J Mol Evol (1999) 49:310–315

JOURNAL OF MOLECULAR EVOLUTION

Molecular Data from the Chloroplast *rpo*C1 Gene Suggest a Deep and Distinct Dichotomy of Contemporary Spermatophytes into Two Monophyla: Gymnosperms (Including Gnetales) and Angiosperms

Tagir Kh. Samigullin,[1] William F. Martin,[2] Aleksey V. Troitsky,[1] Andrey S. Antonov[1]

[1] A.N. Belozersky Institute of Physicochemical Biology, Moscow State University, Moscow 119899, Russia
[2] Institut fuer Genetik, Technische Universitaet Braunschweig, Spielmannstrasse 7, D-38023 Braunschweig, Germany

Received: 20 November 1998 / Accepted: 26 April 1999

Abstract. Partial sequences of the *rpo*C1 gene from two species of angiosperms and three species of gymnosperms (8330 base pairs) were determined and compared. The data obtained support the hypothesis that angiosperms and gymnosperms are monophyletic and none of the recent groups of the latter is sister to angiosperms.

Key words: *Gnetum* — Chloroplast DNA — RNA polymerase — Molecular phylogeny

Introduction

Study of genes from contemporary gymnosperms and their homologues in angiosperms may shed light on the problem of the origin of flowering plants. In this context, the interrelatedness of Gnetales and angiosperms is of particular importance, the more so because recent new data and cladistic analysis of morphological data of extant and extinct species indicate a very close relationship of Gnetopsida to angiosperms (see Crane 1985; Doyle and Donoghue 1987; Doyle 1996, 1998). Though many morphological and molecular data sets tend to suggest that Gnteales are the sisters of angiosperms (reviewed in Doyle, 1998), molecular data derived from the analysis of small and large nuclear and chloroplast rRNAs (Hori et al. 1985; Rakhimova et al. 1989; Troitsky et al. 1991; Shi et al. 1994; Chaw et al. 1997), the internal transcribed spacers cpITS2 and cpITS3 of the chloroplast ribosomal operon (Goremykin et al. 1996), and the mitochondrial gene *cox*I (Bowe and DePamphilis 1997) support, albeit weakly, the monophyly of gymnosperms. New discoveries revealing highly advanced floral types in 90 million-year-old deposits (Crepet 1998) and the recent report of Yurassic angiosperm fossils (Sun et al. 1998) challenge a close gnetalean–angiosperm relationship even more.

When molecular data are used for inferring phylogenetic relationships, different genes may yield different phylogenies, but the reasons for this are not always clear (e.g., Hedges 1994). Clearly, studies of additional genome segments are desirable to reconcile the evolutionary behavior of molecular and morphological characters. Comparison of sequences of 58 proteins encoded in chloroplast genomes of pine, rice, maize, tobacco, marchantia, and a red algal outgroup demonstrated that, using simple methods of inference, only 40 of them permitted obtaining the true, biologically reasonable phylogenies of those species (Goremykin et al. 1997). One such protein is the b′-subunit of chloroplast DNA-dependent RNA polymerase encoded by the chloroplast *rpo*C1 gene. Intron sequences of this gene have already been employed in molecular phylogenetic studies of some groups of angiosperms (Downie et al. 1996), but with limited success, because the gene seems to harbor somewhat too little polymorphism to address phylogenetic questions at lower taxonomic levels. Notably, the pine *rpo*C1 gene contains indels of varying length which are

Correspondence to: T. Samigullin; *e-mail:* samigul@genebee.msu.su

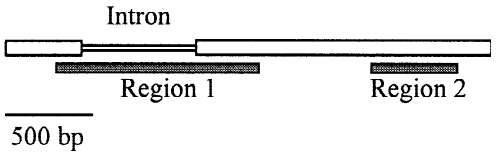

Fig. 1. Location of amplified regions of the *rpo*C1 gene. Sequenced regions are shown as *shaded boxes;* the scale bar indicates 500 base pairs.

absent in the corresponding sequences of angiosperms. Hence, it was of interest to look for these indels in *Gnetum* and other gymnosperm *rpo*C1 genes.

The present paper contains data on nucleotide sequences of two fragments of the *rpo*C1 gene, 1100–1200 and 450–500 base pairs (bp) long, of three species from different classes of extant gymnosperms—Gnetopsida (*Gnetum gnemon*), Ginkgoopsida (*Ginkgo biloba*), and Cycadopsida (*Cycas revoluta* and *Zamia floridiana*)—as well as of two angiosperm species (*Arabidopsis thaliana* and *Allium cepa*). Data of phylogenetic analysis of these sequences together with those of *rpo*C1 gene sequences of four other species of angiosperms and a pine (Pinopsida) suggest that no contemporary gymnosperm is a sister to angiosperms, contrary to predictions (Crane et al. 1995) of anthophyte hypothesis.

