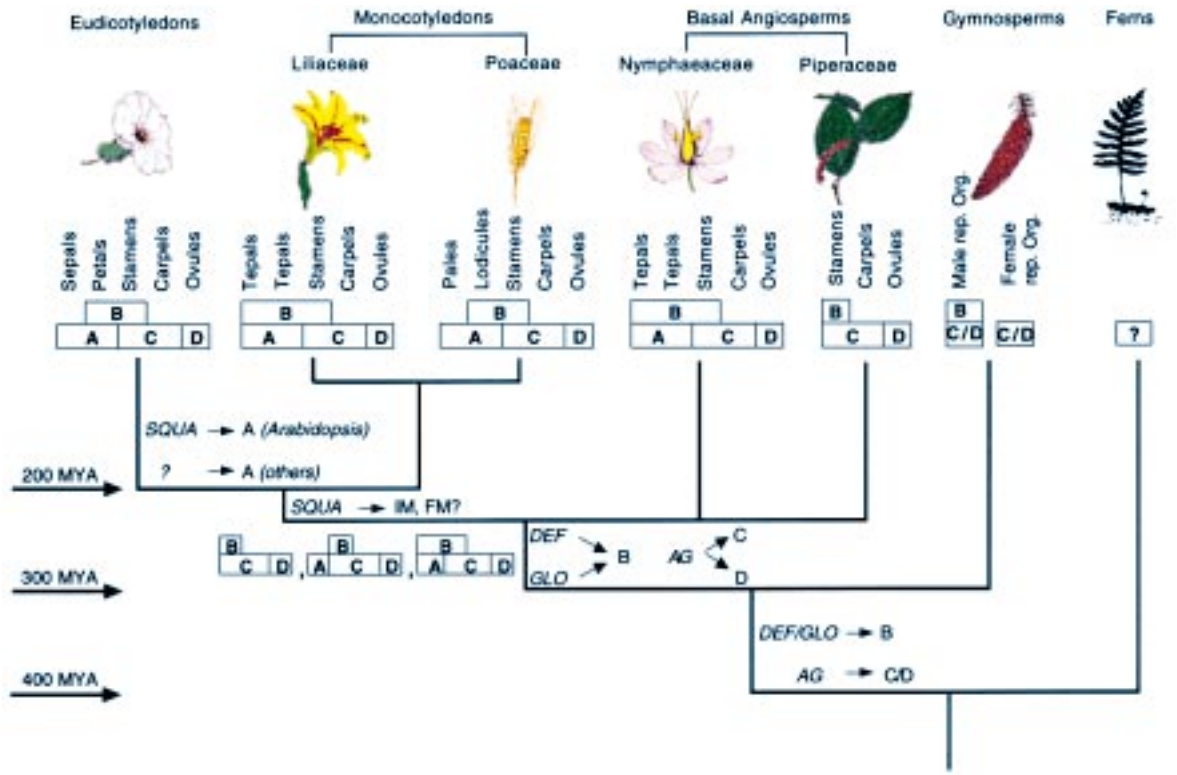

Eudicotyledons
Monocotyledons
Liliaceae
Poaceae
Basal Angiosperms
Nymphaeaceae
Piperaceae
Gymnosperms
Ferns
Sepals
Petals
Stamens
Carpels
Ovules
Tepals
Tepals
Stamens
Carpels
Ovules
Palea
Lodicules
Stamens
Carpels
Ovules
Tepals
Tepals
Stamens
Carpels
Ovules
Stamens
Carpels
Ovules
Male rep. Org.
Female rep. Org.
B
A
C
D
C/D
?
SQUA → A (Arabidopsis)
? → A (others)
200 MYA
SQUA → IM, FM?
DEF
GLO
B
AG
C
D
300 MYA
DEF/GLO → B
AG → C/D
400 MYA

Figure 4 (top). MADS-box genes in the evolution of vascular plants. A phylogenetic tree of major taxa of vascular plants is shown. The ages (in MYA, million years ago) given at some nodes of the tree are rough (and in part controversial) estimates based on different studies. The topology of the tree is also controversial: that Gnetales are more closely related to conifers than to angiosperms could be concluded from molecular data [15, 39, 133, 134], but is in contrast to widely accepted interpretations of morphological data. At the left side of the root and some internal branches of the tree three important stages in the evolution of the megasporangium are schematically depicted. From bottom to top: a sporangium that is not covered by an integument, a condition still found in extant ferns; a sporangium that is covered by an integument (ovule); and a sporangium that, in addition, is surrounded by a carpel. The gene names besides the branches denote gene subfamilies, not single genes. These have been established during the time interval represented by the respective branches of the phylogenetic tree, at the latest. This could be concluded from the presence of respective subfamily members in extant taxa. For example, *AG*-, *AGL2*-, *AGL6*-, *DEF*/*GLO*-, *GGM13*-, *STMADS11*- and *TM3*-like genes have been isolated from angiosperms and gymnosperms, but not from ferns. At the root of the phylogenetic tree, the domain structure of a typical MIKC-type MADS-box gene is shown. Our analyses have demonstrated that the last common ancestor of ferns and seed plants already had at least two genes of that type [85]. Abbreviations of genes or gene subfamilies: *AG*, *AGAMOUS*; *AGL2*, *6*, *12*, *15*, *17*, *AGAMOUS*-like gene 2, 6, 12, 15, 17; *CRM1*, *3*, *6*, *7*, *Ceratopteris* MADS-box gene 1, 3, 6, 7; *DEF*, *DEFICIENS*; *DEF*/*GLO*, a precursor of both *DEF*- and *GLO*-like genes; *GGM4-7*, *10*, *13*, *Gnetum gnemon* MADS-box gene 4–7, 10, 13; *GLO*, *GLOBOSA*; *OPM3*, *4*, *Ophioglossum pedunculosum* MADS-box gene 3, 4; *SQUA*, *SQUAMOSA*; *STMADS11*, *Solanum tuberosum* MADS-box gene 11; *TM3*, *8*, tomato MADS-box gene 3, 8.

Figure 5 (bottom). How the land plants learned the floral ABC. Different states of the ABCD model (some of them hypothetical) of floral organ specification are plotted onto a phylogenetic tree of major taxa of vascular plants. The organs specified by the different homeotic functions are indicated above the models. At the branch leading to the angiosperms (eudicots, monocots and basal angiosperms), different ancestral versions of the ABCD model that might have been present at the base of the angiosperms are shown. These versions have been suggested (from left to right) in this work, or in [6] or [3], respectively. The ages (in MYA, i.e. million years ago) given at some nodes of the tree are rough estimates (as in Figure 4). At the right side of some internal branches of the tree, gene subfamilies, not individual genes, are indicated (e.g., '*SQUA*' means the clade of *SQUA*-like genes). The relationships between representatives of these gene subfamilies and homeotic functions are symbolized by arrows. For example, a *SQUA*-like gene (i.e. *AP1*) provides the A function in *Arabidopsis*. (Note that *SQUA* itself is not an A function gene!). A *DEF*- and a *GLO*-like gene possibly provide the B function in all angiosperms, while *AG*-like genes provide both the C and the D function. The different relationships have been established during the time interval represented by the respective branches of the phylogenetic tree, at the latest. Abbreviations used: A, B, C, D, the floral homeotic functions; *AG*, *AG*-like genes; C/D, a precursor of floral homeotic functions C and D; *DEF*, *DEF*-like genes; *DEF*/*GLO*, a precursor of both *DEF*- and *GLO*-like genes; FM, function in the specification of floral meristems; *GLO*, *GLO*-like genes; IM, function in the specification of inflorescence meristems; *SQUA*, *SQUA*-like genes.

A third unusual observation was made with the intracellular localization of *CRM9* mRNA. In all tissues where *CRM9* expression was detected, *in situ* hybridization studies gave a strong signal in the nucleus, while in the cytoplasm, hybridization signals were much lower, if present at all [22]. Thus it seems that the majority of *CRM9* mRNA is retained in the nucleus and cannot be translated. It could be, therefore, that formation of CRM9 protein is not (only) regulated transcriptionally, but (also) by nuclear export of *CRM9* mRNA. It could also be, however, that *CRM9* represents a nonfunctional gene, or that *CRM9* does not function at the protein level. Whether nuclear export is linked to alternative splicing is unknown so far. It is conceivable, for example, that not all of the different splice variants can be exported, implying that alternative splicing would control nuclear export.

Nuclear retention of mRNA is also not unprecedented in ferns. It has been reported that phytochrome (*PHY1*) mRNA in the fern *Adiantum capillus-veneris* is predominantly nuclear in location in light-grown young leaves (croziers), while the mRNA in dark-grown tissue appears uniformly in both nucleus and cytoplasm [89]. These findings support the view that ferns have included nuclear export of mRNA into their repertoire of gene regulation.

Finally, let us gather the facts and try to reconstruct the last common ancestor of extant ferns and seed plants. Very likely, it had no ovules or floral organs but, like extant ferns, had naked sporangia and an independent gametophytic generation. It already had more than one MIKC-type MADS-box gene, but probably fewer than extant ferns or seed plants. None of the MIKC-type genes was an orthologue of a specific floral homeotic gene. These genes probably had quite a ubiquitous expression during the life cycle of the plant, possibly involving the gametophytic as well as the sporophytic phase. It seems likely that these genes were not organ identity genes, but had more general roles in the transcriptional control of development or cell differentiation, i.e. more comparable to the role of *MCM1* in the life of yeast.

MADS-box genes in gymnosperms

We summarize data suggesting that the last common ancestor of extant gymnosperms and angiosperms, about 300 MYA, already had at least 7 different MADS-box genes, i.e. AG-, AGL2-, AGL6-, DEF/GLO-, GGM13-, STMADS11- *and* TM3-*like genes. Probably, most of these genes were already*

involved in specifying reproductive organs, such as ovules, in the sporophyte. Expression of an ancestral version of the homeotic C and D functions, provided by an AG*-like gene, was probably used to distinguish reproductive from non-reproductive (vegetative) organs. In addition, expression of an ancestral B function, provided by a basal* DEF/GLO*-like gene, was possibly used to distinguish between male and female reproductive organs. Thus, orthologues of floral homeotic genes and a precursor of the ABCD system of floral organ specification (BC/D system) had probably already been established at the evolutionary base of extant seed plants.*

The term 'gymnosperm' (meaning 'naked seed') indicates that we are now dealing with plants that develop seeds, but in which these seeds are not enclosed within a carpel as in angiosperms (see below). Gymnosperms and angiosperms together constitute the taxonomic group of seed plants (spermatopsids, spermatophytes). Seed plants have become the most successful land plants, probably because of the selective advantage the formation of seeds gives these plants over all others [117]. Most likely the reason is that seeds are unrivaled in their capacity to disperse the next generation. Seeds are just ripened ovules, and ovules can be defined as integumented indehiscent megasporangia [117]. They consist of an envelope, the integument(s), with a micropyle, and a megasporangium (the nucellus) inside of which a megagametophyte develops. There is evidence that seed plants evolved from gymnosperm-like plants with a fern-like mode of reproduction called progymnosperms [7]. Therefore, the pollen sacs and nucelli of seed plants are probably homologous to fern sporangia. The transition from the naked dehiscent sporangia of fern-like ancesters to ovules characterizes one of the most important steps in land plant evolution (Figure 4). It involved several key innovations, such as the evolution of heterospory [117]. According to molecular data, the last common ancestor of extant seed plants existed about 300 MYA – recent estimations range from 285 to 348 MYA [40, 108] –, and earliest fossil evidence of gymnosperms dates back about 350–365 MYA [7, 121]. Gymnosperms are, therefore, phylogenetically much older than angiosperms (see below).

Extant gymnosperms comprise four groups: conifers, gnetophytes, cycads and *Ginkgo*. Only a few MADS-box gene cDNAs have been isolated from cycads and *Ginkgo* so far (Figure 3; and our unpublished data), since the focus of MADS-box gene research in gymnosperms has been on conifers (due to their ecological and commercial importance) and gnetophytes (because they are often considered a sister group of the angiosperms).

Conifer MADS-box gene cDNAs have been reported from spruce (*Picea abies*, *Picea mariana*) and pine species (*Pinus radiata*, *Pinus resinosa*) [64, 82–84, 105, 116, 119]. Phylogeny reconstructions revealed that the genes for which full-length cDNAs have been obtained so far all fall into gene clades well known from angiosperms, namely *AG*-, *AGL2*-, *AGL6*-, *DEF*/*GLO*- and *TM3*-like genes [81, 84, 116, 119, 123, 124, 134] (see also Figure 3). However, PCR cloning of a 61 bp segment using degenerate primers targeted to the MADS box suggested the presence of over 27 MADS-box genes within black spruce (*Picea mariana*), including several for which no orthologous angiosperm MADS-box gene has been identified yet [105].

