Commentary

MADS about Gnetales

Michael W. Frohlich*

University of Michigan, Ann Arbor, MI 48109-1057

The evolutionary origin of flowering plants (angiosperms) has long been the most prominent unresolved issue in plant evolutionary biology. Darwin called it an "abominable mystery" because angiosperms seemed to appear suddenly, in considerable diversity, but without obvious antecedents. More than a century later, the problem still remains unsolved.

Flowering plants reproduce by seeds and are clearly related to the seed plants collectively called gymnosperms. Angiosperms and gymnosperms differ by a host of characters, most notably in the reproductive structures, but also in other features, such as the water-conducting cells. These great differences are one cause of the continuing problem: they make homology assessment between flowering plants and gymnosperms difficult and unconvincing, particularly in regard to reproductive structures. Different views on homology have led to grossly different scenarios of angiosperm origins. The second cause is uncertainty in phylogenetic relationships between flowering plants and the gymnosperm groups (both living and fossil) that could be their relatives or ancestors as well as within the flowering plants themselves. Gymnosperm groups vary enormously, so uncertainty on which are related to flowering plants allows many starting points to be imagined for the evolution of flowers. The two problems are related in that differing views on homology influence the analysis of relationships, and knowledge of relationships can suggest what structures are likely homologous.

The paper by Winter *et al.* (1) in the June 22 issue of the *Proceedings* makes a major advance on the second of these critical questions and gives great promise of major advance on the first. Phylogenetic relationships of seed plants have been especially contentious in the last few years, with most morphological analyses pointing in one direction, with most molecular studies giving contrary results, but so weakly that the most recent overall analysis found the morphological results more believable (2). Winter *et al.* come down firmly on the side of the recent molecular analyses, which makes the most widely accepted theory of flower origins very unlikely; but their most exciting results are the identification of putative gymnosperm orthologs to flowering-plant genes that specify flower-organ identity.

Diversity of Seed Plants. Many gymnosperm groups are known from fossils, but only four groups survive to the present: conifers (pine, spruce, larch, etc.; in general, the needle-leaved evergreens); cycads (tropical plants with a single thick stem and large, leathery compound leaves, most familiar as expensive ornamentals); Ginkgo (the tree from China with fan-shaped leaves, now widely planted); and the Gnetales. Gnetales, the least familiar, are bizarre in many ways, starting with the fact that the "G" is silent. They include only three genera: *Ephedra*, a green-twigged shrub of Western Hemisphere and Eurasian semi-deserts, most famous as the source of the cold medicine pseudoephedrine; *Gnetum,* a tropical vine or tree; and *Welwitschia*, with a single species found only in the Namib desert of Southern Africa, which can live over 1,000 years but only makes two leaves that grow from the base as they die at the tips and get shredded lengthwise by the winds. Fossil gymnosperm groups are very diverse, but only some of those pertinent to the origin of flowers will be mentioned below.

History of the Problem. Views on the origin of flowers have shifted dramatically several times. In the early years of this century, Gnetales were thought to comprise the closest living relatives of flowering plants, because of a number of features apparently shared by both groups: the presence of vessels (water-conducting cells with holes all the way through the cell wall); ovules (the structures that grow into seeds) with two layers of tissue (integuments) covering their surface instead of only one layer; and other attributes. The earliest flowers were thought to be very simple, unisexual, and lacking sepals and petals (the typical outer floral organs) and having only stamens (the pollen-producing organs) or carpels (the female organs, containing the ovules).

By the 1930s, the Gnetales were in disfavor as close angiosperm relatives for a variety of reasons: e.g., the simple flowers previously thought to reflect ancestral features turned out to have vestigial structures suggesting reduction from more elaborate predecessors with both stamens and carpels, as well as sepals or petals in the flower. Instead, flowers like *Magnolia*, with many sepals and/or petals and stamens and carpels, were thought to reflect ancestral flower features. Differences between Gnetales and flowering plants were stressed as evidence against a close relationship, e.g., that the vessels in the two groups arose in different ways, suggesting convergent evolution rather than homology. Gnetales ovules are borne directly on a stem tip, whereas flowering-plant ovules are borne on the carpel, which appears to be a modified leaf. People thought flowering plants were only distantly related to any living gymnosperms, with cycads probably the least distant. Possible relationship with fossil groups was stressed, especially Caytoniales. The Caytoniales bore their ovules in cup-shaped organs, bent to face the stalk of the cup; this seemed to be a reasonable antecedent for the typical angiosperm ovule—if the cup contained just one ovule, and the cup got fused to the ovule, then the ovule would have two integuments, the inner one being the usual single covering of gymnosperms and the outer one being the fused-on cup. Furthermore, the structure would be bent over, as in most flowering-plant ovules.

