Letter to the Editor

Gnetum and the Angiosperms: Molecular Evidence that Their Shared Morphological Characters Are Convergent, Rather than Homologous

Andrea Hansen,* Sabine Hansmann,* Tahir Samigullin,† Andrey Antonov,† and William Martin*

*Institut für Genetik, Technische Universität Braunschweig, Braunschweig, Germany; and †A. N. Belozersky Institute of Physicochemical Biology, Moscow State University, Moscow, Russia

The order Gnetales designates a small (in terms of species) group of seed plants that encompasses three contemporary genera: *Gnetum, Welwitschia,* and *Ephedra.* Because members of the Gnetales possess several angiospermlike attributes such as leaf venation, vessels as conductive elements, and double-fertilization-like processes, they have long figured prominently in the yet unresolved issue concerning the evolutionary origins of angiosperms and their flowers (Arber and Parkin 1907; von Wettstein 1907; discussed in detail in Nixon et al. [1994] and Doyle [1996, 1998]). Specifically, due to these and other morphological and anatomical similarities, most current hypotheses for the origin of angiosperms (flowering plants) firmly posit that angiosperms share a more recent common ancestor with members of the Gnetales than they do with any other group of contemporary seed plants (Doyle and Donoghue 1986; Friedman 1990, 1992; Crane, Friis, and Pedersen 1995; Hickey and Taylor 1996). According to this view, some or all of these characters in angiosperms and Gnetales are homologous. Accordingly, a specific yet informal taxonomic designation conceptually unites these two groups of seed plants in order to underscore their suspected relatedness—''anthophytes.'' The hypothesis linking them is known as the ''anthophyte theory'' (Crane, Friis, and Pedersen 1995).

It is a straightforward prediction of the anthophyte theory that molecular sequence data should support a common branch associating members of the Gnetales with angiosperms to the exclusion of any contemporary nongnetalean gymnosperm. It was therefore surprising that in two recent studies encompassing representatives from all three gnetalean genera and from all contemporary gymnosperm groups, the Gnetales showed a tendency to cluster with conifers rather than with angiosperms (Goremykin et al. 1996; Chaw et al. 1997). The support for this position was very weak in analyses of chloroplast internal transcribed spacer (cpITS) noncoding regions from chloroplast DNA (Goremykin et al. 1996), but it was somewhat stronger in analyses of nuclear-encoded 18S rDNA sequences (Chaw et al. 1997). Such findings are quite inconsistent with the expectations crisply formulated from the anthophyte theory (Crane, Friis, and Pedersen 1995) and therefore provide

Abbreviation: BP, boostrap proportion.

Key words: *Gnetum gnemon,* angiosperm evolution, gymnosperms, anthophytes, chloroplast.

justification for reinspecting the merits of the hypothesis more critically, as discussed at length by Doyle (1998) and as succinctly pointed out by Crepet (1998).

We further investigated the suspected relatedness of Gnetales to angiosperms by cloning and sequencing a 9.5-kb portion of the chloroplast DNA from *Gnetum gnemon* and by amplifying and sequencing 990 bp from the coding region of the *rpoC1* gene from *Gnetum* cpDNA. The 9.5-kb segment of *Gnetum* cpDNA was isolated by standard cloning and hybridization techniques (Sambrook, Fritsch, and Maniatis 1989) from a library that we prepared in lambda Dash (Stratagene) from *Mbo*I partially digested *Gnetum* DNA that had previously been enriched for chloroplast DNA using CsCl gradients (Sambrook, Fritsch, and Maniatis 1989). The hybridization probe was radioactively labeled pea chloroplast DNA. Overlapping subclones of the chloroplast region were isolated in p-Bluescript vectors (Stratagene) and sequenced on both strands using exonuclease III deletions (Sambrook, Fritsch, and Maniatis 1989). The sequence was deposited in GenBank (accession number AJ007508). The sequences from the *Gnetum rpoC1* gene were amplified (35 cycles each) with the following nested primers (all 5′–3′): 4/5F (TAYCARATGGGITAYATHAARYT), 4/5R1 (CCYTCYTTICCYTCDATIACRTC), 4/5F2 (CCIGTIRYICAYGTTGGTAY), 4/5R3 (GIARIARRCAIARIACCATCCA), 6F2 (GGIAARMGIGTIGAYTAY), 6R2 (ICCYTGDATIGCYTCYTC), 6F (CIGAYTTYGAYGGIGAYCARATG), and 6R1 (ICCIRVIGTIGTICDDATRTA). The PCR conditions were as follows: region 1—first PCR 4/5F-4/5R1, 3 mM MgCl$_2$, 48°C; second PCR 4/5F2-4/5R3, 2 mM MgCl$_2$, 53°C; region 2— first PCR 6F2-6R2, 2.5 mM MgCl$_2$, 50°C; second PCR 6F-6R1, 3 mM MgCl$_2$, 50°C. The region amplified from the *rpoC1* gene encompassed 990 nt of protein-coding region. The amplification products were subcloned in pBluescript plasmids, and three independent clones each were sequenced.

