Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 7342–7347, June 1999
Evolution

MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants

KAI-UWE WINTER, ANNETTE BECKER, THOMAS MÜNSTER, JAN T. KIM, HEINZ SAEDLER, AND GÜNTER THEISSEN*

Max-Planck-Institut für Züchtungsforschung, Abteilung Molekulare Pflanzengenetik, Carl-von-Linné-Weg 10, D-50829 Cologne, Germany

Communicated by Klaus Hahlbrock, Max Planck Institute for Plant Breeding Research, Cologne, Germany, April 27, 1999 (received for review January 20, 1999)

ABSTRACT **The evolutionary origin of the angiosperms (flowering plants *sensu stricto*) is still enigmatic. Answers to the question of angiosperm origins are intimately connected to the identification of their sister group among extinct and extant taxa. Most phylogenetic analyses based on morphological data agree that among the groups of extant seed plants, the gnetophytes are the sister group of the angiosperms. According to this view, angiosperms and gnetophytes are the only extant members of a clade called "anthophytes" to emphasize their shared possession of flower-like reproductive structures. However, most phylogeny reconstructions based on molecular data so far did not support an anthophyte clade, but also could not clarify the case because support for alternative groupings has been weak or controversial. We have isolated 13 different homologs of MADS-type floral homeotic genes from the gnetophyte *Gnetum gnemon*. Five of these genes fall into monophyletic gene clades also comprising putatively orthologous genes from flowering plants and conifers, among them orthologs of floral homeotic B and C function genes. Within these clades the *Gnetum* genes always form distinct subclades together with the respective conifer genes, to the exclusion of the angiosperm genes. This provides strong molecular evidence for a sister-group relationship between gnetophytes and conifers, which is in contradiction to widely accepted interpretations of morphological data for almost a century. Our phylogeny reconstructions and the outcome of expression studies suggest that complex features such as flower-like reproductive structures and double-fertilization arose independently in gnetophytes and angiosperms.**

In addition to angiosperms, extant seed plants comprise four different groups of gymnosperms, conifers, cycads, gnetophytes (with only three genera, *Gnetum*, *Ephedra*, and *Welwitschia*), and Ginkgo (with the single species *Ginkgo biloba*). Although almost all groups of fossil and living gymnosperms already have been considered as potential angiosperm ancestors (see ref. 1 and refs. therein), there is a century-long tradition interpreting morphological data as evidence for a sister-group relationship between gnetophytes and angiosperms (1–5). These two plant groups often are united together with some mesozoic seed ferns in a clade called anthophytes (4). However, previous phylogeny reconstructions based on different molecular markers obtained from all three plant genomes had difficulties to support this hypothesis. On the contrary, most of the respective phylogenetic trees showed the tendency to place gnetophytes, or Gnetales, as a sister group to conifers rather than to angiosperms or suggested a monophyletic origin of all gymnosperms (6–13). However, in most cases the statistical support for the alternative groupings was relatively weak, and because some phylogenetic trees gave ambiguous results or even weakly supported the anthophyte hypothesis (14–16), the relationship between gnetophytes, angiosperms, and conifers has remained an open question so far.

Because the phylogenetic position of gnetophytes plays a pivotal role for understanding seed plant evolution and the origin of flowers, we wanted to clarify the relationship between gnetophytes and angiosperms by examining the genes responsible for specifying the morphological structures of taxonomic interest. We reasoned that floral meristem and organ identity genes, together with their orthologs from gymnosperms, might be suitable molecular markers for these analyses.

Most floral organ identity genes that could be cloned so far belong to the family of MADS-box genes encoding transcription factors (for recent reviews see refs. 17 and 18). Floral organ identity genes can be subdivided into four different classes, termed A, B, C, and D function genes, whose members provide four different homeotic functions (19, 20). Expression of the A function alone specifies sepal formation within any one of four floral whorls of angiosperm flowers. The combination of A and B function expression specifies the formation of petals, B together with C function expression specifies stamen formation, and expression of the C function genes alone determines the formation of carpels. In many wild-type flowers, the A function is expressed in the first and second floral whorl, the B function is expressed in the second and third whorl, and the C function is expressed in the third and fourth whorl. Therefore, sepals, petals, stamens, and carpels are specified in whorls 1, 2, 3, and 4, respectively. D function genes specify the identity of the ovules that develop within the carpels (20).

MADS-domain proteins from vascular plants share a conserved structural organization, the so-called MIKC-type domain structure, including a MADS (M), intervening (I), keratin-like (K), and C-terminal (C) domain (17, 21–23). The MADS domain is the major determinant of DNA binding, but it also performs dimerization and accessory factor-binding functions (21). The K domain, which has not been found in any of the animal and fungal MADS-domain proteins so far (17, 24), is characterized by a conserved regular spacing of hydrophobic residues, which is proposed to allow for the formation of an amphipathic helix involved in protein dimerization (21, 22).

