Discovery of Paralogous Nuclear Gene Sequences Coding for the Second-Largest Subunit of RNA Polymerase II (*RPB2*) and Their Phylogenetic Utility in Gentianales of the Asterids

Bengt Oxelman and Birgitta Bremer

Department of Systematic Botany, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

Paralogous sequences of the *RPB2* gene are demonstrated in the angiosperm order Gentianales. Two different copies were found by using different PCR primer pairs targeting a region that corresponds to exons 22–24 in the *Arabidopsis RPB2* gene. One of the copies (*RPB2*-d) lacks introns in this region, whereas the other has introns at locations corresponding to those of green plants previously investigated. When analyzed with other available *RPB2* sequences from this region, all 28 *RPB2*-d sequences obtained from the Gentianales and the four sequences from the Lamiales form a monophyletic group, together with a previously published tomato cDNA sequence. The substitution patterns, relative rates of change, and nucleotide compositions of the two paralogous *RPB2* exon regions are similar, and none of them shows any signs of being a pseudogene. Although multiple copies of similar, paralogous sequences can confound phylogenetic interpretations, the lack of introns in *RPB2*-d make a priori homology assessment easy. The phylogenetic utility of *RPB2*-d within the Gentianales is evaluated in comparison with the chloroplast genes *ndhF* and *rbcL.* The hierarchical information in the *RPB2*-d region sequenced is more incongruent with that of the plastid genes than the plastid genes are with each other as determined by incongruence length difference tests. In contrast to the plastid genes, parsimony-informative third codon positions of *RPB2* have a significantly higher rate of change than first and second positions. Topologically, the trees from the three genes are similar, and the differences are usually only weakly supported. In terms of support, *RPB2* gives the highest jackknife support per sequenced nucleotide, whereas *ndhF* gives the highest Bremer support per sequenced nucleotide. The *RPB2*-d locus has the potential to be a valuable nuclear marker for determination of phylogenetic relationships within the euasterid I group of plants.

Introduction

DNA sequencing of PCR-generated fragments has relaxed the requirements for the material used for DNA extraction and thus tremendously enlarged the potential for broadscale systematic studies based on DNA sequence data. In plants, the chloroplast genome and the nuclear ribosomal DNA (rDNA) sequences have been most popular. The chloroplast genome is relatively well known, with several complete sequences (e.g., Shinozaki et al. 1986; Hallick et al. 1993; Maier et al. 1995; Ohyama 1996; Wakasugi et al. 1997). A large fraction of the chloroplast genes are present in a single copy, a fact that simplifies location of orthologous regions for PCR. The nuclear genome has been less utilized as a source of phylogenetic information. Presumably, this is largely due to problems concerning determination of orthology of PCR-generated sequences from the multicopy part of the nuclear genome. In the case of the ribosomal DNA, concerted evolution (e.g., Zimmer et al. 1980) is thought to homogenize the many repeats in a way that facilitates phylogeny reconstruction. Other parts of the nuclear genome are usually much less known, and a large fraction are evidently multicopy gene families, complicating a priori homology assessment of potentially useful sequences (e.g., Mathews, Lavin, and Sharrock 1995; Waters 1995; Doyle, Kanazin, and Shoemaker 1996; Gottlieb and Ford 1996; Morton, Gaut, and Clegg 1996; Clegg, Cummings, and Durbin 1997). Moreover, it is possible that concerted evolution is incomplete in these regions (Wendel and Doyle 1998). Thus, it is generally thought that suitable target regions for phylogenetic studies should be single-copy regions or, minimally, regions in which orthologous copies of multigene families can be unambiguously identified.

In contrast to the chloroplast genome, there is a level in the plant taxonomic hierarchy at which suitable sequence regions of the nuclear genome have not yet been discovered (Soltis and Soltis 1998). At a level spanning approximately genera to families within orders, ribosomal DNA sequences are generally either too conservative (genes) or too divergent (spacer regions). The exon sequences of certain protein-coding genes are potentially useful at this level, but in addition to orthology problems, the exon sequences are often interrupted by large introns (Li and Graur 1991) which may be too divergent to be useful.

Nuclear RNA polymerases in eukaryotes belong to three different classes, referred to as RNA polymerase I, II, and III. Each enzyme is composed of two large (>100 kDa) and several smaller subunits, each of which is typically encoded by a unique single-copy gene. Some genes encode subunits common to two or three of the polymerases (Thuriaux and Sentenac 1992), and a few genes have been reported to be multicopy (e.g., *RPB3* and *RPB5* in *Arabidopsis*; Larkin, Hagen, and Guilfoyle

Abbreviations: Gene nomenclature followed here is composed of RP for RNA polymerase, followed by the letter A for pol I, B for pol II, or C for pol III. Finally, an arabic numeral denotes the subunit number. For example, *RPA2* encodes the second largest subunit of RNA polymerase I, and *RPC1* encodes the largest subunit of RNA polymerase III.

Key words: *RPB2,* single-copy nuclear genes, *ndhF, rbcL,* paralogy, angiosperms, Gentianales.

Address for correspondence and reprints: Bengt Oxelman, Department of Systematic Botany, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden. E-mail: bengt.oxelman@systbot.uu.se.

Mol. Biol. Evol. 17(8):1131–1145. 2000
© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

1999). Multiple copies of *RPB1* have been reported in *Glycine max* (Dietrich, Prenger, and Guilfoyle 1990), but these multiple copies may be due to the polyploid nature of the soybean genome (Kumar and Hymowitz 1989; Denton, McConaughy, and Hall 1998).

*RPB*2 encodes the second-largest subunit of nuclear RNA polymerase II, which is responsible for the transcription of protein-encoding genes (Sawadogo and Sentenac 1990) and is found in all eukaryotes. In the majority of organisms studied so far, *RPB2* has been reported to be single-copy (Thuriaux and Sentenac 1992; Ulmasov, Larkin, and Guilfoyle 1996). In plants, Southern blot experiments have been performed on *Arabidopsis* (Larkin and Guilfoyle 1993), tomato (Warrilow and Symons 1996), and *Rhododendron* (Denton, McConaughy, and Hall 1998), all suggesting just one copy of *RPB2* in these organisms. The large number of specific protein-protein contacts between subunits occurring within a three-dimensional framework is mainly determined by the structure and mutual interactions of the two largest subunits (Wlassoff, Kimura, and Ishihama 1999). Under the assumption that *RPB2* is generally single-copy, it has been suggested that it has great potential for phylogenetic studies (Sidow and Thomas 1993; Denton, McConaughy, and Hall 1998).

The only complete plant *RPB2* sequence published, from *Arabidopsis thaliana* (Larkin and Guilfoyle 1993), contains 3,564 bp of exon sequence with 24 interspersed introns. Denton, McConaughy, and Hall (1998) sequenced 1,668 bp of exon sequence in the 3′ part of the gene and found the intron locations there to be identical in the flowering plants *Arabidopsis* and *Aristolochia,* in *Ginkgo,* and in the liverwort *Marchantia.* Alignments of *RPB2* amino acid sequences from humans, *Drosophila,* fungi, and plants have revealed large parts to be highly conserved (Warrilow and Symons 1996). Denton, McConaughy, and Hall (1998) suggested that *RPB2* could be a valuable nuclear marker for tracing plant phylogeny. Moreover, their results indicated that the PCR approach may be used. With this background, we wanted to explore the potential of *RPB2* as a source for phylogenetic information for resolving family relationships in the Gentianales, an angiosperm order that is well supported by chloroplast DNA sequences (Olmstead et al. 1993; Backlund, Oxelman, and Bremer 2000). The cpDNA sequences provide good internal resolution in many cases (Bremer 1996; Backlund, Oxelman, and Bremer 2000), but the sister group relationship of the recently recognized family Gelsemiaceae (Struwe, Albert, and Bremer 1994) is not settled, and some intrafamilial relationships are still obscure (Backlund, Oxelman, and Bremer 2000).

Specifically, we investigated (1) whether single, orthologous *RPB2* sequences could be obtained using standard PCR methods; (2) the congruence and strength of the phylogenetic signal from *RPB2* sequences compared with sequences from the chloroplast protein-coding genes *rbcL* and *ndhF*; and (3) whether any increase in phylogenetic resolution results from combining *RPB2* sequences with the *rbcL* and *ndhF* sequences.

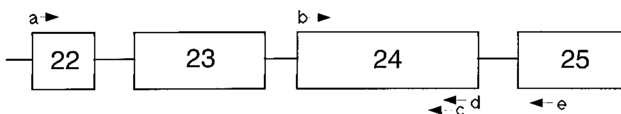


FIG. 1.—Locations and sequences of oligonucleotides used for PCR and sequencing of *RPB2.* Numbers refer to corresponding positions in the *Arabidopsis thaliana* sequence (GenBank accession number Z19121) for the 3′ nucleotide. Arrows indicate approximate locations and directions of primers. Degenerate codes: W = A and T, R = G and A, D = A, G, and T. a = 6935F: TCTGGTGAGGATGTTATCATTGGGAA; b = 7547F: GCTCTGCATAAATGTGGTTACCAGA; c = 7884R: GTTCTTGCAACCTCTGCATTCAAAAGA; d = 7901R: GGACGATATCWGTTTTGTTCTTGCAA; e = 8032R: GCTCCTGGAARAGCARTTTGCADGCATA.

Materials and Methods

Using published tomato (*Solanum*) cDNA and *Arabidopsis* nuclear DNA sequences of the *RPB2* gene, we constructed oligonucleotides for amplification of a region targeting ca. 250 amino acid residues, corresponding to exons 22–24 with intervening introns in *Arabidopsis.* Primer sequences are presented in figure 1. Initially, 6935F and 7884R were used, and the other primers were constructed to amplify templates where the original primer pair failed to produce a clear single product. Here we report previously unpublished DNA sequences of 32 taxa from the orders Gentianales and Lamiales. First, we aligned these with previously published genes coding for the second-largest subunit of RNA polymerase genes to confirm that the obtained sequences were plant *RPB2* genes. Second, we compared these (except *Chironia,* for which there are no available chloroplast DNA sequences) with previously published sequences for the chloroplast *ndhF* and *rbcL* genes (Oxelman, Backlund, and Bremer 1999; Backlund, Oxelman, and Bremer 2000). GenBank accession numbers and voucher information are given in table 1.

PCR reactions were performed using *Taq* polymerase (Advanced Biotechnologies) according to the instructions of the manufacturer on a Perkin-Elmer 480 thermal cycler. The thermal program started with a denaturation step at 95°C for 2 min, followed by 40 cycles of 95°C for 30 s, 55°C for 1 min, and 72°C for 2 min. The program was terminated with a 72°C step for 7 min. Amplification products were purified with the Qiaquick PCR kit according to instructions from the manufacturer (Qiagen). Sequencing reactions were performed using the PCR primers with the PRISM Ready Reaction Dye Deoxy Terminator FS kit (Applied Biosystems) and analyzed on an ABI 377 automated sequencer (Applied Biosystems).

Previously published *RPB2* sequences, as well as homologous parts of *RPA2* and *RPC2* (table 1), were used for manual alignment of the deduced amino acid sequences. This alignment was used for initial determination of that obtained sequences were plant *RPB2* sequences.