Methods

The regions amplified are shown in Fig. 1. Plant DNA was isolated from either fresh or lyophilized leaf tissue by the CTAB method (Murray and Thomson 1980). Two *rpo*C1 regions were amplified using nested PCR; i.e., products of the first PCR were diluted and used as templates in the second PCR with internal primers. The primer sequences (where "I" is inosin) were as follows: 4/5F, 5′-TA(CT)CA(AG)ATGGGITA(CT)AT(ACT)AA(AG)(CT)T-3′; 4/5F2, 5′-CCIGTI(AG)(CT)ICA(CT)GTTGGTA(CT)-3′; 4/5R1, 5′-CC(CT)TC(CT)TTIC C(CT)TC(AGT)ATIAC(AG)TC-3′; 4/5R2, 5′-CAIA(AG)IACCATC CA(CT)T(CT)IGG(CT)TC-3′; 4/5R3; 5′-GIA(AG)IA(AG)(AG)CA IA(AG)IACCATCCA-3′; 6F, 5′-CIGA(CT)TT(CT)GA(CT) GGIGA(CT)CA(AG)ATG-3′; 6F2, 5′-GGIAA(AG)(AC)GIGTIGA(CT) TA(CT)-3′; 6F3, 5′-GGIAT(ACT)CA(AG)GCITT(CT)CA(AG)-3′; 6R1, 5′-ICCI(AG)(AGC)IGTIGTIC(AGT)(AGT)AT(AG)TA-3′; 6R2, 5′-ICC(CT)TG(AGT)ATIGC(CT)TC(CT)TC-3′; intronF, 5′-AGTCTA (GT)(CT)(CT)A(GT)(CT)GCATATA-3′; intronR, 5′-(AT)(CT)TG (AGT)ATITC(AG)TA(CT)T(CT)(AG)AA-3′; and intronR1, 5′-GACA AGAATTTCCATCCA-3′.

Amplifications were performed in 10 m*M* Tris–HCl (pH 8.3), 50 m*M* KCl, 2–3 m*M* $MgCl_2$, a 200 μ*M* concentration of each dNTP, a 2 μ*M* concentration of each primer, containing 0.5 U of Taq polymerase (Perkin Elmer) for 30 cycles of a 50-s denaturation step (94°C), a 40-s annealing step (40–53°C), and a 1-min elongation step (72°C). Primer combinations and exact PCR conditions are available from authors upon request.

PCR products were purified by diafiltration in Microcon 100 columns (Amicon) according to the manufacturer's protocol, cloned using pBluescript KS+ phagemid (Stratagene) as described by Sambrook et al. (1989), and sequenced by the dideoxy method with bacteriophage T7 DNA polymerase (Tabor and Richardson 1987). The *Gnetum rpo*C1 intron was sequenced using primers IntronF and IntronR1 from the cloned region 1 directly, without subcloning. All regions were sequenced from two independent clones; in cases of ambiguity a third clone was sequenced.

The nucleotide sequences have been deposited with the GenBank Data Library under accession numbers AJ012558–AJ012567. Nucleotide sequences of *Marchantia polymorpha, Pinus thunbergii, Oryza sativa, Zea mays, Nicotiana tabacum,* and *Spinacia oleracea* were retrieved from GenBank (accession numbers X04465, D17510, X15901, X86563, S54304, X08671, respectively).

Amino acid sequences were aligned using the MULTALIN (version 5.3.3) program (Corpet 1988). Nucleotide sequences were aligned using the PileUp program of the GCG package (Devereux et al. 1984), version 9.1, and accomplished by eye using the VOSTORG package (Zharkikh et al. 1990). Ambiguously aligned positions were subsequently eliminated from the phylogenetic analysis. Programs from the PHYLIP (Felsenstein 1989) package, Puzzle (Strimmer and von Haeseler 1996), and the PaupSearch program of the GCG package were used for tree reconstruction. *Marchantia* was used as an outgroup.

Results and Discussion

Nucletide sequences of two fragments of the *rpo*C1 gene of five species of plants were determined and combined with analogous data on the homologous gene sequences of six species of angiosperms and gymnosperms from GenBank. Figure 1 illustrates the position of the sequenced fragments in the gene. The first fragment (1160–1250 bp long) is composed of a coding region and an intron. The length of the second fragment is 470–500 bp. We were unable to amplify the second fragment of the *Zamia* gene and replaced it in analysis with a homologous fragment of phylogenetically closely related *Cycas* chloroplast DNA.

Search of Synapomorphic Indels

Indels in protein-coding sequences are a less frequent evolutionary phenomenon than noncoding DNA regions. Hence, they may carry valuable phylogenetic information. To analyze the distribution of indels, the aligned nucleotide sequences of the *rpo*C1 gene were transformed into amino acid sequences. The resulting amino acid alignments are presented in Fig. 2. As follows from Fig. 2A, insertions of seven (positions 65–71) and nine (positions 112–120) amino acid residues are apomorphic for pine.