Phylogeny reconstructions revealed that the MADS-box gene family is composed of several defined gene clades whose members share similar expression patterns and highly related functions. For example, all A, B, C, and D function genes known so far fall into separate clades, namely *SQUAMOSA*- (A function), *DEFICIENS*- or *GLOBOSA*- (B function), and *AGAMOUS*-like genes (C and D function) (17, 23–26). Therefore, the establishment of the mentioned gene clades by gene duplication, diversification, and fixation probably was an im-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AJ132207–AJ132219).

*To whom reprint requests should be addressed. e-mail: theissen@mpiz-koeln.mpg.de.

portant step toward the establishment of the floral homeotic functions (17).

There is evidence from both gene cloning and molecular clock analyses that some of the plant MADS-box gene clades are older than the separation of the lineages that led to extant conifers and angiosperms (27–29). We initiated a screen for orthologs of these genes among the MADS-box gene family of the gnetalean species *Gnetum gnemon* and then used them as molecular markers to clarify the relationship between gnetophytes, conifers, and angiosperms. Our phylogeny reconstructions strongly suggest that gnetophytes are more closely related to conifers than to angiosperms, which is in contradiction to the anthophyte hypothesis. This finding has significant implications for our understanding of flower evolution, which are discussed. Two of the genes introduced here are putative orthologs of floral homeotic B or C function genes, respectively, which is reflected by the expression patterns of these genes.

MATERIALS AND METHODS

Isolation of cDNAs. Partial cDNAs were isolated by 3′ rapid amplification of cDNA ends (RACE) as described generally (23, 30). As template, poly(A)$^+$ RNA isolated from cones of a male and a female *G. gnemon* tree growing in the Botanical Garden of the University of Bochum (Ruhr-Universität, Bochum, Germany) was used. Upstream sequences overlapping with the 3′ fragment were isolated by 5′ RACE, employing a commercially available kit (5′/3′-RACE Kit; Boehringer Mannheim). Sequences of primers used during the RACE procedures can be downloaded from the corresponding author's home page (http://www.mpiz-koeln.mpg.de/~theissen/). For each gene, at least three different cDNA sequences were cloned independently, and both strands were sequenced on automatic sequencers.

Northern Analysis. Gene-specific hybridization probes were obtained from the regions downstream of the MADS-box to avoid cross-hybridization with other gene family members. For the synthesis of probes, linear PCR was employed essentially as described (31), but PCR products of MADS-box gene cDNAs from *Gnetum* were used as templates, and different gene-specific oligonucleotides were used as primers. The probes were hybridized to RNA blots containing 10 μg per lane of total RNA isolated by a standard method (32) from the *Gnetum* trees described above. RNA sources were total male or female cones or young leaves. The filters were hybridized at 65°C in 5× 0.18 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA (SSPE)/5× Denhardt's solution/0.5% SDS/20 μg/ml of herring sperm DNA and washed at 68°C in 0.1× SSPE/0.1% SDS (33).

***In Situ* Hybridization Analysis.** For *in situ* hybridization experiments, PCR fragments of the I, K, or C domains of the *GGM2* and *GGM3* cDNAs (average length of about 200 bp) were introduced into the pGEM-T vector (Promega). To obtain antisense or sense probes, these constructs were linearized in the polylinker region by digestion with *Sac*I or *Nco*I and then used as templates for synthesizing digoxigenin-labeled RNA by using the DIG-RNA-labeling kit (Boehringer Mannheim) and T7 or SP6 polymerase. Plant material was fixed in formalin/acetic acid/alcohol, embedded in Paraplast Plus, and prepared for hybridization as described elsewhere (34). The slides were hybridized at 50°C in a humidified chamber overnight and then washed in 3× SSPE and incubated with 40 μg/ml of RNase A (Boehringer Mannheim) in NTE (500 mM NaCl/10 mM Tris·HCl, pH 7.5/1 mM EDTA) for 30 min at 37°C. The slides were washed twice in 0.3× SSPE for 30 min at 52°C, and immunological detection was carried out as described (34) by using the Anti-DIG Fab fragment antibody (Boehringer Mannheim).