In contrast to many angiosperm flowers, which are hermaphroditic, the investigated conifers are monoecious species that have truly unisexual reproductive axes. The female strobili (or seed cones) are compound axes consisting of two-scaled units with a sterile bract and a seed-bearing (ovuliferous) scale. In contrast, the male strobili are simple structures composed only of microsporophylls [119]. Expression studies indicated that the MADS-box genes identified so far are transcribed in male and female strobili. Some are also expressed in vegetative organs, such as the *AGL6*-like gene *PRMADS3* from Monterey pine (*Pinus radiata*), which is also transcribed in needle primordia [81]. The transcripts of the *TM3*-like gene *DAL3* from Norway spruce (*Picea abies*) were also found in vegetative shoots, but not in embryos, seeds or seedlings [119]. The *AGL6*-like gene *DAL1* is also expressed in vegetative shoots in their first year of development, but not in the epicotyl, including the apical meristem, of the seedling [118]. By *in situ* hybridization, *PRMADS1–3* from Monterey pine were found to be expressed in groups of cells that form ovuliferous scale and microsporophyll primordia [81]. Similarly, expression of the *AG*-like gene, *DAL2*, from Norway spruce was detected in ovuliferous scales, but not in bracts, the cone axis or the apical meristem [120]. Expression of its orthologue from black spruce, *SAG1*, was found to be very similar in female cones [105]. In male cones, *SAG1* expression was detected at a low level in the tissue that makes up the tapetal layer [105]. These data suggest that *AG*-, *AGL2*- and *AGL6*-like genes of conifers are all involved in reproductive structure formation.

To test whether the structural similarity and phylogenetic relatedness between *AG* from the flowering plant *Arabidopsis* and *DAL2* from Norway spruce is coupled to a similarity in function, an analysis of the effect of *DAL2* expression under the control of the constitutive 35S promoter in transgenic *Arabidopsis* plants was made [120]. Most transformants showed phenotypic alterations that seem not very informative with respect to *DAL2* function, such as curled rosette leaves and early flowering. Some transformants, however, had homeotic changes of flower organ identity. In these plants, the sepals had gained female characters at the margins, such as ovule-like structures and papillary cells characteristic of stigma. Petals had obtained male characteristics: they appeared to be transformed into filamentous organs or stamen-like organs with a filament-like proximal part capped with an anther-like structure. The third- and fourth-whorl organs were mainly unaffected by expression of the transgene [120]. The transformants thus resemble *Arabidopsis* plants ectopically expressing *AG* orthologues from angiosperms, such as *AG* itself or *BAG1* from *Brassica napus* [68, 75]. Since the ABC model predicts that the C function antagonizes the A function, the observed phenotype can be expected in case of a loss of the A function or an ectopic expression of the C function in the first and second whorl of the *Arabidopsis* flower. The results obtained with *AG* and *BAG1* have demonstrated that *AG* or its close relatives are sufficient to provide ectopically the homeotic C function. The simplest explanation for the results with *DAL2*, therefore, would be that *DAL2* activity in the perianth organs can functionally substitute for *AG* activity in ectopic expression experiments. This functional substitution would imply several partial functions, i.e. suppressing A gene activity, directing carpel identity to the outermost whorl, and interacting with B class genes (*AP3*, *PI*) in directing stamen identity to the second whorl of organs in transgenic flowers. However, it could also be that expression of *DAL2* results in an extension of *AG* expression into the perianth whorls, for example, because *DAL2* protein is able to activate the *AG* promoter, or because *DAL2* turns off the *Arabidopsis* A function. If so, the homeotic transformation of whorl 1 and whorl 2 organs would be the result of ectopic *AG* expression, or of the formation of functional AG-DAL2 heterodimers rather than *DAL2* alone. In either case, the data indicate that DAL2 is able to interact with components of the regulatory context of *AG*, and that thus these kinds of interactions have been conserved over at least 300 million years (the logic of such conclusions has also been illustrated elsewhere [124]). It might be that MADS, I and K domains of DAL2 are needed for these interactions, explaining why these are so similar to those of AG. However, complementation of an *AG* loss-of-function mutant with the *DAL2* gene could provide a more stringent test for the extent to which *DAL2* is able to substitute the *AG* function in the *Arabidopsis* context. Results very similar to the ones reported here for *DAL2* have also been obtained with its black spruce orthologue, *SAG1* [105].

By definition, C class genes are involved in specifying stamen and carpel identity. Since there are no stamens and carpels in gymnosperms, the question arises as to which function *DAL2/SAG1* fulfills in the conifer context. Note that even successful heterologous transformation studies, as described above, may not always answer such questions! Spruce *DAL2/SAG1* mutants that could give an answer are also not available. Expression studies suggest that *DAL2/SAG1* is involved in the determination of ovuliferous scale or ovule identity, and of male reproductive organ identity. The ability to convert petals into stamens in *Arabidopsis* is consistent with the notion that *DAL2/SAG1* might be able to interact with B class genes in specifying male reproductive organs. The presence of *DEF*/*GLO*-like genes in conifers could be predicted from phylogeny reconstructions [120], but is now also supported by gene cloning (see below). Thus *DAL2/SAG1* might interact with one or several *DEF*/*GLO*-like genes from spruce in order to specify male reproductive organ identity. Expression studies and transgenic experiments both suggest, therefore, that *DAL2/SAG1* function is more similar to that of angiosperm C function than D function genes (whose expression is restricted to ovules, implying that they are not expressed in male reproductive organs, and whose function is in specifying ovule identity). At first glance, this may seem a paradox, since ovules, in contrast to carpels, are present in all gymnosperms and are thus very likely phylogenetically older. Specifying ovules (i.e. D function), therefore, should be a more ancient function of *AG*-like genes than specifying stamens and carpels (i.e. C function). One has to take into consideration, however, that *DAL2/SAG1* function might be ancestral to both C and D functions. So how can the early evolution of *AG*-like gene function in seed plants be conceived? Only one type of *AG*-like gene has been isolated so far from any gymnosperm species (Figure 3). Phylogeny reconstructions suggest that these genes are basal to both the C- and D-function genes from angiosperms (Figure 3). We

suggest, therefore, that it could have been the ancestral function of these genes to distinguish reproductive organs such as male sporophylls and ovuliferous scales, including ovules (where expression is on) from vegetative organs, including cone bracts (where expression of these genes is off). Genes such as *DAL2/SAG1* may still provide such a function today. Later, at the level of angiosperms, a gene duplication and diversification event might have resulted in the fixation of two different genes. While one gene type (C class genes; *AG*-like genes *sensu stricto*) specialized in specifying stamens and carpels, the other (D class genes; *FBP7*/*FBP11*/*AGL11*-like genes) became restricted to specify ovule identity (Figure 5).

Gnetophytes (Gnetales) are an enigmatic group of seed plants with only three genera, *Gnetum*, *Ephedra* and *Welwitschia*. Most phylogenetic analyses of morphological data agree that among the groups of extant seed plants, the gnetophytes are the sister group of the angiosperms [21, 24, 25]. According to this view, angiosperms and gnetophytes are members of a clade called 'anthophytes', to emphasize their shared possession of flower-like reproductive structures [21]. Since answers to the still unresolved question of angiosperm origin are intimately connected to the identification of their sister group among extinct and extant taxa [21, 39], gnetophytes have found much scientific interest. However, some recent phylogeny reconstructions based on molecular data do not support an anthophyte clade; instead, they favor monophyly of extant gymnosperms, albeit with low bootstrap support, implying that gnetophytes are more closely related to conifers than to angiosperms [15, 39].

cDNA sequences of 13 different single-copy MADS-box genes of the gnetophyte *Gnetum gnemon* have been published so far ([133, 134]). Phylogeny reconstructions indicated that seven of them are members of novel gene subfamilies, for which members from dicots have not been published so far (see Figure 3). In one case (*GGM13*), however, a highly related sequence has been isolated recently from a monocotyledonous flowering plant (our unpublished results). Due to the limited knowledge about the number and type of MADS-box genes in any plant species (including *Arabidopsis*) it remains to be seen if orthologues of the other genes are also present in flowering plants. Alternatively, the respective gene clades originated within the gymnosperms after the lineage that led to the angiosperms had already branched off. Most of the genes are expressed in male and/or female strobili, but not in leaves, suggesting that they have functions similar to the floral meristem or organ identity genes of angiosperms ([133], and our unpublished results).

Phylogeny reconstructions revealed that the other six genes (*GGM1*, *2*, *3*, *9*, *11*, *12*) fall into well defined gene clades known already from angiosperms, i.e. *STMADS11* [13], *TM3*-, *DEF*/*GLO*-, *AG*-, or *AGL6*-like genes, respectively ([134], and our unpublished data). They are thus putative orthologues of the respective genes from angiosperms. Among them is *GGM2*, the first *DEF*/*GLO*-like gene (B class gene orthologue) reported from a gymnosperm [133, 134] (Figure 3). The presence of a *DEF*/*GLO*-like gene, however, is not a synapomorphy uniting flowering plants and gnetophytes, since genes belonging to that clade have meanwhile also been found in two conifer species, Norway spruce and Monterey pine [84, 116] (Figure 3). Whether these genes are more closely related to *DEF*- or *GLO*-like genes, or are basal to both (as suggested by Figure 3), could not be clarified unequivocally by the construction of phylogenetic gene trees so far (our unpublished results). Analysis of the exon-intron structure of *GGM2*, however, supports the latter hypothesis (our unpublished results). At this time, therefore, we favor the hypothesis that there was only one *DEF*/*GLO*-like gene in the last common ancestor of extant gymnosperms and angiosperms. The gene duplication that generated distinct *DEF* and *GLO* clades may have happened in the angiosperm lineage after the lineage that led to extant gymnosperms had already branched off (Figures 4 and 5). A close relationship between *GGM2* and the other members of the *DEF*/*GLO* clade is not only supported by phylogeny reconstruction, but also by the presence of a 'paleo AP3 motif' at the C-terminal end of the GGM2 protein and a 'derived PI motif' in a subterminal position (our unpublished results; for the definition of the motifs, see [60]). Moreover, in sequence alignments GGM2 shares a highly specific character state at an indel (insertion-deletion) position with all other DEF- and GLO-like proteins (our unpublished data). However, both features have also been found for *GGM13* (our unpublished results), which according to phylogeny reconstructions is a slightly more distant relative of *DEF*- and *GLO*-like genes (Figure 3).

Analogous to the observations with *DAL2/SAG1*, expression of the *AG*-like gene *GGM3* was found in male as well as female strobili of *Gnetum*, but not in leaves. The *TM3*-like gene *GGM1* showed, as expected, a more ubiquitous expression in male and female strobili and in leaves. However, expression

of the *DEF*/*GLO*-like gene *GGM2* was found to be restricted to male strobili [133, 134]. It could have been an ancient function of members of that gene clade, therefore, to distinguish between male (where gene expression is on) and female reproductive structures (where expression is off) (see Figure 5). It is easy to imagine how the floral homeotic B function of angiosperms evolved from such a gene, since it also distinguishes male reproductive organs (i.e. stamens, expressing B plus C function) from female ones (i.e. carpels, expressing only C function). It also seems plausible that this gene function was recruited to specify petals (expressing A and B function) when these were 'derived' from stamens in some lineages of angiosperms (see below) [3, 60].

Phylogeny reconstructions indicated that in all cases where gene subfamily members are available from angiosperms, gnetophytes and conifers, i.e. within the *AG*-, *AGL6*-, *DEF*/*GLO*- and *TM3*-like genes, the genes from *Gnetum* always form subclades together with conifer genes, to the exclusion of the angiosperm genes (Figure 3). This finding provides molecular evidence for the hypothesis that gnetophytes are more closely related to conifers than to angiosperms (Figure 4). The conclusion is in contradiction to the anthophyte theory and to widely accepted interpretations of morphological data for almost a century [5, 21, 24]. The sister group relationship between gnetophytes and conifers makes it likely that many of the angiosperm-like features of Gnetales, such as the flower-like appearance of reproductive structures, reduced female gametophytes, double-integumented ovules, dicotyledonous seeds, vessels in the secondary wood, net-veined leaves and the presence of double-fertilization, are homoplasies rather than homologous character states. With respect to angiosperm origins, gnetophytes are thus possibly less informative than often thought [21, 24, 25]. It could be, however, that the parallel appearance of the mentioned characters in angiosperms and gnetophytes was facilitated by a common developmental potential that was already present in the last common ancestor of (gnetophytes + conifers) and angiosperms (or even of all extant seed plants, if extant gymnosperms represent a monophyletic group [15]). It seems an exciting hypothesis that a set of MADS-box genes might have been part of the developmental potential facilitating convergent evolution in different seed plant lineages. Therefore, this last common ancestor is of considerable evolutionary interest, so let us try to reconstruct it with respect to some morphological features and MADS-box genes, taking together the data reviewed above. To simplify things, we assume that extant gymnosperms are really a monophyletic group, and that gene types that have been found in angiosperms as well as in gnetophytes or conifers were thus present in the last common ancestor of all extant seed plants.