Frequently, nucleotide sequences of protein-coding genes are recoded to their inferred amino acid sequences before phylogenetic analysis. One reason for this is the presumed saturation of synonymous changes along long branches. On the other hand, substitutions leading to

amino acid changes may be too rare to reveal any information about relationships among closely related taxa, and some synonymous changes may be informative at certain levels of the tree. Agosti, Jacobs, and DeSalle (1996) suggested that nucleic acid sequences and their translated amino acid codings could be combined to address this problem. We performed combined, as well as separate, amino acid and DNA sequence analyses. Intron sequences were removed from the alignments prior to analysis, and the presence/absence of introns was coded as additional characters. The incongruence length difference (ILD; Farris et al. 1994) test was used to assess heterogeneity between the nucleic acid and amino acid matrices. Incongruent signals could help to identify possible cases of long-branch attraction (Felsenstein 1978) and caution against combining nucleic and amino acid codings.

By restricting a tree to *Arabidopsis* II, *Mostuea* II, *Mostuea* II-d, *Kopsia* II, and *Kopsia* II-d, we compared proportion of synonymous changes and transition/transversion ratios along the terminal branches for the two paralogs found in *Mostuea* and *Kopsia.* Taxonwise comparisons of the substitution patterns in the two paralogous *RPB2* genes were performed using Bonferroni-corrected *P*-values from chi-square tests for differences in synonymous/nonsynonymous changes and transitions/transversions for the two paralogous copies in each of the two taxa.

Eighteen land plant *RPB2* sequences were selected for relative rate tests according to the method of Steel et al. (1996) as implemented in the computer program r8s (Sanderson 1997). These 18 sequences were selected because they best represented the land plant *RPB2* sequences available and because they yielded a single most-parsimonious tree. The significance test in the method of Steel et al. (1996) is problematic for two reasons. First, the test variable, which is the difference between the distances from an ingroup sequence to an outgroup, is sensitive to which outgroup is chosen. Second, there is a severe problem arising from performing multiple tests on the same data. On average, we should, with 95% probability, expect 1 significant value out of 12 at the 0.05 level, even if all of the data have been drawn at random from the same probability distribution (see Rice 1989). Moreover, many of the tests will be phylogenetically nonindependent because some of the tests use the same taxa (Sanderson 1998). Because we were primarily interested in comparing whether the paralogous sequences evolved at different rates within the same organism, we decided to ignore the significance values given by the r8s program. Instead, we did all possible pairwise tests from the 18-taxon matrix and determined the location of the test statistics of the comparisons between paralogous sequences in the same organisms (i.e., *Mostuea* and *Kopsia*) in the distribution of test statistics from all comparisons. In this way, we can test the null hypothesis that the rates of evolution in the two paralogous copies are not significantly different, as compared with other plant *RPB2* sequences.

For the comparisons among *RPB2*-d, *ndhF,* and *rbcL,* data matrices for each individual gene were assembled using all ingroup (Gentianales *RPB2*-d except *Chironia*), plus the four Lamiales taxa as outgroup (table 1). Voucher information and GenBank accession numbers for the *ndhF* and *rbcL* sequences can be found in Backlund, Oxelman, and Bremer (2000). To find the most parsimonious trees, the heuristic algorithm of PAUP, version 4.0*d64 (David Swofford, unpublished), was used. One hundred random additions with MULPARS on and tree bisection-reconnection (TBR) branch swapping were performed. To assess clade support, parsimony jackknife analyses (Farris et al. 1996) of the alignments were performed using 1,000 replicates with 36.8% of the characters deleted in each replicate, 5 random additions per jackknife replicate, TBR branch swapping, and MULPARS off. Presence/absence of introns 22–23 and other informative indels were appended as additional characters to the data matrices before analysis. Gap positions were treated as missing data. ILD tests were performed using the partition-homogeneity test in PAUP, version 4.0*d64, with 1,000 randomizations, each with one random addition, MULPARS off, and TBR branch swapping. All uninformative characters were removed prior to the ILD tests (Cunningham 1997). Calculations of the relative substitution rates and transition/transversion ratios were performed with MacClade, version 3.07 (Maddison and Maddison 1997). Overall strength of the phylogenetic signals was measured as total support (t; Bremer 1994) and the sum of jackknife resampling percentage units above 50% (j50). The efficiency of the different regions was assessed by dividing these values by the average number of sequenced nucleotides per region (tEff and jEff, respectively). Mean numbers of changes per site for first, second, and third codon positions were calculated from the inferred number of steps on the most parsimonious trees and divided by the number of variable sites for each position (cf. Olmstead, Reeves, and Yen 1998). The actual rates may be higher due to multiple substitutions along branches, but that does not necessarily affect the relative comparisons in qualitative terms.

Results

Our PCR experiments revealed duplicate copies of *RPB2* in two plants. A characteristic feature of one of the duplicate copies (hereinafter denoted *RPB2*-d) is the lack of introns in the area studied. Amplification with primer pairs 6935F and 7884R gave a ~850-bp band for all taxa except *Mostuea* and *Kopsia,* for which 1,250- and 1,150-bp bands were obtained, respectively. DNA sequencing of the latter two revealed introns in the same positions as in *Arabidopsis,* whereas the other products lacked introns. A new 3′ primer, 7901R, gave ~870-bp products for both *Mostuea* and *Kopsia.* These products both lacked introns. None of these primer pairs gave any product for *Chironia* (or any other herbaceous member of the Gentianaceae tried), but the primer pair 6935F and 8032R gave a ~1,100-bp fragment. The 8032R primer sequence is situated in exon 25 in the *Arabidopsis* sequence, and the *Chironia* product contained an intron between exons 24 and 25 at correspond-

Table 1
List of Taxa, GenBank Accession Numbers, and Literature References or Voucher Information for Previously Unpublished Sequences

Taxon	Gene	GenBank Accession No.	Voucher/Reference
Outgroups			
Saccharomyces cerevisiae	*RPC2*	M38723	James, Whelen, and Hall (1991)
Drosophila melanogaster	*RPC2*	X58826	Seifarth et al. (1991)
D. melanogaster	*RPA2*	X17298	Kontermann et al. (1989)
S. cerevisiae	*RPB2*	M15693	Sweetser, Nonet, and Young (1987)
Schizosaccharomyces pombe	*RPB2*	401013	Kawagishi, Yamagishi, and Ishihama (1993)
D. melanogaster	*RPB2*	X05709	Falkenburg et al. (1987)
Homo sapiens	*RPB2*	X63563	Acker et al. (1992)
Marchantia polymorpha	*RPB2*	AF020844	Denton, McConaughy, and Hall (1998)
Gingko biloba	*RPB2*	AF020843	Denton, McConaughy, and Hall (1998)
Magnolia virginiana	*RPB2*	AF020841	Denton, McConaughy, and Hall (1998)
Solanum lycopersicum	*RPB2*-d	U28403	Warilow and Symons (1996)
Arabidopsis thaliana	*RPB2*	Z19121	Larkin and Guilfoyle (1993)
Kopsia fruticosa	*RPB2*	AF064613	*Bremer 3033* (*UPS*)
Mostuea brunonis	*RPB2*	AF064614	*Thulin 7831* (*UPS*)
Buddleja asiatica	*RPB2*-d	AJ133214	*Bremer 3500* (*UPS*)
	ndhF	AF027277	Oxelman, Backlund, and Bremer (1999)
	rbcL	AJ001758	Oxelman, Backlund, and Bremer (1999)
Retzia capensis	*RPB2*-d	AJ133214	*Källersjö 0401191* (*BOL*)
	ndhF	AF027289	Oxelman, Backlund, and Bremer (1999)
	rbcL	Z29669	Bremer et al. (1994)
Sanango racemosum	*RPB2*-d	AJ133216	*Bremer 3352* (*UPS*)
	ndhF	AF027283	Oxelman, Backlund, and Bremer (1999)
	rbcL	AJ001763	Oxelman, Backlund, and Bremer (1999)
Peltanthera floribunda	*RPB2*-d	AJ133217	*Hammel 19855* (*MO*)
	ndhF	AF027281	Oxelman, Backlund, and Bremer (1999)
	rbcL	AJ001762	Oxelman, Backlund, and Bremer (1999)
Ingroup (Gentianales *RPB2*-d)			
Apocynaceae			
Alstonia scholaris	*RPB2*-d	AJ133218	*Fanning 212* (*FTG*)
	ndhF	AJ011982	Oxelman, Backlund, and Bremer (1999)
	rbcL	X91760	Bremer et al. (1994)
Kopsia fruticosa	*RPB2*-d	AF064609	*Bremer 3033* (*UPS*)
	ndhF	AJ235824	Backlund, Oxelman, and Bremer (2000)
	rbcL	X91763	Sennblad and Bremer (1996)
Periploca graeca	*RPB2*-d	AF064611	*Sennblad 255* (*UPS*)
	ndhF	AJ235825	Backlund, Oxelman, and Bremer (2000)
	rbcL	AJ002889	Sennblad and Bremer (1997)
Stephanotis floribunda	*RPB2*-d	AJ133219	*Sennblad 256* (*UPS*)
	ndhF	AJ235826	Backlund, Oxelman, and Bremer (2000)
	rbcL	000000	Unpublished data
Wrightia arborea	*RPB2*-d	AJ133220	*Leeuwenberg 14225* (*WAG*)
	ndhF	AJ235827	Backlund, Oxelman, and Bremer (2000)
	rbcL	AJ002891	Sennblad and Bremer (1997)
Gelsemiaceae			
Gelsemium sempervirens	*RPB2*-d	AJ133221	*Bremer 3313* (*UPS*)
	ndhF	AJ011984	Oxelman, Backlund, and Bremer (1999)
	rbcL	L14397	Olmstead et al. (1993)
Mostuea brunonis	*RPB2*-d	AF064610	*Thulin 7831* (*UPS*)
	ndhF	AJ235828	Backlund, Oxelman, and Bremer (2000)
	rbcL	L14404	Olmstead et al. (1993)
Gentianaceae			
Anthocleista grandiflora	*RPB2*-d	AJ133222	*Bremer 3098* (*UPS*)
	ndhF	AJ235829	Backlund, Oxelman, and Bremer (2000)
	rbcL	L14389	Olmstead et al. (1993)
Chironia linoides	*RPB2*-d	AF064612	*Bremer 3653* (*UPS*)
Fagraea sp.	*RPB2*-d	AJ133223	*Bremer Su-s 900109* (*UPS*)
	ndhF	AJ235830	Backlund, Oxelman, and Bremer (2000)
	rbcL	L14394	Olmstead et al. (1993)
Potalia resinifera	*RPB2*-d	AJ133224	*Ståhl and Knudsen 3025* (*GB*)
	ndhF	AJ235831	Backlund, Oxelman, and Bremer (2000)
	rbcL	AJ235816	Backlund, Oxelman, and Bremer (2000)
Loganiaceae			
Antonia ovata	*RPB2*-d	*AJ133225*	*Rova 1964* (*GB*)
	ndhF	*AJ235832*	Backlund, Oxelman, and Bremer (2000)
	rbcL	*AJ235817*	Backlund, Oxelman, and Bremer (2000)