Construction of Phylogenetic Trees. Phylogenetic trees were constructed based on a set of MADS-domain protein sequences comprising the most published gymnosperm and fern sequences, an unpublished gymnosperm sequence (DAL13) kindly provided by P. Engström, a representative set of angiosperm sequences, and our *Gnetum* sequences. A comprehensive set of all published MADS-box genes and phylogenetic trees constructed with them are available from the MADS home page (http://www.mpiz-koeln.mpg.de/mads/). Phylogenetic trees were constructed from MADS-domain sequences and from "170 domain" sequences as described elsewhere (17, 23). The "170 domain" encompasses the 60 aa that constitute the MADS domain and the subsequent 110 aa that comprise the I and K domains. Additionally, we used "MIK-domain sets" for phylogeny reconstructions, where we defined the MIK domain by a so-called "regular expression," i.e., an amino acid character pattern defining characteristic amino acid symbols in the MADS and K domains. Using the MIK domains instead of the 170 domains in phylogeny reconstructions reduces noise resulting from length heterogeneities in the I region. Pattern matching was done by using the GCG program FINDPATTERNS. Multiple alignments were compiled by using the GCG program PILEUP with a gap-creation penalty of 12 and a gap-extension penalty of 4 (the default parameters). Phylogenetic trees were constructed with the neighbor-joining algorithm (35), as implemented by the PHYLIP package (36). The neighbor-joining method was chosen because it is known to be quite efficient in obtaining reliable trees from large data sets (ref. 37 and references therein). The trees were evaluated by bootstrap analysis. To further corroborate the tree topology, we computed alignments with all parameter combinations of gap-creation penalties running from 1 to 25 in increments of 1–3 (larger increments at penalties more than 16) and gap-extension penalties running from 1 to 8 in increments of 1.

RESULTS

cDNA Cloning of MADS-Box Genes from *G. gnemon*. We isolated cDNAs of 13 different MADS-box genes from the gnetophyte *G. gnemon* and have named the respective genes *GGM1–13* (*G. gnemon MADS1–13*). Hybridization of DNA gel blots ("Southern blots") containing genomic DNA of an individual *G. gnemon* tree with different probes specific for each of the *GGM* genes under stringent conditions indicated that *GGM1–13* represent 13 different single-copy genes (data not shown).

Some *Gnetum* Genes Are Orthologs of Floral Homeotic Genes. Phylogeny reconstructions with all available MIKC-type MADS domain proteins, or representative subsets thereof, indicate that some of the *Gnetum* genes fall into gene clades well known from angiosperms (Fig. 1). *GGM1* falls into the subfamily of *TM3*-like genes. *GGM2* shows close affinity to a superclade comprising all *DEF*- and *GLO*-like genes, such as the floral homeotic B function genes *DEF* and *GLO* from *Antirrhinum* and *AP3* and *PI* from *Arabidopsis*. *GGM3* is an *AG*-like gene such as the floral homeotic C function genes *AG* and *PLE* from *Antirrhinum* or *Arabidopsis*, respectively. *GGM9* and *GGM11* are putative *AGL6*-like genes. A representative gene tree is shown in Fig. 1, where subfamilies were defined as in some previous publications (17, 23, 24). For simplicity, this tree was constructed with a subset of protein sequences; a tree containing all MADS-domain proteins known from plants so far is accessible via the World Wide Web (http://www.mpiz-koeln.mpg.de/mads/).

A close relationship between *GGM2* and the *DEF*- and *GLO*-like genes is only moderately supported by bootstrap analysis (Fig. 1). However, multiple sequence alignments reveal that GGM2 and GGM13 share a specific gap in the I domain, which is present in all sequences of the DEF/GLO superclade, but absent in all other sequences. Moreover,