Like ferns, the most recent common ancestor of extant seed plants probably had an elaborate two-phase life cycle with a dominating sporophytic generation. In contrast to most ferns, however, the sporophyte produced two types of spores, micro- and megaspores, and the megagametophytes developing from the megaspores were not independent, but remained within the ovules of the sporophyte. After fertilization, the ovules developed into seeds. The sporophyte perhaps had unisexual reproductive axes. Figure 3 and data published elsewhere [134] show that there are five different well-defined clades containing MADS-box gene members from both gymnosperms and angiosperms, indicating that at least five different MIKC-type genes existed in the last common ancestor of contemporary seed plants, namely at least one representative of each of the clades of *AG*-, *AGL2*-, *AGL6*-, *DEF*/*GLO*- and *TM3*-like genes. In addition, there was most likely a sixth gene closely related to *GGM13*, and a seventh gene closely related to *GGM12*, because putative orthologues for these genes also were isolated from angiosperm species ([13], and our unpublished data). Probably most of these genes were already involved in specifying reproductive organs of the sporophyte. The last common ancestor of extant seed plants probably used an ancestral version of the homeotic C and D functions (C/D function), provided by an *AG*-like gene, to distinguish reproductive from non-reproductive organs. In addition, it possibly used an ancestral B function provided by at least one *DEF*/*GLO*-like gene to distinguish between male and female reproductive organs. Thus a precursor of the ABCD system of floral organ specification had probably been established already as a BC/D system at the base of extant seed plants, while it was completely absent in the last common ancestor of ferns and seed plants (Figure 5). The data on MADS-box genes in ferns suggest that there was a relatively small pool of MIKC-type genes in the last common ancestor of ferns and seed plants. Therefore it is likely that descendants of that pool of genes were generated by gene duplication, diversification and fixation, and were recruited in the lineage leading to seed plants to give rise to floral homeotic genes. It is conceivable, therefore, that in the time interval prior to the radiation of extant seed

plants, but subsequent to their divergence from fern-like ancestors, i.e. between 300 and 400 MYA, some if not most clades of MADS-box genes known from angiosperms had been established (Figure 4). There is a striking temporal coincidence between the appearance of these genes and the occurrence of seed plants and the seed habit. For example, the oldest known seed plant (*Elkinsia*) has been preserved in the fossil record of that time interval (Late Devonian, about 365 MYA), and different intermediate stages in the evolution of the ovule have been found in the fossil record of the Lower Carboniferous, about 350 MYA [121]. We are not aware that such a clear coincidence between the appearance of new types (clades) of developmental control genes (such as *AG*-like genes) and the appearance of novel morphological structures (such as ovules and seeds) has ever been reported for the macroevolution of a non-plant system. Since the extant descendants of these genes are expressed in ovules, ovuliferous scales, or seeds, and thus probably are involved in controlling the development of these structures, it seems quite possible that the establishment of the new clades of MADS-box genes at the time of ovule and seed 'invention' was not just a coincidence, but an important functional step in the evolutionary establishment of these structures.

Progymnosperms, i.e. plants that already had gymnospermous wood but still a pteridophytic, free-sporing mode of reproduction, also existed in that critical time interval 300–400 MYA, since their fossils have been found from Middle Devonian to Early Carboniferous (Tournaisian) [7]. It is intriguing to think, therefore, that during this time the establishment of *AG*-like genes in progymnosperms might have been an important aspect to confer ovules to plants that still had a pteridophytic mode of reproduction, but otherwise were already gymnosperm-like [85]. The progenitor of extant seed plants, established at this time, was the starting point for the evolution of the enormous morphological diversity we see in present-day seed plants.

Due to the large morphological gaps between the different seed plant groups (extant and fossil), homologies between their reproductive structures are often difficult to assess [24]. This is especially true for the floral organs of angiosperms compared to the organs of the reproductive units of the gymnosperms. It is one of the reasons why definite answers to the question of what a flower actually is and from which organs of which gymnosperms its organs were derived have been lacking (for a review, see [21]). However, since homologous organs should generally express orthologous developmental control genes, we have good reasons to assume that MADS-box genes are suitable tools to test assumptions about structural and developmental homologies among the reproductive structures within the diverse seed plant groups [24]. For example, some evolutionary models suggest that angiosperm petals are homologous to the outer integument of *Gnetum* reproductive units [24]. If so, orthologues of B function genes such as *GGM2* should be expressed in the outer integument of *Gnetum*, which exists in male as well as female strobili. However, *GGM2* is not expressed in female strobili at all [133, 134]. *GGM2* expression in male strobili is also not in the integuments surrounding the antherophores, but only in the antherophore itself [134]. Expression of the *AG*-like gene *GGM3* in the *Gnetum* outer integuments [133, 134] makes it also appear unlikely that they are homologous to perianth organs of angiosperms, but would be compatible with an alternative model due to which the outer integument of *Gnetum* is homologous to the integument of angiosperm ovules [24] or even to carpels. In line with this, *SAG1*, one of the conifer orthologues of *GGM3*, is especially strongly expressed in the integuments of the ovules [105].

For several reasons, however, these conclusions are still preliminary. For example, orthology between respective genes from gymnosperms and angiosperms should be tested more rigorously, and independent co-option (recruitment) of genes into nonhomologous developmental processes cannot be excluded, so that more genes should be analyzed (for discussion of that problem, see [1]). However, we believe that the strong correlation between MADS-box gene phylogeny and the evolution of certain morphological structures (e.g. ovules) promises that studies such as the ones indicated here will help to clarify the origin of the flower.

It has often been argued that there are insuperable morphological gaps between angiosperms and gymnosperms which are even more difficult to overcome than the gap between ferns and seed plants. With respect to MADS-box genes and the system of reproductive organ specification, we obviously see the opposite: while there are probably no orthologues of floral homeotic genes in ferns, there are clearly some in gymnosperms (Figures 3, 4 and 5). At the level of molecular developmental control, the reproductive units of gymnosperms are thus more similar to the flowers of angiosperms than morphological studies may have suggested.

MADS-box genes in basal angiosperms

Basal angiosperms are crucial for our understanding of flower origin. Although we do not know yet how the 'first flower' looked, we are quite sure that the last common ancestor of extant angiosperms already had at least 9 different MADS-box genes. These were distinct representatives of the clades of DEF- *and* GLO-*like genes, an* AGL15-*like gene, and the set of genes that was already present in the last common ancestor of extant seed plants (besides a* DEF/GLO-*like gene,* AG-, AGL2-, AGL6-, GGM13-, STMADS11- *and* TM3-*like genes).*

Our considerations have now reached the flowering plants. Since flowers are often defined as short, specialized axes bearing closely aggregated sporophylls, gymnosperms and even some pteridophytes (such as clubmosses) may also produce 'flowers'. It is necessary to clarify, therefore, that when we use the term 'flowering plants' within this review, we mean the angiosperms (i.e. flowering plants *sensu stricto*). The term angiosperm means 'vessel seed'. Besides stamens with two pairs of pollen sacs, the most useful diagnostic morphological feature of angiosperms is a carpel enclosing the ovule/seed [21]. The carpel is the morphological basis for fruit development. From the naked sporangia of ferns via the integumented megasporangia (ovules) of gymnosperms (resulting in seeds) to the angiosperm ovules enclosed in carpels (resulting in seeds within fruits) we see a clear macroevolutionary tendency to cover the megasporangium and its derivatives (see Figure 4).

The angiosperm mode of reproduction has proven very succesful, because flowering plants now dominate the vegetation of most ecosystems on land, and they consist of more species than all other groups of land plants combined (about 250 000–300 000) [21]. One probable reason for the angiosperms' success is that fruits provide additional possibilities for an effective distribution of seeds, for example by the help of animals. In many cases animals are also important for outcrossing during sexual reproduction. The capacity to outcross effectively is the second major advantage of the angiosperms. It is facilitated by flower types that efficiently attract diverse pollinators (bees, beetles, birds, etc.), depending on the angiosperm species.

The sudden appearance and considerable diversification of the angiosperms within the fossil record of the Early Cretaceous, about 130–90 MYA, seems still almost the same 'abominable mystery' as it was to Charles Darwin more than a century ago. As already mentioned in the section on gymnosperms, the origin of the flower has also remained a mystery. Homologies between organs within gymnosperm and angiosperm reproductive units are unclear, and the long-standing question of whether angiosperm flowers derive from a simple branch or from multiple branches (euanthial vs. pseudanthial scenario) is still unresolved [21]. We have noticed, however, that according to considerations outlined by Doyle [24], the preliminary expression data of the *Gnetum* genes *GGM2* and *GGM3* (see the gymnosperm section) suggest organ homologies that fit to a pseudanthial rather than an euanthial model of flower origin [134].

Current hypotheses of angiosperm evolution have identified two large clades (monocots and eudicots, see below) embedded within a poorly defined basal assemblage of magnoliid dicots (Magnoliidae) [21], which we call 'basal angiosperms' here. There is a great diversity of floral structure and biology among basal angiosperms. Both large, multiparted bisexual flowers and small, simple, frequently unisexual flowers are widespread, and variation in the number and arrangement of floral parts is extreme [21]. This, and the substantial morphological gap between gymnosperms and angiosperms (see above), has prevented identification of the basic condition of the angiosperm flower. Did the 'first flower' more look like a *Magnolia* flower with its numerous elaborate tepals, or like one of *Sarcandra glabra* with a single bract, stamen and carpel [21]?

Our inability to reconstruct the 'first flower' implies that we do not know the succession of steps in the evolution of the molecular 'control' of flower formation. How did the BC/D system of reproductive organ specification possibly present in gymnosperms change into the ABCD model of floral organ identity? However, educated guesses about plausible intermediate steps and the implications for MADS-box gene phylogeny can be made (Figure 5). A most primitive flower might just have been composed of one or more stamens and carpels (including ovules) without a perianth, such as the flowers of *Sarcandra*. We only need B, C and D function genes expressed in a suitable combinatorial way along a single reproductive shoot axis to specify the respective organs. Perianthless flowers are not prominent among the different suggestions of what the earliest flower might have looked like. However, since the identity of the organs of perianthless flowers could be completely specified with homeotic functions that were possibly present already in gymnosperms (Figure 5), we argue that such simple

flowers should be seriously considered as a plausible model for the 'first flower'. In fact, such flowers might help to bridge the enormous gap between gymnosperm and angiosperm reproductive structures (see the gymnosperm section).

The more convential models assume that the most ancestral flower already had a perianth. One hypothesis suggests that the ancestral condition was a single, petaloid whorl expressing both A- and B-function genes (Figure 5) [3]. According to this hypothesis, the calyx whorl, expressing only the A function, was later added externally to protect flower buds from predation. Other ancestral ABCD models assume that the basal flower had one or more sepaloid perianth whorls specified by A-function genes. Petals, and thus the distinction between corolla and calyx, could have evolved later by the outward extension of B function into the inner of two perianth whorls (Figure 5) [6].

These models provoke several other questions: how were sepals or petals generated? Where did the A function come from? It is widely accepted that, in contrast to the reproductive organs (stamens and carpels), which evolved only once, sterile perianth organs originated several times independently within the angiosperms, although details are unresolved ([29, 60], and references therein). Similar organs, such as petals, are thus not necessarily homologous (in the meaning of 'derived from a common ancestor', i.e. historically orthologous, as defined elsewhere [3]). It has been concluded from morphological evidence that petals have been derived many times independently from stamens, for example, several times within the lower eudicots and at least once at the base of the higher eudicots. Such petals are called andropetals. Among the basal angiosperms, the Nymphaeales probably have andropetals. A second type of petals, bracteopetals, may have been derived from sepals or sterile organs subtending the flowers. Most basal angiosperms are assumed to have bracteopetals, such as the Magnoliales, Piperales and Aristolochiales ([60], and references therein). As outlined elsewhere, however, there are severe conceptional problems with these simple views [3]. For example, historical, positional and process homology (the latter meaning that two structures are specified by the same type of genes) should be distinguished. Moreover, with respect to homology, it would be more appropriate to distinguish between orthology and paralogy (for a detailed discussion of this topic in floral development, see [3]). If not stated otherwise, we use the terms homology, orthology and parology here only in their historical dimension, meaning, for example, that a feature is homologous in different lineages if it was already present in the last common ancestor of these lineages.

The origin of the A function is another notorious problem. It seems that there are different kinds of genes behind it even within the higher eudicots (see below), so a simple and general answer might not be possible. We have noted, however, that in case of *Arabidopsis*, the two A-function genes also function as floral meristem identity genes. Determining floral meristem identity might be a function that is needed earlier than the specification of floral organ identity. This is surely true for ontogeny, but might also be true in the case of evolution, because the earliest flowers did not necessarily have a perianth specified by A-function genes (see above), but very likely already needed some floral meristem identity function to distinguish floral from vegetative tissue. Therefore, we suggest that at least in some cases, the A function could be a derivative of the function determining floral meristem identity (Figure 5). In line with this, A-function genes or their orthologues from other species (such as *SQUAMOSA* from *Antirrhinum*) are often expressed in several whorls of the flower and in non-floral organs, not just in sepals and petals [50, 52].