Table 1
Continued

Taxon	Gene	GenBank Accession No.	Voucher/Reference
Gardneria angustifolia	RPB2-d	AJ133226	*Ludlow et al. 20888* (*UPS*)
	ndhF	AJ235834	Backlund, Oxelman, and Bremer (2000)
	rbcL	AJ235819	Backlund, Oxelman, and Bremer (2000)
Logania vaginalis	*RPB2*-d	AJ133227	*Bremer 3013* (*UPS*)
	ndhF	AJ235837	Backlund, Oxelman, and Bremer (2000)
	rbcL	Z68826	Bremer (1996)
Mitrasacme pilosa	*RPB2*-d	AJ133228	*Muir 1817* (*UPS*)
	ndhF	AJ236058	Backlund, Oxelman, and Bremer (2000)
	rbcL	AJ235821	Backlund, Oxelman, and Bremer (2000)
Mitreola petiolata	*RPB2*-d	AJ133229	*Miller 7570* (*UPS*)
	ndhF	AJ235839	Backlund, Oxelman, and Bremer (2000)
	rbcL	AJ235822	Backlund, Oxelman, and Bremer (2000)
Neuburgia coryocarpum	*RPB2*-d	AJ133230	*Smith 7385* (*S*)
	ndhF	AF027275	Oxelman, Backlund, and Bremer (1999)
	rbcL	AJ001755	Oxelman, Backlund, and Bremer (1999)
Strychnos potatorum	*RPB2*-d	AJ133231	*Bremer 3022* (*UPS*)
	ndhF	AJ23541	Backlund, Oxelman, and Bremer (2000)
Strychnos nox-vomica	*rbcL*	L14410	Olmstead et al. (1993)
Spigelia anthelmia	*RPB2*-d	AJ133232	*Bremer 3377* (*UPS*)
	ndhF	AJ235840	Backlund, Oxelman, and Bremer (2000)
	rbcL	Y11863	Bremer and Thulin (1998)
Usteria guineensis	*RPB2*-d	AJ133233	*Adams 666* (*UPS*)
	ndhF	AJ235842	Backlund, Oxelman, and Bremer (2000)
	rbcL	AJ235823	Backlund, Oxelman, and Bremer (2000)
Rubiaceae			
Chiococca alba	*RPB2*-d	AJ133234	*Bremer 2703* (*UPS*)
	ndhF	AJ130835	Bremer et al. (2000)
	rbcL	L14394	Olmstead et al. (1993)
Gardenia thunbergia	*RPB2*-d	AJ133235	*Gillis 10913* (*FTG*)
	ndhF	AJ235844	Backlund, Oxelman, and Bremer (2000)
	rbcL	X83637	Bremer, Andreasen, and Olsson (1995)
Luculia grandiflora	*RPB2*-d	AJ133241	*Bremer 2713* (*UPS*) *pseduogene
	RPB2-d	00000000	*Bremer 2713* (*UPS*)
	ndhF	AJ011987	Oxelman, Backlund, and Bremer (1999)
	rbcL	X83648	Bremer, Andreasen, and Olsson (1995)
Mussaenda erythrophylla	*RPB2*-d	AJ133236	*Gillis 10838* (*FTG*)
	ndhF	AJ130836	Bremer et al. (2000)
	rbcL	X83652	Bremer, Andreasen, and Olsson (1995)
Oldenlandia corymbosa	*RPB2*-d	AJ133237	*Bremer 3076* (*UPS*)
	ndhF	AJ130837	Bremer et al. (2000)
	rbcL	X83655	Bremer, Andreasen, and Olsson (1995)
Ophiorrhiza mungos	*RPB2*-d	AJ133238	*Bremer 3301* (*UPS*)
	ndhF	AJ130838	Bremer et al. (2000)
	rbcL	X83656	Bremer, Andreasen, and Olsson (1995)
Pinckneya pubescens	*RPB2*-d	AJ133239	*Forbes 81288* (*UC*)
	ndhF	AJ130835	Bremer et al. (2000)
	rbcL	X83661	Bremer, Andreasen, and Olsson (1995)
Rondeletia odorata	*RPB2*-d	AJ133240	*Bremer and Andreasen 3504* (*UPS*)
	ndhF	AJ235845	Backlund, Oxelman, and Bremer (2000)
	rbcL	X83637	Bremer and Thulin (1998)

NOTE.—Plant *RPB2* copies lacking introns in the studied region are denoted with "-d." Intron presence is unknown for the *Magnolia* sequence. Herbarium abbreviations are according to Holmgren, Holmgren, and Barnett (1990).

ing locations but had no introns in the interval equivalent to *Arabidopsis* exons 22–24.

The alignment of the deduced amino acid sequences for the different RNA polymerase genes is shown in figure 2 (based on an initial alignment provided by B. D. Hall). The phylogenetic analysis of this alignment strongly supported the obtained PCR products as plant *RPB2* sequences, regardless of whether amino acid coding, DNA coding, or a combined approach was used (fig. 3). The amino acid coding succeeded in determining that the gene was indeed a polymerase gene and that it came from a plant, but few other branches received substantial support. The basal resolution among the plant *RPB2* sequences was poor. The Lamiales *RPB2*-d and Gentianales *RPB2*-d sequences, both strongly supported by the DNA sequences, grouped together with the *Solanum* cDNA sequence in the most-parsimonious trees of the DNA and combined matrices. Unpublished data confirm that the corresponding genomic *Solanum* DNA sequence lacks introns in the studied region, thus further supporting orthology with the *RPB2*-d sequences obtained for this study. The paralogous *Kopsia* and *Mostuea* fragments group together strongly, but their placement is unresolved with respect to other plant sequenc-

es. Thus, it seems reasonable to postulate a duplication predating the origin of the Gentianales. No significant difference between the nucleic acid and amino acid codings could be detected by the ILD test; only 10 of the 1,000 randomized partitions yielded a sum length larger than the sum of the original partitions.

By restricting a tree to *Arabidopsis* II, *Mostuea* II, *Mostuea* II-d, *Kopsia* II, and *Kopsia* II-d, we compared proportion of synonymous changes and transition/transversion ratios along the terminal branches for the two paralogs in *Mostuea* and *Kopsia* (table 2). No significant differences in substitution patterns could be found.

The relative-rate tests did not reveal any significant differences between the paralogous copies in *Kopsia* (all pairwise comparisons in the 18-taxon tree [fig. 4] are between the 37th and the 99th percentiles) or in *Mostuea* (all pairwise comparisons fall between the 40th and the 98th percentiles).

Alignment (using both the PAM250 cost matrix and CLUSTAL W with default settings) of the *RPB*2-d sequences in Gentianales required the introduction of a few gaps. At amino acid positions 21–30 (fig. 2), all Gentianales *RPB*2-d sequences had at least nine amino acid residues (often AHVPSSLSA; the Gentianaceae had 10 residues: AHVPAGSFNA), whereas the outgroup sequences had only six (often AQGQAT). This region was aligned as follows:

AQ---GQ-AT

AHVPSSL-SA

AHVPAGSFNA

All obtained sequences except those of *Neuburgia* and *Luculia* have reading frames corresponding to possible functionality (i.e., no stop codons). The *Luculia* sequence has a 58-bp deletion spanning the exon 23–24 border, causing a frameshift and possible nonfunctionality. Additional PCR experiments with a reverse primer constructed from other Rubiaceae sequences from the deleted area of the first sequence obtained from *Luculia* revealed a normal (no deletion), otherwise very similar, sequence in *Luculia.* In the 406 nt sequenced, there were seven synonymous changes and one amino acid change. The explanation for the duplicate *RPB2*-d sequence may be the putative hybrid nature of this cultivated plant, although its ploidy is unknown. The *Neuburgia* sequence contains two stop codons if the majority-rule consensus sequence from the individual sequencing reactions is chosen, but in the first case the reverse sequencing reactions are ambiguous at this position, and in the second case there is also a C signal in the first codon position present in all reactions, although with weaker signal than the T.

Examples of most-parsimonious trees from *RPB2, ndhF,* and *rbcL* are shown in figure 5 (strict consensus trees will appear through collapsing of the dotted branches). The highest number of informative characters was found in *ndhF* (table 3), but the *RPB2* sequences were more informative per sequenced nucleotide (32% informative positions, versus 30% for *ndhF* and 15% for *rbcL*). The highest amount of homoplasy was found in *RPB2* (consistency index [CI] = 0.42; table 3). The strongest signal was found in *ndhF,* in terms of both total support and jackknife resampling frequencies (table 3). *RPB2* gave more jackknife support per sequenced nucleotide than the chloroplast genes, but Bremer support per sequenced nucleotide was higher for *ndhF* (table 3). Combination of the genes always yielded fewer trees, and support usually increased. However, jackknife frequencies only marginally increased if *rbcL* and/or *RPB2* were added (table 3), indicating that *ndh*F is making the most significant contribution of the three genes to resampling support. In contrast to jackknife frequencies, total support (t) values were approximately additive to each other when the matrices were combined (table 3). The ILD tests singled out *RPB2* as significantly different from both *rbcL* and *ndhF* ($P = 0.024$ and 0.012, respectively), whereas the incongruence between the latter two was less pronounced ($P = 0.360$). Nevertheless, 11 nodes were strongly supported by all three data sets (fig. 6).

The *ndhF* sequences had the strongest base-compositional bias (67% A+T content, compared with 54% for *RPB2* and 56% for *rbcL*), whereas *RPB2* had the highest transition/transversion ratio (1.85; 1.49 for *rbcL* and 1.23 for *ndh*F). All three genes had bias toward third-position changes (fig. 7). This was most pronounced in *RPB2,* where the CI also was lowest for the third-position changes. The relative number of changes in informative codon positions differed markedly among the three loci (fig. 8). Informative third codon positions were significantly more variable than the other two positions in *RPB2.* A similar pattern is found in *ndhF,* but here only weak significance between third and second positions could be detected ($0.05 > P > 0.01$). The relative variability among codon positions was strikingly different for *rbcL,* conforming to the patterns reported by Olmstead, Reeves, and Yen (1998), but no significant differences could be detected. Changing from nucleic to amino acid coding had great impact on the overall jackknife support on *RPB2* (J50 = 128, compared with 695 in the DNA matrix) and *rbcL* (J50 = 9 instead of 693),

