Molecular Phylogenies in Angiosperm Evolution[1]

William Martin, Derek Lydiate,†[2] Henner Brinkmann,* Gert Forkmann,† Heinz Saedler,† and Rüdiger Cerff**

*Institut für Genetik, Technische Universität Braunschweig; and
†Max-Planck-Institut für Züchtungsforschung

We have cloned and sequenced cDNAs for the glyceraldehyde-3-phosphate dehydrogenase of glycolysis, *gap*C, from a bryophyte, a gymnosperm, and three angiosperms. Phylogenetic analyses are presented for these data in the context of other *gap*C sequences and in parallel with published nucleotide sequences for the chloroplast encoded gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbc*L). Relative-rate tests were performed for these genes in order to assess variation in substitution rate for coding regions, along individual plant lineages studied. The results of both gene analyses suggest that the deepest dichotomy within the angiosperms separates not magnoliids from remaining angiosperms, but monocotyledons from dicotyledons, in sharp contrast to prediction from the Euanthial theory for angiosperm evolution. Furthermore, these chloroplast and nuclear sequence data taken together suggest that the separation of monocotyledonous and dicotyledonous lineages took place in late Carboniferous times [~300 Myr before the present (Mybp)]. This date would exceed but be compatible with the late-Triassic (~220 Mybp) occurrence of fossil reproductive structures of the primitive angiosperm *Sanmiguelia lewisii*.

Introduction

Angiosperms dominate extant floras, yet the origins and early evolution of the group are incompletely understood. Efforts to understand the evolution of flowering plants rely to a large extent on the morphology of reproductive structures, yet these are rare in fossilized form (Stewart 1983, pp. 365–379; Thomas and Spicer 1987, pp. 215–231). The intermediate forms necessary to reconstruct a robust picture of morphological changes surrounding angiospermy have yet to be brought forth from the fossil record, leaving the relationships (1) between angiosperms and other seed plants (Crane 1985; Meyen 1986; Cronquist 1988, pp. 129–157) and (2) within the flowering plants themselves (Burger 1977; Dahlgren and Bremer 1985; Krassilov 1991) a point of controversy. Cronquist (1988, p. 129) has noted that the "abominable mystery" of angiosperm origins "remains scarcely less so to modern students of evolution. It is clear that they [the angiosperms] are vascular plants, related to other vascular plants,

1. Key words: angiosperms, gymnosperms, molecular phylogeny, relative-rate test, glyceraldehyde-3-phosphate dehydrogenase, ribulose-1, 5-bisphosphate carboxylase/oxygenase. Abbreviations: *gap*C = glycolytic glyceraldehyde-3-phosphate dehydrogenase, E.C.1.2.1.12; *rbc*L = large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, E.C.4.1.1.39; *gap*A = Calvin cycle glyceraldehyde-3-phosphate dehydrogenase, E.C.1.2.1.13; Mybp = million years before present.

2. Present address: John Innes Centre for Plant Science Research, Colney Lane, Norwich, Norfolk, NR4 7UJ, England.

Address for correspondence and reprints: William Martin, Institut für Genetik, Technische Universität Braunschweig, Konstantin-Uhde-Str. 5, D-3300 Braunschweig, Germany.

Mol. Biol. Evol. 10(1):140–162. 1993.

0737-4038/93/1001-0009$02.00

. . . and that their immediate ancestors must have been, by definition, gymnosperms. Beyond that, much is debatable."

The accepted time scale of angiosperm evolution is integrally related to questions concerning identification of taxa ancestral to the group (Axelrod 1970; Krassilov 1977; Meeuse 1987, pp. 183–187; Cronquist 1988, pp. 129–157). Generally recognized angiospermous forms appear in the fossil record in lower Cretaceous times, ~120 Myr before the present (Mybp) (Crane et al. 1986), and became the dominant group in fossil floras by ~90 Mybp (Lidgard and Crane 1988). Description of angiosperm evolution in pre-Cretaceous times is at best difficult in the absence of generally recognized angiospermous fossils of pre-Cretaceous age and is further impaired by difficulties in unambiguous definition of "an angiosperm," particularly as applied to fossil forms (Stewart 1983, pp 365–389; Thomas and Spicer 1987, pp. 213–232). Accordingly, the lower Cretaceous is widely viewed as the starting point of angiosperm evolution (Beck 1976, pp. 2–3; Cronquist 1988, pp. 136–138), though proponents of a long pre-Cretaceous history for angiosperms have argued their case (Axelrod 1959; Krassilov 1977; Meeuse 1990, pp. 11–21), albeit often in lack of independent empirical data.

Here we report the cloning and phylogenetic analysis of full-size cDNAs for *gap*C, the nuclear-encoded gene for glycolytic glyceraldehyde-3-phosphate dehydrogenase (Cerff and Chambers 1979) from the bryophyte *Physcomitrella patens,* the gymnosperm *Pinus sylvestris* (Scots pine), and three angiosperms: *Dianthus caryophyllus* (carnation), *Callistephus chinensis* (aster), and *Pisum sativum* (pea). Phylogenetic analysis of *rbc*L sequences from related taxa are presented and compared with the nuclear gene phylogeny. Because the fossil record for the emergence of the first land plants (Gensel and Andrews 1987; Thomas and Spicer 1987, pp. 73–168) and first conifers (Scott and Chaloner 1983; Meyen 1984; Chaloner 1989; Galtier and Rowe 1989) is relatively well documented, gene sequences from a bryophyte and a gymnosperm provide points of calibration with which angiosperm age may be roughly estimated. We also describe the comparative analysis of *gap*C and *rbc*L sequences from a bryophyte, a gymnosperm, and eight angiosperms (including *Magnolia*). These nuclear and chloroplast DNA genes provide congruent pictures of angiosperm evolution. The results support the view that angiosperms arose and diversified long before their extensive appearance in lower-Cretaceous strata. The lower estimates for the monocot-dicot divergence reported by Wolfe et al. (1989) may be because their analysis rested solely on chloroplast DNA sequences from a single dicot species (tobacco), the chloroplast DNA of which has been shown (Wolfe et al. 1987) to evolve at a rate significantly lower than that of other chloroplast sequences. Our estimates of angiosperm age are congruent with the late-Triassic (~220 Mybp) occurrence of the primitive angiosperm *Sanmiguelia lewisii* (Cornet 1989*b*). The phylogenies derived from these nuclear and plastid nucleotide sequence data do not support the widely accepted Euanthial theory for the evolution of angiosperms from strictly magnolean antecedents. Our results are discussed in light of alternative theories concerning angiosperm origins and early evolution and are contrasted to other recent molecular analyses on flowering-plant phylogeny.

Material and Methods

cDNA Cloning

For the bryophyte *Physcomitrella patens,* cultures were grown (Cove and Ashton 1984) and polyA+ mRNA was isolated, according to a method described by Cerff

and Kloppstech (1982) from 2-wk-old caulonema plate cultures. cDNA of *P. patens* was constructed by the method of Lapayre and Amalric (1985), size fractionated on 1% agarose, and cloned into λnm1149 (Murray 1983), according to a method described by Schwarz-Sommer et al. (1987). Fifty thousand recombinants were screened by plaque hybridization at 60°C in 3 × SSPE (Schwarz-Sommer et al. 1985) using a radiolabeled cDNA insert for glycolytic GAPDH from *Magnolia liliiflora* (Martin et al. 1989). Washings were performed at 60°C in 2 × SSPE. Positive clones were subcloned into pUC19 (Vieira and Messing 1982), and terminal regions were sequenced by the dideoxy method (Sanger et al. 1977). Two full-size clones (pPP18 and pPP14) differing in length of their noncoding regions yet otherwise identical were identified, and both strands were sequenced by the chemical degradation method (Maxam and Gilbert 1980) and dideoxy method, for pPPC18 and pPPC14, respectively.

For *Pinus sylvestris,* 20 μg polyA+ mRNA was isolated from 700 μg total RNA provided by Stefan Jansson (Umea). cDNA was constructed with the Pharmacia kit modified through size fractionation of linker-ligated cDNA on 1% agarose according to a method described by Martin et al. (1990) and cloned into λnm1149 (Murray 1983). Four hundred thousand recombinants were obtained from 20 ng of cDNA packaged. Screening was performed as for the *P. patens* library. Five hybridizing cDNAs were subcloned into SK+ plasmids (Stratagene), and terminal sequences were determined by the dideoxy method. Two of these, pPSC15 and pPSC9, were sequenced on both strands by the chemical degradation method (Maxam and Gilbert 1980). Two of the remaining three clones were identical, in terminal regions, to pPSC15, and the other was identical to pPSC9.

mRNA and cDNA for *Dianthus caryophyllus* and *Callistephus chinensis* were prepared from immature flowers by the method described for *P. patens.* One full-size *gap*C clone for each species (pDCC1 for *Dianthus* and pCCC1 for *Callistephus*) with insert >1.2 kb in length was sequenced by the chemical degradation method. The *Pisum* cDNA pPEA1 was isolated from a library described elsewhere (Brinkmann et al. 1989), and both strands were sequenced by the chain-termination method.

Computer Analysis

Sequence handling and alignment were performed with the WISGEN package (Devereux et al. 1984). Numbers of nonsynonymous substitutions per nonsynonymous (K_a) and nondegenerate (K_0) site between sequences were measured with the weighted pathway method of Li et al. (1985). Trees from distance matrices were inferred with the neighbor-joining (NJ) method (Saitou and Nei 1987). Parsimony bootstrap analysis was performed with DNABOOT of the PHYLIP package (Felsenstein 1985). Relative-rate tests were performed for divergence at nondegenerate sites between *gap*C sequences, according to the method described by Li and Tanimura (1987). *rbc*L sequences were either retrieved from the data base or typed in by hand and were analyzed in the same manner as were *gap*C sequences.