It is obvious that studies on *SQUA*-, *DEF*-, *GLO*- and *AG*-like MADS-box genes in basal angiosperms with different floral structures (simple and complex ones) may help to understand the evolution of the ABCD model. It will be interesting to examine, for example, how the independent derivations of petals during angiosperm evolution are reflected in the use of organ identity genes. Have independently derived petals always recruited *DEF*- and *GLO*-like genes to specify petal identity? Or have other types of genes taken over that function in lower angiosperms? If the latter is true, *DEF*- and *GLO*-like genes are not necessarily expressed in the petals of some lower angiosperms! Petal specification genes which are not *DEF*- or *GLO*-like genes may seem more likely in the case of bracteopetals than in the case of andropetals, because bracts or sepals do not express *DEF*- and *GLO*-like genes, while stamens do.

The origin of the first *SQUA*-like gene is another open question. Did it appear already within early angiosperms before the lineages that led to extant monocots and eudicots separated, or even at the gymnosperm level? How was it derived – by gene duplication from an *AGL2*/*AGL6*/*SQUA* ancestral gene, as phylogeny reconstructions may suggest (Figure 3) [85, 97, 123]? Do *SQUA*-like genes also specify sepals

or petals in species other than *Arabidopsis* (e.g., in basal angiosperms), or was the A function a role acquired later by *SQUA*-like genes in the lineage that led to *Arabidopsis*?

During the course of a study devoted to the role of *DEF*- and *GLO*-like genes in petal and stamen evolution, Kramer *et al.* [60] have isolated respective cDNA clones not only from higher and lower eudicots, but also from the Magnoliaceae *Michelia figo* and *Liriodendron tulipifera*, and the Piperaceae *Peperomia hirta* and *Piper magnificum*, which are all basal angiosperms. It was during the course of this work that the already mentioned GLO and DEF specific motifs were detected (see the fern and gymnosperm sections). The study by Kramer *et al.* documents a high frequency of gene duplications within both the *DEF* and *GLO* lineages. The authors assume that at the base of the *DEF*/*GLO* lineage was a *CRM3*-like fern gene containing a paleoAP3 motif [60], but support for this hypothesis by phylogeny reconstructions is weak at best (see the fern section). There is hardly any doubt, however, that a *DEF*/*GLO*-like gene encoding a terminal paleoAP3 motif and a subterminal PI motif was already established when gymnosperms started to diverge (see the gymnosperm section). A key ancestral gene duplication occurred near the base of the angiosperms, resulting in the distinct lineages of *DEF*-like genes (which retained a highly conserved paleoAP3 motif, while the PI motif diverged more strongly) and *GLO*-like genes (which have lost the paleoAP3 motif, but maintained a highly conserved PI motif). A second major duplication event occurred in the *DEF* lineage near the base of the higher eudicot radiation. It resulted in a euAP3 lineage (including *AP3* and *DEF*) in which the euAP3 motif replaces the paleoAP3 motif, and a *TM6* lineage, in which the paleoAP3 motif is maintained. This duplication event may reflect the origin of a petal-specific *DEF* function in the higher eudicot lineage at the time when these plants recruited petals from stamens [60]. Many other independent gene duplication events in the different angiosperm lineages followed the creation of the separate *DEF*, *GLO* and *TM6* clades, resulting in pairs of highly related paralogues within several species such as lily and rice (see Figure 3).

The functions of the different DEF and GLO specific motifs have not yet been reported, nor have the expression patterns of the *DEF*- and *GLO*-like genes from basal angiosperms been published. With these genes in hand, however, the tools are available now to test some of the hypotheses outlined above.

AGL15 is an interesting type of MADS-box gene which is expressed in developing embryos and thus might be involved in 'controlling' embryogenesis [47, 104]. Embryos are formed by all land plants – hence they are also called embryophytes –, making it conceivable that *AGL15*-like genes may even exist in nonseed plants. However, evidence for that is missing: *AGL15*-like genes have been published so far only from Brassicaceae species (Figure 3). The isolation of an *AGL15*-like gene from the basal angiosperm *Magnolia* (Hilde Fischer, personal communication) at least suggests that this type of gene was already present near the base of the flowering plants (Figure 4).

Summing up, we do not know what the first flower looked like, but it probably employed a genetic system for floral reproductive organ specification that had already been established at the gymnosperm level (BC/D system). It is relatively clear that the first flowering plant already had at least nine different MADS-box genes (seven gene types known already from the last common ancestor of extant seed plants, plus an *AGL15*-like gene and separate lineages of *DEF*- and *GLO*-like genes; Figure 4). It is very likely that some of these genes provided part of the molecular basis for the enormous diversification of the flower structure during angiosperm evolution.

MADS-box genes in monocots

Monocots comprise taxa with quite different inflorescences and flowers, such as grasses and lilies. Some data suggest that the B and C functions in these flowers work quite similarly to those of the eudicots. However, the B function in grasses specifies lodicules rather than petals in the second whorl of the flower. In lilies and their close relatives the B function is very likely not only expressed in the second and third, but also in the first floral whorl. Therefore, lilies and their relatives have a simple perianth (perigon) composed of two whorls containing petaloid organs called tepals. In addition to the MADS-box genes that were already present in the last common ancestor of extant angiosperms, the last common ancestor of monocots and eudicots about 200 MYA also already had an AGL17-*and a* SQUA-*like gene. Thus it contained at least 11 different MADS-box genes.*

Among the angiosperms, the monocotyledons (Liliopsida) are defined as a monophyletic group by their single cotyledon and some other features [21]. According to molecular estimates, the monocot lineage

separated from the other angiosperms about 160–200 MYA [40, 135]. However, the oldest known fossils of monocots were deposited just about 90 MYA [33].

Monocots are of great interest here for at least two reasons: the structural diversity of their flowers and inflorescences, and the commercial importance of their flowers, seeds and fruits. The importance of monocots for human culture could be one of the reasons why MADS-box genes have been studied in quite a number of diverse species, including cereal grasses such as maize (*Zea mays* ssp. *mays*), rice (*Oryza sativa*), wheat (*Triticum aestivum*), and sorghum (*Sorghum bicolor*), a lily species (*Lilium regale*), a tulip species (*Tulipa gesneriana*), asparagus (*Asparagus officinalis*), and an orchid species (*Aranda* × *deborah*). Compared to our knowledge about MADS-box genes in nonflowering plants, a great deal is known about these genes in monocots, so that a comprehensive review is not possible here. Therefore, we just give an overview of present knowledge, with a focus on recent progress.

Although the first MADS-box gene cDNA reported from a monocot was from an orchid [65], the majority of MADS research focuses now on the cereals maize and rice [10–12, 17, 18, 31, 32, 42, 53–55, 71, 72, 110, 126]. These two species not only feed the world to a large extent, but are also suitable model systems for plant genetics and molecular biology. Comparatively little is known about MADS-box genes in sorghum [42] and wheat [86].

Like other typical grasses, cereals produce tiny, wind-pollinated flowers that are distinct from the flowers of other taxa. Although the flowers themselves are simplified and small, they are generally assembled into complex higher-order structures (spikelets, inflorescences). Let us take maize as an example. In contrast to rice, which forms hermaphroditic flowers, maize is a monoecious species, i.e. it generates male and female inflorescences separately on the same plant. The male inflorescence (tassel) develops in a terminal position, whereas the female inflorescences (ears) grow in the axils of vegetative leaves. The unisexual flower types of the tassel and ear are both derived from an initially bisexual state through the abortion of pistil primordia in the tassel and stamen primordia in the ear [16]. The three stamens or carpels (pistil) of each maize flower are surrounded by a pair of bract-like organs called palea (inner) and lemma (outer), thus constituting structures called florets. In the flowers of the tassel lodicules are also formed, two knob-like perianth organs which are needed to open the florets at anthesis. Two florets – an upper and a lower one – are together enclosed by another pair of bract-like organs called glumes, thus forming spikelets. In the spikelets of female inflorescences, only the upper floret develops due to abortion of the lower floret tissues at early developmental stages. In the spikelets of male inflorescences, both florets develop to maturity. Spikelets are formed in pairs along the ear and tassel inflorescence, with one spikelet being pedicellate, the other sessile. The inflorescences and flowers of maize and other grasses are thus in some respect similar, in others different from the flowers of other taxa (for reviews about floret, spikelet and inflorescence structures in cereals, see e.g. [16, 19, 109]). It is an interesting question, therefore, how these similarities and dissimilarities are reflected in the structure, expression and function of MADS-box genes.

Since the stamens, carpels and ovules of monocots and eudicots are probably homologous organs, it seems likely that they are specified by orthologous B, C and D class homeotic genes, i.e. *DEF*-, *GLO*- and *AG*-like MADS-box genes. Concerning C class genes, phylogeny reconstructions suggest an orthologous relationship between *ZAG1*/*ZMM2* from maize on the one hand and *AG* from *Arabidopsis* on the other [110, 126]. Employing a reverse genetics approach, a putative null allele of *ZAG1* was identified [71]. Surprisingly, the floral phenotype did not show a homeotic transformation of reproductive organs into nonreproductive ones, which would have been expected from a C function gene. Rather, supernumerary carpels were observed, indicating a loss of floral meristem determinacy. Besides specifying stamen and carpel identity (C function), *AG* also plays a role in establishing floral meristem determinacy. Possibly *ZAG1* has only the latter aspect of the *AG* function. Since *ZAG1* is expressed in stamen and carpel primordia, however, it seems more likely that there is a redundancy in C function between *ZAG1* and *ZMM2* [71, 109] or other *AG*-like genes (see below). Expression patterns suggest that *ZAG1* is more important for carpel development, while *ZMM2* should be more important for stamen development. *ZMM2* mutants and *ZAG1*/*ZMM2* double mutants will show whether these hypotheses are correct. However, even this will probably not be the end of the story. It is well known that, due to the segmental allotetraploid origin of the maize genome, genes that are single copy in diploid plant species are represented by a pair of genes in maize [34, 126]. While most gene pairs in maize trace back to diversification events that started either about 11 or 21 MYA, the *ZAG1*/*ZMM2*

gene pair is much more ancient (60 MYA) [34]. It could well be, therefore, that *ZAG1* and *ZMM2* have duplicate loci in the maize genome, and that even more than two genes are involved in providing the C function and floral meristem determinacy. A putative duplicate locus of *ZMM2* was cloned recently (our unpublished results). In phylogeny reconstructions this gene forms a clade together with *ZMM2* and *OSMADS3* from rice, to the exclusion of *ZAG1*, suggesting an ancestrally orthologous relationship between the gene pair constituted by *ZMM2* and its duplicate locus, and *OSMADS3* (our unpublished results). Interestingly, ectopic expression of *OSMADS3* in tobacco (*Nicotiana tabacum*) driven by the CaMV 35S promoter resulted in phenotypic alterations mimicking the results of ectopic expression of *AG* [55], indicating that *OSMADS3* can substitute a C function gene in transgenic experiments (for a more thorough interpretation of these kinds of experiments, see the gymnosperm section). It seems reasonable to hypothesize that not only *OSMADS3*, but also *ZMM2* and its duplicate locus represent C function genes in grasses.

Thanks to the cloning of the *SILKY1* gene, considerable progress was also made in understanding the B function in maize [109]. The *silky1* mutant has a phenotype strikingly similar to the B function mutants of eudicots. Loss of B function in *Arabidopsis* or *Antirrhinum* leads to a conversion of second-whorl petals into sepals and third-whorl stamens into carpels. In the *silky1* mutant of maize, the stamens in the tassel develop as carpel-like structures, and the lodicules are replaced by palea-like organs. In the ear of a *silky1* mutant, the stamens are converted to carpel-like structures, and the normal program of stamen abortion is bypassed [109]. *SILKY1* is expressed in organ primordia which give rise to lodicules and stamens. According to sequence analysis, *SILKY1* is probably an orthologue of the B class gene *DEF* [109]. The data suggest that with respect to the B function, the ABCD model can be applied to grass species (Figure 5). Moreover, they also suggest that lodicules are homologous to eudicot petals and that the palea is homologous to eudicot sepals (and is not a prophyll, as other interpretations have it). It could also be, however, that orthologous genes have been recruited independently for the specification of organs that are historically not orthologous (in the terminology used by Albert *et al.* [3]).

The mutant phenotype of *OSMADS4*, a *GLO*-like gene from rice, is very similar to that of *silky1* [54], suggesting that, as in higher eudicots, both a *DEF*- and a *GLO*-like gene are necessary to provide the floral homeotic B function in monocots. Whether DEF- and GLO-like proteins interact in monocots in the same way as they do in eudicots is unknown so far.

It is not yet known whether MADS-box genes are involved in providing the A function in monocots (as the *SQUA*-like gene *AP1* does in *Arabidopsis*). Clearly, however, *SQUA*-like genes are present in maize, sorghum and lily (Figure 3) ([10, 31, 42, 72], and our unpublished results). Consequently, this gene clade had already been established in the last common ancestor of monocots and eudicots (Figure 4). Ancestral *SQUA*-like genes could have been involved in specifying inflorescence or floral meristem identity (see the basal angiosperm section). The relatively large number of these genes in some extant species suggests that they later may have been recruited for several different functions (for results supporting that hypothesis see also the eudicot section). For example, cDNAs representing 5 different *SQUA*-like genes have already been isolated from maize ([10, 31, 72], and our unpublished results). For some of them, expression patterns suggest that they are not involved in establishing floral meristem identity, but function at later stages of floret development [10].