→

FIG. 2.—Alignment of deduced amino acid sequences for exons 22–25 (corresponding to the *Arabidopsis* genomic *RPB2* sequence) of *RPA2, RPB2,* and *RPC2* for taxa according to table 1. Primer sites are indicated with >>>> (forward) and <<<< (reverse). Presence (1)/absence (0) values of introns 22, 23, and 24, are appended last in the alignment. SceIII = *Saccharomyces cervisiae RPC2,* DmeIII = *Drospshila melagonaster RPC2,* HsaII = *Homo sapiens RPB2,* DmeII = *Drospshila melagonaster RPB2,* SpoII = *Schizosaccharomyces pombe RPB2,* SceII = *Saccharomyces cervisiae RPB2,* MpoII = *Marchantia polymorpha RPB2,* GbiII = *Ginkgo biloba RPB2,* MviII = *Magnolia virginiana RPB2,* SlyII = *Solanum lycopersicon RPB2*-d, AthII = *Arabidopsis thaliana RPB2,* MbrII = *Mostuea brunonis RPB2,* KfrII = *Kopsia fruticosa RPB2,* PgrII-d = *Periploca graeca RPB2*-d, MbrII-d = *Mostuea brunonis RPB2*-d, KfrII-d = *Kopsia fruticosa RPB2*-d, CliII-d = *Chironia linoides RPB2*-d, DmeI = *Drosophila melagonaster RPA2.*

```
            >>>>>>>>>>
SceIII      GMKVQSGQIYINKSVPTNSADAPNPNNVNVQTQYREAPVIYRGPEPSHIDQVMMSVSDNDQALI
DmeIII      GEQVQNKQIMINKEMPAVTSMNPLQGQ-SAQVPYTAVPISYKGPEPSYIERVMVSANAEEDFLI
HsaII       GVRVSGDDVIIGKTVTLPENEDEL-ESTNRRYTKRDCSTFLRTSETGIVDQVMVTLNQEGYKFC
DmeII       GIRVSGDDVVIGKTITLPENDDEL-DSNTKRFSKRDASTFLRNSETGIVDQVMLTLNSEGYKFC
SpoII       GTRVSGEDIIIGKTAPIPLDHEEL-GQRTQLHAKRDVSTPLRSTESGIVDQVMVTTNQEGLKFV
SceII       GVRVSGEDVIIGKTTPISPDEEEL-GQRTAYHSKRDASTPLRSTENGIVDQVLVTTNQDGLKFV
MpoII       GTRVSGEDVIIGKTTPLPQDETG--SQ-AQRYTKRDQSTCLRHSESGMIDQVLLTTNADGLRFV
GbiII       GTRVSGEDVIIGKTSPISRMKQG--GQ-VARYTKRDQSTSLRHSESGMVDQVLLTTNADGLRFV
MviII       GTRVSGEDVIIGKTIPIAQDDVQ--GQ-VTRYTRRDHSTSLRHSESGMVDQVLLTTNADGLRFV
SlyII       GTRVSGEDVIIGKTTPISQDDAQ--GQ-ASRYTRKDHSTSLRHSETGMVDQVLLTTNADGLRFV
AthII       GTRVSGEDVIIGKTTPISQDEAQ--GQ-SSRYTRRDHSISLRHSETGMVDQVLLTTNADGLRFV
MbrII       ?????????????????SQDDAQ--GT-NSRFTRRDHSTSLRHSETGVVDQVLLTTNADGLKFV
KfrII       ??????????????DPISQEDAQ--GQ-SARFTRRDHSTSLRHSETGIVDQVLLTTNADGLRFV
PgrII-d     ?????????????TPIHQDDAHVP-SSLSARYTKRDHSTSLRHSENGIVDQVLLTTNADGLRFV
MbrII-d     ?????????????????????HAP-GNLNARYTKRDHSTSLRHSETGIVDQVLLTTNADGLRFV
KfrII-d     ??????????????TPITQDDAHVPASSSARYTKRDHSTSLRHS?NGIVDQVLLTTNADGLRFV
CliII-d     ?? ??????????????????????????????????????L?HSETGMVDQVLLTTNADGLRFV
DmeI        GSKLSYGS---------PLYC---YFDGEVATYK---VVKMDEKEDCIVESIRQLGSFDLSP-T

SceIII      KVLLRQNRRPE-----LGDKFSSRHGQKGVCGIIVKQEDMPFNDQGIVPDIIMNPHGFPSRMTV
DmeIII      KILLRQTRIPR------GDKFSSRHGQKGVTGLIVEQEDMPFNDFGICPDMIMNPHGFPSRMTV
HsaII       KIRVRSVRIPQ-----IGDKFASRHGQKGTCGIQYRQEDMPFTCEGITPDIIINPHAIPSRMTI
DmeII       KIRVRSVRIPQ-----IGDKFASRHGQKGTCGIQYRQEDMAFTCEGLAPDIIINPHAIPSRMTI
SpoII       KVRMRSTRIPQ-----IGDKFASRHGQKGTIGMTYRHEDMPFSAQGIVPDIIINPHAIPSRMTV
SceII       KVRVRTTKIPQ-----IGDKFASRHGQKGTIGITYRREDMPFTAEGIVPDLIINPHAIPSRMTV
MpoII       KIRVRSIRIPQ-----IGDKFSSRHGQKGTVGMTYTQEDMPWTQEGITPDIIVNPHAIPSRMTI
GbiII       KIRMRSVRIPQ-----IGDKFSSRHGQKGTVGMTYTQEDMPWTVEGITPDIIVNPHAIPSRMTI
MviII       KVRMRSVRIPQ-----IGDKFSSRHGQKGTVGMTYTQEDMPWTPEGITPDIIVNPHAIPSRMTI
SlyII       KVRVRSVRIPQ-----IGDKFSSRHGQKGTVGMTYTQEDMPWTVEGITPDIIVNPHAIPSRMTI
AthII       KVRVRSVRIPQ-----IGDKFSSRHGQKGTVGMTYTQEDMPWTIEGVTPDIIVNPHAIPSRMTI
MbrII       KIRMKSV??PQ-----IGNKFTSRPGQKGTIGMTYTPED?PWTLEGITPDIIVNPHAIPSRMTI
KfrII       KVRMRSVRLPQ-----IGDKFSSRHGQKGTIGMTYTQEDMPWTLEGVTPDIIVNPHAIPSRMTI
PgrII-d     KVRVRSVRIPQ-----IGDKFSSRHGQKGTVGMTYTQEDMPFTQEGITPDIIVNPHAIPSRMTI
MbrII-d     KVRVRSVRIPQ-----IGDKFSSRHGQKGTVGMTYTQEDMPFTGEGITPDIIVNPHAIPSRMTI
KfrII-d     KVRVRSVRIPQ-----IGDKFSSKHG?KGTIGMTYTQEDMPSTAEGITPDISVNPHAIPSRMTI
CliII-d     KVRVRSVRVPQ-----IGDKFSSRHGQKGTVGMTYTQEDMPFTSEGITPDIIVNPHAIPSRMTI
DmeI        KMVAITLRVPRPAT--IGDKFASRAGQKGICSQKYPAEDLPFTESGLIPDIVFNPHGFPSRMTI

SceIII      GKMIELISGKAGVLNGTLEYGTCFGGSKL----EDMSKILVDQGFNYSGKDMLYSGITGECLQA
DmeIII      GKTLELLGGKAGLLEGKFHYGTAFGGSKV----EDIQAELERHGFNYVGKDFFYSGITGTPLEA
HsaII       GHLIECLQGKVSANKGEIGDATPFNDAVNV---QKISNLLSDYGYHLRGNEVLYNGFTGRKITS
DmeII       GHLIECLQGKLGSNKGEIGDATPFNDAVNV---QKISTFLQEYGYHLRGNEVMYNGHTGRKINA
SpoII       AHLVECQLSKVSALSGFEGDATPFTD-VTV---EAVSKLLRSHGFQSRGFEVMYHGHTGRKLVA
SceII       AHLIECLLSKVAALSGNEGDASPFTD-ITV---EGISKLLREHGYQSRGFEVMYNGHTGKKLMA
MpoII       GQLIECIMGKVAAHMGKEGDATPFTD-VTV---DNISKALHKCGYQMRGFEVMYNGHTGRRLYA
GbiII       GQLIECIMGKVAAHVGKEGDAT PFTD-VTV---DNISKALHKCGYQMRGFETMYNGHTGRKLSA
MviII       GQLIECIMGKVAAHMGKEGDATPFTD-VTV---DNISKALHKCGYQMRGFETMYNGHTGRRLPA
SlyII       GQLIECIMGKVAAHMGKEGDATPFTD-VTV---DNISKALHKCGYQMRGFETMYNGHTGRRLSA
AthII       GQLIECIMGKVAAHMGKEGDATPFTD-VTV---DNISKALHKCGYQMRGFERMYNGHTGRPLTA
MbrII       GQL?ECIMGKVAAHMGKEGDATPFTD-VTV---DNISKALHKCGYQMRGFERMYNGHTGRLMGD
KfrII       GQLIECIMGKVAAHMGKEGDATPFTD-VTV---DNISKALHKCGYQMRGFERMYNGHTGRIMGD
PgrII-d     GQLIECIMGKVAAHMGKEGDATPFTD-VTV---DNISKALHRCGYQMRGFERMYNGHTGRPLTA
MbrII-d     GQLIECIMGKVAAHMGKEGDATPFTD-VTV---DNISKALHKCGYQMRGFERMYNGHTGRPLTA
KfrII-d     GQ?IECIMGKVAAHMGKEG?ATPFTD-VTV---DNISKALHKCGYQMRGFERMYNGHTGRPLTA
CliII-d     G?LIECI?GKVAAHMGKEGDATPFTD-V?V---DNISRA?PKCGY?MRGFERMFNGHTGRPFTA
DmeI        AMMIETMAGKGAAIHGNVYDATPFRFSEENTAIDYFGKMLEAGGYNYYGTERLYSGVDGREMTA
SceIII      YIFFGPIYYQKLKHMVLDKMHARARGPRAVLTRQPTEGRSRDGGLRLGEMERDCVIAYGASQLL
```

```
DmeIII    YIYSGPVYYQKLKHMVQDKMHARARGPKAVLTRQPTQGRSREGGLRLGEMERDCLISYGASMLI
HsaII     QIFIGPTYYQRLKHMVDDKIHSRARGPIQILNRQPMEGRSRDGGLRFGEMERDCQIAHGAAQFL
DmeII     QVFLGPTYYQRLKHMVDDKIHSRARGPVQILVRQPMEGRARDGGLRFGEMERDCQISHGAAQFL
SpoII     QVFLGPTYYQRLKHLVDDKIHARARGPVQILTRQPVEGRSRDGGLRFGEMERDCQISHGCSSVL
SceII     QIFFGPTYYQRLRHMVDDKIHARARGPMQVLTRQPVEGRSRDGGLRFGEMERDCMIAHGAASFL
MpoII     HIFLGPTYYQRLKHMVDDKIH???????????????????????????????????????????
GbiII     MIFLGPTYYQRLKHMVDDKIS???????????????????????????????????????????
MviII     MIFLGPTYYQRLKHMVDDKIH???????????????????????????????????????????
SlyII     MIFLGPTYYQRLKHMVDDKIHSRGRGPVQILTRQPAEGRSRDGGLRFGEMERDCMIAHGAAHFL
AthII     MIFLGPTYYQRLKHMVDDKIHSRGRGPVQILTRQPAEGRSRDGGLRFGEMERDCMIAHGAAHFL
MbrII     VIFLGPTYYQRLKHMVDDKIHSRGRGPVQILTRQPAEGRSRDGGLRFGEMERDCMIAHGAAHFL
KfrII     MIFLGPTYYQRLKHMVDDKIHS?GRGPVQILTRQPAEGRSRDGG?RFGEMERDCMIAHGAAHFL
PgrII-d   TIFLGPTYYQRLKHMVDDKIHSRGRGPVQILTRQPAEGRSRDGGLRFGEMERDCMIAHGAAHFL
MbrII-d   TIFLGPTYYQRLKHMVDDKIHSRGRGPVQILTRQPAEGRSRDGGLRFGEMERDCMIAHGASHFL
KfrII-d   TIFLGPTYYQRLKHMVDDKIHSRGRGPVQILTRQPAEGRSRDGGLRFGEMERDCMIAHGAAHFL
CliII-d   TIFLGPTYY?RLKHMVDDKIHSRGRGPVQILTRQPAEGRSRDGGLRFGEMERDCMIAHGAAHFL
DmeI      DIFFGVVHYQRLRHMVFDKWQVRSTGAVEARTHQPIKGRKRGGGVRFGEMERDALISHGAAFLL
                                                    <<<<<<<<<<
                                                            <<<<<<<<<<
SceIII    LERLMISSDAFEVDVCDKCGLMG-----------------YSGWCTTCKSAENIIK
DmeIII    MERLMISSDAFEVDVCRTCGRMA-----------------YCSWCHFCQSSANVSK
HsaII     RERLFEASDPYQVHVCNLCGIMA-----------IANTRTHTYECRGCRNKTQISL
DmeII     RERLFEVSDPYRVHICNFCGLIA-----------IANLRNNTFECKGCKNKTQISQ
SpoII     RERLFDCSDAYRVIVCDICGLIA-----------IASYKKDSYECRSCQNRTRFSQ
SceII     KERLMEASDAFRVHICGICGLMT----------VIAKLNHNQFECKGCDNKIDIYQ
MpoII     ????????????????????????????????????????????????????????
GbiII     ????????????????????????????????????????????????????????
MviII     ????????????????????????????????????????????????????????
SlyII     KERLFDQSDAYRVHVCERCGLIA-----------IANLKKNSFECRGCKNKTDIVQ
AthII     KERLFDQSDAYRVHVCEVCGLIA-----------IANLKKNSFECRGCKNKTDIVQ
MbrII     KERLFDQSDAYRVHICELCGLVA-----------IA????????????????????
KfrII     KERLFDQSDAYRVHVCQICGLIA-----------IASLKK????????????????
PgrII-d   KERLFDQSDAYRIHVCERCGLIA-----------IANL??????????????????
MbrII-d   KERLFDQSDAYRIHVCERCGLIA-----------IANLKKNSFEC???????????
KfrII-d   KERLFDQSDAYRIHVCERCGLIA-----------IANLKKNSFECRG?????????
CliII-d   KERLFDQSDAYRIHVCERCGLIA-----------IANLKKNSFECRSCKNKTDIVQ
DmeI      QDRLFHNSDKTHTLVCHKCGSILAPLQRIVKRNETGGLSSQPDTCRLCGDNSSVSM

              <<<<<<<<<<<
SceIII    MTIPYAAKLLFQELLSMNIA000
DmeIII    ISMPYACKLLFQELTSMNVV000
HsaII     VRMPYACKLLFQELMSMSIA???
DmeII     VRLPYAAKLLFQELMSMNIA000
SpoII     VYLPYAAKLLFQELMSMNIA000
SceII     IHIPYAAKLLFQELMAMNIT000
MpoII     ???????????????????????
GbiI      ???????????????????????
MviII     ???????????????????????
SlyII     ???????? ???????????????
AthII     VYIPYACKLLFQELMSMAIA111
MbrII     ????????????????????11?
KfrII     ????????????????????11?
PgrII-d   ????????????????????00?
MbrII-d   ????????????????????00?
KfrII-d   ????????????????????00?
CliII-d   VYIP????????????????001
DmeI      IEIPFSFKYLVTELSSVNIN000
```