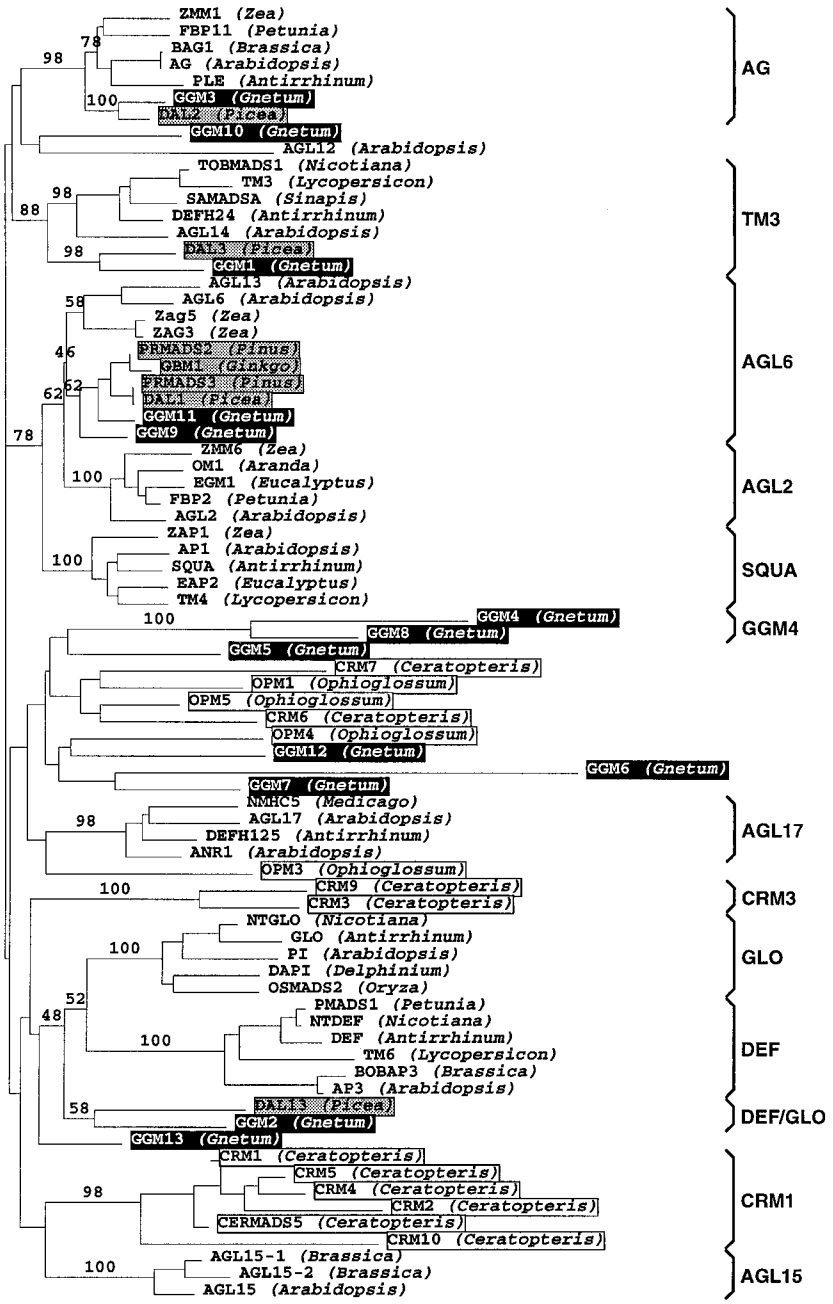

FIG. 1. Phylogenetic tree showing the relationships between a subset of MIKC-type MADS-domain proteins known. The tree was constructed by using the "MIK-domain" sequences and the neighbor-joining algorithm. Genus names of species from which the respective genes were isolated are given in parentheses after the protein names. *Gnetum* proteins are indicated by inverted boxes, and genes from non-gnetalean gymnosperms are indicated by shaded boxes. Proteins from ferns are highlighted by open boxes. Proteins that are not boxed have been derived from angiosperm gene sequences. The numbers next to some nodes give bootstrap percentages, which are shown only for relevant nodes and those defining gene subfamilies. Subfamilies, which generally represent monophyletic gene clades (17, 23), are labeled by brackets at the right margin.

GGM2 contains a "paleoAP3 motif" at its C-terminal end. Such a motif so far has been found only in DEF-like proteins from lower plants and in DEF paralogs from higher plants, called TM6-like proteins (38). Both findings (obvious from sequence alignments accessible via the World Wide Web: http://www.mpiz-koeln.mpg.de/mads/) strongly support the view that *GGM2* is more closely related to *DEF*- and *GLO*-like genes (and perhaps to *GGM13*) than to any other MADS-box genes known.

Bootstrap support for *GGM9* and *GGM11* being *AGL6*-like genes also is quite low (Fig. 1). However, the membership of these genes and some highly related conifer genes in an *AGL2/AGL6/SQUA* superclade (17, 23) (called *AP1/AGL9* group in ref. 26) is quite well supported (Fig. 1).

Because of their membership in defined subclades of the MADS-box gene tree the five genes mentioned above are not just homologs, but are also putative orthologs of the respective genes from angiosperms, meaning that the ancestors of these genes were established during (a) speciation event(s) that separated the lineage(s) that led to extant gymnosperms from the lineage that led to extant angiosperms. The other *GGM* genes do not fall into any of the subfamilies described in the literature (Fig. 1) (17, 23–26).

Expression Patterns of *GGM1*, *2*, *3*, *9*, and *11* Reflect Clade Memberships. Members of any defined MADS-box gene clade from angiosperms generally have very similar expression patterns (17, 24). Although orthology assignments should be based strictly on the fact that two genes originated by specia-

tion, similar expression patterns, therefore, may corroborate hypotheses about orthology if these expression patterns are found for most (or even all) members of the clade of putatively orthologous genes and are rarely (or not at all) found outside the respective gene clade. *GGM1* is a member of the clade of *TM3*-like genes. Many members of this clade belong to the few MADS-box genes known from angiosperms that show a quite ubiquitous expression in both vegetative and reproductive organs (reviewed in ref. 17). In line with this, *GGM1* is exceptional among the *Gnetum* MADS-box genes because it is expressed not only in reproductive cones, but also in vegetative leaves (Fig. 2). In angiosperms, *DEF*- and *GLO*-like B function genes usually are expressed strongly in stamens, the male reproductive organs of flowers, and in petals, but not in carpels (the female reproductive organs) (17). The putative *DEF*/*GLO*-like gene *GGM2* from *Gnetum* also is expressed in the male, but not in the female reproductive cones of *Gnetum* (Fig. 2). (Note that there are no petals in *Gnetum*.) Within the male reproductive units, expression was found in the antherophores, but not in the surrounding envelopes (Fig. 3 *B* and *E*). All known *AG*-like floral homeotic C function genes from angiosperms are expressed in stamens as well as in carpels (17). The *AG*-like gene *GGM3* also is expressed in male as well as in female reproductive cones (Fig. 2). At early stages of development, this gene is expressed in all organs of the reproductive units (nucellus, antherophore, and all envelope organs) (Fig. 3*A*). At late developmental stages, expression is localized in the outer envelope of both male and female reproductive units (Fig. 3 *C* and *D*). Expression of the *AGL6*-like genes *GGM9* and *GGM11* was found in male as well as female cones, but not in leaves (Fig. 2). Expression of the few *AGL6*-like genes from angiosperms that have been characterized also is restricted to inflorescences (22, 39). Within flowers, transcription was found in reproductive as well as sterile organs (39). In summary, the expression patterns of all *GGM* genes considered here are in full agreement with the orthology assignments made by phylogeny reconstructions.