Cladistic Studies. The recent period began in the mid-1980s, when cladistic methodology was first applied to evaluate relationships of flowering plants and gymnosperms, most notably by Crane (3), whose analysis suggested that Gnetales were, after all, the closest living relatives of flowering plants. This was so surprising that it stimulated other studies, such as those by Doyle and Donoghue (4–6) that found the same result. In these morphologically based studies, flowering plants were grouped with Gnetales and two fossil groups, Bennetitales and Pentoxylales, in what was called the Anthophyte clade. Bennetitales reproductive units have flat structures surrounding male organs with ovules in the middle. This arrangement resembles the order of parts in flowers (sepals and/or petals surrounding stamens with carpels in the middle). Gnetales plants are functionally either male or female. Their reproductive units have flat structures (bracts) surrounding either male structures or an ovule, but in

Welwitschia, the male structures surround a sterile ovule. The similarity in overall organization suggested that this pattern evolved only once and was homologous in the three groups. (Male and female structures are borne separately in Pentoxylales.) If true, this would mean a major feature of flowering plants evolved millions of years before the flowering plants themselves and was shared (although in modified form) by the living Gnetales.

Other studies suggested that instead of Gnetales being a sister group of angiosperms, Gnetales were the direct ancestors of flowering plants (7, 8). This Neopseudanthial view leads to radically different homology assessments than those above (9). Further, later analyses of the Anthophyte clade added Caytoniales to the group, next to flowering plants (10), making convergent evolution as reasonable as homology for explaining the similar organization of Anthophyte reproductive units.

These results made Gnetales of enormous interest for understanding the origin of flowering plants, stimulating a renaissance in work on the group (11). Surprising discoveries came quickly: In Gnetales, as in flowering plants, two sperms are delivered to the vicinity of the egg at fertilization, and both sperm nuclei fuse with nuclei from the female parent. In each group, one product is the zygote. In flowering plants, the second sperm typically fuses with two other nuclei, resulting in a triploid nucleus that develops into a food-storage tissue (endosperm) for the seed. In *Gnetum* and *Ephedra*, the second fusion product is diploid and begins to develop as an extra embryo. The evolutionary origin of endosperm in flowering plants and the "double fertilization" that generates it had been utterly mysterious; this suggested that fusion involving the second sperm nucleus could be homologous in Gnetales and flowering plants, perhaps originally generating a second embryo in flowering plants and only later being modified for food storage (12). This discovery also helped support a relationship between Gnetales and flowering plants.

Molecular Studies. A number of molecular studies address the question of relationships between flowering plants and the living gymnosperms. Some studies are consistent with Gnetales as sister to angiosperms (13, 14). Another suggests Gnetales are the most distant seed plants from flowering plants (15). Still others put the living gymnosperms together in a single group (so extant gymnosperms are said to be monophyletic), with flowering plants joined to the base of this gymnosperm group (16–18).

Several analyses, which at first glance may seem to support Gnetales as sister to flowering plants, in fact do not. The massive study of the chloroplast gene rbcL (19) and its subsequent reanalysis as "treezilla" (http://herbaria.harvard.edu/~rice/treezilla/16333.con.asc.html) (20) both portray Gnetales as sister to angiosperms, but this depends on the position of the root of the tree. Here, the root can only be defined by outgroups (organisms more distant than the ones under study, e.g., ferns or other free-spore-producing plants), but none were included in these studies. Cycads were arbitrarily placed at the root to fit the prevailing views of gymnosperm relationships; if the root were placed between Gnetales and flowering plants, then extant gymnosperms would be portrayed as monophyletic. Doyle *et al.*'s (21) combined molecular and morphological study also lacks outgroups. All three of these excellent papers only test relationships within angiosperms; gymnosperms are included only as outgroups.