FIG. 2 *(Continued)*

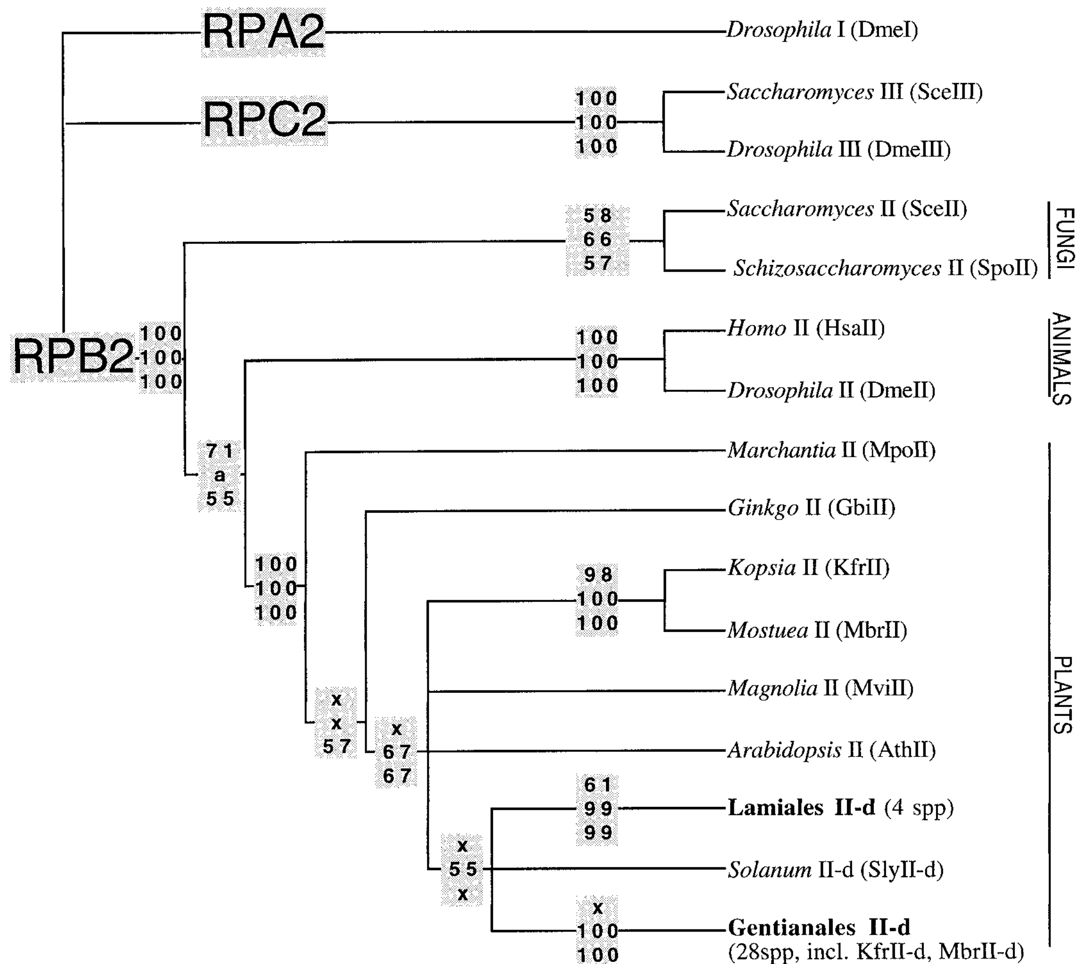


FIG. 3.—Fifty percent jackknife majority-rule tree based on amino acid coding (upper values on branches), nucleic acid coding (middle), and combined coding (lower). Jackknife support of less than 50% is denoted by "x"; "a" indicates that a node is not present in the most-parsimonious tree from the matrix.

but not on *ndhF* (J50 = 898 instead of 852). A possible explanation for this is the low number of informative characters in the *RPB2* and *rbcL* amino acid sequence matrices (32 and 35, respectively) compared with *ndhF* (269). This, in turn, can be explained partly by the larger size of the *ndhF* region sequenced (table 2 and fig. 7) and, perhaps more importantly, by the higher proportion of nonsynonymous changes in *ndhF* (fig. 9).

Discussion

The results presented here for the first time clearly show that there are at least two paralogous *RPB2* sequences occurring in some green plants, as positively demonstrated in *Mostuea* and *Kopsia* of the Gentianales. The tree in figure 3 suggests that the sequences devoid of introns 22–23 form a monophyletic group apart from the intron-containing ones. This means that the duplicate *RPB2* sequences in *Mostuea* and *Kopsia* are unlikely to be due to any recent duplication event. The absence of duplicate *RPB2* loci reported for *Arabidopsis* (Larkin and Guilfoyle 1993), tomato (Warrilow and Symons 1996), and *Rhododendron* (Ericales; Denton, McConaughy, and Hall 1998) suggests that the duplication occurred later than the origin of angiosperms. Ericales and tomato both belong to the asterids, which includes the large euasterid I and II clades (Angiosperm Phylog-

Table 2.
Taxonwise Comparisons of the Substitution Patterns in the Two Paralogous *RPB2* Genes in *Mostuea* and *Kopsia* (II and II-d)

	MOSTUEA			KOPSIA		
	II	II-d	p^a	II	II-d	p^a
Synonymous/nonsynonymous	28/15	20/4	0.20	23/6	30/5	0.75
Transitions/transitions	13/12	25/18	0.86	17/12	25/10	0.48

[a] Bonferroni-corrected probability values from chi-square tests.