Subclades of *Gnetum* and Conifer Genes Reveal That Gnetophytes Are More Closely Related to Conifers than to Angiosperms. It is obvious from the phylogenetic tree (Fig. 1) that in all cases in which gene subfamily members (putative orthologs) 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, whereas angiosperm genes form separate clades. By far, the most plausible explanation for this finding is that the genes from gymnosperms were generated by speciation events that occurred after the lineage that led to extant angiosperms branched off from the lineage that led to extant gnetophytes and conifers. Our data thus strongly support the hypothesis that gnetophytes are more closely related to conifers than to angiosperms.

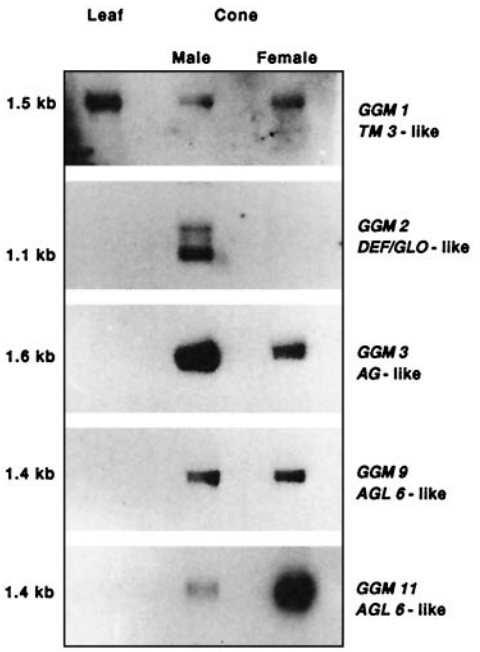

FIG. 2. Northern blot analysis of *GGM* gene expression. The names and subfamily memberships of the respective genes are indicated at the right. At the left, the apparent length of the major band is indicated in kb. RNA sources were young leaves and male or female cones from two individual *G. gnemon* trees as indicated.