Results

cDNA Sequences for *gap*C

Nucleotide sequences for full-size *gap*C cDNAs from *Physcomitrella patens* (a bryophyte), *Pinus sylvestris* (a gymnosperm), *Callistephus chinensis* (aster, Asteridae), *Dianthus caryophyllus* (carnation, Caryophyllidae), and *Pisum sativum* (pea, Rosidae) are shown in figure 1. The cDNAs contain the entire coding regions which are completely colinear, with the exception of a deletion following amino acid 145 in

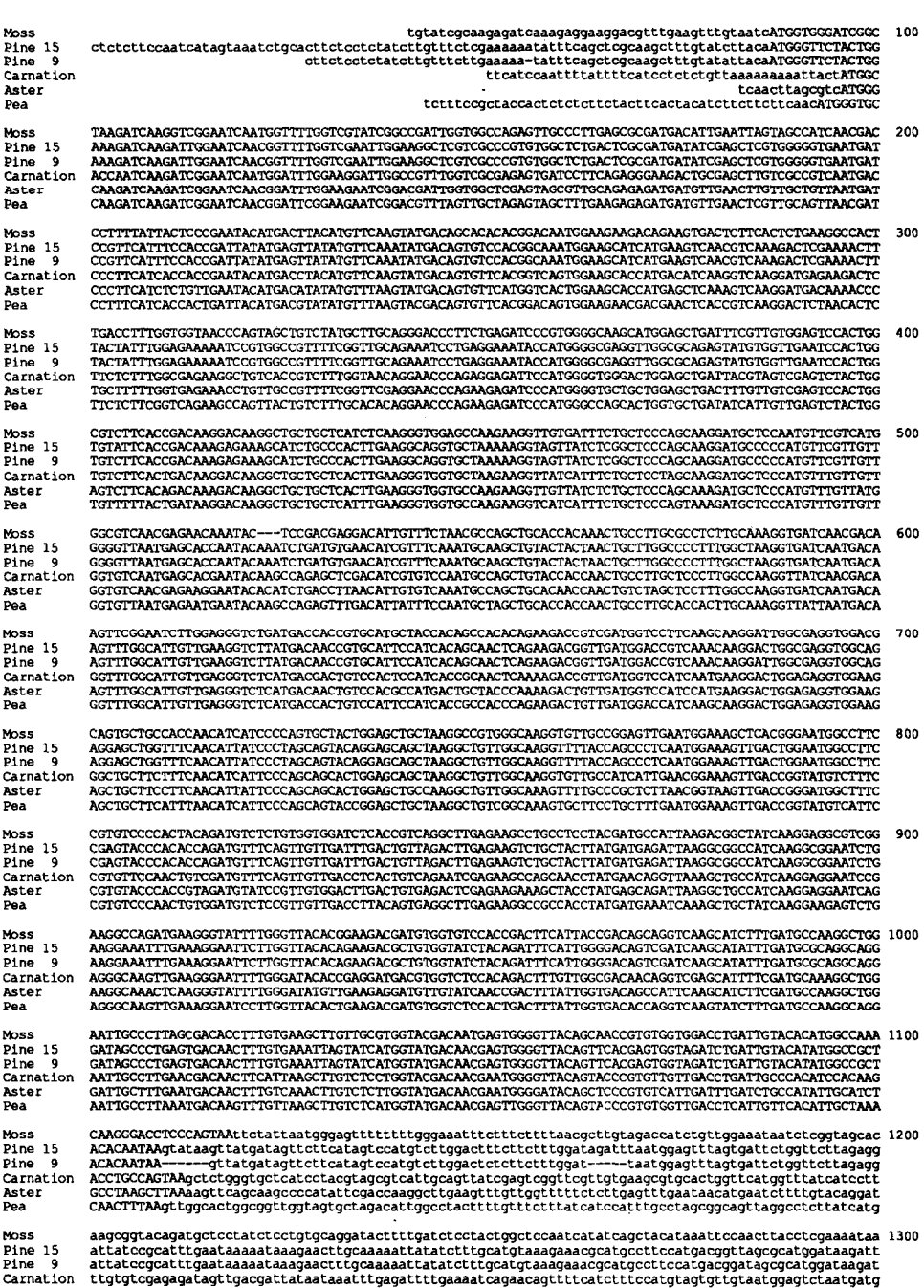


FIG. 1.—Nucleotide sequences for six glycolytic GAPDH cDNAs. Coding regions are shown in uppercase letters, and gaps are shown as dashes. Sequences are of cDNA clones from *Physcomitrella patens* (pPPC18), *Pinus sylvestris* (pPSC15 and pPSC9), *Dianthus caryophyllus* (pDCC1), *Callistephus chinensis* (pCCC1), and *Pisum sativum* (pPEA1) (see Material and Methods).

the moss sequence; to date, this is the only indel found in coding regions of land plant *gap*C sequences. Within the coding region, the two cDNA clones from *Pinus* show only three synonymous substitutions in the coding regions and only few indels in the noncoding regions, suggesting that these are either allelic variants or very recently duplicated genes. For phylogenetic analyses presented here, only the sequence of pPSC9 was used. As judged by their frequency in the cDNA bank, these two clones would appear to represent the major *gap*C protein in pine seedlings. The systematic positions of those plants from which *gap*C sequences were analyzed in this study are given in table 1.

Nucleotide Composition Equilibrium

Some plant genes, such as that for the GAPDH enzyme of the Calvin cycle in chloroplasts—*gap*A (Brinkmann et al. 1987; Quigley et al. 1988)—or chalcone synthase (Niesbach-Klösgen et al. 1987), show an extreme bias for G and C within the coding and flanking regions in graminaceous monocots. To show that coding regions of glycolytic glyceraldehyde-3-phosphate dehydrogenases, *gap*C, in plants are *not* subject to this type of bias, which could potentially influence the evolution of *gap*C genes at nonsynonymous sites, we separately plotted the base composition of the plant *gap*C sequences for first plus second codon positions and for third codon positions (fig. 2). Figure 2 shows that first and second positions of plant *gap*C coding regions are at compositional equilibrium (Prager and Wilson 1988) and that a mild preference for

Table 1
Systematic Position of Species from Which *gap*C Sequences Were Analyzed

Species (gene[a])	Abbreviation[b]	Subclass	Division or Subdivision
Antirrhinum majus	ant	Asteridae	Angiospermae
Petunia hybrida	pet	Asteridae	Angiospermae
Nicotiana tabacum	tob	Asteridae	Angiospermae
Callistephus chinensis	ast	Asteridae	Angiospermae
Sinapis alba	sin	Dilleniidae	Angiospermae
Arabidopsis thaliana	ath	Dilleniidae	Angiospermae
Dianthus caryophyllus	nel	Caryophyllidae	Angiospermae
Mesembryanthemum crystallinum	mes	Caryophyllidae	Angiospermae
Petroselinum hortense	par	Rosidae	Angiospermae
Pisum sativum	pea	Rosidae	Angiospermae
Magnolia liliiflora	mag	Magnoliidae	Angiospermae
Ranunculus acer	ran	Magnoliidae	Angiospermae
Hordeum vulgare (*gap*C1)	bar	Liliidae	Angiospermae
H. vulgare (*gap*C3)	bar	Liliidae	Angiospermae
Zea mays (*gap*C1)	zea	Liliidae	Angiospermae
Z. mays (*gap*C2)	zea	Liliidae	Angiospermae
Z. mays (*gap*C3)	zea	Liliidae	Angiospermae
Pinus sylvestris	pin	Pinidae	Gymnospermae
Physcomitrella patens	mos	Bryidae	Bryophyta

[a] Unless otherwise specifically denoted, all references to *gap*C sequences from *Zea* and *Hordeum* in the text indicate *gap*C1 sequences. Systematic position is based on the scheme of Ehrendorfer (1991).

[b] References and accession numbers for sequences are as follows: ant, X59517; pet, X60346; tob, M14419; ast, present paper; sin, X04302; ath, M64116; nel, present paper; mes, M29956; par, X60344; pea, present paper; mag, X60347; ran, X60345; bar *gap*C1, X60343; bar *gap*C3, Chojecki (1986); zea *gap*C1, X15596; zea *gap*C2 and *gap*C3, Russell and Sachs (1989); pin, present paper; and mos, present paper.

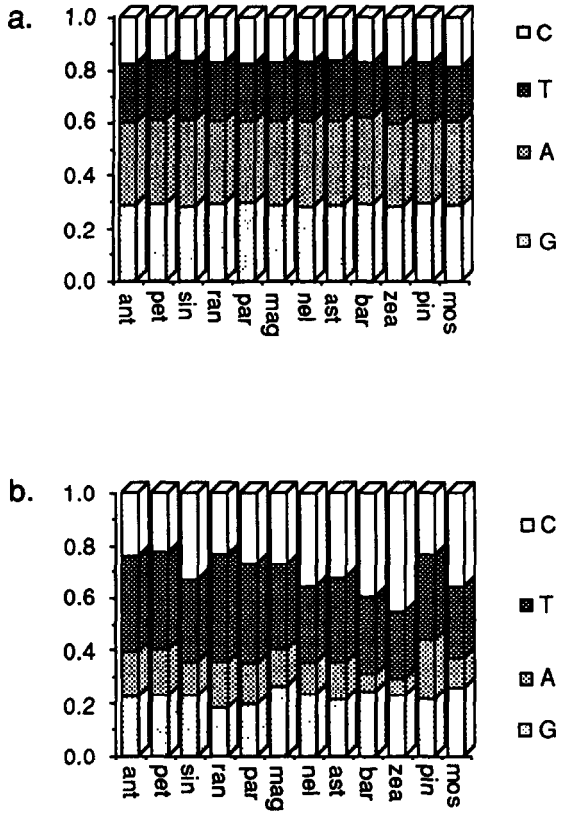


FIG. 2.—Base composition of plant *gap*C sequences demonstrating base-compositional equilibrium (Prager and Wilson 1988) at first and second codon positions (a) and at third codon positions (b). Bars indicate the fraction of bases in each sequence. Species-name abbreviations are as in table 1.