The alignment in positions 65–80 of the second fragment (Fig. 2B) was ambiguous due to the presence of short repeats, and the alignment pattern changed drastically depending on the value of the gap penalty. Positions 116 and 125–126 are of particular interest. They are flanked by relatively conservative sequences, excluding alternative versions of alignment of this region. A deletion in positions 125–126 is synapomorphic for angiosperms. An insertion in position 116 is common to *Pinus* and *Gnetum.*

A

```
            1          11         21         31         41         51         61         71
   ALLIUM   LKRLPSYIAN LLDKPLKELE GLVYCDFSFA RPIAKKPTFL RLRGLFEYEI QSWKYSIPLF FTTQ...... .GFETFRNRE
ARABIDOPSIS LKRLPSYIAN LLDKPLKELE GLVYCDFSFA RPITKKPTFL RLRGSFEYEI QSWKYSIPLF FTTQ...... .GFDIFRNRE
  NICOTIANA LKRLPSYIAN LLDKPLKELE GLVYCDFSFA RPITKKPTFL RLRGLFEYEI QSWKYSIPLF FTTQ...... .GFDTFRNRE
   SPINACIA LKRLPSYIAN FLDKPLKELE GLVYCDFSFA RPIAKKPTFL RLRGLFEYEI QSWKYSIPLF FTTQ...... .GFDTFRNRE
     ORYZA  LKGLPSYIAN LLDKPLKKLE GLVYGDFSFA RPSAKKPTFL RLRGLFEDEI SSCNHSISPF FSTP...... .GFTTFRNRE
       ZEA  LKGLPSYIAN LLDKPLKKLE GLVYGDFSFA RPSAKKPTFL RLRGLFEDEI SSCNHSISPF FSTP...... .GFATFRNRE
 CYCADALES  LKRLPSYIAN LLAKPLKESE GLVYCDLFLA RPIANKPTSL XXXGLFKYEI QSWRDIIPNY FSAR...... .GFEAFRRRE
    GINKGO  SKRLPSYIAN LLAKPLKELE GPVYCDLFIA RPIANKPTSL RSRGTFKYDI QSWGDILPHY LSAQ...... .GFGAFQNRE
     PINUS  LKRLPSYIAN LLAKTRKELE GPVYCDLFLA RPIANKPTLL RSRGTFDYEI QSWREIIPHY LSARPYYLFP RGSGTFKERE
    GNETUM  IKRVPSYIAT LIGKQNSEIK DLVYCNLFLA RPAANKPTIL RFRGLLQHGI TSWMEILVPY ISGW...... .NFVEFQGRE
 MARCHANTIA LKRLPSYIAN LLAKPLKELE SLVYCDLFLA RPITKKPTLL KLQGLFKYED QSWKDIFPRF FSPR...... .GFEVFQNRE

            81         91         101        111        121        131        141        151
   ALLIUM   ISTGAGAIRE QLADLDLRII IENSLVEWKE L......... GDEESAENEW EDRKIRRRKD FLVRRMELAK HFIRTNV
ARABIDOPSIS ISTGAGAIRE QLADLDLRII IENSLVEWKQ L......... GEEGPTGNEW EDRKIVRRKD FLVRRMELAK HFIRTNI
  NICOTIANA ISTGAGAIRE QLADLDLRII IENSLVEWEE L......... GEEGHTGNEW EDRKVGRRKD FLVRRVELAK HFIRTNI
   SPINACIA ISTGAGAIRE QLADLDLRTI IDYSFAEWKE L......... GEEGSTGNEW EDRKVGRRKD FLVRRMELVK HFIRTNI
     ORYZA  IATGAGAIRE QLADLDLRII LENSSVEWKE L......... EDEGYSGDEW EDRKRRIRKV FLIRRMQLAK HFIQTNV
       ZEA  IATGAGAIRE QLADLDLRII IENSLVEWKE L......... EDEGYSGDEW EDRKRRIRKV FLIRRMQLAK HFIQTNV
 CYCADALES  IATGGDAIRE QLTGLDLQTL MNRSYMEWKR L......... GKHKSTGNGW GDRKIKRRKD FSVRRMRLAK HFLQTDI
    GINKGO  IATGGDAIRE QLAGPDLRIL MANSYMEWKI L......... EEQKSTGNEW EDEKIQRRKD FSVRRMELAK HFIQTNI
     PINUS  IATGGDAIGK QLMGLDLQMI IDRSHMEWKN LVELKWNRLE ENQESTVDRW EDEKIRRRKD FLVGRMKLAK HFLRTNI
    GNETUM  LATGGTSIQK QLIGLNLRAL LNHSYMEWRK L......... ..LKNHRIQK RKKKIEKRKN FLVKRIKFAK YLIQAKI
 MARCHANTIA IATGGDAIQK QLTNLNLQNV INLAHLEWKE F......... AEQKSTGNEW EDRKIQRRKD LLVRRIKLAK HFIQTNI
```