The large number of *SQUA*-like genes in maize may seem surprising. However, recent increases in gene number in some clades (by gene duplication, in maize also by ancient allotetraploidy) are a common theme in grasses. For example, 8 different *AGL2*-like genes have been isolated from maize so far (our unpublished results), and for 4 of them putative orthologues from rice or sorghum have already been found (Figure 3). These genes have a broad range of expression patterns, suggesting also a functional diversification [10–12, 123]. For example, *ZMM6* expression is initially restricted to just one primordium out of each pair of developing spikelet primordia, suggesting that this gene is involved in determining the alternative identity of spikelet primordia (pedicellate vs. sessile spikelet) [11, 122]. Expression of *ZMM8* and *ZMM14* is detectable only in the upper, not in the lower, floret of each developing spikelet, suggesting that these genes determine the alternative identity of the upper vs. the lower floret within each spikelet primordium. Alternatively, these genes may be involved in conferring determinacy to the spikelet or upper floret meristem [12]. The timing of *ZMM6* and *ZMM8*/*ZMM14* expression may determine the number of spikelets at a certain position on the inflorescence axis, or the number of florets per spikelet, respec-

tively. These genes could thus have played a role in regulating spikelet or floret number in grasses during evolution [10, 12, 122]. It seems reasonable, therefore, that in the phylogenetic lineage which led to grasses, an ancestral member of the *AGL2* clade has been amplified and its descendants have been recruited for the establishment of novel positional information (concerning spikelet and floret number and arrangement). These positional values are characteristic of grass inflorescences and are not found within the simple inflorescences of some dicotyledonous plants such as *Arabidopsis*.

cDNAs of more than 30 different MADS-box genes, belonging to more than 10 different subfamilies, have been cloned from maize so far ([10, 31, 32, 72, 110, 126], and our unpublished results). Among them is an *AGL17*-like gene, but also several genes for which orthologues from nonmonocots have not been reported yet. In these more than 30 different genes an especially interesting class of sequences is not even included, the Transposed MADS-box elements of *Zea* No. 1 (*TMZ1* elements; also called *ZEM* genes; *ZAG4* is also a representative of this class of sequence elements) [31, 72, 78, 79]. The *TMZ1* elements have many features that are typical of transposons, such as varying copy numbers and genomic locations in different maize lines, 13 bp perfect terminal-inverted repeats (TIRs) at their flanks and 3 bp target sequence duplications. The last two features are both characteristic of the *En*/*Spm* transposon family [31, 79]. The *TMZ1* elements are the only plant sequences containing a MADS box published so far without a MIKC-type domain structure. Remarkably, they contain MADS boxes nearly identical to those of the members of the *AGAMOUS* clade, encoding MADS domains that fall well into the *AG* clade in phylogeny reconstructions based on MADS-domain sequences [123]. These MADS boxes are flanked, however, by sequences that are absolutely unrelated to sequences of *AG*-like genes. The most plausible scenario for the origin and evolution of the *TMZ1* elements thus seems to be that an *En*-like transposable element captured a MADS box of an *AG*-like gene somewhere in the lineage that led to maize, and was then distributed in the genomes of maize and its relatives [31, 79]. *TMZ1* elements have not been reported so far from outside the genus *Zea*.

The Liliaceae have flowers that superficially look very different from the flowers of grasses. For example, often they are very large and have a simple, yet showy perianth (perigon) composed of two whorls of organs, each containing three petaloid tepals. Flowers with a similar perianth structure are also known from some basal angiosperms (such as *Cabomba*, belonging to the Nymphaeales). Floral structures like these could be easily explained by a modified ABCD model in which the expression of the B function has expanded to whorl 1 (Figure 5) [129]. Tulip mutants are known that strongly support this hypothesis. A putative loss-of-B-function mutant is called 'Viridiflora'. It has flowers where the tepals in whorls 1 and 2 are homeotically transformed into sepaloid or leaf-like structures. The 6 stamens of the mutant tulip flower are transformed into carpel-like structures. A putative loss-of-C-function mutant is also known. It has tepal-like structures in whorls 1, 2 and 3, and from the center of the flower a new flower structure arises which again has tepal-like structures in all 3 outer whorls [129]. It seems not unlikely that an *AG*-like gene is affected in the loss-of-C-function mutant. Similarly, a *DEF*- or a *GLO*-like gene may be mutated in the loss-of-B-function mutant. However, one should take into consideration that the perianths of monocots and eudicots possibly evolved independently (see [60], and references therein). But even then a recruitment of *DEF*- and *GLO*-like genes for the specification of perianth organ identity – independently of a very similar event in the eudicots – would seem the most likely scenario to explain the petaloid character of perianth organs in Liliaceae. However, alternative scenarios cannot be excluded yet (see also the basal angiosperm section).

A number of cDNAs representing MADS-box genes from the lily species *Lilium regale* have been cloned recently (Figure 3), and cloning of the respective orthologues from tulip is well underway (our unpublished results). Therefore, the hypotheses outlined above can be rigorously tested soon. Expression of a *DEF*-like gene in both perianth whorls and in the stamen whorls of lily support the hypothesis that the petaloid character of all tepals is due to the expression of B-function genes in perianth whorls 1 and 2 (our unpublished results; see Figure 5).

MADS-box genes in eudicots

The flowers of eudicots are typically composed of a bipartite perianth with sepals in the first whorl and petals in the second whorl, followed by stamen and carpel whorls. The ABCD system of floral organ specification seems highly conserved within the higher

eudicots, except that an A function sensu stricto *may be provided by different genes in different species or may even not exist in some species. In lower eudicots, the temporal and spatial pattern of floral homeotic gene expression is more diverse. For* Arabidopsis *we now have definite proof, provided by mutant analysis, that MADS-box gene function is not restricted to flower development, but reaches from root to fruit development.*

Eudicots are defined by the production of triaperturate or triaperturate-derived pollen [21]. They can be further subdivided into the lower eudicots, comprising the Ranunculidae, basal Hamamelidae and basal Rosidae, and the higher eudicots, made up of the bulk (about 75%) of the angiosperm species, including the major genetic model species such as *Arabidopsis*, *Antirrhinum*, and *Petunia* ([60], and references therein).

The structure, function and phylogeny of eudicotyledonous MADS-box genes has already been extensively reviewed [74, 103, 123, 124]. Therefore, we rather want to concentrate on addressing some general evolutionary issues and on describing recent breakthroughs in understanding.

The gene phylogeny in Figure 3 reveals a complex but intriguing pattern of lineage-specific increases in the number of MADS-box genes during flowering-plant evolution. Most angiosperm subfamilies of MIKC-type genes characterized to date contain at least two members (putative recent paralogues) found in a single eudicotyledonous plant species, documenting continued diversification and fixation of MIKC-type genes during eudicot evolution [103, 123]. The evolution of MADS-box genes obviously did not come to a standstill, even after the establishment of the ABCD model within the eudicots. Since the flower groundplan is quite fixed in eudicots (e.g. most flowers have a bipartite perianth with sepals and petals followed by stamen and carpel whorls) the ABCD functions and the genes encoding them might also be conserved. A number of ongoing studies on MADS-box genes in diverse eudicot species indicates that this is indeed the case for the BCD functions and the respective genes of higher eudicots (e.g. [138], and many unpublished data). In contrast, the A function is less well defined and seems more flexible (e.g. [138], and many unpublished data; see also the basal angiosperm and monocot sections). This holds true even in comparison between *Arabidopsis* and the other model plants.

In *Antirrhinum*, for example, no loss-of-function phenotype has so far been identified which can separate the determination of the first floral whorl (assumed to depend on the A function) from the determination of the flower itself [80]. This suggests that in *Antirrhinum*, the A function is not simply a derivative of the function providing floral meristem identity, as suggested for *Arabidopsis* (see above). It rather seems that these two functions are still fully linked in *Antirrhinum* and thus cannot be separated (which may be the evolutionarily more ancient condition). As suggested by a precursor of the ABC model [111], the production of sepals in the first whorl may be inherent in the establishment of floral meristem identity. Accordingly, in the rare cases where flowers are formed in plants that have the *SQUA* gene mutated – *SQUA* is the *Antirrhinum* orthologue of the A function gene *AP1* from *Arabidopsis* – organ specification defects are not apparent.

The A function is also different in petunia from that in *Arabidopsis*. There is evidence that the petunia orthologue of the non-MADS gene *AP2* from *Arabidopsis* does not perform an A function, although analysis of the blind mutant indicates that such a function exists [67]. Expression of *AP2* and its homologues from petunia suggest an ancestral function in ovule or seed development [52, 67]. Such an origin of the A function is not contradictory to our suggestion that it might have been derived from a function in the specification of floral meristem identity. One possibility is that things could be different for *SQUA*- and *AP2*-like genes. However, *AP2* also works as a floral meristem identity gene, and the function in the specification of floral meristem identity could just represent an intermediate evolutionary step: since ovules are evolutionarily older than flowers, and these are probably older than the eudicot perianth, orthologues of *AP2* may have evolved step by step from genes involved in seed formation, via genes that are also involved in specifying floral meristem identity to genes that also play a role in specifying organ identity in the perianth.

Although the BCD part of the ABCD model can be considered to be relatively strongly conserved, some variations are known, especially with respect to the genes that provide these functions, as has been reviewed already [103, 123]. Recently it has been reported that *DEF* and *GLO* orthologues (i.e. putative B function genes) of some species of the lower eudicot subclass Ranunculidae show expression patterns that significantly deviate from the ones of higher eudicots [59]. While expression of these genes in stamens of the ranunculid species examined is as in higher eudicots, in some species (*Dicentra eximia* and *Papaver*

nudicaule) transcripts or proteins of *DEF*- and *GLO*-like genes accumulate in young petal primordia, but at later stages of develoment expression diminishes, becomes restricted to petal tips and margins, or even completely disappears [59]. It has been suggested that these data indicate that the *DEF*- and *GLO*-like genes of lower eudicots do not function in the same manner as their orthologues from higher eudicots to specify petal identity [59]. It has also been pointed out that these findings may reflect several independent evolutionary derivation events of petals from stamens, during the courses of which *DEF*- and *GLO*-like genes were recruited to new tissue- or cell-type-specific roles in the petals [59]. However, as outlined in the monocot section, *DEF*- or *GLO*-like genes may specify petal or lodicule identity in monocots, a clade which separated from the lineage that led to eudicots earlier than the speciation events that gave rise to lower and higher eudicots. Moreover, during very early developmental stages, the petal primordia of all ranunculid species examined show expression of *DEF*- and *GLO*-like genes [59]. It thus also could be that the developmental program that specifies the identity of petaloid organs is a synapomorphy of all extant monocots and eudicots (or even all extant angiosperms), and that the loss of requirement of *DEF*- and *GLO*-like genes for the maintenance of petal identity during late developmental stages reflects a secondary evolutionary event within some lineages that led to lower eudicots.

Despite the general conservation of the eudicotyledonous flower groundplan, some deviations are also well known here. For example, unisexual flowers evolved several times independently. Thus dioecious plants (where male and female flowers are borne on separate individuals) and monoecious plants (where male and female flowers are borne on the same individual) were established independently in many lineages. Unisexual flowers are usually produced during ontogeny from potentially hermaphroditic flowers by suppression of the development of either male or female organs in a particular whorl. However, the general impression now is that during flower development, organ abortion or suppression events and MADS-box gene expression (and thus floral organ identity) are under independent control mechanisms. For example, in the case of the dioecious plant *Silene latifolia* it has been found that the putative B- and C-type floral homeotic genes are expressed in male flowers (where the gynoecium does not differentiate) and in female flowers (where stamen primordia degenerate during development) in the same whorls with similar timing [44]. Independence of reproductive organ arrest and MADS-box gene expression was also reported for the monoecious plant cucumber (*Cucumis sativus*) [92], and holds also true for the putative C class gene *ZMM2* from the monoecious monocot maize. *ZMM2* is expressed in stamen and carpel primordia throughout their development, though stamen primordia abort in the female inflorescences and carpel primordia in the male ones [10, 12]. The situation is different, however, in the dioecious dicot sorrel (*Rumex acetosa*), because the expression of the putative C function gene becomes undetectable here as soon as the inappropriate set of organs (stamens in female flowers, carpels in male flowers) cease to develop [2]. However, absence of C function gene expression may well be a consequence of the arrest in organ development rather than its cause.

The perianth of *Rumex* is another interesting case: rather than having typical sepals and petals in whorls 1 and 2, respectively, first- and second-whorl organs are both sepaloid, and the second whorl does not express the putative *DEF* orthologues and functional equivalents *RAD1* and *RAD2*, so there is probably no homeotic B function in the second whorl [2]. It is conceivable, therefore, that elimination of B function expression in the second whorl caused the sepaloid phenotype of the petals of *Rumex acetosa* and perhaps also of other species, including many that are wind-pollinated. Note that the situation in *Rumex* thus is somehow opposite to that in the Liliaceae, where we have petaloid organs (tepals) in the first and second whorl, probably due to ectopic expression of B function genes in the first whorl of the flower (see the monocot section).