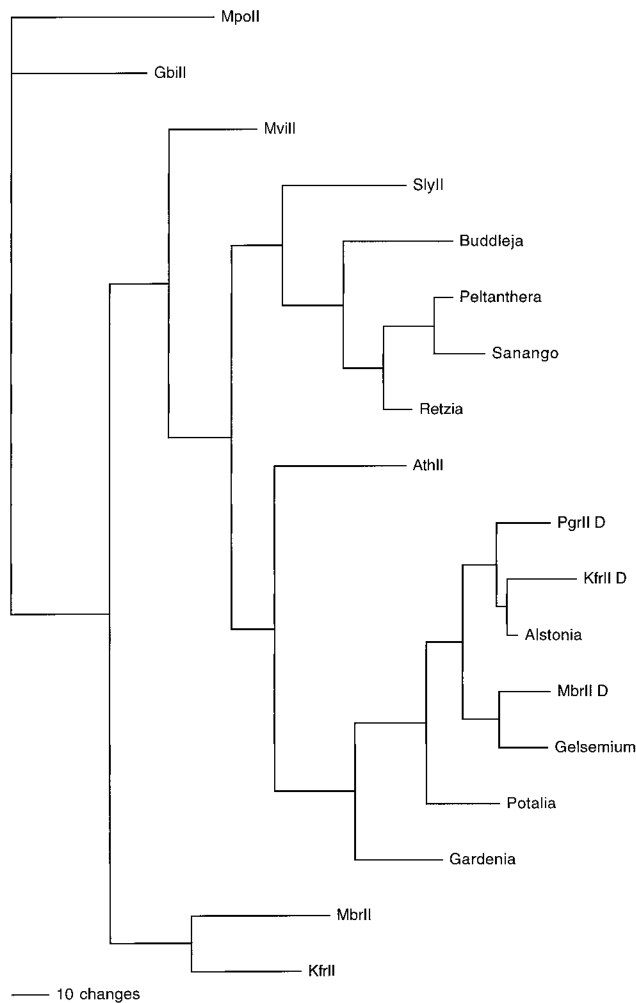
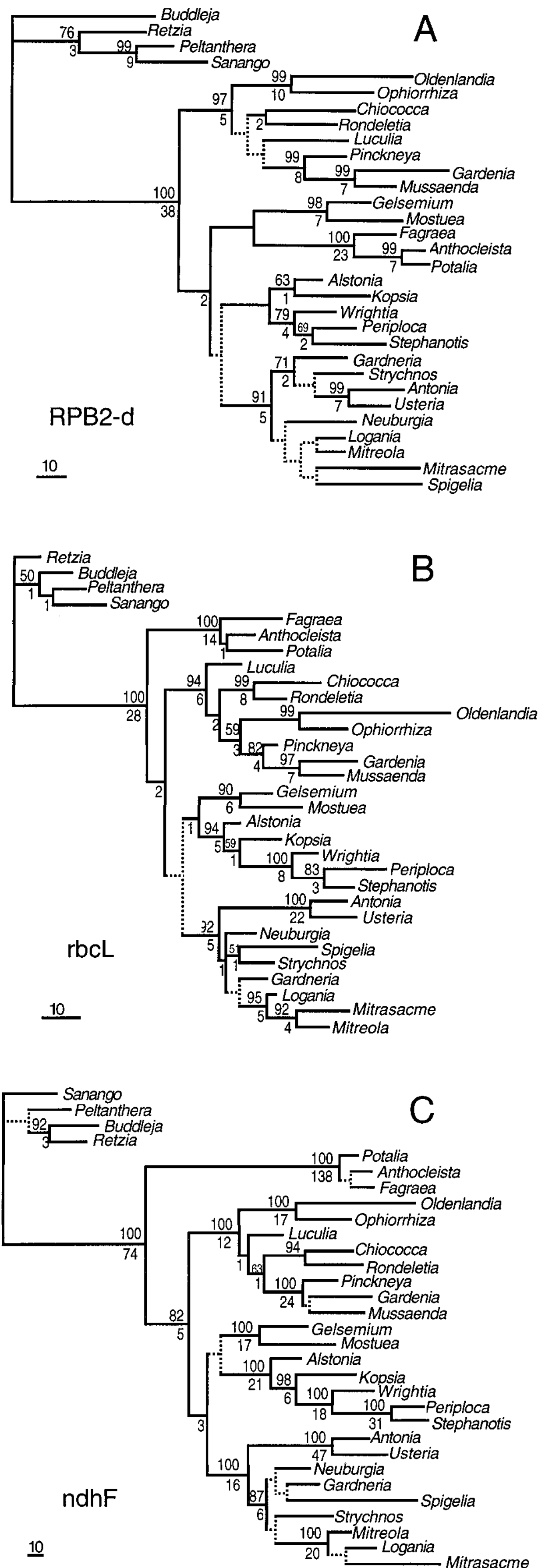


FIG. 4.—Tree used for relative-rate tests. All comparisons between the paralogs in *Kopsia* (Kfr) and *Mostuea* (Mbr) were plotted onto the distribution from all possible triplet tests according to the method of Steel et al. (1996). Note that indels and intron presence/absence were not coded in this analysis, which may explain the slight incongruence with the tree in figure 3. Taxon abbreviations are as in figure 2.

eny Group 1998). The taxa sequenced for this study (Gentianales, Lamiales) and tomato all belong to the large euasterid I clade within the asterids. Recent research has revealed that tomato and *Rhododendron* actually have two *RPB2* genes as well (unpublished data). Taking the tree in figure 4 at face value would suggest a split near the base of the angiosperm clade. However, the support for this scenario is low, and until evidence for the existence of duplicate *RPB2* genes is generated for nonasterid taxa, the origin of a second copy is, in our view, better regarded as unresolved. Current research by Oxelman, Denton, and Hall (unpublished data) indicates that the *RPB2*-d variant is a synapomorphy for the euasterid I clade and that the Ericales may be the only other major angiosperm group with two *RPB2* genes. The phylogenetic relationships among the main asterid clades are not yet unambiguously resolved (e.g., Olmstead et al. 1992, 1993; Angiosperm Phylogeny Group 1998; Soltis, Soltis, and Chase 1999; Albach et al. 2000), so further studies on this issue are clearly desirable. The duplication leading to the two copies found in Gentianales took place at least as early as the origin of the group, but only detailed sequencing studies can shed light on how much earlier and whether there have been multiple duplication events.

An important conclusion that can be drawn from the present study is that great caution is needed when selecting a nuclear, putatively single-copy gene for phylogenetic studies. Despite previous publications based on Southern blot analyses, we show that *RPB2* is not single-copy in all angiosperms. A very substantial number of angiosperms may actually have two copies. The asterids comprise approximately one third of all flowering plant species, and research is ongoing to trace the evolution of *RPB2* in plants.

One caution with regard to the results presented in figure 3 may be that DNA and RNA sequences are intermixed. RNA editing (Bowe and dePamphilis 1996; Malek et al. 1996) could potentially bias the results. However, examination of the sequences deposited in GenBank reveals that DNA exon and mRNA sequences of *Arabidopsis* are identical for the entire *RPB*2, except at a single position (A↔T), indicating that RNA editing is unlikely to cause problems when inferring *RPB2* phylogenies based on both RNA and DNA sequences.



Given the strong bias toward third-position changes in the RNA polymerase data matrix, it is possible that multiple hits along individual branches are a potential problem for phylogeny reconstruction (e.g., Felsenstein 1978) using *RPB2* sequences. To explore this possibility, we conducted the ILD test on the green plant matrix, which would indicate incongruence between the different codings in presence of ''long-branch attraction'' due to excessive synonymous substitutions. Only 10 of 1,000 randomized partitions yielded matrices where the sum of lengths was higher than in the original, thus indicating no significantly conflicting signals. No general conclusions should be drawn from this observation alone, but it seems clear that the long-branch attraction problem, which has been vigorously debated from a theoretical point of view, is in need of more empirical studies. It is possible, however, that most of the long branches in figure 3 are true long branches and that there are no ''true'' short internal branches. It is the mixture of long and short branches that is most likely to create problems in phylogeny reconstruction (see Hillis, Huelsenbeck, and Swofford 1994).

The comparison of *RPB2* with the chloroplast loci *rbcL* and *ndhF* reveals several important observations concerning the properties of the respective genes as phylogenetic markers. Clearly, the most powerful in terms of support of the three is *ndhF.* Moreover, the *ndhF* matrix has the highest CI and is the least biased toward synonymous substitutions of the three. The latter appears to be a desirable property for protein-coding genes, because high proportions of synonymous changes may lead to saturation and obscure historical signal under some circumstances. However, third-position changes in *rbcL* have a higher CI and fewer changes per site than first and second positions, despite the strong bias for synonymous changes in *rbcL* (83%). This observation has been made for other groups as well (Sennblad and Bremer 1997; Olmstead, Reeves, and Yen 1998; Källersjö et al. 1998) and is likely to reflect a general property of *rbcL.*

The ILD tests revealed the results from the *RPB2* matrix to be significantly different from those of the chloroplast genes. Homology of nuclear regions may be confused by recombinant processes such as gene conversion (Wendel and Doyle 1998). In fact, incomplete levels of concerted evolution render such sequence regions less useful for phylogeny reconstruction. However, incongruent patterns may result from other sources, e.g., different substitution patterns like those manifested in this study. Unless homoplasy follows exactly the same stochastic model in the regions compared, ''statistically significant'' ILD will always result, provided that the sequences are long enough. This is because the null

←

FIG. 5.—Examples of most-parsimonious trees from (*A*) *RPB2* sequences, (*B*) *rbcL* sequences, and (*C*) *ndhF* sequences. Dotted lines indicate clades that collapse in strict consensus trees from all most-parsimonious trees found. Jackknife percentages are indicated above branches, and Bremer support values are indicated below. See table 2 for further details.

Table 3
Summary of Results for the *RPB2*-d, *rbcL,* and *ndhF* DNA Sequence Matrices Separately and for All Possible Combinations of These Matrices

Locus	Size[a]	#Inf[b]	Length[c]	CI[d]	RC[e]	J50[f]	jEff[g]	t[h]	tEFF[i]
a) *RPB2*-d	890	285 (1)	1,150 (24)	0.42	0.23	695	0.78	153	0.17
b) *rbcL*	1,402	205 (0)	626 (5)	0.45	0.27	693	0.49	158	0.11
c) *ndhF*	2,202	666 (8)	2,170 (66)	0.51	0.33	852	0.39	501	0.23
a + b	2,292	490 (1)	1,793 (4)	0.42	0.24	678	0.30	297	0.13
a + c	3,092	951 (9)	3,338 (1)	0.47	0.29	959	0.31	698	0.22
b + c	3,604	871 (8)	2,807 (1)	0.49	0.31	923	0.26	678	0.19
a + b + c	4,494	1,156 (9)	3,979 (1)	0.47	0.28	960	0.21	809	0.18

[a] Average length of sequence.
[b] Number of parsimony-informative characters (number of informative indels in parentheses).
[c] Length of most-parsimonious tree(s) (number of optimal trees in parentheses).
[d] Consistency index.
[e] Rescaled consistency index.
[f] Sum of jackknife support above 50%.
[g] J50/size.
[h] Total support.
[i] t/size.

hypothesis tested is that the sequences belong to the same stochastic population. The sample size (i.e., sequence length) needed to yield "significance" will, of course, depend on how different these populations (i.e., genes) actually are, and pairwise comparisons can give clues to the relative magnitudes of these differences. The *RPB2* sequence region is the smallest of the three compared and still stands out as the most incongruent. It is possible that the strong bias for synonymous substitutions and third-position changes in *RPB2* is part of the explanation for this. This may warrant caution with regard to the utility of *RPB2* DNA sequences at higher levels, but on the other hand, visual inspection of the three trees in figure 5 reveals few incongruences with mutual strong support. In fact, the most striking incongruence is the basal position of Gentianaceae (*Anthocleista, Fagraea,* and *Potalia*) in the two chloroplast trees (fig. 5*B* and *C*). This position is in conflict with the tree presented by Backlund, Oxelman, and Bremer (2000), where that position is taken by the Rubiaceae, resulting in congruence with the *RPB2* tree (fig. 5*A*). The chloroplast sequences in this study are a subset of those used by Backlund, Oxelman, and Bremer (2000), so the incongruence must be attributed to the more extensive sampling of that study, in terms of both ingroup and outgroup sequences.