B

```
            1          11         21         31         41         51         61         71
   ALLIUM   AVHVPLSLEA QAXARLLMFS HMNLLSPAIG DPISVPSQDM LIGLYVLTMG NRRGICENRY NPYNCANYQN K.TVDNNNYE
ARABIDOPSIS AVHVPLSLEA QAEARLLMFS HMNLLSPAIG DPISVPTQDM LIGLYVLTSG TRRGICANRY NPCNRKNYQN E.RIYETNYK
  NICOTIANA AVHVPLSLEA QVEARLLMFS HMNLLSPAIG DPISVPTQDM LIGLYVLTSG NHRGICVNRY NPCNRRNYQN QKRSDNSHYK
   SPINACIA AVHVPLSLEA QAEARLLMFS HMNLLSPAIG DPISVPTQDM LIGLYILTSG NRRGICANRY NPWNHKTYQN E.RIDDTNYK
     ORYZA  AVHLPLSLEA QAEARLLMFS HMNLLSPAIG DPICVPTQDM LIGLYVLTIG NRRGICANRY NSCGNYPNQK VNYNNNNP.K
       ZEA  AVHLPLSLEA QAEARLLMFS HMNLLSPAIG DPICVPTQDM LIGLYVLTIG NRLGICANRY NSCGNSPNKK VNYNNNNYYK
 CYCADALES  AVHVPLSLEA QAEARLLMFS HTNLLSPAIG DPISVPTQDM LLGLYILTVG NNQGIYGNRY HPYYS.KY.. .........N
    GINKGO  AVHVPLSLEA QAEARLLMFS HTNLLYPAIG DPISVPTQDM LLGLYILTVE NNQGIYGNRY HPYNS.NK.. .........K
     PINUS  AVHVPLSLEA RAEARLLMFS ETNLLSPAIG DPISIPTQDM LLGLYISTVQ NSQGIYGNRY HPYHS.EN.. .........K
    GNETUM  AVHLPLSIEA ILESRLLMFS HTNLLSPSNG SPITKPTQDM LLGLYILTTE KPRNISQFRC RPSNPTK... .........K
 MARCHANTIA AVHIPLSLEA QAEARLLMLS HKNLLSPATG EPISVPSQDM LLGLYILTIE NNQGIYGNKY NPSKKYDS.. .........K

            81         91         101        111        121        131        141        151
   ALLIUM   YTKEKEPYFG SSYNALGAYR QKRIKLDSPF WLRWR.LDQR VIGL..KEVP IEVQYESFGT YHEIYGHYLI VGSMKKEICCI
ARABIDOPSIS YTKE..PFFC NSYDAIGAYR QKKINLDSPL WLRWQ.LDQR VIAS..REVP IEVHYESFGN YHEIYAHYLI VRSVKKENFCI
  NICOTIANA YTKE..PFFS NSYDAIGAYR QKRINLDSPL WLRWR.LDQR VIAS..RETP IEVHYESLGT FYEIYGHYLI VRSLKKQILFI
   SPINACIA SMKE..PFFC NFYDAIGAYR QKRIHLDSPL WLRWQ.LDQR IIAS..KEAP IEVHYESLGT YHEIYAHYLI IRSVKKEIIDI
     ORYZA  YTKDKESLFS SSYDALGAYR QKQICLDSPL WLRWK.LDQR VIGL..REVP IEVQYESLGT YREIYAHYLV VGNRKKEIRSI
       ZEA  YTKDKEPHFS SSYDALGAYR QKRIGLNSPL WLRWK.LDQR IVGS..REVP IEVQYESFGT YHEIYAHYLV VGNRKKEIRSI
 CYCADALES  IFSCKKPSFY SYDDALGAHW QKRIELDSPL WLRWG.VGLR IITSVDREAP IEVQYESLGI FHEIYEHYRI GKNEVGEILSI
    GINKGO  IFYCKKLSFS SYDDALRAYR EKRIHLYSPL WLRWR.VDLR ITTSVNREAP IEVQYESLGT FREIHEHYRI IRSMMGEILSI
     PINUS  SFSCKKPSFY SYDDVLRAYR QKRIDLYSPL WLRWGEVDLR IITSVNQEAP IEVQYESLGT FHEIHEHYRI RKGRMGEILNI
    GNETUM  FLPEVNLCFC NYDDVFIAYQ KNRVSLKNSL WFRWNVVNGT ILTSVDQEVP IEFQYQSLGT SQQIYEHYTI QRARSGKVLTI
 MARCHANTIA KKFSQIPYFS SYDNVFRALQ QKQIYLHSSL WLRWQ.INLR IITLLNQEGP IEIQYKSFGN SFQIYEHYQL RKNKNQEIIST
```

Fig. 2. Amino acid alignments of *rpo*C1 region 1 **(A)** and region 2 **(B),** created by MULTALIN using symbol comparison table "blosum62," a gap weight of 12, and a gap length weight of 2.

Alignment of introns was dubious due to a high sequence divergence, although some short intron regions are quite conservative (Fig. 3). We failed to find phylogenetically informative indels both in the variable and in the conservative regions of introns. Summing up, analysis of indels in the sequences compared did not discriminate unequivocally between alternative hypotheses about the evolutionary relatedness of Gnetales, other gymnosperms, and angiosperms.