Moreover, there is evidence that an incomplete loss of C function can result in an increase in reproductive organs in an otherwise normal *Arabidopsis* flower [76]. Changes in the strength or spatiotemporal pattern of floral homeotic gene expression may thus have played an important role during the phylogenetic diversification of the eudicotyledonous flower.

To understand better the role of the ABCD genes during flower development, their upstream regulators and their target genes must be identified. Recently, there have been breakthroughs on both topics. One was cloning of the *CURLY LEAF* (*CLF*) gene from *Arabidopsis* [38]. Functional defects in *CLF* lead to leaf curling, which is also caused by ectopic expression of the C function gene *AGAMOUS*. In *clf* mutants, *AG* is indeed ectopically expressed in leaves (and some other parts of the plant), and the available evidence

indicates that it may be the normal function of *CLF* to repress *AG* in inappropriate parts of the plant during relatively late stages of development. However, *CLF* is obviously not involved in initially establishing the spatial pattern of *AG* transcription. It surely came as a surprise that the CLF protein has extensive sequence similarity to the protein product of the *Drosophila* Polycomb-group gene *Enhancer of zeste* (*E*(*z*)) [38]. In *Drosophila*, Polycomb-group genes are part of the 'memory system' that maintains the initial spatial patterns of inactive or active homeotic selector genes through many rounds of cell division. The similarity between CLF and E(Z) suggests that both proteins share a common ancestor and have been conserved since the animal and plant lineages split, and that plants possess Polycomb-group gene functions. This is very remarkable, because the homeotic target genes in plants and animals encode different kinds of transcription factors, i.e. either MADS-domain or homeodomain proteins [124]. It seems likely, therefore, that the use of Polycomb-group proteins as repressors of homeotic genes has evolved independently in plants and animals.

Although candidate genes had been obtained previously, the first gene that was experimentally shown to be a direct target of a floral homeotic gene was identified just recently [106]. A steroid-regulated version of the class B protein AP3 had been expressed in an *ag ap3* mutant plant. The differential display technique was then used to identify an mRNA that was up-regulated by steroid treatment (i.e. after providing the B function). To detect only direct target genes, indirect effects were blocked by a protein synthesis inhibitor. The identified mRNA corresponds to a gene that contains a NAC domain, named after the founding members of this gene family (*NO APICAL MERISTEM* (*NAM*) from petunia, *ATAF1-2*, *CUP-SHAPED COTYLEDONS2* (*CUC2*) from *Arabidopsis*), hence it was called *NAP* (*NAC-LIKE*, activated by *AP3*/*PI*) [106]. The expression pattern of *NAP* and the phenotypes caused by its misexpression indicate that it plays a role in the transition between growth by cell division and cell expansion in stamens and petals [106].

The first MADS-box genes in plants were detected by scientists who were mainly interested in flower development [115, 137]. It is only natural, therefore, that during the early days of plant MADS science, investigations were focused on this topic. However, several studies that demonstrated the transcription of a number of MADS-box genes outside floral organs suggested relatively early that members of this gene family play regulatory roles beyond flower development, i.e. during root, leaf, fruit, seed and embryo development (e.g. [49, 66, 104, 124]). The existence of MADS-box genes in gymnosperms and ferns further demonstrated that the role of these genes in plants is not restricted to flower development. Mutant analysis has provided proof that some *Arabidopsis* MADS-box genes have nonfloral functions. For example, *AGL1* and *AGL5* are a pair of recently duplicated paralogues of *AG*-like genes. These two genes encode functionally redundant proteins that are required for the proper development of the fruit dehiscence zone, because in *agl1 agl5* double mutants, the mature siliques fail to dehisce [63]. Since indehiscent fruits evolved several times independently even within the Brassicaceae (Klaus Mummenhoff, pers. comm.), it will be interesting to find out whether mutations in orthologues of *AGL1*/*AGL5* are involved in some of these evolutionary changes. *AGL8* – now called *FRUITFULL* (*FUL*) – is another *Arabidopsis* gene that is involved in fruit development. The gene is required for the normal pattern of cell division, expansion and differentiation during morphogenesis of the silique [43]. The major part of the silique is provided by the carpel valves, and it could well be that *FUL* is a valve identity gene [63].

The genes discussed above are interesting examples of gene duplications and functional diversifications within gene subfamilies. The class C gene *AG*, the putative class D gene *AGL11*, and the *AGL1*/*5* pair of recent paralogues involved in fruit development are all members of the clade of *AG*-like genes, which was established probably 300–400 MYA (see above). This implies that the gene duplications that led to these genes occurred within the past 400 million years. The place of action (fruits) of some of the descendants of the gene duplication events (*AGL1*/*5*) were not yet established when the first *AG*-like gene appeared, thus documenting involvement of new genes in the appearance of new structures and functions.

An even more striking example for functional change is provided by the *FUL* gene. Together with *AP1* and *CAL*, *FUL* belongs to the clade of *SQUA*-like genes. *SQUA*-like genes are typically expressed in inflorescence or floral meristems, and, accordingly, work as meristem identity genes [123]. Although *ful* single mutants do not have an inflorescence-specific phenotype, it is also one of the first genes expressed after the transition to flowering at the inflorescence apex [63]. And indeed, *ap1 cal ful* triple mutants show an extreme enhancement of the *ap1 cal* phenotype, indicating partial functional redundancy between

AP1 CAL and *FUL* during flower initiation. It seems likely, therefore, that the role in meristems reflects the ancestral function of *SQUA*-like genes, and that *FUL* was recruited later for an additional function during the course of fruit evolution.

Several *Arabidopsis* genes are expressed preferentially or exclusively in roots [104, 141]. One of them, the *AGL17*-like gene *ANR1*, could be identified as a key determinant of developmental plasticity in roots [141]. Transgenic plants in which *ANR1* was repressed no longer responded to NO_3^--rich zones in the soil by lateral root proliferation, which is in contrast to the behavior of wild-type roots. Could there be a better example than that to demonstrate that MADS-box gene function, even within eudicots, reaches far beyond flower development?

Another exciting case is the expression of a *DEF*-like (*NMH7*) and an *AGL17*-like gene (*NMHC5*) during nodule development of alfalfa (*Medicago sativa*; reviewed in [27]). The definition of the exact role of these genes in nodule development, however, awaits the isolation of respective loss-of-function mutants.

Outlook for future studies

We briefly mention major gaps in our understanding of plant MADS-box gene evolution, and discuss some innovative directions of future research. We suggest how cooperation between researchers interested in MADS evodevotics might be facilitated by a MADS homepage provided on the worldwide web.

There is still a long way to go until we will understand the role of MADS-box genes in plant evolution in satisfactory detail. We hope that the previous sections, however, have indicated that we are moving in the right direction. Still there are severe gaps in our knowledge. Concerning the major steps of land plant evolution, we know very little about MADS-box genes in mosses, and nothing about these genes in liverworts and algae, for example. Since liverworts have recently been identified as the earliest land plants [100], and the green charophycean algae are the sister group of all land plants [57], these taxa should be studied with highest priority. We also know nothing about the function of MADS-box genes in any nonseed plant, but generating gene knockouts by homologous recombination in the moss *Physcomitrella patens* might be one way to change that soon.

Gene cloning and sequencing, phylogeny reconstructions, expression studies and mutant analysis have made it possible to correlate MADS-box gene phylogeny with the evolution of plant morphology. For example, the contribution of gene duplications and gene recruitments to the evolution of the gene networks controlling plant morphology thus became obvious. Comparison of mutant phenotypes between different taxa (e.g. floral homeotic mutants of *Arabidopsis* and *Antirrhinum*) which are caused by orthologous genes indicated the equivalence of gene functions in often distantly related species. The expression of orthologous genes in heterologous backgrounds (e.g. conifer *AG*-like genes in *Arabidopsis*) indicated to what extent genes from other taxa can substitute for the homologous function within the host plant. However, the conditions of the experimental design have to be carefully taken into consideration. For example, functional substitution in mis- or overexpression studies might be less demanding than the complementation of loss-of-function mutants (see the gymnosperm section). However, experiments using the *DEF* gene to complement *AP3* mutants have already been succesful, although the donor and acceptor species (*Antirrhinum* and *Arabidopsis*, respectively) are relatively distantly related eudicots [51, 107].

The replaceability of gene functions by heterologous transgenes is a stringent test for the conservation of gene function. Therefore, more and more of these kinds of experiments will probably be carried out, involving more and more species. For an optimal use of such experiments it is essential that the phylogenetic relationships between the genes of interest are clarified. For example, interspecific comparisons between orthologous genes are generally more useful than comparisons between non-orthologous genes, and the complementation of mutants should also be tried preferably with orthologous genes. To facilitate the determination of the phylogenetic relationship between any MADS-box gene of interest and the published set of genes, a 'MADS homepage' has been established and made accessible via the worldwide web (URL: http://www.mpiz-koeln.mpg.de/mads/). Among other things, it contains a list of published MADS-box genes from which information about these genes can be accessed. Phylogenetic trees which clarify the evolutionary relationship between a new gene of interest and all published genes will be made on demand. The only condition for having such an analysis carried out is that a permission is given to present the phylogenetic tree including the position of the new gene (and indicating the species it was isolated from) on the MADS homepage. The gene sequence itself will be kept confi-

dential if this is requested. (For details, see the MADS homepage.) The MADS homepage may thus become an information node at which scientists who have isolated a new MADS-box gene might be able to identify colleagues who have cloned orthologues from other species.

Another way to minimize the problem of identifying orthologous genes is to study monophyletic groups of closely related species. Especially the phylogenetic surroundings of *Arabidopsis* and *Antirrhinum* seem interesting in that respect, not only because so much is known about the model plants that can be compared to other species, but also because substantial morphological variation exists among the respective taxa [28]. Studies in which the role of MADS-box genes in the variability of floral morphology is investigated have been initiated for relatives of both *Arabidopsis* and *Antirrhinum* [8, 102]. These studies address an important issue: that the function of flowering-plant MADS-box genes is largely known from studying mutants. It is not yet known whether variability at the same loci that, upon mutation, change the flower structure or traits such as time to flowering, also underlies natural variability in these traits. It is essential, therefore, to determine the amount of natural variability at MADS-box gene loci at the molecular level in natural populations of plants. Such studies have just begun, using the *Arabidopsis* genes *CAL*, *AP3* and *PI* as model systems [98, 99].

During the course of this review, changes in MADS-box gene function during evolution have become obvious. Changes in gene expression pattern, often caused by mutations in *cis*-regulatory elements within promoter regions [23], may be a major reason for many of these changes. Therefore, comparative studies of promoters of orthologous genes with different expression patterns would be very interesting. Unfortunately, we know only very little about promoter functions of MADS-box genes. However, detailed studies on the promoters of the *AG* and *AP3* genes from *Arabidopsis* published recently [48, 113, 127] indicate that the situation is significantly improving.

Other changes in the connections of the MADS-box genes to other genes within the gene networks have probably also taken place during evolution. For example, target genes may have changed (e.g. when the B function genes started to specify not only reproductive organs, but also petals), and also protein-protein interactions may be phylogenetically dynamic (e.g. while the DEF- and GLO-like proteins of eudicots only form heterodimers, the single ancestor of both proteins may have formed homodimers, or heterodimers with other proteins). Thus the biochemical properties of the proteins encoded by MADS-box genes have probably changed during evolution. In order to work properly, a MADS-domain protein, like most other transcription factors, generally has to fulfil several subfunctions: it must form homo- or heterodimers, or even higher-order protein complexes; it must bind to DNA; and it must activate or repress the basal transcriptional machinery. To study protein-protein and protein-DNA interactions, a number of *in vitro* and *in vivo* assays are available and have already been applied to MADS-domain proteins from the genetic model plants. These techniques include electrophoretic mobility shift assays (EMSAs), the yeast two-hybrid system, DNA footprinting assays and random DNA binding site selection experiments (e.g. [30, 49]). In order to find out how the interaction with DNA and proteins changed during MADS-domain protein evolution, all these techniques should be applied to series of orthologous proteins from distantly related species.

There is obviously still a long way to go until we understand the role of MADS-box genes in plant evolution in satisfactory detail. But every step on the way will be a great intellectual pleasure.

Acknowledgements

We thank Peter Engström, Hilde Fischer, Dirk Gassen, Mitsuyasu Hasebe, Katrin Henschel, Vivian Irish, Elena Kramer, Bill Martin, Aidyn Mouradov, Jens Pahnke and Marty Yanofsky for communicating unpublished data or manuscripts in press. We are indebted to Aidyn Mouradov, Peter Engström and Mitsuyasu Hasebe, who provided us with constructive criticism on preliminary versions of the gymnosperm or the fern section, respectively. Many thanks also to Kirsten Bomblies and two anonymous reviewers for their very helpful comments on the manuscript. We also would like to thank Wolfram Faigl and Susanne Werth for skillful technical assistance. Part of this work was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) to G.T. (grant Th 417/3-1). One of us (A.B.) was supported by a fellowship from the DFG (Graduiertenkolleg ‘Molekulare Analyse von Entwicklungsprozessen bei Pflanzen’). A.K. was supported by a Japanese Society for the Promotion of Science (JSPS) Postdoctoral Fellowship for Research Abroad.