Which of the analyses are the strongest? Strength of phylogenetic inference is most commonly judged by bootstrap analysis: the data (e.g., columns in the alignment) are resampled (with replacement) to make multiple data sets, each with the same total number of characters (columns) as the real data set, but with some characters appearing multiple times and others being absent, as controlled by a random number generator. One then sees how many of these resampled data sets yield the same phylogenetic groupings as the real data, expressed as a percentage. If the percentage is very high, then those groupings in the real data set probably do not depend on the chance co-occurrence of particular characters, and so are more strongly supported. A

second test of support is used in parsimony analyses, in which the phylogenetic tree with the smallest number of steps is taken as the one with the best support. One can also consider trees that are a few steps longer than the shortest. The number of extra steps it takes to find trees that lack a particular grouping is the "Bremer support," with large numbers implying strong support. Bootstrap and Bremer support do correlate, although not precisely, and strongly supported groups have a maximum bootstrap value of 100%, whereas Bremmer support values can increase indefinitely.

Unfortunately, the molecular studies are contradictory and all are weak. No bootstrap support is given in ref. 13, and in ref. 14, it is <50%, with Bremer support of only 1. Several contradictory molecular trees are shown in ref. 15. In refs. 16–18, the bootstrap supports are 63%, 58%, and 56%, barely high enough to define the groups.

The morphological studies generally lack bootstrap analyses, largely because of missing data from fossils, which causes artifactually low values. In ref. 10, it is 46%. The morphology-based trees all show similar results, and Doyle (2, 22) finds them more believable than results from the molecular studies.

Why are the molecular phylogenetic studies so weak? The immediate cause is too much homoplasy (i.e., parallel or convergent evolution), likely related to "long branch attraction," i.e., the relatively high probability of change on long branches allows character states that correctly indicate relationship to be overwhelmed by convergent changes that suggest erroneous relationships. In this tree, branches leading back to the divergence of extant gymnosperm groups are very long, while internodes connecting extant gymnosperms to each other and to flowering plants—which must be resolved to obtain an informative phylogenetic tree—are much shorter.

Why should the results of Winter *et al.* give strong evidence when the much larger traditional molecular data sets are inconclusive? Their work is based on nuclear genes that are part of gene families, which have arisen through paralogous gene duplications. Paralogs typically evolve independently after the duplication (or at least after sufficient sequence divergence inhibits gene conversion), causing a branch in the phylogenetic tree of genes at each duplication event. These branches can cut the long internodes that generate long-branch attraction. This is especially important in placing the root of the seed plants, because the longest branch of all may be the one to the living outgroups. The numerous MADS-box genes have multiple duplications, reducing long branch attraction and allowing each duplicated paralog to help root the others (23).

Plant molecular taxonomy has mostly been done with chloroplast genes (especially rbcL, the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase) and nuclear ribosomal RNA genes. The reason is partly historical; these were the easiest subjects to study in the early days of molecular taxonomy, before PCR became the method of choice. They are also convenient targets of PCR, because they are multicopy, but they do not persist as long-lived paralogs that could cut long branches. I note that these genes have not changed function in a billion years. Thus, sequence changes are likely to be nearly neutral, with the sequence doing a random walk in the space of acceptable sequence variants. Even a long random walk permits the same state to recur, simulating evidence of relationship. Homeotic genes may undergo directional selection, because they change function or change protein partners or DNA-binding specificity, thus shifting the neutral zone in which the random walk can occur. Furthermore, directional selection may be most frequent in the course of substantial morphological evolution, as in the evolution of novel features of major taxonomic groups. Although this is speculative, I suggest that it should reduce homoplasy and so make homeotic genes more useful for phylogenetics.

Current Result. The analysis of Winter *et al.* covers five homeotic genes. In each case, the Gnetales gene pairs with a conifer gene rather than with the flowering-plant genes. Three of these are supported at high bootstraps (e.g., 100% or 98%),

constituting very strong evidence that Gnetales are not sister to the angiosperms. This evidence is enough to refute the Anthophyte and Neopseudanthial theories. Attempts to escape from this conclusion—by claiming that undiscovered (or lost) angiosperm genes would insert next to the Gnetales genes—fail against so many genes supporting the same conclusion. Without genes from the other two gymnosperm groups, one cannot determine the actual living sister group of flowering plants; whether extant gymnosperms are monophyletic is an open question.