Phylogenetic Tree Reconstruction

In phylogenetic tree reconstruction experiments, we concatenated the two fragments, and highly divergent regions were excluded from the analysis (alignment available upon request). Of 1654 bp used for analysis, 632 bp belong to the intron.

Figure 4 shows a 50% majority rule consensus neighbor-joining (NJ) tree based on Tamura (1992) distances,

A

```
ALLIUM       AACTGTCATCCCATTCAATCTGATTGGGATGCCCCCGACTCTGACATGG.TTCTTGGAAGGAGTAACATGGAGCTCAGA
ARABIDOPSIS  AACTCTCATCCCCTTCAATCAATTTGGGATGCCCTCGA.TCTGACATGT.CTCTTGAAAAGAGTAACATGAAGCTCAGA
NICOTIANA    AACTGTCATCCCAGTCAATCCAATTGGGATGCCCTGGAC.CTGACATGTGCACTTGGAAAAGAGTAACATGAAGCTCAGA
SPINACIA     AACTGTCATCCCATTCAATCCAAAGGGGATGCCCTGG.CTCTGACATGT.CTCTGGGGAGGAGAAACATGAAGCTCAGA
CYCADALES    AACTGTCATCCTATTCAATCTCATTGGGATGCCTTTAGCTCTGACATGT.CTCTCAGGAGAAATAGCATGAAGCTCAAA
GINKGO       AACTGTCATCCTATTTAATCTCATTGGGATGCCTTTAGCTCTGACATGT.CTTTCGGGATAAGTAACATGAAGCTCAGA
PINUS        AACTGTCATCCTATTAAATCTCATTGGGATGCCTTTAGCTCTGACATGA.CCCTCCAGAGGAGTAGCATGAAGCTCAGA
GNETUM       AACTGTCATCCTTTT.....TTTTTAGGATGTTCTTAAGTCTGACATGT.TAATCCTTATGAGTAGCATGAAGCTTAGA
MARCHANTIA   AACTGTTATTTTAATTTACTTTTAATAAATACCTCAAATTTTGACATTA.CTTAATAAAAATATGAAATGAAGCTCAAA
```

B

```
ALLIUM       AGAATTATGGGTGTATTCAATATTCCCAAAT.....GAAGGGGGAATTGATCCATGGTCGA....TTCTGTATTATTAGT
ARABIDOPSIS  AGAATTGTGGGTGTATTCAATACTCCCAAAAAAAAAAAAAAAGGCGATTGGTCTATGGTCGA....TTCAG..TTGATAGG
NICOTIANA    AGAATTG.GGGTGTAGTCAATACTCCCAAAT.....AAAAAGGGAATTGATCTATGGTCGA....TTTCGTATTTCTATT
SPINACIA     AGAATTATGGGTGTATTCAATACTCCCAAAT.....GAAAAGGGAATTGATCCATGGTCGA....TTTCGT.TTTCTAGT
CYCADALES    AAAATCACAAGCATCTTGAAGAGATGTTGA......GAAAGAGGAATTGGTCCAGGGTTGATATATTCCAT.TTGCTAAT
GINKGO       AGAATCATAAGCATCTTAGAGAGATGTTGA......GAAAGAGGAATTGGTCCAGGGTTGATATCTTTCAT.TTGCTAAT
PINUS        AGAATTATAAGCATCTTTTCAAGATGCTGA......GAAAGAGGAATTAGTCCAGGGTTTAGATATTATTT.TTGCTAAT
GNETUM       AGA.TTTCAAAAA.....AAAAAAAGTTGA......AAAAGAGATTTTAGTCTAAGGTTGAGATATTTCAT.CTGCTTTT
MARCHANTIA   AAAAAAAACAAT.TATCATATTATTTTTTGTATTTGTTTAGAGGATTTAATCTAGGGTT.............TTGCTATT
```

Fig. 3. Nucleotide alignments of variable **(A)** and concervative **(B)** fragments of the intron. *Zea* and *Oryza* are not shown because of their lack of the *rpo*C1 intron.

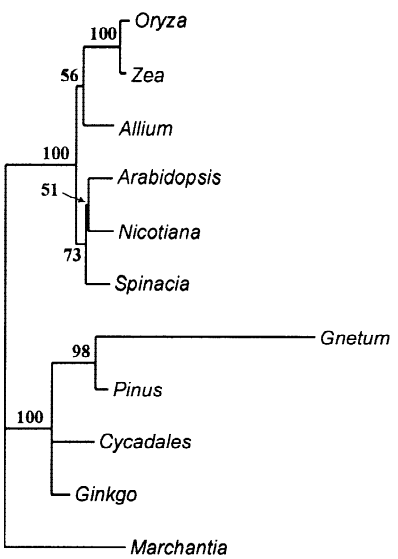

Fig. 4. The 50% majority-rule consensus neighbor-joining tree, based on nucleotide sequences of concatenated regions 1 and 2 of the *rpo*C1 gene. Tamura distances and 1000 bootstrap replicates were used.

which take into account possible variations in the rate of accumulation of substitution in different sites and allow deviations in GC content. In Fig. 4, gymnosperms and angiosperms appear as two monophyletic clades with strong support. This finding is at odds with the hypothesis of a sister relationship between angiosperms and Gnetales.

