Deep-Level Diagnostic Value of the rDNA-ITS Region

*Mark A. Hershkovitz**† and Louise A. Lewis‡*

*NCBI (Genbank), National Library of Medicine, National Institute of Health; †Laboratory of Molecular Systematics, Smithsonian Institution; and ‡Department of Biology, University of New Mexico

The similarity of certain reported angiosperm rDNA internal transcribed spacer (ITS) region sequences to those of green algae prompted our analysis of the deep-level phylogenetic signal in the highly conserved but short 5.8S and hypervariable ITS2 sequences. We found that 5.8S sequences yield phylogenetic trees similar to but less well supported than those generated by a ca. 10-fold longer alignment from rDNA-18S sequences, as well as independent evidence. We attribute this result to our finding that, compared to 18S, the 5.8S has a higher proportion of sites subject to vary and greater among-site substitution rate homogeneity. We also determined that our phylogenetic results are not likely affected by intramolecular compensatory mutation to maintain RNA secondary structure nor by evident systematic biases in base composition. Despite historical homology, there appears to be no ITS2 primary sequence similarity shared between fungi, green algae, and angiosperms. ITS2 sequences within each of these groups, however, share sufficient similarity to cluster correctly on the basis of alignability. Our results indicate that ITS region sequences can diagnose organismal origins and phylogenetic relationships at many phylogenetic levels and provide a useful paradigm for molecular evolutionary study.

Introduction

Diagnosing the organismal identity or affinities of a DNA sequence can be especially challenging in the absence of adequate evolutionary characterization of phylogenetic signal in the molecule. The internal transcribed spacer (ITS) region separating 18S and 26S nuclear ribosomal DNA (rDNA), which includes two spacers (ITS1 and ITS2) and the intervening 5.8S coding sequence, has become well characterized across interspecific-intergeneric-level divergences (reviewed in Baldwin et al. 1995) but is rarely compared at deeper levels. The spacers are reputedly poorly conserved at deeper levels, e.g., above the family level in angiosperms (Baldwin et al. 1995). The 5.8S coding region, by contrast, is little used in phylogenetics because of its substantial conservation at deeper levels (Troitsky and Bobrova 1986; Troitsky et al. 1991; Suh, Thien, and Zimmer 1992), which, considering its small size (164–165 bp), translates to little informative variation relative to more exploited sequences, e.g., 18S rDNA. For example, a 5.8S analysis that included sequences from three fungi, one green alga, a moss, and three angiosperms (Troitsky and Bobrova 1986) produced the "correct" topology among the land plants but an unresolved trichotomy among land plants, the green alga, and fungi. Casting further suspicion on 5.8S phylogenetic signal are the potential for character nonindependence because of secondary structural constraints (Wheeler and Honeycutt 1988) and mutational saturation such as characteristic of 5S rDNA (Steele et al. 1991).

In the course of analyzing conserved sequence and structural motifs in angiosperm ITS sequences (Hershkovitz and Zimmer 1996), we discovered (via FASTA searches) similarities between 5.8S/ITS sequences reported for eight species of monkeyflower (*Mimulus guttatus* complex; Scrophulariaceae; Ritland and Straus 1993; Ritland, Ritland, and Straus 1993) and those of green algae. Subsequently, a typical angiospermous ITS region was recovered from *Mimulus guttatus* (A. Liston, personal communication), confirming our suspicion that the reported sequences were contaminants, but also prompting this reconsideration of deep-level phylogenetic signal in the ITS region. In this study, we compare 5.8S phylogenetic trees with those derived using rDNA-18S sequences and independent evidence and compare the amount of sequence variation, relative base substitution patterns, and among-site rate heterogeneity in 5.8S and 18S. We also compare conservation patterns in ITS2 sequences from fungi, green algae, and angiosperms and determine the degree to which ITS2 sequences from these taxa cluster on the basis of pairwise alignability.

Methods

Taxon Sampling

The 5.8S analysis included representative angiosperms, green algae (including two of the putative *Mimulus* sequences), available representatives of other chlorophytes, and fungal and protistan outgroups (table 1). Phylogenetic signal of the 5.8S data was compared with that of an 18S data set that was similarly representative taxonomically but differing in composition because of sequence availability.

We limited the spacer analysis to ITS2 because of sequence availability and because we found that it contained greater sequence conservation than ITS1. The sample comprised 2 of the putative *Mimulus* sequences, 5 fungi, 4 chlorophycean (here, for simplicity, including trebouxiophycean—see Friedl 1995) green algae, 2 conifers, and 14 diverse angiosperms, including representatives of 6 families of the monophyletic (Cronquist 1981; Chase et al. 1993) angiosperm order Caryophyllales. The Caryophyllales were included to demonstrate interfamilial phylogenetic signal. ITS2 conservation in angiosperms is considered in detail elsewhere (Hershkovitz and Zimmer 1996).

Key words: ITS, 5.8S, rDNA evolution, eukaryotes, maximum-likelihood, among-site rate variation.

Address for correspondence and reprints: Mark A. Hershkovitz, NCBI (Genbank), National Library of Medicine, National Institute of Health, Bldg. 38A, 8600 Rockville Pike, Bethesda, Maryland 20984. E-mail: hershkov@ncbi.nlm.nih.gov.

Mol. Biol. Evol. 13(9):1276–1295. 1996

Table 1
Taxa, Acronyms, and Genbank Accessions

Acronym	Taxon	Family	Broader Affiliation	Genbank	Source
1. 5.8S sequences					
Angiosperms[a]					
ARABI.....	*Arabidopsis thaliana*	Brassicaceae	Magnoliopsida; eudicot	X52320	DNA
CANEL.....	*Canella winterana*	Winteraceae	Magnoliopsida; magnoliid dicot	L03844	DNA
CUCUR	*Cucurbita melo*	Cucurbitaceae	Magnoliopsida; eudicot	M36377	DNA
LACTU.....	*Lactuca sativa*	Asteraceae	Magnoliopsida; eudicot	L13957	DNA
LYCOP.....	*Lycopersicon esculentum*	Solanaceae	Magnoliopsida; eudicot	X52265	DNA
POTAM	*Potamogeton natans*	Potamogetonaceae	Liliopsida	—[b]	RNA
SILEN.....	*Silene dioica*	Caryophyllaceae	Magnoliopsida; eudicot	X86830	DNA
TRITI	*Triticum vulgare*	Poaceae	Liliopsida	M10469	RNA
VICIA.....	*Vicia faba*	Fabaceae	Magnoliopsida; eudicot	M10471	RNA
Gnetophytes					
EPHED.....	*Ephedra kokonika*	Ephedraceae	Gnetopsida, Ephedrales	X15676	RNA
GNETU	*Gnetum gnemon*	Gnetaceae	Gentopsida, Gnetales	—[b]	RNA
Conifers[c]					
PICEA.....	*Picea mexicana*	Pinaceae	Coniferopsida, Coniferales	U24251	DNA
PINUS.....	*Pinus contorta*	Pinaceae	Coniferopsida, Coniferales	U23956	DNA
TAXUS.....	*Taxus baccata*	Taxaceae	Coniferopsida, Taxales	X93991	DNA
Ferns					
MARSI.....	*Marsilea quadrifolia*	Marsileaceae	Filicopsida; Marsiliales	X15939	RNA
OSMUN.....	*Osmunda regalis*	Osmundaceae	Filicopsida; Filicales	X63199	RNA
Moss					
MNIUM	*Mnium rugicum*	Mniaceae	Bryopsida	X13432	RNA
Chlorophycean, trebouxiophycean, micromonadophycean, green algae[d]					
CHLOR	*Chlorella ellipsoidea*	?[e]	?	D13340	DNA
CHLAM.....	*Chlamydomonas reinhardtii*	Chlamydomonaceae	Chlorophyceae; Volvocales;	X65621	DNA
GONIU.....	*Gonium pectorale*	Volvocaceae	Chlorophyceae; Volvocales	U23534	DNA
MIURF.....	*"Mimulus nasutus"*	?[f]	?	L02799	DNA
MIURH.....	*"Mimulus guttatus"*	?[f]	?	L02801	DNA
PANDO	*Pandorina morum*	Volvocaceae	Chlorophyceae; Volvocales	U23530	DNA
SPERM.....	*Spermatozopsis similis*	?[g]	Chlorophyceae; Volvocales?	X69488	DNA
TETRA.....	*Tetraselmis striata*	Chlorodendraceae	Micromonadophyceae	X65967	DNA
Ulvophycean green alga					
CLADO	*Cladophora albida*	Cladophoraceae	Ulvophyceae	—[h]	DNA
Fungi[i]					
GLOMU.....	*Glomus mossae*	Endogonaceae	Zygomycotina; Endogonales	U16756	DNA
BIPOL.....	*Bipolaris oryzae*	Pleosporaceae	Ascomycotina; Loculoascomycetes	X78122	DNA
SACCH.....	*Saccharomyces cervesiae*	Saccharomycetaceae	Ascomycotina; Hemiscomycetes	K01048	RNA
SCLER.....	*Sclerotinia sclerotiorum*	Sclerotiniaceae	Ascomycotina; Discomycetes	M96382	DNA
HETER.....	*Heterobasidion annosum*	Schizophyllaceae	Basidiomycotina; Aphyllophorales	X70024	DNA
VOLVA	*Volvariella volvacea*	Pluteaceae	Basidiomycotina; Agaricales	U15973	DNA
Brown alga					
SCYTO.....	*Scytosiphon lomentaria*	Scytosiphonaceae	Phaeophyceae	D16558	DNA
Red algae					
CYANI.....	*Cyanidium caldarium*	Porphyridiaceae	Porphyridiales	X65077	DNA
SARCO.....	*Sarcodiotheca furcata*	Solieriaceae	Gigartinales	U21346	DNA
Dinoflagellates					
CRYPT.....	*Crypthecodinium cohnii*	Crypthecodiniaceae	Dinophyceae; Peridiniales	M25116	RNA
PRORO.....	*Prorocentrum micans*	Prorocentraceae	Dinophyceae; Prorocentrales	M14649	DNA
Oomycete					
PHYTO.....	*Phytophthora megasperma*	Pythiaceae	Peronosporales	X75632	DNA
2. 18S sequences					
Angiosperms[a]					
DRIMY18 ...	*Drimys winteri*	Winteraceae	Magnoliopsida; magnoliid dicot	L24089	DNA
GLYCI18	*Glycine max*	Fabaceae	Magnoliopsida; eudicot	X02623	RNA
GOSSY18 ...	*Gossypium hirsutum*	Malvaceae	Magnoliopsida; eudicot	L24145	DNA
HYDRO18...	*Hydrocotyle sibthorpioides*	Apiaceae	Magnoliopsida; eudicot	X16605	RNA
LYCOP18 ...	*Lycopersicon esculentum*	Solanaceae	Magnoliopsida; eudicot	X51576	DNA
ORYZA18 ...	*Oryza sativa*	Poaceae	Liliopsida	X00755	RNA
PRUNU18 ...	*Prunus persica*	Rosaceae	Magnoliopsida; eudicot	L28749	DNA
RIBES18.....	*Ribes aureum*	Grossulariaceae	Magnoliopsida; eudicot	L28143	DNA
SINAP18	*Sinapis alba*	Brassicaceae	Magnoliopsida; eudicot	X66325	DNA
SPINA18	*Spinacia oleracea*	Chenopodiaceae	Magnoliopsida; eudicot	L24420	DNA