The debate over the past few years on the issue of whether one should combine different data sets or not has been intense. Most authors seem to have reached the

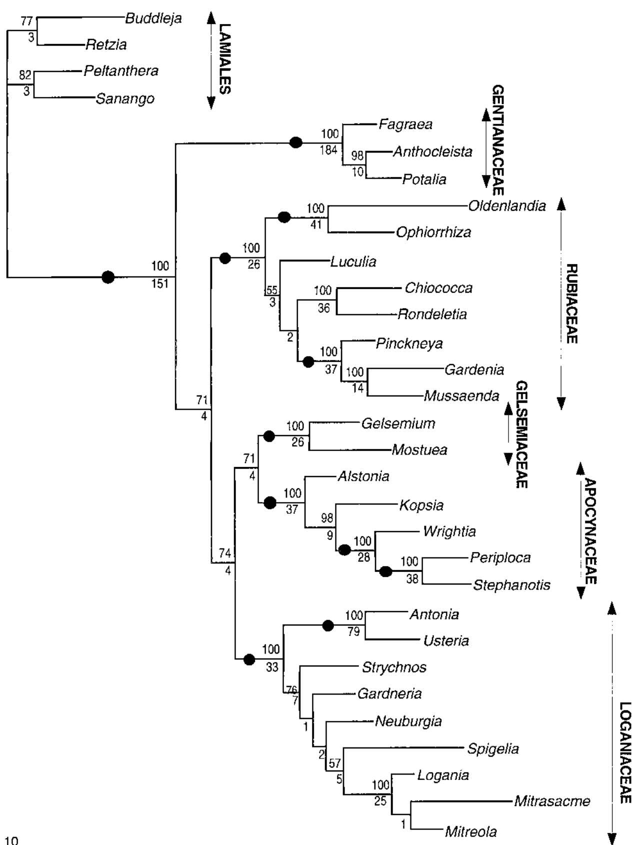


FIG. 6.—Single most-parsimonious tree obtained from the combined *ndhF, rbcL,* and *RPB2* DNA matrix. Dots indicate nodes that are strongly supported by each of the three data sets.

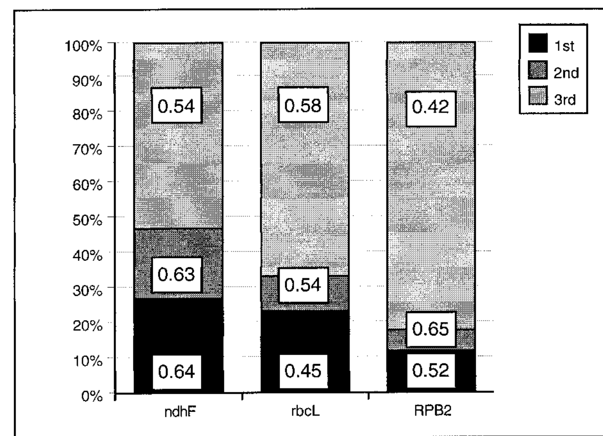


FIG. 7.—Bar chart indicating the proportion of changes in first, second, and third codon positions in the *ndhF, rbcL,* and *RPB2* trees in figure 4. Boxed numbers represent consistency indices for the respective categories.

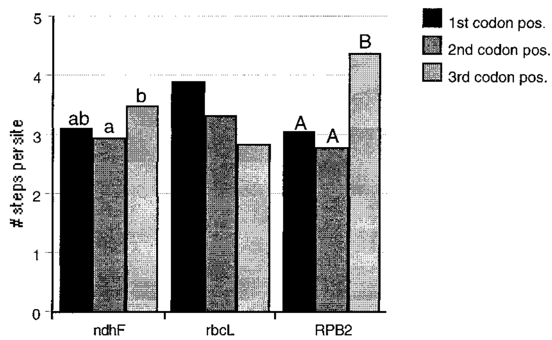


FIG. 8.—Mean number of inferred steps per site by codon position for *ndhF, rbcL,* and *RPB2*-d for parsimony-informative sites. Different lowercase letters indicate significant differences at the 0.05 level; different uppercase letters indicate significant differences at the 0.001 level (Bonferroni-corrected Mann-Whitney *U*-tests).

conclusion that data sets should be combined if they can be assumed to have been generated by a common branching process. As discussed above, incongruence may be observed even if this assumption is correct, and unfortunately, we are unaware of any analytical methods that distinguish between apparent and real incongruence. Given the historical nature of the processes that generated the patterns in the sequences, it is doubtful that such distinction is even theoretically possible. In light of this, it seems difficult to define clear-cut rules for when data should be combined and when they should not. In this case, we saw no serious topological differences among the data sets. Although gene conversion cannot be ruled out, we are unable to detect it in this case.

There can be no definite conclusions drawn about the functional significance of the findings in this study. The lack of premature stop codons, the absence of substitution patterns typical of silent sequences (e.g., an increase in A+T-content), and the strong bias in favor of third-codon-position substitutions all over the tree may indicate functionality of *RPB2*-d, but further studies of the RNA transcript must be performed to confirm this (Oxelman, Denton, and Hall [unpublished data] have demonstrated that both the ''normal'' *RPB2* and the *RPB2*-d gene are transcribed in *Gardenia* and tomato). The presence of intron 24 in the *Chironia* sequence excludes cDNA integration as a possible origin of the duplicate variant. Oxelman, Denton, and Hall (unpublished data) have shown for a large number of euasterid I taxa that six introns are absent (introns 18–23 in the *Arabidopsis* sequence). The 58-bp deletion in *Luculia* suggests, however, that this sequence is a pseudogene. The position of *Luculia* agrees with that reported in other analyses (Bremer and Jansen 1991; Bremer and Struwe 1992; Bremer, Andreasen, and Olsson 1995; Bremer 1996; Andersson and Rova 1999), indicating a recent loss of functionality.

Multiple copies of similar, paralogous sequences can confuse phylogenetic interpretations. However, under the assumption that loss of introns has happened in orthologous copies of *RPB2,* a priori homology assessment is easy, even from the size of the PCR product. Indeed, unpublished data strongly corroborate the suggestion made here about orthology of *RPB2*-d variants. Moreover, sequencing of highly variable introns that may be difficult to align at this level of divergence is avoided. We suggest that the *RPB2*-d locus has the potential to be a valuable nuclear marker for determination of phylogenetic relationships within the euasterid I group of angiosperms, a major group of flowering plants for which there are numerous phylogenetic problems.

Acknowledgments

This study was financed by Swedish Natural Science Research Council B-BU 1487-322 to B.B. Benja-

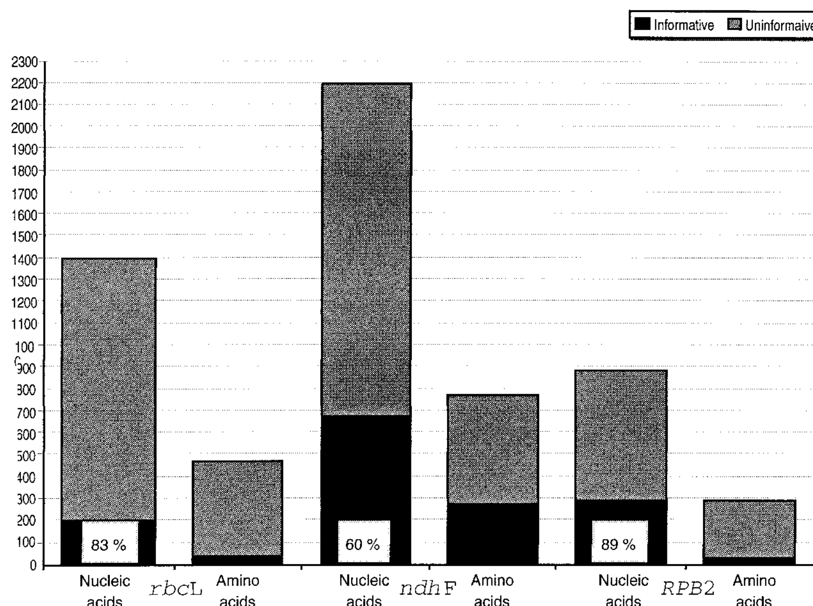


FIG. 9.—Sequence lengths and proportions of informative characters for the *ndhF, rbcL,* and *RPB2* matrices. Left bar, nucleic acid sequence; right bar, inferred amino acid sequence. The boxed number indicates the proportion synonymous changes for each gene.

min D. Hall generously shared the amino acid alignment of various RNA polymerase genes and made numerous valuable comments on a previous draft of the manuscript. Skillful technical assistance from Nahid Heidari and Jessica Rönnholm is gratefully acknowledged. We thank David Swofford for permitting us to use PAUP 4.0*d64. Helpful comments by Pamela Soltis and anonymous reviewers on this manuscript are gratefully acknowledged.

LITERATURE CITED

Acker, J., M. Wintzerith, M. Vigneron, and C. Kedinger. 1992. Primary structure of the second largest subunit of human RNA polymerase II (or B). J. Mol. Biol. **226**:1295–1299.

Agosti, D., D. Jacobs, and R. DeSalle. 1996. On combining protein sequences and nucleic acid sequences in phylogenetic analysis: the homeobox protein case. Cladistics **12**:65–82.

Albach, D. C., P. S. Soltis, D. E. Soltis, and R. G. Olmstead. 2000. Phylogenetic analysis of the Asteridae s.l. using sequences of four genes. Ann. Mo. Bot. Gard. (in press).

Andersson, L., and J. H. E. Rova. 1999. The *rps16* intron and the phylogeny of the Rubioideae (Rubiaceae). Plant Syst. Evol. **214**:161–186.

Angiosperm Phylogeny Group. 1998. An ordinal classification for the families of flowering plants. Ann. Mo. Bot. Gard. **85**:531–553.

Backlund, M., B. Oxelman, and B. Bremer. 2000. Phylogenetic relationships within the Gentianales based on *ndhF* and *rbcL* sequences, with particular reference to the Loganiaceae. Am. J. Bot. (in press).

Bremer, B. 1996. Phylogenetic studies within Rubiaceae and relationships to other families based on molecular data. Opera Bot. Belg. **7**:33–50.

Bremer, B., K. Andreasen, and D. Olsson. 1995. Subfamilial and tribal relationships in the Rubiaceae based on *rbcL* sequence data. Ann. Mo. Bot. Gard. **82**:383–397.

Bremer, B., and R. K. Jansen. 1991. Comparative restriction site mapping of the chloroplast DNA implies new phylogenetic relationships within the Rubiaceae. Am. J. Bot. **78**: 198–213.

Bremer, B., R. K. Jansen, B. Oxelman, M. Backlund, H. Lantz, and K.-J. Kim. 1999. More characters or more taxa for a robust phylogeny—case study from the Coffee family (Rubiaceae). Syst. Biol. **48**:413–435.

Bremer, B., R. G. Olmstead, L. Struwe, and J. A. Sweere. 1994. *rbcL* sequences support exclusion of *Retzia, Desfontainia,* and *Nicodemia* from Gentianales. Plant Syst. Evol. **190**:213–230.

Bremer, B., and L. Struwe. 1992. Phylogeny of the Rubiaceae and the Loganiaceae: congruence or conflict between morphological and molecular data? Am. J. Bot. **79**:1171–1184.

Bremer, B., and M. Thulin. 1998. Collapse of Isertieae, reestablishment of Mussaendeae, and a new genus of Sabiceeae (Rubiaceae); phylogenetic relationships based on *rbcL* data. Plant Syst. Evol. **211**:71–92.