Chloroplast DNA is extremely AT-rich in noncoding regions and at synonymous sites (Lockhart et al. 1994). Furthermore, synonymous sites in protein-coding regions from the large single-copy region of chloroplast DNA are saturated with substitutions, or very nearly so, in comparisons across angiosperms and gymnosperms (Goremykin et al. 1996). Therefore, we excluded all third codon positions from all analyses here in an effort to reduce that bias. This left 660 first and second codon positions from the *rpoC1* gene, 162 from the *psaC* gene, and 652 from the *ccsA* gene. These data were combined with the sequences encoding the five tRNA genes shown in figure 1*a,* the chloroplast 23S rRNA, 439 bases determined from the chloroplast 16S rRNA gene, the 5S rRNA, the 4.5S rRNA, and the reasonably well-con-
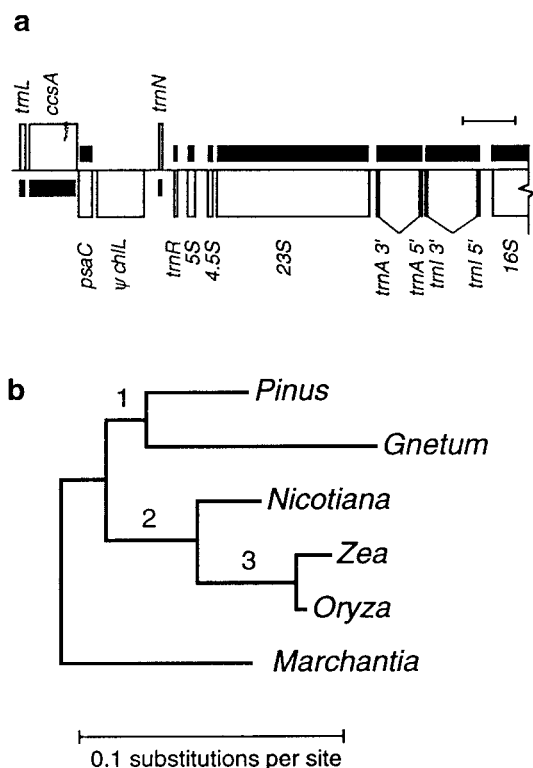
a

b

0.1 substitutions per site

Fig. 1.—Organization and phylogeny of a segment of *Gnetum gnemon* chloroplast DNA. *a,* A segment of chloroplast DNA cloned from *G. gnemon.* Open boxes indicate locations of genes, and solid bars indicate sequences used for phylogenetic analysis. Introns in the *trnA* and *trnI* genes are indicated. Scale bar: 1 kb. Homologs from chloroplast DNA were taken from the corresponding sequenced chloroplast genomes (reviewed in Stoebe, Martin, and Kowallik 1998). In the case of 18S rDNA, the available sequences from *Gnetum leyboldii* (L24045), *Pinus elliottii* (D38245), tomato (X51576), and the marchantiopsid *Lophocolea heterophylla* (X89872) were used for lack of homologs from *G. gnemon, Pinus thunbergii, Nicotiana,* and *Marchantia,* respectively. *b,* The result obtained with DNAML of PHYLIP for the concatenated 9,149-site data. Branches 1–3, discussed in the text, are indicated. Bootstrap proportions are 0.94 for branch 1 in the 9,149-site data and 1.00 for branches 2 and 3 (see also table 1).

Table 1
Bootstrap Proportions for Branch 1 with the Complete 9,149-Site Data and with Increasingly Less Polymorphic Subsets Thereof

	No. of Sites Analyzed (max. no. of states at sites analyzed)				
	9,149 (6)	8,962 (5)	8,796 (4)	8,573 (3)	8,020 (2)
ML[a]	94	100	99	99	100
MP[b]	100	100	100	96	95
NJ					
TN,[c] $a = 0.5$. . .	90	97	94	99	99
TN, $a = 1$	93	98	95	99	100
TN, $a = 2$	95	99	98	99	100
TN, $a = 4$	94	98	94	99	100
Kimura[d]	90	96	93	99	99
MLD[e]	90	96	94	99	99

[a] Maximum likelihood (DNAML).

[b] Parsimony (DNAPARS).

[c] Neighbor joining, Tajima-Nei distance for corresponding values of *a.*

[d] Neighbor joining, Kimura distance.