G and C within the coding region, as reflected at third positions of the barley and maize sequences, is not sufficient to have introduced bias at first and second codon positions, which contain most of the nondegenerate and nonsynonymous sites. The equilibrium found for the GAPDH of glycolysis (*gap*C) does not, however, hold for Calvin cycle GAPDH (*gap*A) genes, where third codon positions contain 97% G+C in maize (Brinkmann et al. 1987).

Relative-Rate Tests with *gap*C Sequences

The mRNA for *gap*C, a slowly evolving glycolytic protein (Fothergill-Gilmore 1986), satifies several criteria required of a molecular marker able to address plant phylogeny. *gap*C contains ~1 kb of coding region; this is considerably longer than markers such as 55 RNA (Hori et al. 1985) or cytochrome *c* (Boulter et al. 1972) or shorter peptides (Boulter et al. 1979). As is true of the larger rRNA species, *gap*C evolves at a conservative pace, yet, in contrast to rRNA, *gap*C genes do not belong to the middle repetitive fraction of the genome and thus are not subject to the influences of concerted evolution for such DNA (Dover 1987, 1989). Nucleotide sequences for *gap*C are known for fungi and several metazoa whose fossil histories are well documented, allowing determination and comparison of substitution rates for *gap*C from

the three eukaryotic kingdoms, for the purpose of linking the geologic time scale of animal evolution with that of its elusive angiosperm counterpart.

To examine constancy of substitution rate for plant *gap*C cDNAs, we translated and aligned nucleotide sequences, leaving an average of 664 nondegenerate and 760 nonsynonymous nucleotide sites for comparison. To determine whether significant variation in substitution rate among higher-plant *gap*C sequences could be detected, we measured numbers of substitutions per site at fourfold degenerate sites in the coding region, K_0, and subjected them to the relative-rate test (Li and Tanimura 1987), using both yeast and bryophyte sequences as the outgroup. The results of these tests are summarized in tables 2 and 3. With yeast as the outgroup, no significant (at the 5% level) differences between plant and animal *gap*C sequences could be detected in substitution rate at nondegenerate sites (table 2), suggesting that the substitution rate of *gap*C has remained relatively constant since the divergence of these lineages. The larger number of negative values for $K_{13}-K_{23}$ (69 of 96 values; table 2) indicates that the plant *gap*C sequences may be evolving more slowly than their animal counterparts. Using the bryophyte *gap*C sequence as the outgroup, we tested the constancy of rate within seed plant *gap*C sequences (table 3). Table 3 reveals that only the *gap*C sequence from *Petroselinum* (parsley) is evolving at a rate significantly higher than that of other spermatophyte *gap*C sequences. *gap*C of *Hordeum* (barley), a graminaceous monocot, appears to be evolving at a slower rate than its counterparts from the dicotyledonous species surveyed, except pea. Thus, relative-rate tests reveal (1) that *gap*C sequences in plants do not evolve more rapidly than their animal counterparts and (2) that, with the exception of *Petroselinum*, *gap*C in the different plant lineages surveyed is not evolving at significantly different rates. This conservative mode of evolution for *gap*C is also observed in bacterial *gap*C sequences (Nelson et al. 1991) and renders these valuable markers in molecular phylogenetic studies.

*rbc*L Sequence Analyses

Parallel to the *gap*C analyses, we studied a gene of the chloroplast DNA, i.e., the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbc*L) from the same species or closely related taxa for which *rbc*L sequences are currently available: *Pseudotsuga menziesii* (Douglas fir), a second Pinaceae instead of *Pinus*, *Marchantia polymorpha* (liverwort), a second bryophyte instead of *Physcomitrella*, *Spinacia oleracea* (spinach) instead of *D. caryophyllus* as a second Caryophyllidae, and *Oryza sativa* (rice) as a second graminaceous monocot instead of *Hordeum*. *gap*C is nuclear encoded in plants such that, if *gap*C gene duplication events occurred prior to speciation of the taxa under study, a distorted picture of plant evolution potentially could be inferred from the cDNA sequences. Because there is only one copy of *rbc*L in the chloroplast DNA, *rbc*L gene duplication prior to separation of the higher-level taxa considered (Nei 1987, pp. 287–290) can be effectively ruled out. As for *gap*C, we first performed a relative-rate test for values of K_0 derived from the aligned *rbc*L nucleotide sequences, using the *Chlamydomonas* outgroup to examine constancy of substitution rate for this gene. The *rbc*L gene from maize has accumulated a higher number of substitutions at nondegenerate sites than have the other angiosperm sequences under study, whereas the gene from rice has not, suggesting that the increase of rate observed in the maize *rbc*L gene may have occurred subsequent to the separation of these graminaceous monocots (table 4). The *Marchantia rbc*L gene is evolving at a slower rate than the other *rbc*L sequences considered. The significant slowdown observed for the lineage of the *Nicotiana* (tobacco) chloroplast genome on the whole

Table 2

Relative-Rate Test between 16 Plant and Six Animal *gap*C Sequences at Nondegenerate Sites, Using Yeast as the Outgroup

		Species 1															
	Yeast	Ant	Pet	Tob	Ran	Par	Nel	Ast	Pea	Sin	Mag	Mes	Ath	Bar	Zea	Pin	Mos
Yeast	...	2,686	2,666	2,743	2,679	2,837	2,780	2,577	2,628	2,513	2,762	2,730	2,598	2,590	2,732	2,679	2,612
Species 2:																	
Nem	2,694	−8	−28	49	−15	143	86	−117	−66	−181	68	36	−96	−104	38	−15	−82
Dro2	2,887	−201	−221	−144	−208	−50	−107	−310	−259	−374	−125	−157	−289	−297	−155	−208	−275
Dro	2,899	−213	−233	−156	−220	−62	−119	−322	−271	−386	−137	−169	−301	−309	−167	−220	−287
Chk	2,682	4	−16	61	−3	155	98	−105	−54	−169	80	48	−84	−92	50	−3	−70
Hum	2,733	−47	−67	10	−54	104	47	−156	−105	−220	29	−3	−135	−143	−1	−54	−121
Rat	2,670	16	−4	73	9	167	110	−93	−42	−157	92	60	−72	−80	62	9	−58

NOTE.—Data in the top row are K_{13}, i.e., the divergence between the respective plant (species 1) and yeast (species 3) at nondegenerate sites (i.e., K_0) × 10^4. Data in the leftmost column are K_{23}, i.e., the divergence between the respective animal (species 2) and yeast at nondegenerate sites × 10^4. All other data in the matrix are the difference K_{13}–K_{23} at nondegenerate sites × 10^4; negative values reflect a higher rate for the animal *gap*C sequence tested in the comparison. Absolute values of K_{13}–K_{23} greater than 0.0200 are underlined and were tested for significance. No differences were significant at the 5% level. Abbreviations are as in table 1, except for dro and dro2 (two *gap*C genes of *Drosophila:* Tso et al. 1985*b*), chk (chicken; Dugaiczyk et al. 1983), hum (human; Tso et al. 1985*a*), nem (nematode; Yarbrough et al. 1987), and rat (Tso et al. 1985*a*).

Table 3
Relative-Rate Test among Seed Plant *gap*C Sequences, Using Bryophyte as the Outgroup

	Species 1												
	Ant	Pet	Tob	Ran	Par	Nel	Ast	Pea	Sin	Mag	Bar	Zea	Pin
Mos	1,461	1,607	1,520	1,430	1,727	1,564	1,535	1,381	1,397	1,406	1,384	1,482	1,347
Species 2:													
Ant													
Pet	−146												
Tob	−59	87											
Ran	31	177	90										
Par	−266	−120	−207	−297*									
Nel	−103	43	−44	−134	163								
Ast	−74	72	−15	−105	192	29							
Pea	80	226	139	49	346*	183	154						
Sin	64	210	123	33	330*	167	138	−16					
Mag	55	201	114	24	321*	158	129	−25	−9				
Bar	77	223	136	46	343*	180	151	−3	13	22			
Zea	−21	125	38	−52	245	82	53	−101	−85	−76	−98		
Pin	114	260	173	83	380**	217	188	34	50	59	37	135	

NOTE.—Data in the top row are K_{13}, i.e., the divergence between the respective species 1 and the bryophyte (*Physcomitrella*) outgroup (species 3) at nondegenerate sites (i.e., K_0) × 10^4. Data in the matrix indicate the difference K_{13}–K_{23} at nondegenerate sites × 10^4; negative values reflect a higher rate for species 2 in the comparison. Absolute values of K_{13}–K_{23} greater than 0.0200 are underlined and were tested for significance. Species name abbreviations are as in table 1.

* $P < .05$.

** $P < .01$.

(Wolfe et al. 1987) is reflected neither in the *rbc*L gene of *Nicotiana* nor in that of the confamilial species *Petunia* (table 4).