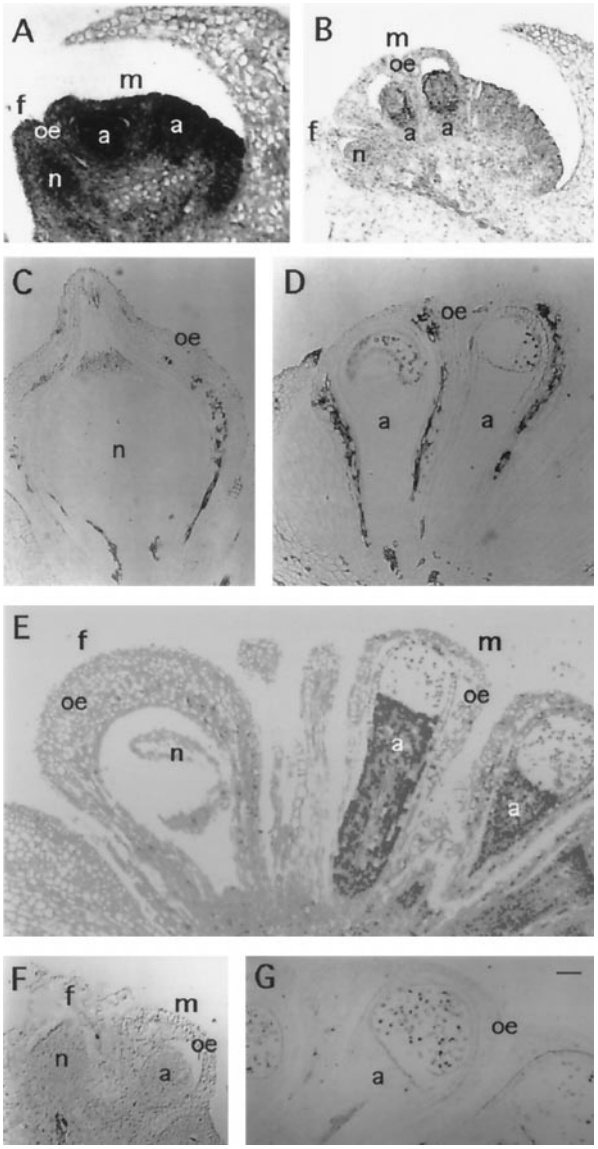

FIG. 3. Expression patterns of *GGM2* and *GGM3* as determined by *in situ* hybridization. In *A–E*, digoxigenin-labeled antisense probes were used, which detect *GGM2* (*B* and *E*) or *GGM3* (*A*, *C*, and *D*) transcripts, respectively. Using sense probes of *GGM3* as control did not result in visible signals (*F* and *G*). All sections are longitudinal ones. (*A*, *B*, and *F*) Sections of a node of a male strobilus at an early developmental stage. (*C*) Section through a sterile female reproductive unit of a male strobilus at a relatively late developmental stage. (*D*) Section through male reproductive units at a relatively late developmental stage. (*E*) Section of a node of a male strobilus at a relatively late developmental stage. (*G*) Section through male reproductive units at a relatively late developmental stage. (Bar = 100 μm.) a, antherophore; f, female reproductive unit (sterile); m, male reproductive unit; n, nucellus; oe, outer envelope.

To critically evaluate the statistical significance of our finding, we have constructed phylogenetic trees for each of the relevant gene clades individually, using single members from other gene clades as outgroups. It turned out that bootstrap support for *Gnetum*-conifer clades within the individual gene subfamily trees is 100% (*AG* and *TM3* clade) or at least above the values given in Fig. 1 (*AGL6* and *DEF/GLO* clade) (66% in both cases). In addition to the trees based on the well defined "MIK-pattern" (see *Materials and Methods*), trees have been calculated by using alignments that are based on the MADS domain (60 aa) and on the "110 domain," the 110 aa directly downstream of the MADS domain including the amino acids of the I and the K domain. The trees obtained gave the same results with respect to the subclade structures mentioned above. Because inaccurate sequence alignments are one of the most serious reasons for errors in phylogeny reconstructions based on sequence data (40), we also have constructed phylogenetic trees based on alignments for which we systematically varied the alignment-parameters gap weight and gap-length weight (see *Materials and Methods*). These trees have been compared with the tree shown in Fig. 1. All important subfamilies as well as all *Gnetum*-conifer clades are strongly supported over a wide range of parameters tested (data not shown). Therefore, our findings are insensitive to using different data subsets, sequence domains, and alignment parameters.

In addition to MADS-box genes, we used all currently available sequences from orthologs of the non-MADS-type floral meristem identity genes *FLORICAULA/LEAFY* to reconstruct the relationship between conifers, gnetophytes, and angiosperms. In the obtained phylogenetic tree (accessible via the World Wide Web: http://www.mpiz-koeln.mpg.de/~theissen), the respective genes from gnetophytes and conifers also form a highly supported clade that excludes the angiosperm genes, thus supporting our conclusions concerning the relationships between these taxa based on MADS-box genes.

DISCUSSION

Our conclusions concerning the evolutionary relationships between the taxa in question are based on five different genes of four different MADS-box gene subfamilies (plus the *FLORICAULA/LEAFY* data), all showing essentially the same result. In addition, our results are robust with respect to using different gene subsets, sequence domains, and alignment parameters, and at least some of the essential clades have reasonably high bootstrap support (Fig. 1). The robustness of our results with respect to parameterization of the alignment computation (generally a potential reason of significant errors; see ref. 40) and to potential sampling errors can be attributed to the fact that alignments using the M, I, and K domains can be computed with very little ambiguity, owing to the strong conservation of the domain structure of MIKC-type genes (17, 23, 26). Moreover, given the high sequence similarity between *Gnetum* and conifer sequences relative to the average similarity between plant MADS-domain proteins, it seems unlikely that undetected angiosperm genes exist that would dissect the different *Gnetum*-conifer clades and, thus, would indicate alternative relationships such as *Gnetum*-angiosperm clades. Finally, our results are in agreement with most other recent phylogeny reconstructions based on molecular markers (6–13), including the tree we calculated with the sequences of the *FLORICAULA/LEAFY* orthologs, although all the former work did not lead to final conclusions concerning the phylogenetic relationships of the taxa in question. Taken together, we suggest that within this work an important aspect of seed plant phylogeny has been identified correctly.

The phylogenetic relationship between gnetophytes, angiosperms, and conifers is a serious but interesting case of conflict between morphological and molecular data. We assume that certain morphological characters erroneously have been classified as being homologous, resulting in a phylogenetic interpretation that now turns out to be irreconcilable with molecular data. The most striking features of gnetophytes, which have been interpreted as synapomorphies of the "anthophytes," are the flower-like appearance of reproductive structures, vessels in the secondary wood, and a kind of double-fertilization (5). More recent investigations, however, revealed that gnetalean wood shares many more features with the wood of conifers than with the wood of angiosperms and that gnetalean and angiosperm vessels have independent evolutionary origins (41). Additionally, the second fertilization event of gnetophytes does not lead to the formation of a triploid endosperm as in angiosperms, but to a diploid product that expresses the developmental program of an embryo (42). Thus, it is clearly different from the second fertilization event of angiosperms and, therefore, may have an independent origin. Our data strongly support the view that at least some of the morphological or physiological features that are similar in angiosperms and in Gnetales represent analogies rather than homologies. This does not exclude, however, that the parallel appearance of these characters was facilitated by a common developmental potential that already was present in the last common ancestor of angiosperms, gnetophytes, and conifers (and probably all lineages leading to extant seed plants) 300–400 million years ago. A common set of developmental control genes, including representatives of the subfamilies of MADS-box genes presented here, may have contributed significantly to that developmental potential.