[e] Neighbor joining, ML distance from PHYLIP.

served *trnI* and *trnA* introns shown (fig. 1*a*). Together, this makes 7,314 nucleotide sites for analysis from the chloroplast DNA of *G. gnemon.* To these data, we added 1,835 sites from previously published nuclear 18S rDNA sequences (Chaw et al. 1997). The homologous regions from the conifer *Pinus thunbergii,* three angiosperms, and a liverwort as the outgroup were aligned using Pileup of the GCG Package (Genetics Computer Group 1994). The final alignment contains 9,149 sites, of which 6,362 are invariant and 1,167 are autapomorphic. The data are available via anonymous ftp from 134.169.70.80/usr/local/ftp/pub/incoming/Gnetum.

Neighbor-joining (Saitou and Nei 1987) using the Tajima-Nei (Tajima and Nei 1984), Kimura (1980), and maximum-likelihood (Felsenstein 1993) distance estimates, in addition to maximum-likelihood (using empirical base frequencies and a transition/transversion ratio of 2) and parsimony analyses, were performed with the PHYLIP package (Felsenstein 1993). With these methods, the *Gnetum* sequences do not branch with the angiosperm homologs sampled, but rather with their ho-

mologs from the conifer *Pinus* (fig. 1*b*). Branches 2 and 3 (fig. 1*b*) had bootstrap proportions (BPs) of 100/100 in all analyses here. Branch 1, which unites *Gnetum* with *Pinus,* was found to have a BP > 0.9 in all analyses (table 1). In all cases except for the Tajima-Nei distance estimates with $a = 2$ or $a = 4$, the alternative trees placed *Pinus,* not *Gnetum,* as a sister to angiosperms.

Although we excluded third codon positions of protein-coding regions from analysis for these data, there was still the possibility that other variable or poorly alignable sites may have biased this result. To check this, we systematically excluded highly variable and highly gapped sites from the analysis. The variability of sites was scored by simply counting the number of different nucleotide states that occurred at a given site in the alignment, whereby a gap is counted as being different from every other base at the site, including other gaps (Goremykin, Hansmann, and Martin 1997). For the six-species case, a maximum of six states can occur at a given site. Excluding six-state, five-state, four-state, and three-state sites from analysis did not alter the result (table 1). The computer program SORTAL, written in C++ for UNIX environments, which counts the states per site and sorts the data by positional variability for convenient exclusion of highly variable, hence potentially biased, sites, is available from S.H. or W.M. on request. The sorted alignment is also available from the ftp site.

We also examined the presumably conservatively evolving informative sites to see whether they were consistent with the computer analyses. Among the 450 non-autapomorphic sites at which only two nucleotide states occur in these data, 124 support branch 3 in figure 1*b,* 124 support branch 2, and 69 support branch 1. Thus, 70% of the informative sites are consistent with figure 1*b.* The remaining 30% of the informative sites are not consistent with the topology in figure 1*b.* Of the 450 informative sites, only 28 suggest a common branch for

Gnetum and angiosperms. In contrast, 69 informative sites group *Gnetum* with the conifer.

Criticism may arise because only one representative of the Gnetales was analyzed here. However, numerous phylogenetic analyses have been conducted in the past years that include *Gnetum, Ephedra,* and *Welwitschia,* both using morphological characters and using molecular data (reviewed by Doyle 1996, 1998). The overwhelming majority of those studies strongly indicate the Gnetales to be a monophylum, a premise that we accept here as being true, such that one member of the group is sufficient for our test, which is thus contingent upon that premise. Criticism may also arise because we did not include *Ginkgo,* cycads, more closely related pteridophyte outgroups, or putatively more primitive angiosperms in our analysis. Obviously, it will be important to obtain sequence data from these groups in future studies. However, by utilizing the vast information contained within sequenced plastid genomes (Martin et al. 1998), it is possible to investigate the question of whether molecular data reflect closer affinities between *Gnetum* and angiosperms, or between *Gnetum* and conifers, since the reference genomes necessary for that test, including that of a bryophyte outgroup, are already available for analysis.

Thus, notwithstanding the need to analyze more data and species and notwithstanding the need to use other methods of analysis, these findings rather straightforwardly suggest that Gnetales, represented here by *Gnetum,* are not sisters of angiosperms. From this, it would follow that the anthophyte theory is incorrect. From this, it would follow, as the simplest explanation, that the characters shared by *Gnetum* and angiosperms, such as double-fertilization-like processes (Friedman 1992, 1992; Crane, Friis, and Pedersen 1995), leaf venation patterns, flowerlike reproductive structures (Crane, Friis, and Pedersen 1995; Hickey and Taylor 1996), and vessels (Carlquist 1996) arose independently in these lineages. Patterns of character evolution among seed plants are notoriously complex (Crepet 1998). Indeed, the current data suggest that homologies between reproductive organs of angiosperms and gnetophytes that have been suspected under the anthophyte theory and have been claimed to receive strong independent support from molecular data (Crane, Friis, and Pedersen 1995) are apparently not as sound as previously asserted and that these assumptions deserve continued critical reinspection (Doyle 1998; Crepet 1998).