Table 1
Continued

Acronym	Taxon	Family	Broader Affiliation	Genbank	Source
Gnetophyte					
EPHED18....	*Ephedra sinica*	Ephedraceae	Gnetopsida, Ephedrales	D38242	RNA
Conifer					
PINUS18	*Pinus griffithii*	Pinaceae	Coniferopsida	X75080	DNA
Cycadophyte					
ZAMIA18....	*Zamia pumila*	Zamiaceae	Cycadopsida	M20017	DNA
Ferns					
OSMUN18 ...	*Osmunda cinnamomea*	Osmundaceae	Filicopsida; Filicales	U18516	DNA
SALVI18	*Salvinia natans*	Salviniaceae	Filicopsida; Salviniales	X90413	DNA
Moss					
MNIUM18 ...	*Mnium hornum*	Mniaceae	Bryopsida	X80985	DNA
Chlorophycean, trebouxiophycean, micromonadophycean green algae[d]					
CHLAM18 ...	*Chlamydomonas reinhardtii*	Chlamydomonaceae	Chlorophyceae; Volvocales	M32703	DNA
CHLOR18 ...	*Chlorella ellipsoidea*	?[e]	?	X63520	DNA
SPERM18....	*Spermatozopsis similis*	?[g]	Chlorophyceae; Volvocales?	X65557	DNA
TETRA18....	*Tetraselmis striata*	Chlorodendraceae	Micromonadophyceae	X70802	DNA
VOLVO18 ...	*Volvox carteri*	Volvocaceae	Chlorophyceae; Volvocales	X53904	DNA
Ulvophycean green alga					
CLADO18 ...	*Cladophora albida*	Cladophoraceae	Ulvophyceae	Z35317	DNA
Fungi[i]					
GIGAS18	*Gigaspora margarita*	Gigasporaceae	Zygomycotina; Gigasporales	X58726	DNA
GLOMU18 ...	*Glomus intraradices*	Endogonaceae	Zygomycotina; Endogonales	X58725	DNA
OMPHA18 ...	*Omphalina umbellifera*	Tricholomataceae	Basidiomycotina, Agaricales	U23543	DNA
PLEUR18	*Pleurotus ostreatus*	Polyporaceae	Basidiomycotina; Aphyllophorales	U23544	DNA
SCLER18	*Sclerotinia sclerotiorum*	Sclerotiniaceae	Ascomycotina; Discomycetes	L37541	DNA
Brown alga					
SCYTO18....	*Scytosiphon lomentaria*	Scytosiphonaceae	Phaeophyceae	L43066	DNA
Red algae[k]					
PORPH18....	*Porphyridium aerugineum*	Porphyridiaceae	Porphyridiales	L27635	RNA
RHODE18 ...	*Rhodella maculata*	Porphyridiaceae	Porphyridiales	U21217	RNA
SARCO18....	*Sarcodiotheca furcata*	Solieriaceae	Gigartinales	U43553	DNA
Dinoflagellate					
CRYPT18....	*Crypthecodinium cohnii*	Crypthecodiniaceae	Dinophyceae; Peridiniales	M64245	DNA
Oomycete					
PHYTO18....	*Phytophthora megasperma*	Pythiaceae	Peronosporales	X54265	DNA
3. ITS2 sequences					
Angiosperms[a]					
AMARA.....	*Amaranthus retroflexus*	Amaranthaceae	Magnoliopsida; eudicot; Caryophyllales	L48798	DNA
ARABI	*Arabidopsis thaliana*	Brassicaceae	Magnoliopsida; eudicot	X52320	DNA
CANEL......	*Canella winterana*	Winteraceae	Magnoliopsida; magnoliid dicot	L03844	DNA
CUCUR	*Cucurbita melo*	Cucurbitaceae	Magnoliopsida; eudicot	M36377	DNA
LACTU......	*Lactuca sativa*	Asteraceae	Magnoliopsida; eudicot	L13957	DNA
LYCOP......	*Lycopersicon esculentum*	Solanaceae	Magnoliopsida; eudicot	X52265	DNA
MAIAN......	*Maianthemum racemosum*	Liliaceae	Liliopsida	U24041	DNA
MOLLU	*Mollugo verticillata*	Molluginaceae	Magnoliopsida; eudicot; Caryophyllales	L48799	DNA
PHYTL......	*Phytolacca americana*	Phytolaccaceae	Magnoliopsida; eudicot; Caryophyllales	L48800	DNA
SILEN.......	*Silene dioica*	Caryophyllaceae	Magnoliopsida; eudicot; Caryophyllales	X86830	DNA
TALIN	*Talinum paniculatum*	Portulacaceae	Magnoliopsida; eudicot; Caryophyllales	L48801	DNA
TETRG......	*Tetragonia tetragonioides*	Aizoaceae	Magnoliopsida; eudicot; Caryophyllales	L48802	DNA
TRITU	*Triticum urartu*	Poaceae	Liliopsida	X66108	DNA
VICIF	*Vicia faba*	Fabaceae	Magnoliopsida; eudicot	X17535	DNA
Conifers					
PICEA.......	*Picea mexicana*	Pinaceae	Coniferopsida, Coniferales	U24251	DNA
PINUS.......	*Pinus contorta*	Pinaceae	Coniferopsida, Coniferales	U23956	DNA
Chlorophycean, trebouxiophycean, micromonadophycean green algae[d]					
CHLOR	*Chlorella ellipsoidea*	?[e]	?	D13340	DNA
GONIU......	*Gonium pectorale*	Volvocaceae	Chlorophyceae; Volvocales	U23534	DNA
MIURF......	"*Mimulus nasutus*"	?[f]	?	L02799	DNA
MIURH......	"*Mimulus guttatus*"	?[f]	?	L02801	DNA
PANDO	*Pandorina morum*	Volvocaceae	Chlorophyceae; Volvocales	U23530	DNA
SPERM......	*Spermatozopsis similis*	?[g]	Chlorophyceae; Volvocales?	X69488	DNA
SPONG......	*Spongiochloris spongosa*	Chlorococcaceae	Chlorococcales; Chlorophyceae	U34776	DNA

Table 1
Continued

Acronym	Taxon	Family	Broader Affiliation	Genbank	Source
Fungi[i]					
BIPOL.......	*Bipolaris oryzae*	Pleosporaceae	Ascomycotina; Loculoascomycetes	X78122	DNA[j]
GIGAS	*Gigaspora margarita*	Gigasporaceae	Zygomycotina; Gigasporales	X84233	DNA
HETER......	*Heterobasidion annosum*	Schizophyllaceae	Basidiomycotina; Aphyllophorales	X70024	DNA
SACCH......	*Saccharomyces cerevesiae*	Saccharomycetaceae	Ascomycotina; Hemiscomycetes	K01048	RNA
VOLVA	*Volvariella volvacea*	Pluteaceae	Basidiomycotina; Agaricales	U15973	DNA

NOTE.—Within sequence data subsets, taxa are grouped by broader affiliation. The acronyms are those used in tables 2 and 4 and figures 1–8. The classification is intended only as a guide to the diversity—see footnoted references for additional information on presumed interrelationships within groups represented by three or more samples.