Trees

For the construction of phylogenetic trees we chose the NJ (Saitou and Nei 1987) method, since it has been shown in computer simulation to be more efficient in recovering the correct topology than are many are other molecular sequence and distance matrix methods, under a variety of sequence parameters (including unequal rates of evolution in different lineages) (Saitou and Nei 1987; Saitou and Imanishi 1989; Jin and Nei 1990). Different evolutionary models deal differently with the contribution of substitutions at twofold-degenerate sites to the total proportion of nonsynonymous substitutions (Li et al. 1985; Nei and Gojobori 1986). Because the effect of an accurate distance measure has been shown in computer simulation to be critical for recovery of the correct topology (Jin and Nei 1990; Nei 1991), we thus chose divergence at nondegenerate sites as the distance measure for trees presented here. From the matrices of K_a and K_0 values (Li et al. 1985) for *gap*C and *rbc*L for the nine higher-plant taxa under study, NJ trees were constructed in order to contrast the pictures of plant evolution reflected in the sequences of both a nuclear gene (*gap*C) and chloroplast gene (*rbc*L). The NJ trees for values of K_0 are shown in figure 3. The NJ tree for *gap*C values of K_a yielded a branching order identical to that shown for K_0, whereas in the *rbc*L tree of K_a values, monocotyledons branched below the gymnosperm, perhaps because of the higher substitution rate in the maize *rbc*L gene (table 4), which is more pronounced in values of K_a than for divergence at nondegenerate sites (data not shown). With the exception of the position of *Magnolia* within the dicotyledons surveyed, NJ trees for K_0 values of both *gap*C and *rbc*L yield the same topology, suggesting that no

Table 4
Relative-Rate Test among Plant *rbc*L Sequences, Using *Chlamydomonas* as the Outgroup

	Species 1								
	Mpo	Fir	Spi	Pea	Mag	Tob	Pet	Ric	Zea
C. reinhardii	669	793	823	727	727	813	784	781	894
Species 2:									
Mpo									
Fir	−124								
Spi	−154*	−30							
Pea	−58	66	96						
Mag	−58	66	96	0					
Tob	−144	−20	10	−86	−86				
Pet	−115	9	39	−57	−57	29			
Ric	−112	12	42	−54	−54	32	3		
Zea	−225**	−101	−71	−167*	−167*	−81	−110	−113	

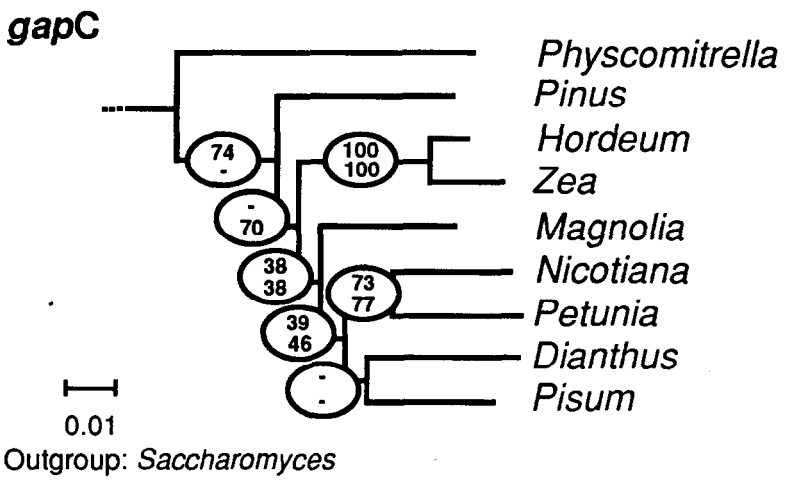
NOTE.—Data in the top row are K_{13}, i.e., the divergence between the respective species 1 and the *C. reinhardii* outgroup (species 3) at nondegenerate sites (i.e., K_0) × 10^4. Data in the matrix indicate the difference K_{13}–K_{23} at nondenegerate sites × 10^4; negative values reflect a higher rate for species 2 in the comparison. Absolute values of K_{13}–K_{23} greater than 0.0100 are underlined and were tested for significance. Species-name abbreviations are as in table 1, except for fir (*Pseudotsuga menziesii;* Douglas fir), mpo (*Marchantia polymorpha*), mag (*Magnolia macrophylla*), spi (*Spinacia oleracea;* spinach), and ric (*Oryza sativa;* rice).

* $P < .05$.

** $P < .01$.

*gap*C sequences included in figure 3 arose through duplication prior to the separation of those monocotyledons and dicotyledons surveyed. A composite matrix of K_0 values from concatenated *gap*C and *rbc*L sequences was also used to construct an NJ tree for the taxa shown (fig. 4), in order to increase the total number of sites for comparison. The branching order of the composite tree is identical to that obtained with the *rbc*L data alone (fig. 3). *Magnolia* is not borne on the deepest branch within the angiosperms surveyed, in either *rbc*L or *gap*C trees (see Discussion).

There is a single gene for *gap*C in *Arabidopsis* (Shih et al. 1991), yet in maize there are three separate genes for *gap*C, termed "*gap*C1," "*gap*C2," and "*gap*C3" (Russell and Sachs 1989). The relatively good correlation between the *gap*C (nuclear) and *rbc*L (chloroplast) trees (fig. 3) suggested that *gap*C comparisons in most cases were not paralogous. To examine the possibility that *gap*C comparisons between monocotyledons and dicotyledons may involve genes that arose through duplication prior to the separation of these taxa, we constructed a tree from the available sequence data for *gap*C2 and *gap*C3 from maize, in the context of all higher-plant *gap*C cDNA sequences currently available (fig. 5). Indeed, a duplication that gave rise to *gap*C3 of maize occurred before the divergence of barley and maize. Yet, if the *gap*C family of maize did arise prior to the monocot-dicot separation, we must expect one of the members of the maize *gap*C genes to share a common branch with a dicotyledonous sequence, a result that is not found, indicating that the *gap*C sequences compared between monocotyledons and dicotyledons here (60 pairwise comparisons) did not arise through duplication prior to the separation of these taxa. Similar results were previously obtained in the analysis of *gap*C and chalcone synthase genes (Martin et al. 1989). These data and those derived from rRNA genes (Troitsky et al. 1991) lend strength to the molecular evidence for divergence of monocotyledons and dicotyledons prior to the divergence of magnoliids from other angiosperms.



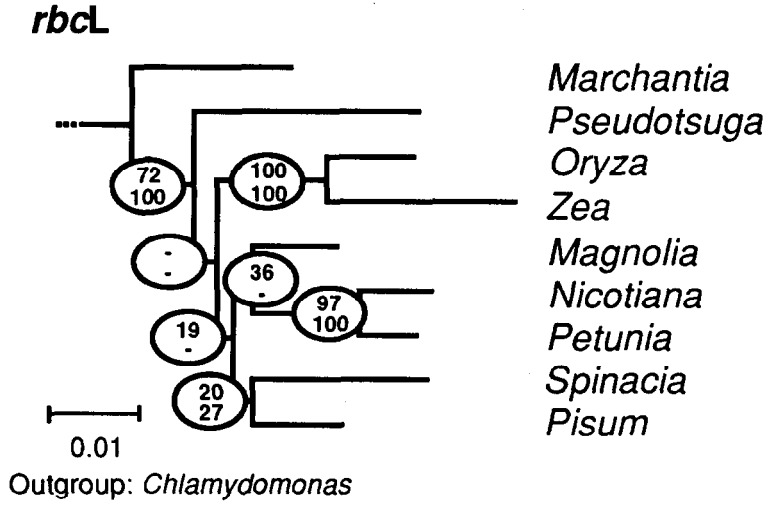


FIG. 3.—Phylogenetic trees inferred by the NJ method for divergence at nondegenerate sites in a nuclear gene, *gap*C, and a chloroplast gene, *rbc*L. Scale bars indicate 0.01 substitutions/site. Branch lengths are drawn to scale. Accession numbers and/or references for *gap*C sequences not presented in this paper are given in table 1, except for *Saccharomyces* (J10324). Accession numbers for *rbc*L sequences are as follows: *Marchantia* (liverwort), X10647 and X04465; *Pseudotsuga* (Douglas fir), X52937; *Oryza* (rice), D00207; *Zea* (maize), M11592; *Magnolia* (Golenberg et al. 1990); *Nicotiana* (tobacco), M16867; *Petunia*, X04976; *Spinacia*, J10443; *Pisum* (pea), X03853; and *Chlamydomonas*, J10399. Systematic positions of the species from which *gap*C sequences were studied are given in table 1; for *rbc*L, this information is given in table 1 and in the text. Ovals containing numbers attached to internal branches contain the results of bootstrap parsimony analysis (DNABOOT of PHYLIP; Felsenstein 1985) for the corresponding gene (*gap*C or *rbc*L) and indicate the number of times, in 100 replicates, that sequences belonging to the species borne on that branch occurred on that branch. Two numbers are given in each oval. The upper number indicates the number of times, of 100, that the branch was found in parsimony bootstrap analysis of nucleotide

Bootstrap parsimony analysis (Felsenstein 1985) was performed with various subsets of the *gap*C and *rbc*L nucleotide sequences, including complete sequences, first positions, second positions, and first plus second positions. The results of analyses with first plus second codon positions and with complete sequences are given in figures 3 and 4. Probably because of the relatively starlike phylogeny and low divergence of genes under study, only a few clades were detected in >95/100 replicates in any data set. Parsimony methods do not have a very high probability of recovering the correct topology when the total divergence between sequences is low, because the number of informative sites becomes very small (Sourdis and Nei 1988). Distance methods, however, can also utilize the information contained in noninformative sites (Nei 1991, pp. 90–128). The NJ method has been shown to have a higher probability of recovering the correct topology than does the parsimony method, when the total divergence between sequences is low (Sourdis and Nei 1988), as is the case for spermatophyte *gap*C and *rbc*L divergence at nondegenerate sites presented here. Divergence at synonymous sites between both *gap*C and *rbc*L sequences analyzed was high (≥1 substitution/site in most cases; data not shown). One of the assumptions of the parsimony algorithm used (Felsenstein 1985) is that the probability of base change along a given branch is low, a condition not fulfilled for synonymous sites in the sequences analyzed. It is therefore not surprising that parsimony bootstrap trees, produced with data sets containing third codon positions, supported fewer clades than were supported with the matrix method for both *gap*C and *rbc*L (figs. 3 and 4). The parsimony trees for *rbc*L sequences detected the presumably true branching order of *Marchantia* (bryophyte), *Pseudotsuga* (gymnosperm), and angiosperms (see Discussion) in <19/100 replicates, with either first plus second positions or complete sequences (fig. 3), whereas the expected topology was found in the NJ tree. This suggests to us that divergence between *rbc*L sequences of major spermatophyte groups may be too low at nonsynonymous sites—yet too high at synonymous sites—to resolve early phases of seed plant evolution with parsimony methods.