Finally, these findings are consistent with the arguments of Carlquist (1996), who argued quite resolutely that despite the presence of vessels in gnetalean wood, its overall anatomy shares many more features with the wood of conifers than it does with the wood of angiosperms. Therefore, in light of these results, and in light of previous hypotheses linking Gnetales to conifers (discussed in Doyle 1996), it is possible that wood anatomical characters might hold particular promise for linking contemporary seed plant groups with fossil forms. If so, molecular data might be able to provide interpretative aids for the weighting of characters preserved in fossil material for the purpose of making sense of seed plant phylogeny.

LITERATURE CITED

ARBER, E. A. N., and J. PARKIN. 1907. On the origin of angiosperms. J. Linn. Soc. **38**:29–80.

CARLQUIST, S. 1996. Wood anatomy of primitive angiosperms: new perspectives and syntheses. Pp. 68–90 *in* L. J. HICKEY and D. W. TAYLOR, eds. Flowering plant origin, evolution and phylogeny. Chapman and Hall, New York.

CHAW, S.-M., A. ZHARKIKH, H.-M. SUNG, T.-C. LAUM, and W.-H. LI. 1997. Molecular phylogeny of gymnosperms and seed plant evolution: analysis of 18S rRNA sequences. Mol. Biol. Evol. **14**:56–68.

CRANE, P. R., E. M. FRIIS, and K. R. PEDERSEN. 1995. The origin and early diversification of angiosperms. Nature **374**: 27–33.

CREPET, W. L. 1998. The abominable mystery. Science **282**: 1653–1654.

DOYLE, J. A. 1996. Seed plant phylogeny and the relationships of Gnetales. Int. J. Plant Sci. **157**(Suppl.):3–39.

———. 1998. Molecules, morphology, fossils and the relationship of angiosperms and Gnetales. Mol. Phylogenet. Evol. **9**:448–462.

DOYLE, J. A., and M. J. DONOGHUE. 1986. Seed plant phylogeny and the origin of the angiosperms: an experimental cladistic approach. Bot. Rev. **52**:321–431.

FELSENSTEIN, J. 1993. PHYLIP (phylogeny inference package). Version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.

FRIEDMAN, W. 1990. Double fertilization in *Ephedra,* a non-flowering seed plant: its bearing on the origin of angiosperms. Science **247**:951–954.

———. 1992. Evidence for a pre-angiosperm origin of endosperm: implications for the evolution of flowering plants. Science **255**:336–339.

GENETICS COMPUTER GROUP. 1994. Program manual for version 8. Madison, Wis.

GOREMYKIN, V. V., V. K. BOBROVA, J. PAHNKE, A. V. TROITSKY, A. S. ANTONOV, and W. MARTIN. 1996. Noncoding sequences from the slowly evolving chloroplast inverted repeat in addition to *rbc*L data do not support gnetalean affinities of angiosperms. Mol. Biol. Evol. **13**:383–396.

GOREMYKIN, V. V., S. HANSMANN, and W. MARTIN. 1997. 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.

HICKEY, L. J., and D. W. TAYLOR. 1996. The origin of the angiosperm flower. Pp. 176–231 *in* L. J. HICKEY and D. W. TAYLOR, eds. Flowering plant origin, evolution and phylogeny. Chapman and Hall, New York.

KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

LOCKHART, P. J., M. A. STEEL, M. D. HENDY, and D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. **11**:605–612.

MARTIN, W., B. STOEBE, V. GOREMYKIN, S. HANSMANN, M. HASEGAWA, and K. V. KOWALLIK. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature **393**: 162–165.

NIXON, K. C., W. L. CREPET, D. STEVENSON, and E. M. FRIIS. 1994. A reevaluation of seed plant phylogeny. Ann. Mo. Bot. Gard. **81**:484–533.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SAMBROOK, J., E. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor, New York.

STOEBE, B., W. MARTIN, and K. V. KOWALLIK. 1998. Distribution and nomenclature of protein-coding genes in 12 se-quenced chloroplast genomes. Plant Mol. Biol. Reptr. **16**: 243–255.

TAJIMA, F., and M. NEI. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. **1**: 269–285.

VON WETTSTEIN, R. 1907. Handbuch der Systematischen Botanik. Volume 2. Franz Deuticke, Vienna.

GEOFFREY MCFADDEN, reviewing editor