[a] Cronquist (1981); Chase et al. (1993).

[b] Not currently accessioned in Genbank; transcribed from Troitsky et al. (1991).

[c] Page (1990).

[d] Mattox and Stewart (1984); Friedl (1995).

[e] *Chlorella* is polyphyletic; species identification and affinities remain problematic; Huss and Sogin (1990); Wilcox et al. (1992); Friedl (1995).

[f] Organismal origin and precise phylogenetic relationships not inferrable from present data.

[g] Position among green algae remains unresolved; Steinkötter et al. (1994).

[h] Not currently accessioned in Genbank; from Bakker, Olson, and Stam (1995; fig. 1, "*C. albida* Roskoff").

[i] Gargas et al. (1995).

[j] The Genbank documentation erroneously indicates RNA.

[k] Garbary and Gabrielsen (1990).

Multiple Sequence Alignments

Multiple alignments were constructed with the aid of the PILEUP program of GCG version 8.1 (Genetics Computer Group 1994), CLUSTALW (Thompson, Higgens, and Gibson 1994), and CLUSTALV (Higgins, Bleasby, and Fuchs 1992): for (1) total 5.8S data; (2) green algal ITS2 sequences; and (3) green algal, fungal, and seed plant ITS2 sequences. Alignments and boundary approximations were adjusted manually using the GDE version 2.2 (Smith 1994) multiple sequence editor. For the 18S sequences, we constructed a manual alignment for the region spanning positions 24–1,783 of the *Oryza sativa* sequence (table 1). We truncated the alignment because the 5′ and 3′ regions were unavailable for many of the sequences. The broader ITS2 alignments were derived using CLUSTALV, which applies the "fast, dirty" algorithm used in CLUSTALW. We used a 10-base window to detect conserved motifs and high (8–15) gap-opening combined with low (2–4) gap-extension penalties to extend unconserved sequence as necessary to accommodate conserved motif alignment. To avoid input-order bias, we interspersed sequences from the different major groups. We excluded CLADO from the alignment because we found no primary sequence similarity to green algal or any other samples. A FASTA search inputting the CLADO ITS2 retrieved no other ITS2 sequences in the database.

Phylogenetic/Comparative Sequence Analysis

5.8S Sequences

Phylogenetic analyses of a 5.8S data set modified from the alignment were carried out using maximum likelihood (ML) and maximum parsimony (MP) procedures in prerelease test versions d27–d45 of PAUP* 4.0 (Swofford 1996). Delimitation of the 5.8S sequence was approximate because of apparent inconsistencies in the reported boundaries among the taxa (see Results). We excluded the highly diverged 5.8S sequence of the dinoflagellate *Crypthecodinium* (table 1) from the final analysis but determined in preliminary analyses that it always paired with the *Prorocentrum* (PRORO) sequence. Using ML under the six-parameter general time-reversible model with empirically determined base frequencies, we obtained preliminary estimates of the relative substitution rates and parameters describing among-site rate heterogeneity (α, the shape parameter of the gamma distribution, Yang 1994; and ρ, the proportion of invariant sites; Swofford et al. 1996) over MP trees. An initial ML tree was generated using the initial rate parameter estimates and the heuristic search procedure in PAUP* with the tree-bisection/reconnection (TBR) swapping algorithm. Substitution rate and heterogeneity parameters were reestimated over the initial ML tree, and these were used to perform TBR swapping on the initial tree to generate a second and final ML tree. This recursive procedure was repeated, but the resulting ML topology was the same. Parameters were reestimated over the final tree and used to estimate pairwise distances. We also compared base compositional bias in angiosperms and green algae. Finally, the data set was analyzed and bootstrapped (Felsenstein 1985) 1,000 times using MP, and we derived MP and ML trees constrained for a topology supported by independent evidence.

18S Sequences

The 18S sequences were examined primarily to compare their phylogenetic signal and evolutionary pattern with those of 5.8S. Phylogenetic analysis was limited to the MP procedure above, including bootstrapping and determining the parsimony score for a tree constrained for a plausible independently supported topology. We estimated, as above, substitution rates and site-to-site rate heterogeneity using ML over the parsimony tree and calculated pairwise distances using the ML parameters.

ITS2 Sequences

Similarities among groups of ITS2 sequences were estimated using the guide tree feature in CLUSTALW, which yields a neighbor-joining tree based on similarity scores for pairs of sequences aligned optimally for given gap-opening and gap-extension penalties (see Hershkovitz and Zimmer 1996). Nominally, the guide tree topology functions as a phylogenetic template for successive sequence alignment, but we found that the guide tree provides a phenogram for comparing, without imposition of a fixed multiple alignment, a set of divergent and length-variable sequences. The most similar sequences form terminal bifurcations in the tree, and the internal branches cluster mutually more similar sequences. The branch lengths, however, are distance distorted in that the pairwise optimal alignments are noncommutative. For example, an optimal multiple alignment of the sequences AGAGAA, AGGAA, and AAGAA would not maintain the optimal pairwise alignments, but the guide tree branch lengths are based on the optimal pairwise scores. Nonetheless, the procedure will cluster sequences sharing alignable motifs relative to those that do not. For the ITS2 sequences, we generated trees using the "slow, accurate" algorithm, and, as in the multiple alignment described above, high (15–50) gap-opening combined with low (0–1) gap-extension penalties.

Results

5.8S and 18S Sequences

Sequence Alignment

The 5.8S sequences are substantially alignable (fig. 1). In the case of apparent alignment ambiguities, we optimized the alignment for groups of sequences that were otherwise most similar. Aside from the 5′ and 3′ ends, the 5.8S sequences show few indels, particularly within chlorophytes and fungi. A length-variable region between positions 121 and 150 generally aligns better within versus between major taxonomic groups. In the empirically deduced yeast 5.8S secondary structure model (Yeh and Lee 1991), the length-variable region corresponds to a small stem structure that begins with the G residue at position 118 in the alignment and extends to C at 149. This stem is reproduced in other taxa folded according to the yeast model (CHLAM, Thompson and Herrin 1994, Fig. 7c; CLADO, Bakker, Olsen, and Stam 1995, Fig. 2) or other criteria (Suh, Thien, and Zimmer 1992; Ritland and Straus 1993; Troitsky and Bobrova 1986).

The 5′ and 3′ 5.8S boundaries reported for the various sequences are inconsistent. This might be attributable to a paucity of RNA-based sequences, misalignment with existing RNA-based sequences (e.g., CLADO; Bakker, Olsen, and Stam 1995; cf. Thompson and Herrin 1994), or clerical error in Genbank documentation (e.g., GONIU, cf. Coleman, Suarez, and Goff 1994; CANEL, cf. Suh, Thien, and Zimmer 1992). Another factor might be developmentally regulated 5.8S rRNA length variability. In yeast, a long form, possibly derived from an alternative pre-rRNA processing pathway (Henry et al. 1994; Lygerou et al. 1994; cf. Lindahl, Archer, and Zengel 1994; van Nues et al. 1994), extends 6–7 bases 5′ from the short form. The *Saccharomyces* (SACCH) sequence in figure 1 represents the short (cf. van Nues et al. 1994; fig. 2A) and apparently more abundant (Henry et al. 1994) form; the long form extends to the 5′ end of the figure 1 alignment or 1 base upstream. Long 5.8S forms reportedly exist in other eukaryotes as well (Henry et al. 1994).

Another potential alignment problem is sequencing error, particularly in the RNA-based sequences. Although these are useful for delineating 5.8S boundaries, they lack complementary-strand confirmation. Thompson and Herrin (1994), for example, found 12 errors in the RNA-based *Chlamydomonas* sequence in Genbank.

Properties of 5.8S Sequences

Table 2 shows that the observed proportion of variable sites in 5.8S is roughly twice that for 18S, although the absolute number of variable 5.8S sites is much less. For both data sets, the observed proportion of variable sites increases as more distantly related taxa are included. When estimated simultaneously, α and ρ both generally increase as more distantly related taxa are included (table 3). When each parameter is estimated without allowance for the other, α decreases and ρ increases with the inclusion of more distantly related taxa. For 5.8S, the α and ρ estimates for the fungal plus chlorophyte subset differs markedly from those of the next larger and smaller taxon subsets. Obviously, the proportion of invariant sites cannot decrease as more distantly related taxa are excluded. In order to examine the effect of scoring 18S regions unalignable with angiosperms as "missing," we estimated parameters over an 18S parsimony tree with the embryophytes pruned. This removes ca. 30 of the most variable positions from the alignment (table 2), although the fungi and protist sequences still have many sites scored as missing. In any case, the among-site rate heterogeneity estimates for this taxon complement are $\alpha/\rho = 0.58/0.26$, $\alpha = 0.33$, and $\rho = 0.53$ (cf. table 3).