Therefore, we believe that the genes we have discussed here are not only useful markers to determine the deep branching of the seed plant phylogenetic tree, but are also helpful tools to test assumptions about structural and developmental homologies among the reproductive structures of the diverse seed plant groups (4). Some examples are known now in which orthologous developmental control genes do not specify homologous structures or, likewise, in which the development of homologous organs is not controlled by orthologous genes; such cases seem to be rare, however (43). In most cases it can be expected, therefore, that homologous organs express orthologous developmental control genes (44), so the expression of such genes can be used with some confidence to make inferences about organ homology. For example, because the organs of the outer envelope of *Gnetum* reproductive units express *GGM3*, an ortholog of floral homeotic C and D function genes (Fig. 3 *A*, *C*, and *D*), but not the putative B function gene ortholog *GGM2* (Fig. 3 *B* and *E*), these organs appear not to be homologous to petals (which express B, but not C or D function genes). The flower-like appearance of the reproductive units of *Gnetum* is based largely on the presence of integuments (or envelopes) that resemble a floral perianth. The hypothesis that the respective integument organs are not homologous to the perianth organs of angiosperms (but possibly to the integument organs of angiosperm ovules) suggests that the flower-like appearance of the reproductive units of *Gnetum* is also a case of parallel or convergent evolution rather than common ancestry with angiosperms. This hypothesis thus is in full agreement with our conclusion concerning the relationship between gnetophytes and angiosperms and our assumption about the analogous character of some other morphological similarities between angiosperms and gnetophytes. The hypothesis that the organs of the outer envelope of *Gnetum* are not homologous to petals also is in contrast to a version of the euanthial model of flower origin (4). Euanthial models assume that the flower was derived from a single plant axis with sporophylls on it (4).

We believe, however, that although gnetophytes and angiosperms are more distantly related than often assumed, the genes we are studying still might be helpful to clarify flower

origin. For example, the presence of orthologs of floral homeotic B and C function genes in gymnosperms such as *Gnetum* suggests that the system for the specification of reproductive organ identity in angiosperms was recruited from a similar system that already was present in the last common ancestor of all extant seed plants about 300 million years ago. Because the C function genes of angiosperms specify the identity of reproductive organs (stamens and carpels, respectively) and because their ortholog from *Gnetum* also is expressed in both male and female reproductive units (Fig. 3 *A*, *C*, and *D*), it may have been the function of the expression of ancestral C function genes to distinguish between reproductive organs (where expression is on) and nonreproductive organs (where expression is off). Because the B function genes of angiosperms specify stamens (male organs), but not carpels (female organs), and because the ortholog from *Gnetum* also is expressed exclusively in male reproductive units (Fig. 3 *B* and *E*), it may have been the function of the expression of ancestral B function genes to distinguish between male reproductive organs (where expression is on) and female reproductive organs (where expression is off). Differential expression of B function genes thus may represent the primary sex-determination mechanism of all seed plants.

We thank Thomas Stützel (Lehrstuhl Spezielle Botanik, Ruhr-Universität Bochum, Germany) for plant material from *G. gnemon*. We also thank the Automatic DNA Isolation and Sequencing team of our institute for sequencing the cDNA clones. We are indebted to Bill Martin, Aidyn Mouradov, Jens Sundström, and Peter Engström for providing data or information before publication. Financial support from the Deutsche Forschungsgemeinschaft to G.T. (Grant Th 417/3–1) and to A.B. (via Graduiertenkolleg "Molekulare Analyse von Entwicklungsprozessen bei Pflanzen") is highly acknowledged.