If Gnetales are not sister to the flowering plants, why do morphological analyses give that result? Blaming chance would beg the question, suggesting the alternative that character states are miscoded. This problem arises from the difficulty in determining homologies, which is critical for defining characters and states. The wrong morphological analysis can lead to tendentious character-state coding; to avoid this, the vessels of Gnetales and angiosperms were called the same state (implying homology), despite old work by Thompson (24) suggesting that they arose in different ways. The difference is now supported by Carlquist (25), so Winter *et al.* suggest these vessels represent different character states (an inference that is based partly on their own phylogenetic tree). Winter *et al.* stress the differences between "double fertilization" in Gnetales and flowering plants, but in the absence of the tree, one could equally well stress the similarities. It was the Anthophyte clade that suggested homology between angiosperm flowers and Gnetalean reproductive units; with that tree debunked, the inference vanishes. Other similarities could represent convergences or primitive features retained by Gnetales and angiosperms.

Some may be disconcerted by the repeated, dramatic shifts in phylogenetic interpretation and character-state codings in this story, but such is normal in this type of analysis, which aims not simply to describe what happens (as in much of biochemistry and molecular biology) but rather to infer historical events that cannot be observed directly. Some might claim there are hints of circularity between phylogenetic analysis and character-state coding, but I think this only shows that the fundamental criterion for such endeavors is the internal self-consistency of the observations, the interpretations, and the resulting theories (including the view as to which data are crucial and need to be explained). The criterion of self-consistency allows major shifts between what amount to miniparadigms.

The New Genes. This said, there is another aspect of Winter *et al.* that is even more exciting, i.e., their discovery of so many MADS-box genes in *Gnetum* and the expression patterns they report and inferences that begin to draw. MADS-box transcription factors form a sizeable family in plants, with many subfamilies (26). Most show specific expression patterns, suggesting involvement in the development of various structures. All but one of the most famous plant homeotic genes—the A, B, and C class genes—are MADS genes of, respectively, subfamilies SQUA, DEF + GLO, and AG (ref. 1, Fig. 1). The major flower organs are specified by combinations of the ABC genes: sepals (class A), petals (classes A and B), stamens (classes B and C), and carpels (class C). The recently proposed D class genes specify ovules (27). The expression patterns Winter *et al.* report for *Gnetum* MADS genes fit beautifully with the angiosperm genes of the same subfamilies and will serve to illuminate both developmental mechanisms in gymnosperms and the evolution of development across seed plants. As Winter *et al.* note, this shows these gene subfamilies antedate the divergence of flowering plants and conifers+Gnetales.

The identification of class B MADS-box genes in *Gnetum* and *Picea* (a conifer) is especially important. Although placement of the gymnosperm DEF/GLO genes in the tree is weak, the identification of the "paleo AP3 motif" (28) and the shortened I segment in GGM2 convincingly make it a B class gene. The shortened I segment in GGM13 suggests that it may be the

dimerization partner of GGM2, because this shortened segment appears important for heterodimerization of the flowering plant B class genes (DEF and GLO) (29).

Winter *et al.* note that GGM3 is expressed in the flattened outer structures of *Gnetum* reproductive units, suggesting that these outer structures cannot be petal or sepal homologs, as in the defunct Anthophyte theory, but this would be consistent with homology to the outer integument of flowering plant ovules. This mirrors the ideas of 90 years ago, when Gnetales were first thought to be related to angiosperms, as well as the recent Neopseudanthial theory, and questions ovule origin from a Caytoniales-like ancestor. Of course, the homology could be very distant, and perhaps Caytoniales also expressed C class genes in their cup-shaped organs (although they are extinct, so we will never know). There is also another possibility. Plants have repeatedly incorporated structures adjacent to the reproductive organs into increasingly elaborate reproductive complexes. This is likely mediated by expression of reproduction-specific genes in these newly incorporated structures. Perhaps C class gene expression served such a role for these flattened structures of Gnetales. Various genes specific to integuments of flowering plant ovules are being cloned. When their homologs are studied in Gnetales, this question may be resolved.