The 5.8S divergences appear to be roughly twice those of 18S (table 2). The 18S divergences are increasingly underestimated with increasing distance from angiosperms, however, because nonangiosperm 18S regions not alignable with angiosperms were scored as missing. We presume that poor alignability reflects extreme sequence divergence.

In both data sets, C–T transition rates are ca. twice that of A–G transitions and five times that of each transversion (table 3). The G–T rate among the 5.8S sequences is ca. half that of other transversions.

We examined base compositional patterns because we noticed that many A residues shared between the reported *Mimulus* sequences and green algae aligned with G residues in angiosperms. This suggested the possibility that similarities between the *Mimulus* and green algal sequences represented a convergent increase in adenosine. Table 4 shows base composition for subsets of positions varying between the nine angiosperm and six chlorophycean algae. These comparisons are not statistically rigorous because they ignore phylogenetic

```
                    1                                                                                                      90
                    |                        I           II                    III  IV                   IV'      III'      |
1   ARABI    taaA AACGACTCTC GGCAACGGAT ATCTCGGCTC TCGCATCGAT GAAGAACGTA GCGAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA
2   LYCOP    acca aacgac.... .......... .......... .......... .......... .......... .......... .......... ..........
3   CUCUR    taaC A......... .......... .......... .......... .......... .......... .......... .......... G......C..
4   triti     CAC -.c....... .......... .......... .......... .......... .......... ....C..... .......... .......C..
5   LACTU    caCA A......... .......... ..T....... A......... .......... ..A....... .......... .......... ..........
6   SILEN    atta A......... .......... .......... .......... .......... .......... .......... .......... ..........
7   vicia    atAG A.T....... .......... ....A..... .T........ .......... .......... .......... .......... ..........
8   CANEL    TCAA G.T....... .......... .......... .......... .......... ..A....... .......... .......... .......C..
9   potam      AT C.T....... .......... ....T..... .......... .......... .......... .......... .....G.... ..........
10  ephed      CT T......... .......... .......... .......... .......... .......... .......... .......... ..........
11  gnetu      CC C........- .A........ .......... ...TT-.... .......... .......... .......... .......... ..........
12  PINUS     *GA A......... .T........ .......... .T.TTA.... .......... .......... .......... .......... ..........
13  PICEA         *.T.......            .......... .T.TCAAC.. .......... .......... .......... .......... ..........
14  TAXUS      CT TGGC...... .......... .......... ...-CA.... .......... .......... ..........            ..........
15  marsi          .G....... A......... ....T..... .T...A.... ........C. .......... ....G..A.. .......... ...T...C..
16  osmun          ......... A......... ....T..... .T...A.... ........C. .......... ....G..... ....C..... ...T...C..
17  mnium          .TA..C... A......... ....T..... .T...A.... ........C. .......... ....G..... .......... ...T...C..
18  CHLAM    ccaa gacA...... AA........ ....T..... ...G...... ........C. .......... ....G..... .......... ..ATA.....
19  GONIU    ccaa G..A...... AA........ ....T..... ..AG...... ........C. .......... ....G..... .......... ..ATA.....
20  PANDO    ccaa G..A...... AA........ ....T..... ..AG...... ........C. .......... ....G..... .......... ..ATA.....
21  MIURH    caGA G..A...... AA........ ....T..... .....A.... ........C. .......... ....G..... .......... ...T......
22  MIURF    caAA G..A...G.G TA........ ....T..... ...T.A.... ........C. .......... ....G..... .......... ...T......
23  SPERM    caaa gacA...... AA........ ....T..... ...T.A.... ........C. .......... ....G..... .......... ...T......
24  TETRA      *A G..A...... AA........ ....T..... .TA..A.... ........C. .......... ....G..... .......... ...T......
25  CHLOR    aaaT G..A...... AA........ ....T..... C......... ........C. .......... ....G..... .......... ...T......
26  CLADO    tgat agtaaca..g ta........ T...T..... C.A....... ........C. ..A..GC... ...GG..... .......... ...T......
27  HETER    ctta t..A...T.. AA........ C......... .......... ........C. .......... ...AG..A.. .......... ...T.A....
28  VOLVA    taaa ta.A...T.. AA........ C...T..... .......... ........C. .......... ...AG..A.. .......... ...T.A....
29  SCLER    tagt taAA...T.. AA........ C...T..T.. .G........ ........C. .......... ...AG..A.. .......... ...T.A....
30  sacch    tatt aaAA...T.. AA........ C...T..T.. .......... ........C. .......... ....G..A.. .......... ...T......
31  BIPOL    ttat ta.A...T.. AA........ C...T..T.. .G........ ........C. .......... ....G..... .......... ...T.A....
32  GLOMU    aaaa gaTC...T.. AA........ C...T..... .......... ........C. .......... ...AG..... .........T ...TTT....
33  CYANI    CGAA AGAAG..... AA..G..... .......... ..TTGA.... ........C. -.A...A... .A.TG..... ....C....A GT.TT.....
34  SARCO    aaat ta.A....AT .A..G..... G...T..... CA...A.... ..C.....C. .....T.... .A.-G....A C...C..... ..-.......
35  PHYTO      AT AG.A...T.. A...G..... G...A..... G.A....... ........CT ....C..... ....G..A.. C......... G..T.A....
36  SCYTO    GTTG T.AA...T.. A......... G...T..... C.A..A.... ........C. .......... ....G..T.. C..C...... .....A....
37  crypt          ..A...T.. A...G.T... TC..T..T.. AGA.C..... ....GG.ACT ......-.T. .-.TGG.A.. ....-..... GCAT...G..
38  PRORO    gtat t..A...T.. A......... G......... GAA..A.... ....GG..C. ........T. ...AGC.A.. .......... ...T......


             91                                                                            170
             | II'             I'          V ....                   ....V'                   |
1   ARABI    ACCATCGAGT CTTTGAACGC AAGTTGCGCC C---CAAGCC TTCT-GGCCG ----AGGGCA CGTCTGCCTG GGTGTCACAA
2   LYCOP    .......... .......... .......... C---GAAGCC ATTT-GGCCG ----A..... .......... ..C...ACG-
3   CUCUR    ....C..... .......... .......... C---GGAGCC TTCT-GGCCG ----A..... .......... ..C...ACGC
4   triti    .......... .......... .......... C---GAGGCC ACTC-GGCCG ----A..... ..C....... ..C...ACGC
5   LACTU    .......... T......... .......... C---GAAGCC ATCC-GGCTG ----A..... ..C....... ..C...ACGC
6   SILEN    .......... T......... .......... C---GAAGC- -TTC-GGCTG ----A..... .......... ..C...ACgc
7   vicia    .......... .......... .......... C---GATGCC ATTA-GGTTG ----A..... .......... ......ACAT
8   CANEL    .......... .......... .......... C---GAGGCC ACTA-GGCTG ----A....T ..C......- ..C...ACgc
9   potam    .......... T......... .......... C---TAAGCT TCCG-GGCCG ----A..... A.C....G.. ..C...AAGC
10  ephed    .T........ .......... ......-... C---GAAGCC --TC-CGCCA ----A..... .......... ..C...GCAA
11  gnetu    .T......A. .......... ......-... CTCCG-AGCC --TA-GGCCG ----A..... .......... ..C...GCAA
12  PINUS    .T........ T......... ..T....... C---GAGGCC --TC-GGTCG ----A..... ......T... ..C...GCAT cc
13  PICEA    .T........ T......... ..T....... C---GAGNCC --TC-GGTCG ----A..... ......T... ..C...GCAT CC
14  TAXUS    .TG....... .......... ......GC.. G---GAG-C- --TC-GGCCG ----A..... .......T.. ..C..-GCAC
15  marsi    .T......A. .......... ..-...-... C---GAGGC- --TC-GTCCG ----A..... .......... A.C...CAC
16  osmun    .TA....... .......... ......-... C---GCGGC- --TC-GTCCA ----A..... T.C....... A.C...CTC
17  mnium    .T........ .......... .......... C---GAGGC- --TC-GTCCG ----A..... TT..C.TTA. A.C...ACC
18  CHLAM    .T......A. .......... .TA......T C---GAGGC- -TTC-GGCCA ----A.A... T........C A.C...GGGT Ta
19  GONIU    .T......A. .......... .TA......T C---GAGGC- -TTC-GGCCA ----A.A... T........C A.C...GGGT T
20  PANDO    .T......A. .......... ..A......T C---GAGGC- -TTC-GGCCA ----A.A... T........C A.C...GGGT TA (+26 bases)
21  MIURH    ........A. .......... .TA......T C---GAGGC- -TTC-GGCCA ----A.A... T........C A.C...GGGT Ta
22  MIURF    ......A.A. .......... .TA......T C---AAGGC- -TTC-GGCCG ----A.A... T........C A.C...GGGT Ta
23  SPERM    ........A. .......... .TC......T C---GAGGC- -TTC-GGCCG ----A.A... T..T.....C A.C...GG-T TT
24  TETRA    ........A. .......... .TA......T C---GAGGC- -CTC-GGCCA ----A.A... ..C......C A.A...GGTT TT
25  CHLOR    .T......A. .......... .TA......T C---GAGAC- -CTC-GGTCA ----A.A... T.....G..C A.C...GGCT CA
26  CLADO    .T......A. T......... .CA......T C---AAGTC- -TAC-GGACT ----T.A... T........C A.C...GTTT TAAATGGC
27  HETER    .T......A. .......... .CC....... CTTT--GGT- ATT----CCG A---A..-.. ..C...TT.. A.....GT-- GAa
28  VOLVA    .T......A. .......... .CC......T CTTT--GGCC ATT----CCG A---A.A... T.C...TT.. A.....ATC- gaa
29  SCLER    .T......A. .......... .CA....... CCTT--GGT- ATT----CCG G---G..... T.C...TTC. A.C...ATTT caa
30  sacch    .T......A. .......... .CA....... CCTT--GGT- ATT----CCA G---G..... T.C...TT.. A.C...ATTT cct
31  BIPOL    .T......A. .......... .CA....... CTTT--GGT- ATT----CCA A---A..... T.C...TTC. A.C...ATTT gta
32  GLOMU    .T......A. .......... ..A......T CCCT--GGT- ATT----CCG G---G.A.T. T.C...TT.. A.G...GTTa gaa
33  CYANI    .T...T..A. .......... .........T TTC--AGGAG AATTTTATTT TCCTG.A... ......TTG. A.....TCTt tga
34  SARCO    .T......A. T......... ...--....T CGC---GGTA A-TC------ --CTGCA.-- T.....TT.. A.....Cgca caa
35  PHYTO    .T......AA T......... .TA....A.T TCCG--GGTT AGTC---CTG ----G.A.T. T.C...TA.C A.....CGTA cat
36  SCYTO    .T....A.AA .......... .T-C....-T TCCG--GGAT ATGC---CTG ----G.A... T.CT..T.G. A.....TGTT GAC
37  crypt    .TTGAGAGC. TC..C..T.. .TAC..---C TCC---AGC- -TGA---CT- ---------- T...A.TTG. ATAA..TCCC TCA
38  PRORO    ....ATAG.G AC.......T .TAC.....T TTCG--GGAT ATCC---CTG ----AA.... T.C....T.C A...tctatt ctt
```