Discussion

We have cloned and analyzed cDNAs for a conservatively evolving enzyme of glycolysis, *gap*C, and have analyzed published sequences for *rbc*L, a conservatively evolving chloroplast gene, in order to retrace angiosperm phylogeny from molecular sequences. The conservative mode of plant *gap*C sequence evolution is reflected both in the well-conserved nucleotide composition at first and second codon positions (fig. 2a) and in the very low fluctuation in substitution rate at nondegenerate sites (tables 2 and 3). We tested, using yeast as the outgroup (Gouy and Li 1989), plant and animal *gap*C sequences for differences in substitution rate at nondegenerate sites (table 2) and found no differences that were significant at the 5% level, indicating that the conservative rate of *gap*C evolution is maintained within these eukaryotes. Having shown (1) that *gap*C and *rbc*L yield very similar pictures of higher-plant evolution

sequence data sets from which all third codon positions had been deleted. This procedure approximates the removal of synonymous sites (Martin et al., accepted). The lower number indicates the number of times, of 100, that the branch was found in parsimony bootstrap analysis of complete nucleotide sequences. Dashes contained within ovals indicate that the branch was found in less than 19/100 parsimony bootstrap replicates. Sequences indicated as outgroups (*Saccharomyces* and *Chlamydomonas*) were defined as the outgroup in parsimony runs.

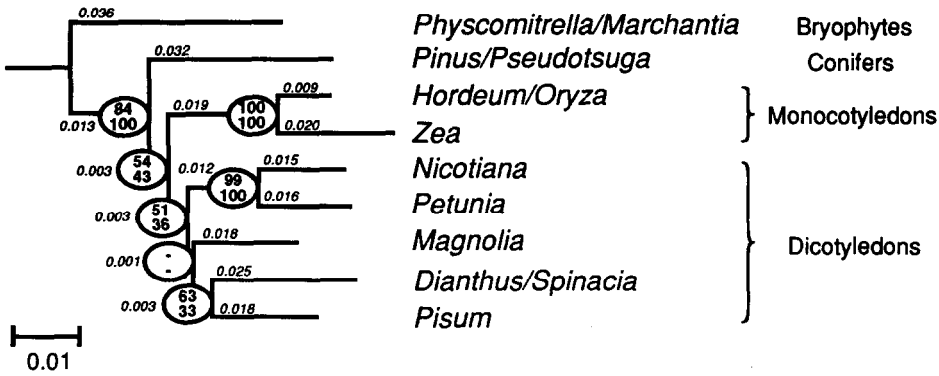


FIG. 4.—NJ tree for concatenated *gap*C/*rbc*L sequences from nine higher plants. The outgroup was yeast/*Chlamydomonas*. Branch lengths are indicated in substitution per nondegenerate site. Ovals attached to internal branches contain the results of bootstrap parsimony analysis as described in the legend to fig. 3. The sequence indicated as the outgroup (*Saccharomyces*/*Chlamydomonas*) was defined as the outgroup in parsimony runs.

for the taxa from which both sequences are known (fig. 3) and (2) that nuclear gene duplication for *gap*C did not occur prior to the separation of monocotyledons and dicotyledons studied here (see Results), we wished to compare and contrast these genes' molecular phylogenies against prevalent views of angiosperm evolution that are derived from both paleobotanical evidence and deduction from extant forms.

The origins and early course of angiosperm evolution remain obscure because of gaps in and difficulties of interpretation of the angiosperm fossil record (Stewart 1983, pp. 365–369). The lower Cretaceous is widely accepted as the starting point of angiosperm evolution (Beck 1976; Lidgard and Crane 1988; Les et al. 1991), yet *interpretation* of fossil plant forms is, for various reasons, to a large extent subjective (Krassilov 1977; Thomas and Spicer 1987, pp. 215–231; Meeuse 1990, pp. 11–52). The most unifying character shared by angiosperms that might define them as a group, double fertilization, also occurs in the gymnosperm *Ephedra* (Krassilov 1991; Friedman 1992) and cannot be applied to fossil forms (Thomas and Spicer 1987, p. 217), complicating the definition of "an angiosperm."

A number of theories exist that (1) address the origins of flowering plants and (2) should be testable with molecular phylogenetic methods. The most widely recognized of these theories, the Euanthial theory (Arber and Parkin 1907), assumes that the angiosperm flower derived from bisexual strobilus-like gymnosperm reproductive structures bearing many spirally arranged ovulate and pollen organs. Under the Euanthial theory, polymeric strobiloid flowers of the Magnoliaceae were interpreted (Bessey 1915) as the most primitive among extant angiosperms, more simple forms being derived through reduction and fusion of floral parts. The Pseudanthial theory (von Wettstein 1924, pp. 515–541; Zimmermann 1930, pp. 324–330), by contrast, assumes that angiosperm flowers derived from unisexual gymnosperm reproductive structures and views small, simple, unisexual flowers, such as those found among extant Hamamelididae and Piperales, as the most primitive. Proponents of polyphylesis for angiosperms (Meeuse 1990, pp. 11–52; Krassilov 1991) stipulate that flowering

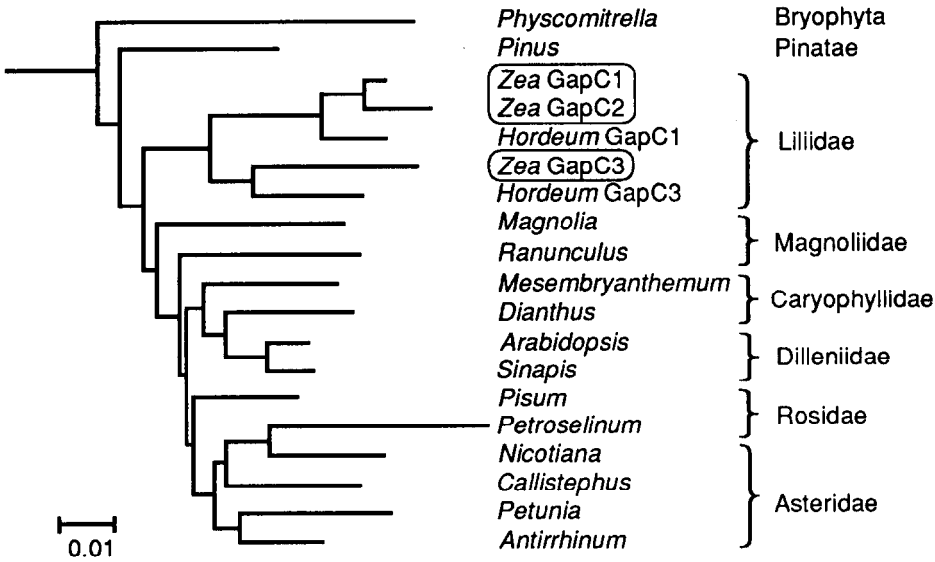


FIG. 5.—NJ tree for a matrix of K_0 values (average for 450 sites) from partial *gap*C sequences (696 total sites/sequence) from 19 plants, using yeast as the outgroup. The length of sequences compared was limited by the shortest sequence, *gap*C3. Subclasses of angiosperms studied are indicated as in table 1. The scale bar indicates 0.01 substitutions/nondegenerate site.

plants derived from separate and distinct gymnospermous clades, the fossil histories of which can be traced into Jurassic and Triassic strata; double fertilization is considered either a parallelism in modern angiosperms (Meeuse 1990, pp. 85–89) or as not unique to the group (Krassilov 1991). Still other authors homologize the angiosperm female reproductive structure with its male counterpart from gymnosperms rather than with the female, thus invoking gamoheterotrophy (sexual transmutation) in angiosperm origins (Meyen 1986), a process known to occur in mutants of *Antirrhinum* (Sommer et al. 1990) and implicated in the evolution of maize (Iltis 1983).

Thus, although alternative theories exist, the Euanthial theory constitutes the basis of most modern angiosperm taxonomic systems (see Friis et al. 1987). We performed the experiments described here to test a clear prediction of this theory: if all angiosperms derived from magnolean ancestors, then we must expect the genes of extant magnoliids to be borne on the deepest branch within the angiosperm portion of the corresponding tree, to the exclusion of other flowering plants. For neither *gap*C nor *rbc*L of *Magnolia* is this the case (figs. 3 and 4). Furthermore, both *Magnolia* and *Ranunculus* branch within the dicotyledonous portion of the *gap*C tree (fig. 5), whereas the deepest bifurcation found for angiosperm sequences is that between monocots and dicots surveyed (see Results). The nuclear (*gap*C) and cpDNA (*rbc*L) gene trees yield the same result in this respect, suggesting a nonoutgroup status of magnoliids relative to other angiosperms. This finding is substantiated by the analysis of the complete *gap*C gene family of maize (fig. 5; also see Results). This important result is incongruent with predictions derived from the Euanthial theory, although few species have been studied for both genes. Moreover, within the dicotyledons, the affinities (depicted in fig. 5) between angiosperm subclasses within the dicotyledons are, in general, consistent with those found in modern phylogenetic systems (see Steb-

bins 1974, pp. 246–353; Takhtajan 1980; Ehrendorfer 1991), although the extent of gene duplications prior to speciation within the dicotyledonous *gap*C sequences analyzed is difficult to assess at present. The positions of the *Nicotiana* and/or *Petroselinum gap*C sequences appear anomalous and may reflect paralogy similar to that found within the *gap*C gene families of maize and barley. Further molecular studies are necessary to identify paralogous *gap*C comparisons and to provide improved levels of resolution concerning angiosperm phylogeny. Nonetheless, both a basal position of magnoiliids within dicotyledons and derivation of ranunculiids from these are implicit in figure 5, as is an origin of Asteridae from rosiid ancestors, as postulated elsewhere (Takhtajan 1980; Ehrendorfer 1991).