Among the angiosperms, monocots and dicots form two clusters, with bootstrap values of 56 and 73%, respectively. Among gymnosperms, *Gnetum* and *Pinus* branch robustly together (with a bootstrap value of 98%); the relationship among this clade, *Ginkgo*, and Cycadales is unresolved. It is noteworthy that the curiously close relationship between *Gnetum* and *Pinus* deduced from analysis of nucleotide substitution is in accord with the occurrence of a common insertion mentioned above.

The monophyly of gymnosperms is further supported using other tree reconstruction algorithms—the maximum-likelihood (ML) and maximum-parsimony (MP) methods (Figs. 5 and 6). In both trees angiosperms subdivide into monocot and dicot clades. The topology of the ML tree (Fig. 5) is almost-identical to that of the NJ tree. However, in this case the Cycadales, *Ginkgo*, and (*Gnetum*, *Pinus*) branches are well separated. One cannot help but note that the *Gnetum* branch is relatively long in both trees.

In the MP tree (Fig. 6), *Gnetum* occupies a basal position among gymnosperms and *Pinus* appears as a sister group to the Cycadales + *Ginkgo* clade. This MP tree is a single one with the smallest number of steps (1643). To evaluate the stability of its topology, we also analyzed 16 trees with a number of steps up to 1655. One of these trees, in which *Gnetum* joins *Pinus* and *Ginkgo* joins Cycadales, needs 8 additional steps. For the grouping of *Gnetum* with *Marchantia*, *Pinus* being the sister group to the *Ginkgo* + *Cycadales* clade, 9 additional steps are necessary (score, 1652). A tree in which Cycadales is a sister group to *Ginkgo* + *Pinus*, and *Gnetum* combines with *Marchantia*, is 11 steps longer (score, 1654). The three shortest suboptimal trees (one of 1654 and two of 1650 steps) had, in the main features, the same topology as the minimal tree (Fig. 5), differing only in the relative position of dicot species. It should be noted that, in none of the trees with a length of 1643–1655 steps, does *Gnetum* appear as a sister group to angiosperms.

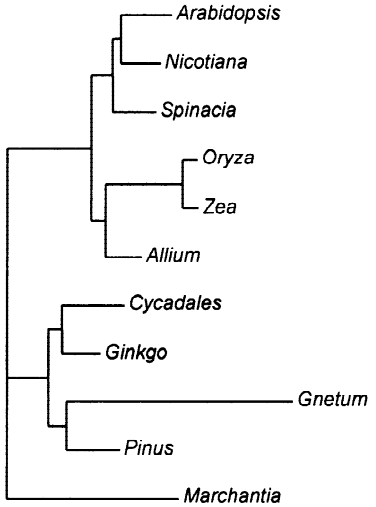

Fig. 5. Maximum-likelihood tree constructed from nucleotide sequences of concatenated regions 1 and 2 of the *rpo*C1 gene.

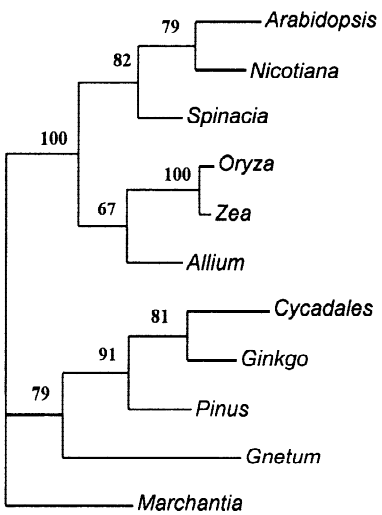

Fig. 6. Maximally parsimonious tree of 1643 steps constructed from nucleotide sequences of concatenated regions 1 and 2 of the *rpo*C1 gene. Bootstrap values of the 50% majority-rule consensus parsimonious tree, which has the same topology, are indicated. One hundred bootstrap replicates were used.

The basal position of *Gnetum* in the MP tree may be explained in the following way. The MP method gives reliable results only if the rate of accumulation of substitutions is relatively equal in all the branches of the tree (Felsenstein 1978), which is more likely for closely related taxa. However, the length of the *Gnetum* branch is substantially longer than other branches, indicating a higher rate of accumulation of nucleotide substitutions. It cannot be excluded that such a higher rate of evolution is characteristic not only for the *rpo*C1 gene but for the whole *Gnetum* chloroplast genome.