FIG. 1.—Provisional alignment of 5.8S rDNA sequences. Taxon acronyms, definitions, and sequence documentation are given in table 1. Acronyms in lowercase letters denote RNA-based sequences. Upper- and lowercase bases denote, respectively, the 5.8S and flanking ITS regions according to Genbank documentation, except for the SCYTO sequence, in which the boundaries were not indicated. Where sequence data are available the 5′-end of the alignment extends to include the 5′-most bases reported for the CYANI sequence, and the 3′ ends are, except for the green algae, extended to include the 3′-most bases reported within each major taxonomic group. The anomalous 3′ extensions for PANDO (not shown) and CLADO are presumably erroneously included in the ITS2 sequence. The first 3 bases of the CANEL sequence are actually ITS1 (Suh, Thien, and Zimmer 1992). The SACCH 5.8S is RNA based, but the flanking sequence is DNA based (table 1). The 5′ flanking sequence, however, is part of a "long" 5.8S form (see text). Position 1 in the alignment corresponds to character 1 in the phylogenetic analysis data set (appendix). The length-variable region (ca. 120–150) is aligned optimally within major groups but arbitrarily between them. In the yeast 5.8S secondary structure model, the superposed regions I–V pair with I′–V′. Pairing relationships in the length-variable region (V–V′) probably varies among the taxa.