Bremer, K. 1994. Branch support and tree stability. Cladistics **10**:295–304.

Bowe, L. M., and C. W. dePamphilis. 1996. Effects of RNA editing and gene processing on phylogenetic reconstruction. Mol. Biol. Evol. **9**:1159–1166.

Clegg, M. T., M. P. Cummings, and M. L. Durbin. 1997. The evolution of plant nuclear genomes. Proc. Natl. Acad. Sci. USA **94**:7791–7798.

Cunningham, C. W. 1997. Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. **14**: 733.

Denton, A. L., B. L. McConaughy, and B. D. Hall. 1998. Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. Mol. Biol. Evol. **15**: 1082–1085.

Dietrich, M. A., J. P. Prenger, and T. J. Guilfoyle. 1990. Analysis of the genes encoding the largest subunit of RNA polymerase II in *Arabidopsis* and soybean. Plant Mol. Biol. **15**:207–223.

Doyle, J. J., V. Kanazin, and R. C. Shoemaker. 1996. Phylogenetic utility of histone H3 sequences in the perennial relatives of soybean (*Glycine:* Leguminosae). Mol. Phylogenet. Evol. **6**:438–447.

Falkenburg, D., B. Dworniczak, D. M. Faust, and E. K. Bautz. 1987. RNA polymerase II of *Drosophila.* Relation of its 140,000 Mr subunit to the beta subunit of *Escherichia coli* RNA polymerase. J. Mol. Biol. **195**:929–937.

Farris, J. F., V. A. Albert, M. Källersjö, D. Lipscomb, and A. G. Kluge. 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics **12**:99–124.

Farris, J. F., M. Källersjö, A. G. Kluge, and C. Bult. 1994. Testing significance of incongruence. Cladistics **10**:315–320.

Felsenstein, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. **27**:401–410.

Gottlieb, L. D., and V. S. Ford. 1996. Phylogenetic relationships among the sections of *Clarkia* (Onagraceae) inferred from the nucleotide sequences of *PgiC.* Syst. Bot. **21**:45–62.

Hallick, R. B., L. Hong, R. G. Drager, M. R. Favreau, A. Monfort, B. Orsat, A. Spielmann, and E. Stutz. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res. **15**:3537–3544.

Hillis, D. M., J. P. Huelsenbeck, and D. L. Swofford. 1994. Hobgoblin of phylogenetics. Nature **369**:363–364.

Holmgren, K. P., H. N. Holmgren, and L. C. Barnett. 1990. Index Herbariorum 1: the herbaria of the world. 8th edition. Regnum Veg. **120**:1–693.

James, P., S. Whelen, and B. D. Hall. 1991. The *RET1* gene of yeast encodes the second-largest subunit of RNA polymerase III. Structural analysis of the wild-type and *ret1-1* mutant alleles. J. Biol. Chem. **266**:5616–5624.

Källersjö, M., J. S. Farris, M. W. Chase, B. Bremer, M. F. Fay, C. J. Humphries, G. Petersen, O. Seberg, and K. Bremer. 1998. Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants, and flowering plants. Plant Syst. Evol. **213**:259–287.

Kawagishi, M., M. Yamagishi, and A. Ishihama. 1993. Cloning and sequence determination of the *Schizosaccharomyces pombe RPB2* gene encoding the subunit 2 of RNA polymerase II. Nucleic Acids Res. **21**:469–473.

Kontermann, R., S. Sitzler, W. Seifarth, G. Petersen, and E. K. Bautz. 1989. Primary structure and functional aspects of the gene coding for the second-largest subunit of RNA polymerase III of *Drosophila.* Mol. Gen. Genet. **219**: 373–380.

Kumar, P. S., and T. Hymowitz. 1989. Where are the diploid (2n=2x=20) genome donors of *Glycine* Willd. (Leguminosae, Papillionoideae)? Euphytica **40**:221–226.

Larkin, R., and T. Guilfoyle. 1993. The second largest subunit of RNA polymerase II from *Arabidopsis thaliana.* Nucleic Acids Res. **21**:1038.

Larkin, R., G. Hagen, and T. J. Guilfoyle. 1999. *Arabidopsis thaliana* RNA polymerase II subunits related to yeast and human *RPB5.* Gene **231**:41–47.

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

MADDISON, W. P., and D. R. MADDISON. 1997. MacClade. Version 3.07. Sinauer, Sunderland, Mass.

MAIER, R. M., K. NECKERMANN, G. L. IGLOI, and H. KOSSEL. 1995. Complete sequence of the maize chloroplast genome—gene content, hotspots of divergence and fine-tuning of genetic information by transcript editing. J. Mol. Biol. **5**: 614–628.

MALEK, O., K. LATTIG, R. HIESEL, A. BRENNICKE, and V. KNOOP. 1996. RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO J. **6**:1403–1411.

MATHEWS, S., M. LAVIN, and R. A. SHARROCK. 1995. Evolution of the phytochrome gene family and its utility for phylogenetic studies of angiosperms. Ann. Mo. Bot. Gard. **82**: 296–321.

MORTON, B. R., B. S. GAUT, and M. T. CLEGG. 1996. Evolution of alcohol dehydrogenase genes in the palm and grass families. Proc. Natl. Acad. Sci. USA **93**:11735–11739.

OHYAMA, K. 1996. Chloroplast and mitochondrial genomes from a liverwort, *Marchantia polymorpha*—gene organization and molecular evolution. Biosci. Biotechnol. Biochem. **1**:16–24.

OLMSTEAD, R. G., B. BREMER, K. M. SCOTT, and J. D. PALMER. 1993. A parsimony analysis of the Asteridae sensu latu based on *rbcL* sequences. Ann. Mo. Bot. Gard. **80**:700–722.

OLMSTEAD, R. G., H. J. MICHAELS, K. M. SCOTT, and J. D. PALMER. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbcL.* Ann. Mo. Bot. Gard. **79**:249–265.

OLMSTEAD, R. G., P. A. REEVES, and A. C. YEN. 1998. Patterns of sequence evolution and implications for parsimony analysis of chloroplast DNA. Pp. 164–187 *in* D. E. SOLTIS, P. S. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants II: DNA sequencing. Kluwer Academic Publishers, Dordrecht, the Netherlands.

OXELMAN, B., M. BACKLUND, and B. BREMER. 1999. Relationships of the Buddlejaceae s. l. investigated using parsimony jackknife and branch support analysis of chloroplast *ndhF* and *rbcL* sequence data. Syst. Bot. **24**:164–182.

RICE, W. R. 1989. Analyzing tables of statistical tests. Evolution **43**:223–225.

SANDERSON, M. J. 1997. A non-parametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. **14**:1218–1231.

———. 1998. Estimating rate and time in molecular phylogenies: beyond the molecular clock? Pp. 242–264 *in* D. E. SOLTIS, P. S. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants II: DNA sequencing. Kluwer Academic Publishers, Dordrecht, the Netherlands.

SAWADOGO, M., and A. SENTENAC. 1990. RNA polymerase B (II) and general transcription factors. Annu. Rev. Biochem. **59**:711–754.

SEIFARTH, W., G. PETERSEN, R. KONTERMANN, M. RIVA, J. HUET, and E. K. BAUTZ. 1991. Identification of the genes coding for the second-largest subunits of RNA polymerases I and III of *Drosophila melanogaster.* Mol. Gen. Genet. **228**:424–432.

SENNBLAD, B., and B. BREMER. 1997. Is there a justification for differential *a priori* weighting in coding sequences?—a case study from *rbcL* and Apocynaceae. *In* B. SENNBLAD, ed. Phylogeny of the Apocynaceae s.l. Ph.D. dissertation, Uppsala University, Uppsala, Sweden.

———. 1996. The familial and subfamilial relationships of Apocynaceae and Asclepiadaceae evaluated with *rbcL* data. Plant Syst. Evol. **202**:153–175.

SHINOZAKI, K., M. OHME, M. TANAKA et al. (22 co-authors). 1986. The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. EMBO J. **5**:2043–2049.

SIDOW, A., and W. K. THOMAS. 1993. A molecular evolutionary framework for eukaryotic model organisms. Curr. Biol. **4**:596–603.

SOLTIS, D. E., and P. S. SOLTIS. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. Pp. 1–42 *in* D. E. SOLTIS, P. S. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants II: DNA sequencing. Kluwer Academic Publishers, Dordrecht, The Netherlands.

SOLTIS, P. S., D. E. SOLTIS, and M. W. CHASE. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature **402**:402–404.

STEEL, M., A. COOPER, and D. PENNY. 1996. Estimating the time to divergence for groups of taxa. Syst. Biol. **45**:127–134.

STRUWE, L., V. A. ALBERT, and B. BREMER. 1994. Cladistics and family level classification of the Gentianales. Cladistics **10**:175–206.

SWEETSER, D., M. NONET, and R. A. YOUNG. 1987. Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proc. Natl. Acad. Sci. USA **84**:1192–1196.

THURIAUX, P., and A. SENTENAC. 1992. Yeast nuclear RNA polymerases. Pp. 1–48 *in* E. W. JONES, J. R. PRINGLE, and J. R. BROACH, eds. The molecular and cellular biology of the yeast *Saccharomyces.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

ULMASOV, T., R. M. LARKIN, and T. J. GUILFOYLE. 1996. Association between 36- and 13.6-kDa α-like subunits of *Arabidopsis thaliana* RNA polymerase II. J. Biol. Chem. **271**: 5085–5094.

WAKASUGI, T., T. NAGAI, M. KAPOOR et al. (15 co-authors). 1997. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division. Proc. Natl. Acad. Sci. USA **11**:5967–5972.

WARRILOW, D., and R. H. SYMONS. 1996. Sequence analysis of the second largest subunit of tomato RNA polymerase II. Plant Mol. Biol. **30**:337–342.

WATERS, E. R. 1995. An evaluation of the usefulness of the small heat shock genes for phylogenetic analysis in plants. Ann. Mo. Bot. Gard. **82**:278–295.

WENDEL, J. F., and J. J. DOYLE. 1998. Phylogenetic incongruence: window into the genome history and molecular evolution. Pp. 265–296 *in* D. E. SOLTIS, P. S. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants II: DNA sequencing. Kluwer Academic Publishers, Dordrecht, The Netherlands.

WLASSOFF, W. A., M. KIMURA, and A. ISHIHAMA. 1999. Functional organization of the two large subunits of the fission yeast *Schizosaccharomyces pombe* RNA polymerase II. J. Biol. Chem. **274**:5104–5113.

YANO, R., and M. NOMURA. 1991. Suppressor analysis of temperature-sensitive mutations of the largest subunit of RNA polymerase I in *Saccharomyces cerevisiae*: a suppressor gene encodes the second-largest subunit of RNA polymerase I. Mol. Cell. Biol. **11**:754–764.

ZIMMER, E. A., S. L. MARTIN, S. M. BEVERLEY, Y. W. KAN, and A. C. WILSON. 1980. Rapid duplication and loss of genes coding for the a chains of hemoglobin. Proc. Natl. Acad. Sci. USA **77**:2158–2162.

PAMELA SOLTIS, reviewing editor

Accepted March 27, 2000