Previous molecular phylogenetic studies that addressed the question of magnolean ancestry for angiosperms have brought forth evidence suggesting that the deepest branch within extant angiosperms separates monocotyledons from their dicotyledonous counterparts, not magnoliids from other angiosperms. Our previous analysis of two nuclear genes (Martin et al. 1989), as well as analyses of nuclear (Zimmer et al. 1989; Troitsky et al. 1991) and chloroplast (Troitsky et al. 1991) rRNA genes, are congruent, in this respect, with the data for *gap*C and *rbc*L protein-coding sequences presented here. Thus, to date, molecular phylogenetic analyses of nucleotide sequences do not provide support for an ancestral position of magnoliids within angiosperm taxonomic systems (Takhtajan 1980; Cronquist 1988, pp. 195–207). The finding that magnolean representatives do not assume an outgroup position within molecular phylogenetic trees does not argue directly against a strobiloid nature of the earliest angiosperm reproductive structures but would suggest that, if such were the case, reduction and fusion processes took place independently within the graminaceous monocotyledonous and dicotyledonous lineages surveyed, subsequent to the separation of magnoliids from other angiosperms. This finding may, on the one hand, evoke vehement criticism, as did our previous study (see Crane et al. 1989), yet may, on the other hand, be welcomed as supportive evidence for the arguments of those who interpret the angiosperm evolution as having occurred from other than strictly magnolean ancestors (Burger 1977; Krassilov 1977, 1991; Meeuse 1987, pp. 183–187; see Stewart 1983, pp. 381–389). Recent molecular analyses suggest that perhaps *no* extant gymnosperm group may represent a sister group to angiosperms (Troitsky et al. 1991; Hasebe et al. 1992, and accepted). Furthermore, recent analyses of *rbc*L sequences from the aquatic magnoliid *Nelumbo* (Les et al. 1991) are not incongruent with the extremely ancient separation of this species from other angiosperms, suggested by Meeuse (1990, pp. 65–85).

The time of angiosperm origins relates to the question of which gymnospermous lineages shared the most recent common ancestors with angiosperms and thus bears on the nature of the structure(s) from which the flowers of angiosperms may derive. Estimates of angiosperm age as deduced from molecular sequences have been reported (Ramshaw et al. 1972; Martin et al. 1989; Wolfe et al. 1989) and are controversial. The estimates of Ramshaw et al. (1972) were based on cytochrome *c* amino acid sequences and obtained divergence times between angiosperms that well exceeded the age of land plants, probably because of quite large stochastic errors resulting from the small number of sites ($\sim$100) used for comparison and from the extremely low divergence observed between plant cytochrome *c* sequences (Boulter et al. 1972). We previously estimated angiosperm age, on the basis of nuclear sequence data, at >300 Mybp—i.e., roughly at a time during which the major groups of gymnosperms diverged from one another. Wolfe et al. (1989) used chloroplast DNA sequence data to estimate

angiosperm age at ~200 Mybp. Their analysis relied entirely on *Nicotiana* as the single dicotyledonous representative, although it had previously been shown (Wolfe et al. 1987) that the rate of nucleotide substitution in the *Nicotiana* lineage is significantly slower than that in other angiosperms analyzed. Therefore, it may have been preferable to use a different dicot for analysis (Wolfe et al. 1989). Conflicting estimates of angiosperm age that are based on molecular sequence data have been discussed (Clegg 1990).

In light of this, we compared divergence at nondegenerate sites in *gap*C and *rbc*L sequences (table 5) from the nine plants for which both sequences were analyzed (in fig. 4) and used these to estimate approximate divergence times for the taxa in question. We wished to compare molecular estimates of divergence time with the age of those events that can be relatively reliably dated on the basis of the fossil record of land plant evolution. For this, we first tested the concatenated *gap*C/*rbc*L sequences, for significant difference in substitution rate at nondegenerate sites; these analyses showed that the concatenated maize sequence is evolving at a significantly higher rate ($P < 0.05$) and that the concatenated *Magnolia* sequence is evolving more slowly (though not significantly so) than other angiosperm sequences studied (data not shown). For estimation of divergence time, we used both the calculation method of Li and Graur (1991, pp. 117–118), which assumes constant substitution rate in individual lineages, and the method of Li and Tanimura (1987), which does not assume constant rate in individual lineages.

We used a divergence time (T_1) of 450 Mybp for bryophytes and spermatophytes, which likely represents an upper bound for the divergence of bryophytes and spermatophytes. The justification for this time is as follows: The first land plants may be of Ordovician age, as documented by conductive elements (Gensel and Andrews 1987),

Table 5
Divergence at Nondegenerate Sites in *gap*C and *rbc*L Sequences

	n; $K_0 \pm$ SE[a]	
	*gap*C (L_0 = 648)	*rbc*L (L_0 = 927)
Individual sequences:		
Bryophyte-spermatophyte	8; 0.1374 ± 0.0157	8; 0.0475 ± 0.0073
Conifer-angiosperm	7; 0.0879 ± 0.0123	7; 0.0488 ± 0.0075
Monocot-dicot[b]	10; 0.0850 ± 0.0120	10; 0.0454 ± 0.0072
Dicot subclasses[b]	9; 0.0711 ± 0.0110	9; 0.0320 ± 0.0061
	*gap*C/*rbc*L (L_0 = 1,575)	*gap*C/*rbc*L (L_0 = 1,575)[c]
Concatenated sequences:		
Bryophyte-spermatophyte	8; 0.0841 ± 0.0076	7; 0.0831 ± 0.0076
Conifer-angiosperm	7; 0.0652 ± 0.0065	6; 0.0631 ± 0.0065
Monocot-dicot[b]	10; 0.0630 ± 0.0065	5; 0.0582 ± 0.0063
Dicot subclasses[b]	9; 0.0486 ± 0.0057	9; 0.0486 ± 0.0057

[a] n = number of comparisons; L_0 = number of nondegenerate sites in an average comparison; K_0 = average divergence at nondegenerate sites of n comparisons across the given node; and SE = mean standard error of individual K_0 values across the node.

[b] Applies only to those subclasses surveyed in fig. 4.

[c] Excludes comparisons involving maize, since this lineage is evolving at a significantly higher rate than the other *gap*C/*rbc*L sequences studied (see text).

and various spores indicative of land plant life are known from Llandoverian (~430 Mybp) strata (Thomas and Spicer 1987, pp. 6–17). Vascular plants, more advanced than bryophytes, had become well established (Gensel and Andrews 1987) by the mid-Silurian (~420 Mybp). Stomatopores are known from *Cooksonia* specimens (Edwards et al. 1986) of the basal Devonian (~410 Mybp). Chloroplast DNA sequence data suggest that the lineage that gave rise to bryophytes and ferns may have separated very early from that which gave rise to spermatophytes (Yoshinaga et al. 1992).

For the divergence of conifers from angiosperms, we used a divergence time (T_2) of 330 Mybp, since conifers are extremely unlikely candidates for angiosperm antecedents (Doyle and Donoghue 1986; Cronquist 1988, pp. 139–152), making them a good gymnosperm outgroup to the angiosperms. The earliest seed plants are of Upper Devonian–Lower Carboniferous age (~360 Mybp; Chaloner 1989; Galtier and Rowe 1989); conifers can be traced, in the fossil record (Scott and Chaloner 1983; Meyen 1984), into the Upper Carboniferous (~320 Mybp) and thus must have been distinct from angiosperm progenitors before that time. The Pinaceae themselves can be traced into Upper Triassic strata (Thomas and Spicer 1987, pp. 169–192).

The results of our molecular estimates for angiosperm age are shown in table 6. We note that the molecular estimates in table 6 that are obtained, on the basis of the conifer-angiosperm calibration point (and vice versa), for times of bryophyte-spermatophyte divergence agree reasonably well with the corresponding times documented in fossil material. Using the calibration point of 450 Mybp for bryophyte-spermatophyte divergence, we obtain estimates for conifer-angiosperm divergence (~330 Mybp) that are relatively close to the upper bound from fossil evidence, with an uncertainty of ~50 Mybp, suggesting that the molecular and paleobotanical data are coherent with regard to bryophyte-spermatophyte and conifer-angiosperm separation. In our opinion, these data suggest that, at least for the *gap*C and *rbc*L data set at hand, the molecular estimates derived, by the same means, for *angiosperm* age (~300 ± 40 Mybp) should

Table 6
Approximate Divergence Time between Land Plant Groups

Comparison	Evidence from Fossil Record (Mybp)	Molecular Estimates from *gap*C/*rbc*L Sequences[a] (Mybp)			
		I	II	III	IV
Bryophyte-spermatophyte	450[b]	NA	441 ± 40	NA	464 ± 40
Conifer-angiosperm	330[b]	337 ± 41	NA	320 ± 41	NA
Monocot-dicot[c]	(see text)	311 ± 34	304 ± 34	292 ± 34	301 ± 34
Dicot subclasses[c]	(see text)	260 ± 33	254 ± 33	266 ± 33	272 ± 33

[a] Four estimates from molecular sequence data are given: I, the method of Li and Graur (1991, pp. 117–118) using the known divergence time (T_1) as 450 Mybp for bryophytes and spermatophytes; II, the method of Li and Graur (1991, pp. 117–118) using the known divergence time (T_2) as 330 Mybp for conifers and angiosperms; III, the method of Li and Tanimura (1987) using the known divergence time (T_1) as 450 Mybp for bryophytes and spermatophytes; and IV, the method of Li and Tanimura (1987) using the divergence time of 330 Mybp for conifers and angiosperms. Estimates III and IV are based on branch lengths in fig. 4, whereby the maize and *Magnolia* lineages were not considered because they have the longest and shortest branches, respectively; branch lengths within the dicots were averaged to yield a mean dicot subclass branch length for calculation. Standard errors attached to molecular estimates of divergence time were derived by converting average standard errors of K_0 values into approximate time by the procedure indicated (I–IV). The largest standard errors for the various molecular estimates (I–IV) of divergence time for given taxa were conservatively attached to all estimates. NA = not applicable.

[b] For a justification of divergence times used, see text.

[c] Applies only to those subclasses surveyed in fig. 4.

not be wholly anomalous. Criticisms of our previous estimates of angiosperm age rested on (1) lack of gymnosperm sequences (Crane et al. 1989), (2) gene duplication arguments (Clegg 1990), or (3) nucleotide biases (Drouin and Dover 1990) found in chalcone synthase, *gap*A, and the chloroplast GAPDH enzyme (Brinkmann et al. 1987; Quigley et al. 1988, 1989). These criticisms have been addressed here in detail (see Results).