Table 2
Divergence Matrices for 5.8S and 18S rDNA Sequences

	Sites	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5.8S sequences																
1 ARABI.....	162	—	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2 LYCOP.....	161	0.03	—	.	.	.	.	.	.	.	.	.	.	.	.	.
3 CUCUR	162	0.05	0.04	—	.	.	.	.	.	.	.	.	.	.	.	.
4 TRITI.....	161	0.08	0.04	0.07	—	.	.	.	.	.	.	.	.	.	.	.
5 LACTU	162	0.08	0.05	0.08	0.07	—	.	.	.	.	.	.	.	.	.	.
6 SILEN	159	0.05	0.02	0.05	0.05	0.03	—	.	.	.	.	.	.	.	.	.
7 VICIA	162	0.08	0.06	0.11	0.10	0.11	0.07	—	.	.	.	.	.	.	.	.
8 CANEL	161	0.12	0.07	0.10	0.04	0.08	0.07	0.09	—	.	.	.	.	.	.	.
9 POTAM	162	0.13	0.12	0.14	0.12	0.13	0.11	0.18	0.14	—	.	.	.	.	.	.
10 EPHED.....	158	0.08	0.06	0.11	0.09	0.12	0.07	0.12	0.13	0.17	—	.	.	.	.	.
11 GNETU	157	0.08	0.06	0.11	0.09	0.13	0.09	0.10	0.11	0.17	0.06	—	.	.	.	.
12 PINUS	158	0.14	0.11	0.17	0.13	0.16	0.10	0.13	0.18	0.24	0.12	0.10	—	.	.	.
13 PICEA	151	0.17	0.14	0.20	0.15	0.19	0.13	0.13	0.18	0.23	0.12	0.12	0.04	—	.	.
14 TAXUS.....	156	0.16	0.13	0.18	0.13	0.19	0.13	0.18	0.17	0.24	0.10	0.13	0.14	0.15	—	.
15 MARSI.....	155	0.20	0.17	0.20	0.15	0.23	0.17	0.20	0.20	0.25	0.17	0.17	0.17	0.19	0.18	—
16 OSMUN.....	156	0.23	0.19	0.21	0.15	0.24	0.19	0.23	0.21	0.27	0.17	0.21	0.21	0.25	0.22	0.07
17 MNIUM.....	157	0.23	0.20	0.23	0.18	0.26	0.20	0.22	0.22	0.28	0.22	0.25	0.21	0.22	0.22	0.10
18 CHLAM.....	157	0.28	0.22	0.28	0.23	0.29	0.23	0.28	0.27	0.33	0.21	0.21	0.25	0.30	0.28	0.16
19 GONIU.....	160	0.29	0.23	0.30	0.24	0.31	0.24	0.31	0.29	0.36	0.23	0.23	0.27	0.32	0.34	0.18
20 PANDO	160	0.28	0.22	0.28	0.23	0.30	0.23	0.30	0.27	0.34	0.22	0.22	0.25	0.30	0.32	0.17
21 MIURH	160	0.24	0.19	0.25	0.19	0.26	0.20	0.26	0.24	0.30	0.21	0.21	0.22	0.26	0.27	0.14
22 MIURF.....	160	0.29	0.24	0.30	0.24	0.32	0.25	0.32	0.29	0.35	0.29	0.23	0.25	0.30	0.33	0.19
23 SPERM.....	159	0.24	0.20	0.26	0.20	0.27	0.21	0.26	0.25	0.32	0.23	0.19	0.20	0.23	0.27	0.13
24 TETRA.....	160	0.28	0.25	0.31	0.21	0.30	0.25	0.27	0.26	0.34	0.24	0.25	0.23	0.28	0.31	0.15
25 CHLOR	160	0.30	0.25	0.31	0.24	0.32	0.26	0.29	0.28	0.36	0.23	0.23	0.24	0.27	0.31	0.17
26 CLADO	159	0.49	0.23	0.53	0.48	0.47	0.44	0.50	0.50	0.58	0.44	0.42	0.46	0.48	0.53	0.39
27 HETER.....	157	0.22	0.46	0.27	0.22	0.27	0.24	0.26	0.28	0.32	0.18	0.19	0.25	0.28	0.29	0.17
28 VOLVA	160	0.26	0.24	0.31	0.25	0.31	0.27	0.28	0.32	0.31	0.26	0.27	0.31	0.36	0.36	0.18
29 SLCER.....	160	0.32	0.27	0.35	0.29	0.35	0.31	0.31	0.35	0.35	0.30	0.31	0.34	0.38	0.38	0.19
30 SACCH	158	0.25	0.30	0.28	0.24	0.28	0.24	0.25	0.27	0.28	0.24	0.25	0.27	0.32	0.30	0.15
31 BIPOL	160	0.26	0.24	0.29	0.24	0.29	0.25	0.27	0.30	0.30	0.24	0.26	0.28	0.32	0.34	0.18
32 GLOMU.....	160	0.33	0.34	0.39	0.34	0.40	0.35	0.36	0.36	0.39	0.31	0.31	0.38	0.41	0.40	0.26
33 CYANI.....	146	0.47	0.47	0.49	0.51	0.53	0.48	0.48	0.54	0.68	0.47	0.43	0.45	0.50	0.50	0.38
34 SARCO	141	0.42	0.44	0.51	0.47	0.50	0.43	0.46	0.54	0.51	0.37	0.38	0.45	0.49	0.53	0.33
35 PHYTO	143	0.46	0.47	0.48	0.47	0.46	0.44	0.46	0.52	0.53	0.45	0.49	0.52	0.59	0.60	0.38
36 SCYTO.....	145	0.46	0.46	0.52	0.45	0.52	0.47	0.46	0.53	0.52	0.46	0.50	0.51	0.52	0.55	0.32
37 PRORO	147	0.54	0.54	0.59	0.53	0.56	0.54	0.55	0.60	0.61	0.55	0.67	0.61	0.65	0.66	0.40
18S sequences																
1 PRUNU18....	1733	—	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2 SPINA18	1726	0.03	—	.	.	.	.	.	.	.	.	.	.	.	.	.
3 DRIMY18....	1723	0.03	0.04	—	.	.	.	.	.	.	.	.	.	.	.	.
4 GOSSY18....	1717	0.03	0.03	0.04	—	.	.	.	.	.	.	.	.	.	.	.
5 RIBES18.....	1725	0.03	0.03	0.03	0.03	.	.	.	.	.	.	.	.	.	.	.
6 SINAP18	1736	0.03	0.04	0.04	0.03	0.04	—	.	.	.	.	.	.	.	.	.
7 GLYCI18	1742	0.03	0.03	0.03	0.03	0.03	0.04	—	.	.	.	.	.	.	.	.
8 HYDRO18 ...	1609	0.04	0.04	0.04	0.04	0.04	0.05	0.04	—	.	.	.	.	.	.	.
9 ORYZA18 ...	1744	0.04	0.04	0.05	0.04	0.05	0.06	0.05	0.06	—	.	.	.	.	.	.
10 LYCOP18....	1735	0.03	0.03	0.04	0.03	0.03	0.03	0.04	0.04	0.04	—	.	.	.	.	.
11 EPHED18....	1712	0.09	0.08	0.09	0.09	0.08	0.09	0.09	0.10	0.09	0.08	—	.	.	.	.
12 ZAMIA18....	1738	0.08	0.07	0.08	0.07	0.07	0.08	0.08	0.09	0.08	0.07	0.08	—	.	.	.
13 PINUS18	1737	0.09	0.08	0.09	0.09	0.08	0.09	0.09	0.10	0.09	0.08	0.07	0.07	—	.	.
14 SALVI18	1739	0.10	0.09	0.10	0.10	0.10	0.11	0.10	0.11	0.11	0.10	0.11	0.09	0.10	—	.
15 OSMUN18 ...	1675	0.10	0.09	0.10	0.09	0.09	0.10	0.10	0.11	0.10	0.10	0.10	0.08	0.09	0.05	—
16 MNIUM18 ...	1706	0.09	0.08	0.10	0.09	0.09	0.10	0.09	0.11	0.09	0.09	0.09	0.08	0.09	0.04	0.04
17 CHLAM18 ...	1642	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.13	0.13	0.13	0.13	0.13	0.11	0.12
18 VOLVO18 ...	1642	0.13	0.13	0.13	0.13	0.12	0.14	0.13	0.14	0.13	0.13	0.13	0.13	0.13	0.11	0.10
19 SPERM18....	1640	0.13	0.12	0.13	0.12	0.13	0.13	0.13	0.14	0.13	0.13	0.13	0.13	0.13	0.10	0.10
20 TETRA18....	1634	0.13	0.12	0.13	0.13	0.13	0.13	0.13	0.14	0.13	0.13	0.12	0.12	0.13	0.10	0.11
21 CHLOR18....	1631	0.12	0.11	0.12	0.11	0.12	0.12	0.11	0.12	0.11	0.11	0.11	0.11	0.12	0.09	0.09
22 CLADO18 ...	1628	0.27	0.26	0.27	0.27	0.26	0.27	0.27	0.30	0.26	0.26	0.26	0.26	0.27	0.25	0.27
23 SCLER18	1434	0.15	0.15	0.16	0.15	0.15	0.16	0.15	0.17	0.15	0.16	0.16	0.16	0.17	0.15	0.15
24 OMPHA18 ...	1473	0.18	0.18	0.19	0.18	0.19	0.19	0.18	0.21	0.19	0.18	0.20	0.19	0.20	0.17	0.18
25 PLEUR18....	1485	0.18	0.18	0.19	0.17	0.18	0.18	0.18	0.20	0.18	0.18	0.19	0.18	0.20	0.16	0.17
26 GIGAS18	1554	0.16	0.15	0.16	0.16	0.16	0.15	0.16	0.18	0.17	0.16	0.15	0.16	0.17	0.14	0.15
27 GLOMU18 ...	1527	0.18	0.17	0.18	0.18	0.18	0.17	0.17	0.20	0.18	0.18	0.17	0.18	0.18	0.16	0.17
28 SARCO18....	1455	0.34	0.35	0.35	0.35	0.35	0.35	0.35	0.43	0.35	0.35	0.36	0.33	0.36	0.33	0.36
29 RHODE18 ...	1484	0.18	0.19	0.19	0.19	0.18	0.19	0.19	0.22	0.19	0.19	0.20	0.19	0.21	0.18	0.20
30 PORPH18....	1480	0.17	0.17	0.17	0.17	0.17	0.18	0.17	0.20	0.17	0.17	0.17	0.17	0.17	0.15	0.17
31 PHYTO18....	1497	0.18	0.18	0.19	0.18	0.18	0.18	0.19	0.21	0.19	0.18	0.19	0.19	0.20	0.17	0.18
32 SCYTO18....	1493	0.15	0.16	0.16	0.15	0.16	0.15	0.16	0.18	0.16	0.15	0.15	0.15	0.16	0.14	0.15
33 CRYPT18....	1495	0.23	0.23	0.23	0.22	0.23	0.24	0.23	0.26	0.23	0.23	0.25	0.22	0.24	0.22	0.23

NOTE.—Taxon acronyms are listed in table 1. Divergences were calculated using substitution rates and among-site rate heterogeneity (α and ρ) corrections (se table 3) estimated using maximum likelihood (ML) over (for 5.8S) the ML tree (fig. 2) or (for 18S) the parsimony tree (fig. 5). The number of sites refers t positions scored as nucleotides. Distances from positions scored as missing or gaps were extrapolated from nucleotide positions. This extrapolation probabl underestimates true distance because poorly aligning, hence more extremely divergent, regions were scored as missing.

Table 2
Extended

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
—	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
0.11	—	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
0.20	0.22	—	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
0.21	0.24	0.01	—	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
0.20	0.22	0.01	0.01	—	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
0.16	0.18	0.03	0.04	0.05	—	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
0.24	0.24	0.08	0.09	0.10	0.05	—	.	.	.	.	.	.	.	.	.	.	.	.	.	
0.18	0.18	0.06	0.06	0.07	0.03	0.05	—	.	.	.	.	.	.	.	.	.	.	.	.	
0.17	0.21	0.09	0.08	0.09	0.05	0.10	0.07	—	.	.	.	.	.	.	.	.	.	.	.	
0.19	0.21	0.06	0.07	0.08	0.05	0.11	0.08	0.08	—	.	.	.	.	.	.	.	.	.	.	
0.42	0.42	0.23	0.24	0.25	0.24	0.24	0.26	0.27	0.22	—	.	.	.	.	.	.	.	.	.	
0.20	0.22	0.20	0.22	0.22	0.19	0.25	0.20	0.19	0.18	0.27	—	.	.	.	.	.	.	.	.	
0.19	0.20	0.17	0.19	0.19	0.16	0.22	0.16	0.18	0.14	0.25	0.05	—	.	.	.	.	.	.	.	
0.22	0.21	0.21	0.25	0.26	0.22	0.28	0.24	0.23	0.21	0.28	0.09	0.09	—	.	.	.	.	.	.	
0.18	0.18	0.16	0.18	0.19	0.16	0.21	0.18	0.17	0.14	0.22	0.10	0.10	0.03	—	.	.	.	.	.	
0.17	0.17	0.17	0.20	0.20	0.17	0.22	0.19	0.18	0.16	0.27	0.09	0.08	0.05	0.05	—	.	.	.	.	
0.25	0.26	0.21	0.23	0.21	0.21	0.27	0.23	0.21	0.19	0.29	0.17	0.15	0.15	0.14	0.19	—	.	.	.	
0.41	0.38	0.40	0.43	0.41	0.39	0.46	0.37	0.38	0.41	0.45	0.40	0.41	0.43	0.36	0.44	0.42	—	.	.	
0.32	0.35	0.36	0.39	0.37	0.33	0.40	0.34	0.37	0.31	0.46	0.36	0.31	0.40	0.38	0.38	0.39	0.42	—	.	
0.41	0.45	0.30	0.30	0.32	0.29	0.37	0.33	0.29	0.27	0.32	0.28	0.25	0.32	0.26	0.31	0.31	0.48	0.41	—	
0.36	0.36	0.33	0.34	0.36	0.30	0.34	0.28	0.28	0.29	0.48	0.31	0.25	0.27	0.26	0.28	0.40	0.43	0.38	0.28	
0.42	0.48	0.45	0.45	0.47	0.37	0.42	0.39	0.34	0.41	0.57	0.38	0.37	0.40	0.40	0.40	0.57	0.68	0.57	0.42	0.36