Conclusions

In summary, we were unable to find any evidence in the higher plant *rpo*C1 genes that *Gnetum* might be a sister group of angiosperms. Hence, the results obtained in this study do not support the anthophyte concept, which posits a closer relationship between angiosperms and Gnetales than to any contemporary gymnosperm group (Crane 1985; Donoghue and Doyle 1989; Doyle and Donoghue 1992; Loconte and Stevenson 1990; Hamby and Zimmer 1992; Chase et al. 1993; Doyle et al. 1994; Nixon et al. 1994; Albert et al. 1994; Crane et al. 1995; Doyle 1996, 1998).

One could contend that analysis of the same regions for other Gnetalean representatives might reveal other results. That could be, but we consider it very unlikely that this will occur, because previous molecular studies provide quite robust support for the view that the Gnetales is a monophylum.

Despite claims to the contrary (Crane et al. 1995), the anthophyte concept does not have truly firm support from any molecular data (Doyle 1998). Data on *rbc*L gene evolution are contradictory (Hasebe et al. 1992; Nixon et al. 1994) and must be analyzed with great care because the synonymous substitutions in this gene are saturated (Goremykin et al. 1996). The protein encoded by *rbc*L proved to be 1 of the 12 chloroplast proteins which could not be used to construct a "true" phylogenetic tree based on pine, rice, maize, tobacco, marchantia, and porphyra sequences (Goremykin et al. 1997). Hamby and Zimmer (1992) have analyzed 18S rRNA using *Equisetum* and *Psilotum* as outgroups and were unable to prove unequivocally the close relatedness of *Gnetum* and angiosperms: a variant of a tree with Gnetales as a sister group to all other seed plants is one step shorter than a tree in which Gnetales is a sister group of angiosperms. Doyle et al. (1994) did not include nonseed plant outgroups in their analysis of rDNA evolution; their MP trees with rather dissimilar topologies differed in only one step. At the same time, the results obtained in the present study are in accord with the results of analysis of small and large nuclear and chloroplast rRNAs (Hori et al. 1985; Rakhimova et al. 1989; Troitsky et al. 1991), the 5′-terminal region of the 25S rRNA gene (Shi et al. 1994), cpITS2 and cpITS3 of the chloroplast ribosomal operon (Goremykin et al. 1996), complete sequences of nuclear 18S rDNA (Chaw et al. 1997), and the mitochondrial gene *cox*I (Bowe and DePamphilis 1997).

Unfortunately, due to some misunderstanding in review articles by Crane et al. (1995) and Sytsma and Hahn (1994), papers by Chase et al. (1993) and Doyle et al. (1994) are cited as molecular evidence of Gnetales being a sister group of angiosperms. In fact, the trees inferred from *rbc*L data (Chase et al. 1993) are unrooted and the root between *Gnetum* and other Gymnospermae was introduced by authors based on the paper by Doyle et al. (1994). However, in that paper, the phylogenetic tree based on rRNA sequences is also unrooted.

Gnetales as "living fossils" exhibit a unique mosaic of angiospermous and gymnospermous characters. Recent intriguing paleontological discoveries of extant angio-

sperms (Crepet and Nixon 1998) and vertebrates (Clack 1998; Daeschler and Shubin 1998; Ji et al. 1998) suggest that such a mosaicism is a common phenomenon for major taxa at the time of their origin. Shubin (1998) use a "cut-and-paste" metaphor for describing the independent appearance of key characters in different groups at different times. If the independent evolution of key morphological characters is a common theme in seed plant evolution, the significance of molecular data for phylogenetic reconstruction will increase accordingly. However, analysis of several types of macromolecules is needed to avoid the possible discordance and false conclusions due to the potentially possible specific nature of evolution of particular genes under study and cases of homoplasy on the molecular level.

Acknowledgments. This work was supported by RFBR-DAAD Grant 96-04-00127G, RFBR Grants 96-04-50932 and 96-15-97970, and grants to W.M. from the DFG.

References

Albert VA, Backlund A, Bremer K, et al. (1994) Functional constraints and *rbc*L evidence for the land plant phylogeny. Ann Mo Bot Gard 81:534–567

Bowe LM, DePamphilis CW (1997) Conflict and congruence among three plant genetic compartments: Phylogenetic analysis of seed plants from gene sequences of *cox*I, *rbc*L and 18S rDNA. Am J Bot 84(6; Suppl):178–179

Chase MH, Soltis DE, Olmstead RG, et al. (1993) Phylogenetics of seed plants: An analysis of nucleotide sequence from the plastid gene rbcL. Ann Mo Bot Garden 80:528–580

Chaw S-M, Zharkikh A, Sung H-M, Lau T-S, Li W-H (1997). Molecular phylogeny of gymnosperms and seed plant evolution: Analysis of 18S rRNA sequences. Mol Biol Evol 14:56–68

Clack JA (1998) A new Early Carboniferous tetrapod with a *mélange* of crown-group characters. Nature 394:66–69

Crane PR (1985) Phylogenetic analysis of seed plants and the origin of angiosperms. Ann Mo Bot Gard 72:716–793

Crane PR, Friis EM, Pederson KR (1995) The origin and early diversification of angiosperms. Nature 374:27–33