The question of angiosperm age hinges on interpretation, both of fossil material (Cleal 1989) and, to some extent, of molecular data as well. For some fossil specimens, notably *Sanmiguelia* (Brown 1956), the pre-Cretaceous age, rather than the morphology, argued against angiospermous character for the taxon (Daghlian 1981). ["Although no conclusive statement can be made concerning the affinities of *Sanmiguelia* until structurally preserved specimens are collected, . . . an outright rejection of this problematic fossil as a monocotyledon or monocotyledon precursor may be unwarranted. The Triassic age rather than the morphological considerations appears to be the main obstacle to accepting *Sanmiguelia* as a possible angiosperm" (Daghlian 1981, p. 523).] Our data and resulting estimates of angiosperm age are quite compatible with the interpretation given by Cornet (1986, 1989)—i.e., that his recent fossil finds represent well-preserved angiosperm reproductive structures of *S. lewisii* from late-Carnian deposits, ~220 Myr of age. Cornet's interpretation of these finds has been debated. ["Cornet's third conclusion, that *Sanmiguelia* is an angiosperm, is the most difficult to assess because this is a complex issue which hinges on the correct interpretation of specific details in the fossil material" (Crane 1988, p. 778).] Our estimates for angiosperm age are compatible with the general view (Krassilov 1977) that pre-Cretaceous age would represent no argument against the angiospermous nature of a given specimen. The biology of the earliest angiosperms cannot be directly inferred from molecular sequences, and the evolutionary mechanisms that gave rise to the success of angiosperms, as measured in species number, remain elusive. Whereas it is widely believed that early angiosperms underwent adaptive radiation (see Thomas and Spicer 1987, pp. 232–245), nonadaptive morphological changes may be very important in evolution (Nei 1987, pp. 414–431). It is not known to what extent nonadaptive changes may have been instrumental in angiosperm evolution. Molecular analyses are congruent in some major points, as discussed above, and are inconsistent with two prevalent basic postulates concerning flowering-plant evolution: (1) Creataceous angiosperm age and (2) the Euanthial theory. Of course, more species and, more important, more genes need to be studied before a robust picture of land plant evolution will begin to emerge. Molecular phylogenies can contribute to our understanding of plant evolution and yield insights into questions historically reluctant to relinquish answers.

Acknowledgments

We wish to thank the Gesellschaft für Biotechnologische Forschung, Braunschweig, for permission to use their computer facilities, and we wish to thank Doug Soltis and Uta-Regina Böhle for critical comments on the manuscript. This work was supported in part by Deutsche Forschungsgemeinschaft grant Ma 1426/1-1 to W.M.

LITERATURE CITED

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

AXELROD, D. I. 1959. Poleward migration of early angiosperm flora. Science **130**:203–207.
———. 1970. Mesozoic paleogeography and early angiosperm history. Bot. Rev. **36**:227–319.
BECK, C. B. 1976. Origin and early evolution of angiosperms: a perspective. Pp. 1–10 *in* C. B. BECK, ed. Origin and early evolution of angiosperms. Columbia University Press, New York.
BESSEY, C. E. 1915. The phylogenetic taxonomy of flowering plants. Ann. Mo. Bot. Garden **2**: 109–164.
BOULTER, D., D. PEACOCK, A. GUISE, J. T. GLAEVES, and G. ESTABROOK. 1979. Relationships between the partial amino acid sequences of plastocyanin from ten families of flowering plants. Phytochemistry **18**:603–608.
BOULTER, D., J. A. M. RAMSHAW, E. W. THOMPSON, M. RICHARDSON, and R. H. BROWN. 1972. A phylogeny of higher plants based on the amino acid sequences of cytochrome *c* and its biological implications. Proc. R. Soc. Lond. [Biol.] **181**:441–455.
BRINKMANN, H., R. CERFF, M. SALOMON, and J. SOLL. 1989. Cloning and sequence analysis of cDNAs encoding the cytosolic precursors of subunits GapA and GapB of chloroplast glyceraldehyde-3-phosphate dehydrogenase from pea and spinach. Plant Mol. Biol. **13**:81–94.
BRINKMANN, H., P. MARTINEZ, F. QUIGLEY, W. MARTIN, and R. CERFF. 1987. Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. J. Mol. Evol. **26**:24–33.
BROWN, R. W. 1956. Palm-like plants from the Dolores Formation (Triassic), southwestern Colorado. U.S. Geol. Surv. Prof. Pap. **274H**:205–209.
BURGER, W. C. 1977. The Piperales and the monocots: alternate theories for the origin of the monocot flower. Bot. Rev. **43**:345–393.
CERFF, R., and S. CHAMBERS. 1979. Subunit structure of higher plant glyceraldehyde-3-phosphate dehydrogenases (EC 1.2.1.12 and 1.2.1.13). J. Biol. Chem. **254**:6094–6098.
CERFF, R., and K. KLOPPSTECH. 1982. Structural diversity and differential light control of mRNAs coding for angiosperm glyceraldehyde-3-phosphate dehydrogenases. Proc. Natl. Acad. Sci. USA **79**:7624–7628.
CHALONER, W. 1989. A missing link for seeds? Nature **340**:185.
CHOJECKI, J. 1986. Identification and characterisation of a cDNA clone for cytosolic glyceraldehyde-3-phosphate dehydrogenase in barley. Carlsberg Res. Commun. **51**:203–210.
CLEAL, C. 1989. Evolution in hidden forests. Nature **339**:16.
CLEGG, M. 1990. Dating the monocot-dicot divergence. Trend Ecol. Evol. **5**:1–2.
CORNET, B. 1986. The leaf venation and reproductive structures of a Late Triassic angiosperm, *Sanmiguelia lewisii*. Evol. Theory **7**:231–309.
———. 1989. Late Triassic angiosperm-like pollen from the Richmond rift basin of Virginia, USA. Paleontographica **213**:37–87.
COVE, D. J., and N. W. ASHTON. 1984. The hormonal regulation of gametophytic regulation in bryophytes. Pp. 177–201 *in* A. F. DYER and J. G. DUCKETT, eds. Experimental biology of bryophytes. Academic Press, London.
CRANE, P. 1985. Phylogenetic analysis of seed plants and the origin of angiosperms. Ann. Mo. Bot. Garden **72**:716–793.
CRANE, P. R. 1988. Review of Cornet, B., The leaf venation and reproductive structures of a Late Triassic angiosperm, *Sanmiguelia lewisii*. Taxon **36**:788–789.
CRANE, P. R., E. M. FRIIS, and K. R. PEDERSEN. 1986. Lower Cretaceous angiosperm flowers: fossil evidence on early radiation of Dicotyledons. Science **232**:852–854.
CRANE, P. R., M. J. DONOGHUE, J. A. DOYLE, and E. M. FRIIS. 1989. Critique of Martin *et al.*, 1989. 'Angiosperm Origins' Nature **342**:131.
CRONQUIST, A. 1988. The evolution and classification of flowering plants. New York Botanical Garden, New York.
DAGHLIAN, C. P. 1981. A review of the fossil record of monocotyledons. Bot Rev. **47**:517–555.
DAHLGREN, R., and K. BREMER. 1985. Major clades of the angiosperms. Cladistics **1**:349–368.

DEVEREUX, J., P. HAEBERLI, and O. SMITHIES. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.

DOVER, G. A. 1987. DNA turnover and the molecular clock. J. Mol. Evol. **26**:47–58.

———. 1989. Linkage disequilibrium and molecular drive in the rDNA multigene family. Genetics **122**:249–252.

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.

DROUIN, G., and G. A. DOVER. 1990. Independent gene evolution in the potato actin gene family demonstrated by phlyogenetic procedures for resolving gene conversions and the phylogeny of angiosperm actin genes. J. Mol. Evol. **31**:132–150.

DUGAICZYK, A., J. A. HARON, E. M. STONE, O. E. DENNISON, K. N. ROTHBLUM, and R. J. SCHWARTZ. 1983. Cloning and sequencing of a deoxyribonucleic acid copy of glyceraldehyde-3-phosphate dehydrogenase messenger ribonucleic acid isolated from chicken muscle. Biochemistry **22**:1605–1613.

EDWARDS, D., U. FANNING, and J. B. RICHARDSON. 1986. Stomata and sterome in early land plants. Nature **323**:438–440.

EHRENDORFER, F. 1991. Spermatophyta Samenpflanzen. Pp. 699–826 *in* P. SITTE, H. ZIEGLER, F. EHRENDORFER, and A. BREZINSKY, eds. Lehrbuch der Botanik. Gustav Fischer, Stuttgart.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

FOTHERGILL-GILMORE, L. A. 1986. The evolution of the glycolytic pathway. Trends Biochem. Sci. **11**:47–51.

FRIEDMAN, W. E. 1992. Evidence for a pre-angiosperm origin of endosperm: implications for the evolution of flowering plants. Science **255**:336–338.

FRIIS, E. M., W. G. CHALONER, and P. R. CRANE. 1987. Introduction to angiosperms. Pp. 1–16 *in* E. M. FRIIS, W. G. CHALONER, and P. R. CRANE, eds. The origins of angiosperms and their biological consequences. Cambridge University Press, Cambridge.

GALTIER, J., and N. P. ROWE. 1989. A primitive seed-like structure and its implications for early gymnosperm evolution. Nature **340**:225–227.

GENSEL, P. G., and H. N. ANDREWS. 1987. The evolution of early land plants. Am. Sci. **75**: 478–489.

GOLENBERG, E. M., D. E. GIANASSI, M. T. CLEGG, C. J. SMILEY, M. DURBIN, D. HENDERSON, and G. ZURAWSKI. 1990. Chloroplast DNA sequence from a Miocene *Magnolia* species. Nature **344**:656–658.

GOUY, M., and W.-H. LI. 1989. Molecular phylogeny of the kingdoms Animalia, Plantae, and Fungi. Mol. Biol. Evol. **6**:109–122.

HASEBE, M., M. ITO, R. KOFUJI, K. IWATSUKI, and K. UEDA. 1992. Phylogeny of land plants inferred from *rbc*L gene sequences. Plant Mol. Evol. Newslett. **1**:5–6.