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
—	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.10	—	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.10	0.01	—	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.09	0.05	0.05	—	.	.	.	.	.	.	.	.	.	.	.	.	.
0.09	0.07	0.07	0.07	—	.	.	.	.	.	.	.	.	.	.	.	.
0.08	0.05	0.06	0.05	0.04	—	.	.	.	.	.	.	.	.	.	.	.
0.24	0.23	0.23	0.22	0.23	0.20	—	.	.	.	.	.	.	.	.	.	.
0.14	0.16	0.16	0.17	0.16	0.14	0.29	—	.	.	.	.	.	.	.	.	.
0.17	0.19	0.19	0.20	0.19	0.18	0.33	0.10	—	.	.	.	.	.	.	.	.
0.16	0.18	0.18	0.19	0.19	0.17	0.31	0.09	0.03	—	.	.	.	.	.	.	.
0.14	0.16	0.16	0.16	0.15	0.14	0.27	0.09	0.11	0.11	—	.	.	.	.	.	.
0.16	0.20	0.19	0.19	0.19	0.17	0.32	0.11	0.14	0.14	0.05	—	.	.	.	.	.
0.33	0.38	0.38	0.40	0.39	0.38	0.52	0.37	0.35	0.35	0.33	0.35	—	.	.	.	.
0.18	0.22	0.22	0.23	0.21	0.20	0.31	0.22	0.22	0.22	0.20	0.21	0.32	—	.	.	.
0.15	0.18	0.18	0.17	0.18	0.15	0.23	0.20	0.23	0.22	0.20	0.20	0.33	0.17	—	.	.
0.16	0.20	0.20	0.19	0.17	0.17	0.32	0.20	0.21	0.21	0.18	0.21	0.38	0.25	0.21	—	.
0.14	0.18	0.18	0.18	0.15	0.15	0.26	0.16	0.19	0.18	0.16	0.18	0.38	0.22	0.19	0.12	—
0.21	0.24	0.25	0.23	0.23	0.22	0.34	0.24	0.25	0.26	0.23	0.26	0.46	0.27	0.27	0.23	0.24

Table 3
Properties of 5.8S and 18S rDNA Sequences

	Taxa Subsets								Relative Substitution Rates							
	1	2	3	4	5	6	Var	%Var	AC	AG	AT	CG	CT	α/ρ[a]	α[b]	ρ[c]
5.8S Sequences...	x	x	x	x	x	x	113	0.66	2.5	6.2	2.6	1.9	11.4	0.62/0.00	0.62	0.30
	x	x	x	x	x		90	0.53	2.4	5.0	2.3	1.7	9.7	1.05/0.26	0.49	0.43
	x	x	x	x			81	0.48	1.8	4.3	3.0	1.8	9.7	0.52/0.00	0.52	0.43
	x	x	x				76	0.45	2.1	4.4	2.7	1.9	10.6	0.50/0.00	0.50	0.44
	x	x					64	0.38	2.1	3.1	1.8	1.2	9.9	0.82/0.23	0.45	0.49
	x						53	0.31	2.5	4.3	3.0	1.9	13.3	0.96/0.33	0.40	0.54
18S Sequences ...	x	x	x	x	x	x	864	0.48	0.8	2.6	1.0	1.1	4.1	0.48/0.22	0.32	0.48
	x	x	x	x	x		617	0.34	0.8	2.6	1.0	1.0	5.7	0.55/0.33	0.29	0.57
	x	x	x	x			570	0.31	0.7	2.5	1.0	1.0	5.6	0.61/0.40	0.27	0.61
	x	x	x				470	0.26	0.7	2.5	1.0	0.9	5.7	0.68/0.49	0.22	0.67
	x	x					370	0.20	0.6	2.3	1.2	1.0	6.3	0.66/0.56	0.17	0.73
	x						304	0.17	0.5	2.4	1.6	0.9	7.3	0.68/0.62	0.11	0.77

NOTE.—Based on sequences listed in table 1. Taxon subsets are: 1—seed plants; 2—seedless embryophytes; 3—Chlorophyceae (including Trebouxiophyceae and Micromonadophyceae); 4—*Cladophora*; 5—fungi; 6—protists. Reversible substitution rates (relative to a G–T rate = 1) and among-site rate heterogeneity parameters (a, the gamma distribution shape parameter, and r, the proportion of constant sites) were estimated using maximum likelihood (ML) over the figure 2 ML tree or portions thereof, with the excluded taxa pruned. The observed number of variable sites (Var) includes only sites with nucleotide polymorphisms. The observed proportions of variable sites (%Var) are rough estimates because they were calculated over the total alignment length (170 sites for 5.8S and 1,813 sites for 18S) and ignore variability scored as missing data in poorly alignable regions.

[a] Estimated simultaneously with substitution rates.

[b] Estimated with substitution rates shown and ρ set to 0.

[c] Estimated with substitution rates shown, without γ rates.

structure within the sample (for which reason we do not include variance statistics) and are based on low numbers of sites. Nevertheless, these comparisons show that a high A bias in Chlorophyceae occurs at 22 sites (lines 5 and 6) that are invariant among the angiosperm sample and, therefore, not likely subject to nonstationarity effects within angiosperms. Sites observed to vary among the angiosperm sample may indeed have a higher A content (line 4), but these sites are conservedly low A in chlorophycean algae. Base compositions in the *Mimulus* sequences fit the algae pattern. These comparisons illustrate that base compositional biases in specific angiosperms or chlorophycean algae are not likely to obscure phylogenetic affinities.

We explored the justification for differential weighting of the 5.8S position for apparent compensatory mutations (Wheeler and Honeycutt 1988) but concluded that the impact of reweighting would be negligible. The experimentally determined pre-rRNA 5.8S secondary structure for yeast (Yeh and Lee 1991; cf. Thompson and Herrin 1994) includes 54 positions paired in *cis* (fig. 1). These positions are highly conserved among eukaryotes, and the instances of variation appear mostly uncorrelated. Position 114 (C/T) in chlorophytes and fungi, however, usually complements position 139 (G/A). Positions 116 and 137 appear correlated in fungi, but no compensatory correlation is evident for chlorophytes. Otherwise, pairing in the proposed yeast model is in *trans*: the 5′ and 3′ ends pair with, respectively, the ITS1-3′ and 26S-5′ ends, and the terminal two bases of the 5.8S are paired with ITS2: 28–29. This pattern is evident in models proposed for diverse eukaryotes (Coleman, Suarez, and Goff 1994; Fritz et al. 1994; Holmberg, Melander, and Nygard 1994; Suh, Thien, and Zimmer 1992).

Phylogenetic Trees

The consensus of 5.8S topologies derived heuristically using ML (fig. 2; −ln likelihood = 1,803.4397)

Table 4
Base Composition of Angiosperms, Chlorophyceae, and Reported *Mimulus* 5.8S rDNA Sequences

		Angiosperms				Chlorophyceae				"*Mimulus*"			
Class of Sites	n	A	C	G	T	A	C	G	T	*A*	*C*	*G*	*T*
1. Variable in either angio[a] or chloro[b]....	56	0.14	0.36	0.27	0.23	0.26	0.27	0.18	0.29	0.23	0.29	0.19	0.30
2. Variable in angio..................	32	0.18	0.40	0.10	0.31	0.11	0.32	0.21	0.36	0.12	0.30	0.22	0.37
3. Variable in chloro..................	20	0.08	0.46	0.24	0.23	0.20	0.33	0.13	0.34	0.13	0.38	0.18	0.33
4. Variable in angio, constant in chloro ..	24	0.23	0.36	0.10	0.31	0.14	0.27	0.27	0.32	0.16	0.27	0.25	0.32
5. Constant in angio, variable in chloro ..	12	0.08	0.42	0.33	0.17	0.32	0.25	0.18	0.25	0.20	0.38	0.21	0.21
6. Constant in angio and chloro, but varying between them..................	10	0.00	0.10	0.80	0.10	0.70	0.20	0.10	0.00	0.65	0.20	0.10	0.05
7. Variable in both angio and chloro.....	8	0.06	0.53	0.10	0.32	0.02	0.45	0.06	0.46	0.00	0.50	0.13	0.38

NOTE.—The data represent only the sequences included in table 1. The sites examined are derived from the figure 1 alignment, as classified here. Thus, the values for Chlorophyceae for Class 2 (variable in angiosperms) represent the base frequencies of sites in Chlorophyceae that align with variable sites in angiosperms. n = Number of sites in each class.

[a] Angiosperms.

[b] Chlorophyceae, including Trebouxiophyceae and Micromonadophyceae.