Note added in proof

According to Dr Mushigian, one of the authors of Ref. 87, UspA now appears to be an ATP-binding, not a DNA-binding protein, so that the homology between a part of the MADS-domain and a stretch of the UspA protein discussed above and elsewhere [87] is doubtful.

We apologize for not citing all of the relevant papers of our colleagues because of space constraints. Please note, however, that a more comprehensive list of publications about MADS-box genes is provided at the 'MADS homepage' (URL: http://www.mpiz-koeln.mpg.de/mads/).

References

1. Abouheif E, Akam M, Dickinson WJ, Holland PWH, Meyer A, Patel NH, Raff RA, Roth VL, Wray GA: Homology and developmental genes. Trends Genet 13: 432–433 (1997).
2. Ainsworth C, Thangavelu M, Crossley S, Buchanan-Wollaston V, Parker J: Male and female flowers from the dioecious plant *Rumex acetosa* show different patterns of MADS-box gene expression. Plant Cell 7: 1583–1598 (1995).
3. Albert VA, Gustafsson MHG, Di Laurenzio L: Ontogenetic systematics, molecular developmental genetics, and the angiosperm petal. In: Soltis DE, Soltis PS, Doyle JJ (eds), Molecular Systematics of Plants II, pp. 349–374. Kluwer Academic Publishers, Boston, MA (1998).
4. Angenent GC, Colombo L: Molecular control of ovule development. Trends Plant Sci 1: 228–232 (1996).
5. Arber EAN, Parkin J: Studies on the evolution of the angiosperms: the relationship of the angiosperms to the Gnetales. Ann Bot 22: 489–515 (1908).
6. Baum DA: The evolution of plant development. Curr Opin Plant Biol 1: 79–86 (1998).
7. Beck CB: Origin and Evolution of Gymnosperms. Columbia University Press, New York (1988).
8. Bowman JL, Smyth DR: Patterns of petal and stamen reduction in Australian species of *Lepidium* L. (Brassicaceae). Int J Plant Sci 159: 65–74 (1998).
9. Bradley D, Carpenter R, Sommer H, Hartley N, Coen E: Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena* locus of *Antirrhinum*. Cell 72: 85–95 (1993).
10. Cacharrón J: MADS-Box-Gene in *Zea mays*: Vergleichende Expressionsuntersuchungen an Modellen paraloger und orthologer Genpaare. Ph.D. thesis, Mathematisch-Naturwissenschaftliche Fakultät der Universität zu Köln, Germany (1998).
11. Cacharrón J, Fischer A, Saedler H, Theissen G: Expression patterns of MADS-box genes in maize as studied by *in situ* hybridization. Maize Genet Coop Newsl 69: 37–38 (1995).
12. Cacharrón J, Saedler H, Theissen G: Expression of the 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 (1999).
13. Carmona MJ, Ortega N, Garcia-Maroto F: Isolation and molecular characterization of a new vegetative MADS-box gene from *Solanum tuberosum* L. Planta 207: 181–188 (1998).
14. Chasan R: Ceratopteris: a model plant for the 90s. Plant Cell 4: 113–115 (1992).
15. Chaw S-M, Zharkikh A, Sung H-M, Lau T-C, Li W-H: Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol Biol Evol 14: 56–68 (1997).
16. Cheng PC, Greyson RI, Walden DB: Organ initiation and the development of unisexual flowers in the tassel and ear of *Zea mays*. Am J Bot 70: 450–462 (1983).
17. Chung Y-Y, Kim S-R, Finkel D, Yanofsky MF, An G: Early flowering and reduced apical dominance result from ectopic expression of a rice MADS box gene. Plant Mol Biol 26: 657–665 (1994).
18. Chung Y-Y, Kim S-R, Kang H-G, Noh Y-S, Park MC, Finkel D, An G: Characterization of two rice MADS box genes homologous to *GLOBOSA*. Plant Sci 109: 45–56 (1995).
19. Clifford HT: Spikelet and floral morphology. In: Soderstrom TR, Hilu KW, Campbell CS, Barkworth ME (eds), Grass Systematics and Evolution, pp. 21–30. Smithsonian Institution Press, Smithsonian Institution, Washington, DC (1987).
20. Coen ES, Meyerowitz EM: The war of the whorls: genetic interactions controlling flower development. Nature 353: 31–37 (1991).
21. Crane PR, Friis EM, Pedersen KR: The origin and early diversification of angiosperms. Nature 374: 27–33 (1995).
22. Di Rosa A: Molekularbiologische Untersuchungen zum Ursprung homöotischer Gene in Pflanzen am Beispiel der MADS-Box-Genfamilie aus dem Farn *Ceratopteris richardii*. PhD thesis, Mathematisch-Naturwissenschaftliche Fakultät der Universität zu Köln, Germany (1998).
23. Doebley J, Lukens L: Transcriptional regulators and the evolution of plant form. Plant Cell 10: 1075–1082 (1998).
24. Doyle JA: Origin of the angiosperm flower: a phylogenetic perspective. Plant Syst Evol (Suppl) 8: 7–29 (1994).
25. Doyle JA: Seed plant phylogeny and the relationships of Gnetales. Int J Plant Sci 157 (Suppl): S3–S39 (1996).
26. Doyle JJ: Evolution of a plant homeotic multigene family: towards connecting molecular systematics and molecular developmental genetics. Syst Biol 43: 307–328 (1994).
27. Doyle JJ: Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. Trends Plant Sci 3: 473–478 (1998).
28. Endress PK: Evolution and floral diversity: the phylogenetic surroundings of *Arabidopsis* and *Antirrhinum*. Int J Plant Sci 153: S106–S122 (1992).
29. Endress PK: Floral structure and evolution of primitive angiosperms: recent advances. Plant Syst Evol 192: 79–97 (1994).
30. Fan H-Y, Hu Y, Tudor M, Ma H: Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. Plant J 12: 999–1010 (1997).
31. Fischer A, Baum N, Saedler H, Theissen G: Chromosomal mapping of the MADS-box multigene family in *Zea mays* reveals dispersed distribution of allelic genes as well as transposed copies. Nucl Acids Res 23: 1901–1911 (1995).
32. Fischer A, Saedler H, Theissen G: Restriction fragment length polymorphism-coupled domain-directed differential display: a highly efficient technique for expression analy-

sis of multigene families. Proc Natl Acad Sci USA 92: 5331–5335 (1995).
33. Gandolfo MA, Nixon KC, Crepet WL, Stevenson DW, Friis EM: Oldest known fossils of monocotyledons. Nature 394: 532–533 (1998).
34. Gaut BS, Doebley JF: DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA 94: 6809–6814 (1997).
35. Gehring WJ: The homeobox in perspective. Trends Biochem Sci 17: 277–280 (1992).
36. Gifford EM, Foster AS: Morphology and Evolution of Vascular Plants, 3rd ed. Freeman, New York (1988).
37. Gilbert SF, Opitz JM, Raff RA: Resynthesizing evolutionary and developmental biology. Dev Biol 173: 357–372 (1996).
38. Goodrich J, Puangsomlee P, Martin M, Long D, Meyerowitz EM, Coupland G: A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. Nature 386: 44–51 (1997).
39. Goremykin V, Bobrova V, Pahnke J, Troitsky A, Antonov A, Martin W: Noncoding sequences from the slowly evolving chloroplast inverted repeat in addition to *rbcL* data do not support Gnetalean affinities of angiosperms. Mol Biol Evol 13: 383–396 (1996).
40. Goremykin VV, Hansmann S, Martin WF: Evolutionary analysis of 58 proteins encoded in six completely sequenced chloroplast genomes: revised molecular estimates of two seed plant divergence times. Plant Syst Evol 206: 337–351 (1997).
41. Gould SJ: Ontogeny and phylogeny – revisited and reunited. BioEssays 14: 275–279 (1992).
42. Greco R, Stagi L, Colombo L, Angenent GC, Sari-Gorla M, Pè ME: MADS box genes expressed in developing inflorescences of rice and sorghum. Mol Gen Genet 253: 615–623 (1997).
43. Gu Q, Ferrándiz C, Yanofsky MF, Martienssen R: The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. Development 125: 1509–1517 (1998).
44. Hardenack S, Ye D, Saedler H, Grant S: Comparison of MADS box gene expression in developing male and female flowers of the dioecious plant white campion. Plant Cell 6:1775–1787 (1994).
45. Hasebe M, Banks JA: Evolution of MADS gene family in plants. In: Iwatsuki K, Raven PH (eds), Evolution and Diversification of Land Plants, pp. 179–197. Springer-Verlag, Tokyo (1997).
46. Hasebe M, Wen C-K, Kato M, Banks JA: Characterization of MADS homeotic genes in the fern *Ceratopteris richardii*. Proc Natl Acad Sci USA 95: 6222–6227 (1998).
47. Heck GR, Perry, SE, Nichols, KW, Fernandez DE: AGL15, a MADS domain protein expressed in developing embryos. Plant Cell 7: 1271–1282 (1995).
48. Hill TA, Day CD, Zondlo SC, Thackeray AG, Irish VF: Discrete spatial and temporal *cis*-acting elements regulate transcription of the *Arabidopsis* floral homeotic gene *APETALA3*. Development 125: 1711–1721 (1998).
49. Huang H, Tudor M, Weiss CA, Hu Y, Ma H: The *Arabidopsis* MADS-box gene *AGL3* is widely expressed and encodes a sequence-specific DNA-binding protein. Plant Mol Biol 28: 549–567 (1995).
50. Huijser P, Klein J, Lönnig W-E, Meijer H, Saedler H, Sommer H: Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus*. EMBO J 11: 1239–1249 (1992).
51. Irish VF, Yamamoto YT: Conservation of floral homeotic gene function between *Arabidopsis* and *Antirrhinum*. Plant Cell 7: 1635–1644 (1995).
52. Jofuku KD, den Boer BGW, Van Montagu M, Okamuro JK: Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. Plant Cell 6: 1211–1225 (1994).
53. Kang H-G, An G: Isolation and characterization of a rice MADS box gene belonging to the *AGL2* gene family. Mol Cell 7: 45–51 (1997).
54. Kang H-G, Jeon J-S, Lee S, An G: Identification of class B and class C floral organ identity genes from rice. Plant Mol Biol 38: 1021–1029 (1998).
55. Kang H-G, Noh Y-S, Chung Y-Y, Costa MA, An K, An G: Phenotypic alterations of petal and sepal by ectopic expression of a rice MADS box gene in tobacco. Plant Mol Biol 29: 1–10 (1995).
56. Kappen C, Ruddle FH: Evolution of a regulatory gene family: *HOM*/*HOX* genes. Curr Opin Gen Dev 3: 931–938 (1993).
57. Kenrick P, Crane PR: The origin and early evolution of plants on land. Nature 389: 33–39 (1997).
58. Kofuji R, Yamaguchi K: Isolation and phylogenetic analysis of MADS genes from the fern *Ceratopteris richardii*. J Phytogeogr Taxon 45: 83–91 (1997).
59. Kramer EM, Irish VF: Evolution of genetic mechanisms controlling petal development. Nature 399: 144–148 (1999).
60. Kramer EM, Dorit RL, Irish VF: Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. Genetics 149: 765–783 (1998).
61. Krüger J, Aichinger C, Kahmann R, Bölker M: A MADS-box homologue in *Ustilago maydis* regulates the expression of pheromone-inducible genes but is nonessential. Genetics 147: 1643–1652 (1997).
62. Kyozuka J, Harcourt R, Peacock WJ, Dennis ES: Eucalyptus has functional equivalents of the *Arabidopsis AP1* gene. Plant Mol Biol 35: 573–584 (1997).
63. Liljegren SJ, Ferrándiz C, Alvarez-Buylla ER, Pelaz S, Yanofsky MF: *Arabidopsis* MADS-box genes involved in fruit dehiscence. Flowering Newsl 25: 9–19 (1998).
64. Liu J-J, Podila GK: Characterization of a MADS box gene (Accession No. Y09611) from immature female cone of red pine (PGR 97-032). Plant Physiol 113: 665 (1997).
65. Lu Z-X, Wu M, Loh C-S, Yeong C-Y, Goh C-J: Nucleotide sequence of a flower-specific MADS box cDNA clone from orchid. Plant Mol Biol 23: 901–904 (1993).
66. Ma H, Yanofsky MF, Meyerowitz EM: *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. Genes Dev 5: 484–495 (1991).
67. Maes T, Van de Steene N, Van Montagu M, Gerats T: The *AP2*-like genes of *Petunia hybrida*. Flowering Newsl 25: 35–40 (1998).
68. Mandel MA, Bowman JL, Kempin SA, Ma H, Meyerowitz EM, Yanofsky MF: Manipulation of flower structure in transgenic tobacco. Cell 71: 133–143 (1992).
69. Martin WF: Is something wrong with the tree of life? BioEssays 18: 523–527 (1996).
70. McGinnis W, Kuziora M: The molecular architects of body design. Scient Am 270 (2): 36–42 (1994).
71. Mena M, Ambrose BA, Meeley RB, Briggs SP, Yanofsky MF, Schmidt RJ: Diversification of C-function activity in maize flower development. Science 274: 1537–1540 (1996).
72. Mena M, Mandel MA, Lerner DR, Yanofsky MF, Schmidt RJ: A characterization of the MADS-box gene family in maize. Plant J 8: 845–854 (1995).