HASEBE, M., R. KOFUJI, M. ITO, M. KATO, K. IWATSUKI, and K. UEDA. Phylogeny of gymnosperms inferred from *rbc*L gene sequences. Bot. Mag. (Tokyo) (accepted).

HORI, H., B.-L. LIM, and S. OSAWA. 1985. Evolution of green plants as deduced from 5S rRNA sequences. Proc. Natl. Acad. Sci. USA **82**:820–823.

ILTIS, H. 1983. From teoscinte to maize: the catastrophic sexual transmutation. Science **222**: 886–894.

JIN, L., and M. NEI. 1990. Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol. Biol. Evol. **7**:82–102.

KRASSILOV, V. A. 1977. The origin of angiosperms. Bot. Rev. **43**:143–176.

———. 1991. The origin of angiosperms: new and old problems. Trends Ecol. Evol. **6**:215–220.

LAPEYRE, B., and F. AMALRIC. 1985. A powerful method for the preparation of cDNA libraries: isolation of cDNA encoding a 100 kD nucleolar protein. Gene **37**:215–220.

LES, D. H., D. K. GARVIN, and C. F. WIMPEE. 1991. Molecular evolutionary history of ancient aquatic angiosperms. Proc. Natl. Acad. Sci. USA **88**:10119–10123.

LI, W.-H., and D. GRAUR. 1991. Fundamentals of molecular evolution. Sinauer, Sunderland, Mass.

LI, W.-H., and M. TANIMURA. 1987. The molecular clock runs more slowly in man than in apes and monkeys. Nature **326**:93–96.

LI, W.-H., C.-I. WU, and C.-C. LUO. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150–174.

LIDGARD, S., and P. R. CRANE. 1988. Quantitative analyses of the early angiosperm radiation. Nature **331**:344–346.

MARTIN, W., A. GIERL, and H. SAEDLER. 1989. Molecular evidence for pre-Cretaceous angiosperm origins. Nature **339**:46–48.

MARTIN, W., T. LAGRANGE, Y.-F. LI, C. BISANZ-SEYER, and R. MACHE. 1990. Hypothesis for the evolutionary origin of the chloroplast ribosomal protein L21 of spinach. Curr. Genet. **18**:553–556.

MARTIN, W., C. C. SOMERVILLE, and S. LOISEAUX-DE GOER. Molecular phylogenies of plastid origins and algal evolution. J. Mol. Evol. (accepted).

MARTINEZ, P., W. MARTIN, and R. CERFF. 1989. Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolyic glyceraldehyde-3-phosphate dehydrogenase from maize. J. Mol. Biol. **208**:551–565.

MAXAM, A. M., and W. GILBERT. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.

MEEUSE, A. D. J. 1987. All about angiosperms. Eburon, Delft.

———. 1990. Flowers and fossils. Eburon, Delft.

MEYEN, S. V. 1984. Basic features of gymnosperm systematics and phylogeny as evidenced by the fossil record. Bot. Rev. **50**:1–111.

———. 1986. Origin of the angiosperm gynoecium by gamoheterotropy. Bot. J. Linnean Soc. **97**:171–178.

MURRAY, N. 1983. Phage lambda and molecular cloning. Pp. 395–432 *in* R. W. HENDRIX, J. W. ROBERTS, F. W. STAHL, and R. A. WEISBERG, eds. Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

———. 1991. Relative efficiencies of different tree-making methods for molecular data. Pp. 90–128 *in* M. M. MIYMOTO and J. L. CRACRAFT, eds. Phylogenetic analysis of DNA sequences. Oxford University Press, Oxford.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. Mol. Biol. Evol. **3**:418–426.

NIESBACH-KLÖSGEN, U., E. BARZEN, J. BERNHARDT, W. RHODE, ZS. SCHWARZ-SOMMER, H. J. REIF, U. WIENAND, and H. SAEDLER. 1987. Chalcone synthase hgenes in plants: a tool to study evolutionary relationships. J. Mol. Evol. **26**:213–225.

NELSON, K., T. S. WHITTAM, and R. K. SELANDER. 1991. Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. Proc. Natl. Acad. Sci. USA **88**:6667–6671.

PRAGER, E. M., and A. C. WILSON. 1988. Ancient origin of lactalbumin from lysozyme: analysis of DNA and amino acid sequences. J. Mol. Evol. **27**:326–335.

QUIGLEY, F., W. MARTIN, and R. CERFF. 1988. Intron conservation across the prokaryote-eukaryote boundary: structure of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. Proc. Natl. Acad. Sci. USA **85**:2672–2676.

QUIGLEY, F., W. MARTIN, and R. CERFF. 1988. Intron conservation across the prokaryote-eukaryote boundry: structure of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. Proc. Natl. Acad. Sci. USA **85**:2672–2676.

RAMSHAW, J. A. M., D. L. RICHARDSON, B. T. MEATYARD, R. H. BROWN, M. RICHARDSON, E. W. THOMPSON, and D. BOULTER. 1972. The time of origin of the flowering plants determined by using amino acid sequence data of cytochrome *c*. New Phytol. **71**:773–779.

RUSSELL, D. A., and M. M. SACHS. 1989. Differential expression and sequence analysis of the maize glyceraldehyde-3-phosphate dehydrogenase gene family. Plant Cell **1**:793–803.

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

SAITOU, N., and T. IMANISHI. 1989. Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. Mol. Biol. Evol. **6**:514–525.

SANGER, F., S. NICKLEN, and A. R. COULSON. 1977. DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

SCHWARZ-SOMMER, ZS., H. CUYPERS, A. GIERL, P. A. PETERSON, and H. SAEDLER. 1985. Plant transposable elements generate the DNA sequence diversity needed in evolution. EMBO J. **4**:591–597.

SCHWARZ-SOMMER, ZS., N. SHEPARD, E. TACKE, A. GIERL, W. RHODE, L. LECLERCQ, M. MATTES, R. BERNTGEN, P. A. PETERSON, and H. SAEDLER. 1987. Influence of transposable elements on the structure and function of the *A1* gene of *Zea mays*. EMBO J. **6**:287–294.

SCOTT, A. C., and W. G. CHALONER. 1983. The earliest conifer from the Westphalian B of Yorkshire. Proc. R. Soc. Lond. [Biol.] **220**:163–182.

SHIH, M.-C., P. HEINRICH, and H. M. GOODMAN. 1991. Cloning and chromosomal mapping of nuclear genes encoding chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenase from *Arabidopsis thaliana*. Gene **104**:133–138.

SOMMER, H., J.-P. BELTRAN, P. HUISER, H. PAPE, W.-E. LÖNNING, H. SAEDLER, and ZS. SCHWARZ-SOMMER. 1990. *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus:* the protein shows homology to transcription factors. EMBO J. **9**:605–613.

SOURDIS, J., and M. NEI. 1988. Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. Mol. Biol. Evol. **5**:298–311.

STEWART, W. N. 1983. Paleobotany and the evolution of plants. Cambridge University Press, Cambridge.

TAKHTAJAN, A. L. 1980. Outline of the classification of flowering plants (Magnoliophyta). Bot. Rev. **46**:225–359.

THOMAS, B. A., and R. A. SPICER. 1987. The evolution and paleobiology of land plants. Croom Helm, London.

TROITSKY, A. V., Y. F. MELEKHOVETS, G. M. RAKHIMOVA, V. K. BOBROVA, K. M. VALIEJO-ROMAN, and A. S. ANTONOV. 1991. Angiosperm origin and early seed plant evolution deduced from rRNA sequence comparisons. J. Mol. Evol. **32**:253–261.

TSO, J. Y., X.-H. SUN, T.-H. KAO, K. REECE, and R. WU. 1985*a*. Isolation and characterisation of rat and human glyceraldehyde-3-phosphate dehydrogenase cDNAs: genomic complexity and molecular evolution of the gene. Nucleic Acids Res. **13**:2485–2502.

TSO, J. Y., X.-H. SUN, and R. WU. 1985*b*. Structure of two unlinked *Drosophila melanogaster* glyceraldehyde-3-phosphate dehydrogenase genes. J. Biol. Chem. **260**:8220–8228.

VIEIRA, J., and J. MESSING. 1982. The pUC plasmids, an M13 mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.

VON WETTSTEIN, R. 1924. Handbuch der systematischen Botanik, 3. Aufl. Franz Deutige, Leibzig.

WOLFE, K. H., M. GOUY, Y. W. YANG, P. SHARP, and W.-H. LI. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA **86**:6201–6205.

WOLFE, K. H., W.-H. LI, and P. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84**: 9054–9058.

YARBROUGH, P. O., M. A. HAYDEN, L. A. DUNN, P. S. VERMESCH, M. R. KLASS, and R. M. HECHT. 1987. The glyceraldehyde-3-phosphate dehydrogenase gene family in the nematode

Caenorhabditis elegans: isolation and characterisation of one of the genes. Biochim. Biophys. Acta **908**:21–33.

YOSHINAGA, K., Y. KUBOTA, T. ISHII, and K. WADA. 1992. Nucleotide sequence of *atpB, rbcL, trnR, dedB,* and *psaI* chloroplast genes from a fern *Angiopteris lygodiifolia:* a possible emergence of Spermatophyta lineage before the separation of Bryophyta and Pteridophyta. Plant Mol. Biol. **18**:79–82.

ZIMMER, E. A., R. K. HAMBRY, M. L. ARNOLD, D. A. LEBLANC, and E. L. THERIOT. 1989. Ribosomal RNA phylogenies and flowering plant evolution. Pp. 205–215 *in* B. FERNHOLM, K. BREMER, and H. JÖRNVALL, eds. The hierarchy of life. Elsevier, Amsterdam.

ZIMMERMANN, W. 1930. Die Phylogenie der Pflanzen: ein Überblick über Tatsachen und Probleme. Gustav Fischer, Jena.

MASATOSHI NEI, reviewing editor

Received June 12, 1992; revision received August 18, 1992

Accepted August 19, 1992