Crepet WL (1998) The abominable mystery. Science 282:1653–1654

Crepet WL, Nixon KC (1998) The new fossil flowers of magnoliid affinity from the Late Cretaceous of New Jersey. Am J Bot 85(9): 1273–1288

Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16(22):10881–10890

Daeschler EB, Shubin N (1998) Fish with fingers? Nature 391:133

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Donoghue MJ, Doyle JA (1989) Phylogenetic analysis of angiosperms and the relationships of Hamamelidae. In: Crane PR, Balckmore S (eds) *Evolution systematics, and fossil history of the Hamamelidae,* Vol 1. Clarendon Press, Oxford, pp 17–45

Downie SR, Katz-Downie DS, Cho K-J (1996) Phylogenetic analysis of Apiaceae subfamily Apioideae using nucleotide sequences from the chloroplast *rpo*C1 intron. Mol Phylogenet Evol 6:1–18

Doyle JA (1996) Seed plant phylogeny and the relations of *Gnetales.* Int J Plant Sci 157(6; Suppl):S3–S39

Doyle JA (1998) Molecules, morphology, fossils and the relationship of angiosperms and Gnetales. Mol Phylogenet Evol 9:448–462

Doyle JA, Donoghue MJ (1987) The importance of fossils in elucidating seed plant phylogeny and macroevolution. Rev Palaebot Palynol 50:63–95

Doyle JA, Donoghue MJ (1992) Fossils and seed plant phylogeny reanalized. Brittonia 44:89–106

Doyle JA, Donoghue MJ, Zimmer EA (1994) Integration of morphological and ribosomal RNA data on the origin of angiosperms. Ann Mo Bot Gard 81:419–450

Felsenstein J (1978) Cases in which parsimony and compatibility methods will be positively misleading. Syst Zool 28:49–62

Felsenstein J (1989) PHYLIP—Phylogeny inference package (version 3.2). Cladistics 5:164–166

Goremykin V, Bobrova VK, Pahnke J, Troitsky AV, Antonov AS, Martin WF (1996) Noncoding sequences from the slowly evolving chloroplast inverted repeat in addition to *rbc*L data do not support Gnetalian affinities of angiosperms. Mol Biol Evol 13(2):383–396

Goremykin V, Hansmann S, Martin WF (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

Hamby KR, Zimmer EA (1992) Ribosomal RNA as a phylogenetic tool in plant systematics. In: Soltis PS, Soltis DE, Doyle JJ (eds) *Molecular systematics of plants.* Chapmen and Hall, New York, pp 50–91

Hasebe M, Ito M, Kofuji R, Iwatsuki K, Ueda K (1992) Phylogeny of gymnosperms inferred from *rbc*L gene sequences. Bot Mag Tokyo 105:673–679

Hedges SB (1994) Molecular evidence for the origin of birds. Proc Natl Acad Sci USA 91:2621–2624

Hori H, Lim BL, Osawa S (1985) Evolution of green plants as deduced from 5S rRNA sequences. Proc Natl Acad Sci USA 82:820–832

Ji Q, Currie PJ, Norell MA, Ji S-A (1998) Two feathered dinosaurus from northeastern China. Nature 393:753–761

Lokonte H, Stevenson DW (1990) Cladistics of the spermatophyta. Brittonia 42:197–211

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Nixon KC, Crepet WL, Stevenson D, Friis EM (1994) A reevaluation of seed plant phylogeny. Ann Mo Bot Gard 81:483–533

Rakhimova GM, Troitsky AV, Klikunova IN, Antonov AS (1989) Phylogenetic analysis of 18S rRNA partial sequences from 14 higher plant species. Molekularnaya Biologiya 23:830–842 (in Russian)

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Shi SH, Qu LH, Xi BQ, Chang HT (1994) Sequencing and comparative analyses of the large-subunit ribosomal RNA 5′ terminal region for four species of gymnosperms and their phylogenetic significance. Cathaya 6:35–42

Shubin N (1998) Evolutionary cut and paste. Nature 394:12–13

Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Sun G, Dilcher DL, Zheng S, Zhou Z (1998) In search of the first flower: A Jurassic angiosperm, *Archaefructus,* from Northeast China. Science 282:1692–1695

Sytsma KJ, Hahn WJ (1994) Molecular systematics: 1991–1993. Prog Botany 55:307–333

Tabor S, Richardson CC (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc Natl Acad Sci USA 84:4767–4771

Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content bias. Mol Biol Evol 9:678–687

Troitsky AV, Melekhovets YF, Rakhimova GM, Bobrova VK, Valiejo-Roman KM, Antonov AS (1991) Angiosperm origin and early stages of seed plant evolution deduced from rRNA sequence comparison. J Mol Evol 32:255–261

Zharkikh AA, Rzhetsky A Yu, Morozov PS, Sitnikova TL, Krushkal JS (1990) A package of microcomputer programs for sequence analysis and construction of phylogenetics. Gene 101:217–218