73. Meyerowitz EM: Plants and the logic of development. Genetics 145: 5–9 (1997).
74. Meyerowitz EM: Genetic and molecular mechanisms of pattern formation in *Arabidopsis* flower development. J Plant Res 111: 233–242 (1998).
75. Mizukami Y, Ma H: Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic *Arabidopsis* plants alters floral organ identity. Cell 71: 119–131 (1992).
76. Mizukami Y, Ma H: 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 (1995).
77. Monod J: Le hasard et la nécessité. Éditions du Seuil, Paris (1970).
78. Montag K, Salamini F, Thompson RD: *ZEMa*, a member of a novel group of MADS box genes, is alternatively spliced in maize endosperm. Nucl Acids Res 23: 2168–2177 (1995).
79. Montag K, Salamini F, Thompson RD: The ZEM2 family of maize MADS box genes possess features of transposable elements. Maydica 41: 241–254 (1996).
80. Motte P, Wilkinson M, Schwarz-Sommer Z: Floral meristem identity and the A function in *Antirrhinum*. Flowering Newsl 25: 41–43 (1998).
81. Mouradov A, Glassick TV, Hamdorf BA, Murphy LC, Marla SS, Yang Y, Teasdale R: Family of MADS-box genes expressed early in male and female reproductive structures of Monterey pine. Plant Physiol 117: 55–61 (1998).
82. Mouradov A, Glassick T, Teasdale R: Isolation and characterization of a new MADS-box cDNA from *Pinus radiata* (Accession No. U76726) (PGR 97-027). Plant Physiol 113: 664 (1997).
83. Mouradov A, Glassick T, Vivian-Smith A, Teasdale R: Isolation of a MADS box gene family from *Pinus radiata* (accession No. U42399 and U42400) (PGR 96-002). Plant Physiol 110: 1047 (1996).
84. Mouradov A, Hamdorf B, Teasdale RD, Kim J, Winter K-U, Theissen G: A *DEF*/*GLO*-like MADS-box gene from a gymnosperm: *Pinus radiata* contains an orthologue of angiosperm B class floral homeotic genes. Dev Genet (in press) (1999).
85. Münster T, Pahnke J, Di Rosa A, Kim JT, Martin W, Saedler H, Theissen G: Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. Proc Natl Acad Sci USA 94: 2415–2420 (1997).
86. Murai K, Murai R, Ogihara Y: Wheat MADS box genes, a multigene family dispersed throughout the genome. Genes Genet Syst 72: 317–321 (1997).
87. Mushegian AR, Koonin EV: Sequence analysis of eukaryotic developmental proteins: ancient and novel domains. Genetics 144: 817–828 (1996).
88. Okada K, Shimura Y: Genetic analyses of signalling in flower development using *Arabidopsis*. Plant Mol Biol 26: 1357–1377 (1994).
89. Okamoto H, Silverthorne J, Wada M: Spatial patterns of phytochrome expression in young leaves of the fern *Adiantum capillus-veneris*. Plant Cell Physiol 38: 1397–1402 (1997).
90. Olson EN, Perry M, Schulz RA: Regulation of muscle differentiation by the MEF2 family of MADS box transcription factors. Dev Biol 172: 2–14 (1995).
91. Pellegrini L, Tan S, Richmond TJ: Structure of serum response factor core bound to DNA. Nature 376: 490–498 (1995).
92. Perl-Treves R, Kahana A, Rosenman N, Xiang Y, Silberstein L: Expression of multiple *AGAMOUS*-like genes in male and female flowers of cucumber (*Cucumis sativus* L.). Plant Cell Physiol 39: 701–710 (1998).
93. Philippe H, Chenuil A, Adoutte A: Can the Cambrian explosion be inferred through molecular phylogeny? Development (Suppl): 15–25 (1994).
94. Pichersky E, Soltis D, Soltis P: Defective chlorophyll *a/b*-binding protein genes in the genome of a homosporous fern. Proc Natl Acad Sci USA 87: 195–199 (1990).
95. Pryer KM, Smith AR, Skog JE: Phylogenetic relationships of extant ferns based on evidence from morphology and *rbcL* sequences. Am Fern J 85: 205–282 (1995).
96. Purugganan MD: The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. J Mol Evol 45: 392–396 (1997).
97. Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky M: Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. Genetics 140: 345–356 (1995).
98. Purugganan MD, Suddith JI: Molecular population-genetics of the *Arabidopsis* cauliflower regulatory gene: nonneutral evolution and naturally occuring variation in floral homeotic function. Proc Natl Acad Sci USA 95: 8130–8134 (1998).
99. Purugganan MD, Suddith JI: Molecular population genetics of floral homeotic loci: departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana*. Genetics 151: 839–848 (1999).
100. Qiu Y-L, Cho Y, Cox JC, Palmer JD: The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394: 671–674 (1998).
101. Raff RA: The Shape of Life. Genes, Development, and the Evolution of Animal Form. University of Chicago Press, Chicago, IL (1996).
102. Reeves PA, Olmstead RG: Evolution of novel morphological and reproductive traits in a clade containing *Antirrhinum majus* (Scrophulariaceae). Am J Bot 85: 1047–1056 (1998).
103. Riechmann JL, Meyerowitz EM: MADS domain proteins in plant development. Biol Chem 378: 1079–1101 (1997).
104. Rounsley SD, Ditta GS, Yanofsky MF: Diverse roles for MADS box genes in *Arabidopsis* development. Plant Cell 7: 1259–1269 (1995).
105. Rutledge R, Regan S, Nicolas O, Fobert P, Coté C, Bosnich W, Kauffeldt C, Sunohara G, Séguin A, Stewart D: 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 (1998).
106. Sablowski RWM, Meyerowitz EM: A homolog of *NO APICAL MERISTEM* is an immediate target of the floral homeotic genes *APETALA3*/*PISTILLATA*. Cell 92: 93–103 (1998).
107. Samach A, Kohalmi SE, Motte P, Datla R, Haughn GW: Divergence of function and regulation of class B floral organ identity genes. Plant Cell 9: 559–570 (1997).
108. Savard L, Li P, Strauss SH, Chase MW, Michaud M, Bousquet J: Chloroplast and nuclear gene sequences indicate Late Pennsylvanian time for the last common ancestor of extant seed plants. Proc Natl Acad Sci USA 91: 5163–5167 (1994).
109. Schmidt RJ, Ambrose BA: The blooming of grass flower development. Curr Opin Plant Biol 1: 60–67 (1998).
110. Schmidt RJ, Veit B, Mandel MA, Mena M, Hake S, Yanofsky MF: Identification and molecular characterization of *ZAG1*,

the maize homolog of the *Arabidopsis* floral homeotic gene *AGAMOUS*. Plant Cell 5: 729–737 (1993).
111. Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H: Genetic control of flower development by homeotic genes in *Antirrhinum majus*. Science 250: 931–936 (1990).
112. Shore P, Sharrocks AD: The MADS-box family of transcription factors. Eur J Biochem 229: 1–13 (1995).
113. Sieburth LE, Meyerowitz EM: Molecular dissection of the *AGAMOUS* control region shows that *cis* elements for spatial regulation are located intragenically. Plant Cell 9: 355–365 (1997).
114. Slack JMW, Holland PWH, Graham CF: The zootype and the phylotypic stage. Nature 361: 490–492 (1993).
115. Sommer H, Beltrán J-P, Huijser P, Pape H, Lönnig W-E, Saedler H, Schwarz-Sommer Z: *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 (1990).
116. Sundström J, Carlsbecker A, Svensson M, Svenson M, Urban J, Theissen G, Engström P: MADS-box genes active in developing pollen cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. Dev Genet (in press).
117. Stewart WN, Rothwell GW: Paleobotany and the Evolution of Plants, 2nd ed. Cambridge University Press, Cambridge, UK (1993).
118. Tandre K: Molecular approaches to the developmental biology of Norway spruce, *Picea abies*. Acta Universitatis Upsaliensis, Uppsala, Sweden (1997).
119. Tandre K, Albert VA, Sundas A, Engström P: Conifer homologues to genes that control floral development in angiosperms. Plant Mol Biol 27: 69–78 (1995).
120. Tandre K, Svenson M, Svensson ME, Engström P: Conservation of gene structure and activity in the regulation of reproductive organ development of conifers and angiosperms. Plant J 15: 615–623 (1998).
121. Taylor TN, Taylor EL: The Biology and Evolution of Fossil Plants. Prentice Hall, Englewood Cliffs, NJ (1993).
122. Theissen G, Cacharrón J, Fischer A, Saedler H: MADS-box genes in the evolution of maize (*Zea mays* ssp. *mays*). In: Hong JC, Kim MJ (eds), Proceedings of the 2nd Korea-Germany Joint Symposium in Plant Biotechnology, pp. 1–12. Gyeongsang National University of Korea, Chinju, Korea (1994).
123. Theissen G, Kim J, Saedler H: 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 (1996).
124. Theissen G, Saedler H: MADS-box genes in plant ontogeny and phylogeny: Haeckel's 'biogenetic law' revisited. Curr Opin Genet Dev 5: 628–639 (1995).
125. Theissen G, Saedler H: Molecular architects of plant body plans. Prog Bot 59: 227–256 (1998).
126. Theissen G, Strater T, Fischer A, Saedler H: Structural characterization, chromosomal localization and phylogenetic evaluation of two pairs of *AGAMOUS*-like MADS-box genes from maize. Gene 156: 155–166 (1995).
127. Tilly JJ, Allen DW, Jack T: The CArG boxes in the promoter of the *Arabidopsis* floral organ identity gene *APETALA3* mediate diverse regulatory effects. Development 125: 1647–1657 (1998).
128. Tröbner W, Ramirez L, Motte P, Hue I, Huijser P, Lönnig WE, Saedler H, Sommer H, Schwarz-Sommer Z: *GLOBOSA*: a homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. EMBO J 11: 4693–4704 (1992).
129. van Tunen AJ, Eikelboom W, Angenent GC: Floral organogenesis in *Tulipa*. Flowering Newsl 16: 33–37 (1993).
130. Veuthey A-L, Bittar G: Phylogenetic relationships of fungi, plantae, and animalia inferred from homologous comparison of ribosomal proteins. J Mol Evol 47: 81–92 (1998).
131. Weigel D, Meyerowitz EM: Genetic hierarchy controlling flower development. In: Bernfeld M (ed), Molecular Basis of Morphogenesis, pp. 91–105. John Wiley, New York (1993).
132. Weigel D, Meyerowitz EM: The ABCs of floral homeotic genes. Cell 78: 203–209 (1994).
133. Winter K-U: Charakterisierung von MADS-Box-Genen der Gymnosperme *Gnetum gnemon* L. Diploma thesis, Universität Bonn, Germany (1997).
134. Winter K-U, Becker A, Münster T, Kim JT, Saedler H, Theissen G: MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. Proc Natl Acad Sci USA 96: 7342–7347 (1999).
135. Wolfe KH, Gouy M, Yang Y-W, Sharp PM, Li W-H: Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA 86: 6201–6205 (1989).
136. Yabana N, Yamamoto M: *Schizosaccharomyces pombe map1*$^+$ encodes a MADS-box-family protein required for cell-type-specific gene expression. Mol Cell Biol 16: 3420–3428 (1996).
137. Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldman KA, Meyerowitz EM: The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature 346: 35–39 (1990).
138. Yu D, Kotilainen M, Pöllänen E, Mehto M, Elomaa P, Helariutta Y, Albert VA, Teeri TH: Organ identity genes and modified patterns of flower development in *Gerbera hybrida* (Asteraceae). Plant J 17: 51–62 (1999).
139. Zachgo S, de Andra Silva E, Motte P, Tröbner W, Saedler H, Schwarz-Sommer Z: Functional analysis of the *Antirrhinum* floral homeotic *DEFICIENS* gene in vivo and in vitro by using a temperature-sensitive mutant. Development 121: 2861–2875 (1995).
140. Zachgo S, Saedler H, Schwarz-Sommer Z: Pollen-specific expression of *DEFH125*, a MADS-box transcription factor in *Antirrhinum* with unusual features. Plant J 11: 1043–1050 (1997).
141. Zhang H, Forde BG: An *Arabidopsis* MADS box gene that controls nutrient-induced changes in root architecture. Science 279: 407–409 (1998).