1. Crane, P. R., Friis, E. M. & Pedersen, K. R. (1995) *Nature (London)* **374,** 27–33.
2. Arber, E. A. N. & Parkin, J. (1907) *J. Linn. Soc. Bot.* **38,** 29–80.
3. Arber, E. A. N. & Parkin, J. (1908) *Ann. Bot.* **22,** 489–515.
4. Doyle, J. A. (1994) *Plant Syst. Evol.* **8,** Suppl., 7–29.
5. Doyle, J. A. (1996) *Int. J. Plant Sci.* **157,** Suppl. 6, S3–S39.
6. Hasebe, M., Kofuji, R., Ito, M., Kato, M., Iwatsuki, K. & Ueda, K. (1992) *Bot. Mag. Tokyo* **105,** 673–679.
7. Goremykin, V., Bobrova, V., Pahnke, J., Troitsky, A., Antonov, A. & Martin, W. (1996) *Mol. Biol. Evol.* **13,** 383–396.
8. Barnabas, S., Krishan, S. & Barnabas, J. (1995) *Bioscience* **20,** 259–272.
9. Fischer, H., Chen, L. & Wallisch, S. (1996) *J. Mol. Evol.* **43,** 399–404.
10. Chaw, S.-M., Zharkikh, A., Sung, H.-M., Lau, T.-C. & Li, W.-H. (1997) *Mol. Biol. Evol.* **14,** 56–78.
11. Malek, O., Lättig, K., Hiesel, R., Brennicke, A. & Knoop, V. (1996) *EMBO J.* **15,** 1403–1411.
12. Shutov, A. D., Braun, H., Chesnokov, Y. V., Horstmann, C., Kakhovskaya, I. A. & Bäumlein, H. (1998) *J. Mol. Evol.* **47,** 486–492.
13. Schneider-Poetsch, H. A. W., Kolukisaoglu, Ü., Clapham, D. H., Hughes, J. & Lamparter, T. (1998) *Physiol. Plant.* **102,** 612–622.
14. Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y.-L., *et al.* (1993) *Ann. Missouri Bot. Gard.* **80,** 528–580.
15. Sanderson, M. J. (1997) *Mol. Biol. Evol.* **14,** 1218–1231.
16. Stevanovic, S., Jager, M., Deutsch, J., Broutin, J. & Masselot, M. (1998) *Am. J. Bot.* **85,** 688–697.
17. Theissen, G., Kim, J. & Saedler, H. (1996) *J. Mol. Evol.* **43,** 484–516.
18. Riechmann, J. L. & Meyerowitz, E. M. (1997) *Biol. Chem.* **378,** 1079–1101.
19. Weigel, D. & Meyerowitz, E. M. (1994) *Cell* **78,** 203–209.
20. Angenent, G. C. & Colombo, L. (1996) *Trends Plant Sci.* **1,** 228–232.
21. Shore, P. & Sharrocks, A. D. (1995) *Eur. J. Biochem.* **229,** 1–13.
22. Ma, H., Yanofsky, M. F. & Meyerowitz, E. M. (1991) *Genes Dev.* **5,** 484–495.
23. Münster, T., Pahnke, J., Di Rosa, A., Kim, J. T., Martin, W., Saedler, H. & Theissen, G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 2415–2420.
24. Theissen, G. & Saedler, H. (1995) *Curr. Opin. Genet. Dev.* **5,** 628–639.
25. Doyle, J. J. (1994) *Syst. Biol.* **43,** 307–328.
26. Purugganan, M. D., Rounsley, S. D., Schmidt, R. J. & Yanofsky, M. (1995) *Genetics* **140,** 345–356.
27. Tandre, K., Albert, V. A., Sundas, A. & Engström, P. (1995) *Plant Mol. Biol.* **27,** 69–78.
28. Mouradov, A., Glassik, T. V., Hamdorf, B. A., Murphy, L. C., Marla, S. S., Yang, Y. & Teasdale, R. (1998) *Plant Physiol.* **117,** 55–61.
29. Purugganan, M. D. (1997) *J. Mol. Evol.* **45,** 392–396.
30. Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8998–9002.
31. Fischer, A., Baum, N., Saedler, H. & Theissen, G. (1995) *Nucleic Acids Res.* **23,** 1901–1911.
32. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
33. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
34. Jack, T., Brockman, L. L. & Meyerowitz, E. M. (1992) *Cell* **68,** 683–697.
35. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
36. Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package) (Department of Genetics, University of Washington, Seattle), Version 3.5c.
37. Zhang, J. & Nei, M. (1996) *Genetics* **142,** 295–303.
38. Kramer, E. M., Dorit, R. L. & Irish, V. F. (1998) *Genetics* **149,** 765–783.
39. Mena, M., Mandel, M. A., Lerner, D. R., Yanofsky, M. F. & Schmidt, R. J. (1995) *Plant J.* **8,** 845–854.
40. Goldman, N. (1998) *BioEssays* **20,** 287–290.
41. Carlquist, S. (1996) *Int. J. Plant Sci.* **157,** Suppl. 6, S58–S76.
42. Friedman, W. E. & Carmichael, J. S. (1996) *Int. J. Plant Sci.* **157,** Suppl. 6, S77–S94.
43. Wray, G. A. & Abouheif, E. (1998) *Curr. Opin. Genet. Dev.* **8,** 675–680.
44. Albert, V. A., Gustafsson, M. H. G. & Di Laurenzio, L. (1998) in *Molecular Systematics of Plants II*, eds. Soltis, D. E., Soltis, P. S. & Doyle, J. J. (Kluwer, Boston), pp. 349–374.