The story of extant seed plant phylogeny is reaching its dénouement. Analysis of the evolution of development among the major seed plant groups is just beginning. The next few years will be very interesting.

1. Winter, K.-U., Becker, A., Münster, T., Kim, J. T., Saedler, H. & Theissen, G. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 7342–7347.
2. Doyle, J. A. (1998) *Mol. Phyl. Evol.* **9**, 448–462.
3. Crane, P. R. (1985) *Ann. Mo. Bot. Gard.* **72,** 716–793.
4. Doyle, J. A. & Donoghue, M. J. (1986) *Bot. Rev.* **52,** 321–431.
5. Doyle, J. A. & Donoghue, M. J. (1992) *Brittonia* **44,** 89–106.
6. Doyle, J. A. & Donoghue, M. J. (1993) *Paleobiology* **19,** 141–167.
7. Nixon, K. C., Crepet, W. L., Stevenson, D. & Friis, E. M. A. (1994) *Ann. Mo. Bot. Gard.* **81,** 484–533.
8. Hickey, L. J. & Taylor, D. W. (1996) in *Flowering Plant Origin, Evolution and Phylogeny*, eds. Taylor, D. W. & Hickey, L. J. (Chapman & Hall, New York), pp. 176–231.
9. Doyle, J. A. (1994) *Plant Syst. Evol. Suppl.* **8,** 7–29.
10. Doyle, J. A. (1996) *Int. J. Plant Sci.* **157,** S3–S39.
11. Friedman, W. E., ed. (1996) *Int. J. Plant Sci.* **157,** Suppl.
12. Friedman, W. E. & Carmichael, J. S. (1996) *Int. J. Plant Sci.* **157,** S77–S94.
13. Hamby, R. K. & Zimmer E. A. (1992) in *Molecular Systematics of Plants*, eds. Soltis, P. S., Soltis, D. E. & Doyle J. J. (Chapman & Hall, New York), pp. 50–91.
14. Stefanović, S., Jager, M., Deutsch, J., Broutin, J. & Masselot, M. (1998) *Am. J. Bot.* **85,** 688–697.
15. Albert, V. A., Backlund, A., Bremer, K., Chase, M. W., Manhart, J. R., Mishler, B. D. & Nixon, K. C. (1994) *Ann. Mo. Bot. Gard.* **81,** 534–567.
16. Hasebe, M., Kofuji, R., Ito, M., Kato, M. & Ueda, K. (1992) *Bot. Mag. Tokyo* **105,** 673–679.
17. Goremykin, V., Bobrova, V., Pahnke, J., Troitsky, A., Antonov, A. & Martin, W. (1996) *Mol. Biol. Evol.* **13,** 383–396.
18. Chaw, S. M., Zharkikh, A., Sung, H. M., Lau, T. C. & Li, W. H. (1997) *Mol. Biol. Evol.* **14,** 56–68.
19. 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. Mo. Bot. Gard.* **80,** 528–580.
20. Rice, K. A., Donoghue, M. J. & Olmstead, R. G. (1996) *Syst. Biol.* **46,** 554–563.
21. Doyle, J. A., Donoghue, M. J. & Zimmer, E. A. (1994) *Ann. Mo. Bot. Gard.* **81,** 419–450.
22. Doyle, J. A. (1998) *Annu. Rev. Ecol. Syst.* **29,** 567–599.
23. Donoghue, M. J. & Mathews, S. (1998) *Mol. Phylogenet. Evol.* **9,** 489–500.
24. Thompson, W. P. (1918) *Bot. Gaz.* **65,** 83–90.
25. Carlquist, S. (1996) *Int. J. Plant Sci.* **157,** S58–S76.
26. Theissen, G., Kim, J. T. & Saedler, H. (1996) *J. Mol. Evol.* **43,** 484–516.
27. Colombo, L., Franken, J., Van Went, J., Angenent, H. J. M. & VanTunen, A. J. (1995) *Plant Cell* **7,** 1859–1868.
28. Kramer, F. R., Dorit, R. L. & Irish, V. F. (1998) *Genetics* **149,** 765–783.
29. Huang, H., Tudor, M., Su, T., Zhang, Y., Hu, Y. & Ma, H. (1996) *Plant Cell* **8,** 81–94.