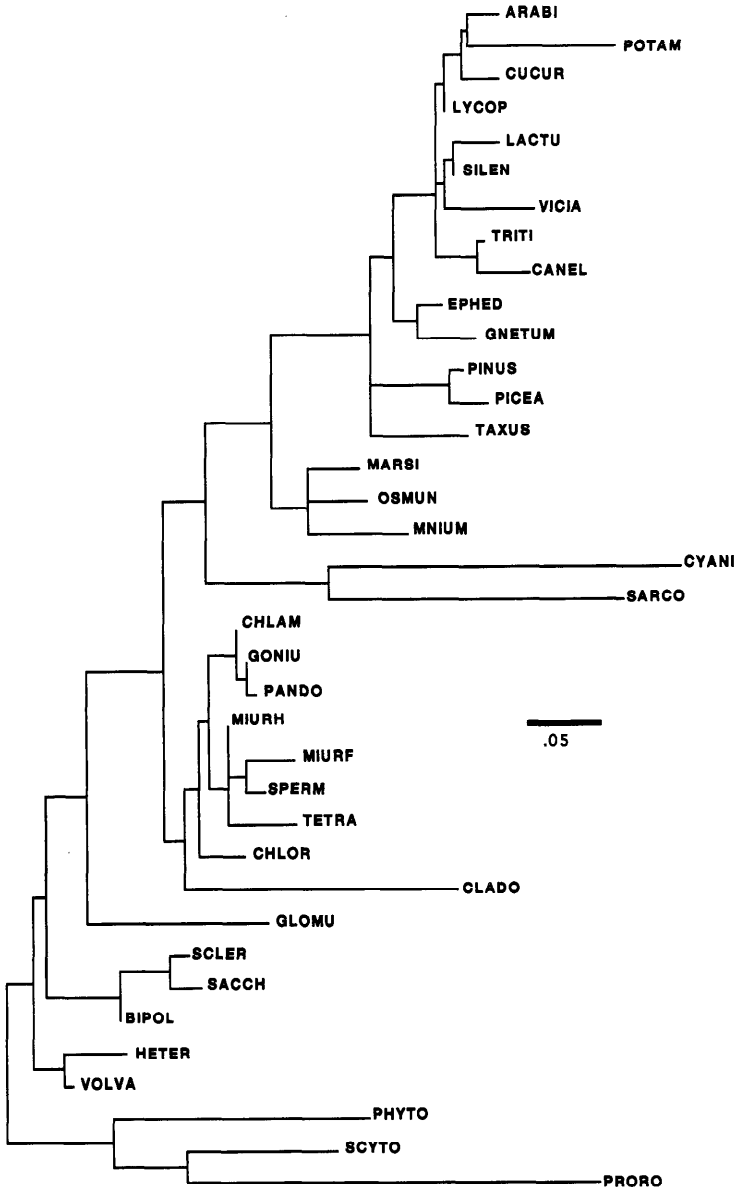

FIG. 2.—Maximum-likelihood (ML) tree for the 5.8S data set (appendix 1) modified from the 5.8S alignment (fig. 1). Taxon acronyms are listed in table 1. The tree (−ln likelihood = 1,803.4408) was derived using the heuristic ML procedure (stepwise taxon addition, tree bisection–reconnection branch swapping) in PAUP* with estimated substitution and rate-heterogeneity parameters on line 1 of table 3. Missing data (including gap scores) are ignored in calculating branch lengths. Branch lengths are equal to the expected number of substitutions per site (see scale bar).

conflict with morphological and/or other molecular evidence in the arrangement of angiosperms (Chase et al. 1993), monophyly of seedless embryophytes (MARSI, OSMUN, MNIUM; Garbary, Renzaglia, and Duckett 1992; Kranz et al. 1995), inclusion of red algae (CYANI, SARCO) among chlorophytes (Delwiche, Kuhsel, and Palmer 1995), paraphyly of fungi (Bruns et al. 1992), and relationships among fungi (Gargas et al. 1995). The inclusion of red algae among chlorophytes is perhaps the most striking anomaly, but character state reconstruction using parsimony under ACCTRAN option in PAUP* indicates that of the 11 apomorphies potentially supporting such a relationship, 10 subsequently reverse.

Figure 3 shows one of 1,170 377-step trees derived using MP. Like the ML tree, the MP trees conflict with broader evidence in the arrangement of angiosperms, monophyly of seedless embryophytes, and relationships among fungi. In addition, the chlorophycean alga *Chlorella* (CHLOR) erroneously grouped with the ulvophycean *Cladophora* (CLADO; Mishler et al. 1994; Friedl 1995, 1994). Chlorophytes are monophyletic in the MP trees, but the bootstrap support is low. The best ML score among MP trees is 1,806.4889, while the parsimony score for the ML tree (fig. 2) is 383.

Using the rate parameters from the ML tree above, the ML score for a 5.8S tree constrained for independently supported groupings shown in fig. 4 is

1,820.0009. With parameters reestimated over this constrained topology, the score improves only slightly to 1,819.7321, while the unconstrained topology becomes 1,803.7472. The best parsimony tree consistent with the figure 4 constraints has length 387.

The 18S parsimony tree (fig. 5) also contains several groupings contrary to broader evidence: the angiosperm arrangement, monophyly of gymnosperms (EPHED, PINUS, ZAMIA; Doyle, Donoghue, and Zimmer 1994), monophyly of seedless embryophytes, and polyphyly of the stramenopiles (CRYPT, PHYTO, SCYTO; Leipe et al. 1994). The 18S tree has more bootstrap recoveries ≥70% (at 16/32 nodes) than the 5.8S parsimony tree (6/36 nodes). One of these high recoveries (PINUS plus EPHED; 77%) supports a likely incorrect node. We did not explore secondary structural constraints on 18S sequence evolution across this data set, although such have been proposed in other 18S analyses (Wheeler and Honeycutt 1988). The parsimony score for a tree constrained for independently supported relationships (analogous to fig. 4) was 2,380.

ITS2 Sequences

Alignment

Figure 6 shows the CLUSTALV alignment of green algal, fungal, and angiosperm ITS2 sequences. The alignment was not adjusted a posteriori; it demonstrates mechanical detection of conserved motifs in highly divergent sequences rather than optimal multiple or pairwise alignment. As with 5.8S, the ITS2/26S boundary is not always clear, but even fewer RNA-based sequences exist for the latter. We presume that the 3′ end of the alignment approximates the boundary. There are no unambiguous conserved motifs shared between the algae, fungi, and plants, although a G-rich region is evident at positions 190–210, usually containing one or two YGG trinucleotides. Its conservation appears bona fide among angiosperms (Liu and Schardl 1994; Hershkovitz and Zimmer 1996), as does the alignment to conifers, but the homology is less clear in the other groups, especially the fungi. Within major groups, motif conservation is amply evident, especially among Caryophyllales, for which many of the gaps are nearly perfectly aligned, e.g., between positions 70 and 250. The most alignable regions among Caryophyllales superimpose over motifs shared with other angiosperms, e.g., position 75–97 and positions 131–141. Both regions are shared by conifers, as well, and the latter corresponds to the angiosperm-conserved motif previously identified by Liu and Schardl (1994). The conifers share broad regions of alignability, but the taxa are too closely related to infer broadly conserved conifer-specific motifs.

Because one issue of the present paper is the identity of the reported *Mimulus* sequences, the chlorophycean sequences are aligned more optimally in figure 7. Most striking is the similarity between the reported *Mimulus* sequences (MIURF, MIURH, and MIURM) and the chlorococcalean (SPONG) and volvocalean (GONIU and PANDO) taxa. Worth reiterating, however, is the apparent lack of primary sequence conservation between these green algae and the ulvophycean CLADO, as evidenced by comparison (not shown) and failure of this sequence to retrieve algal or any other ITS sequences from the databases.

The alignment of the fungi appears less cohesive than for the other major groups, although the 5′ end appears conserved. Greater similarity is evident within the fungal classes, i.e., the basidiomycete (VOLVA and HETER) and ascomycete (BIPOL and SCLER) pairs. The zygomycete (GLOMU) most resembles the ascomycetes. Conservation within the fungal groups would likely be better characterized by incorporating more of the available sequences, but consideration of fungi here is primarily for perspective in relation to green algae and plants.

Guide Tree Analysis

Figure 8 shows the CLUSTALW guide tree generated using gap-opening/-extension penalties of 25/0.3. Caryophyllales and green algae, including the putative *Mimulus* sequences, each appear distinct. The *Mimulus* sequences (MIURF and MIURH) cluster with the chlorococcalean *Spongiochloris* (SPONG) and, in turn, with two volvocaleans, *Pandorina* (PANDO) and *Gonium* (GONIU). The close relationship between the MIURF and *Spermatozopsis* (SPERM) in the 5.8S sequences is not evident in ITS2, although sampling among chlorophycean algae is sparse for both sequences, and no *Spongiochloris* 5.8S sequence is available. Separation of other taxonomic groups is less pronounced, although the tree groups the basidiomycete, ascomycete, and conifer pairs. Mild perturbation of the gap penalties causes all of the "resolution" to disappear. Still, we found no guide tree with comparable internal branch resolution that showed intermixing of algal, fungal, and seed plant sequences.

Discussion

We undertook the present analysis because phylogenetic signal in the ITS region had not been characterized sufficiently to diagnose the organismal origin of a suspected contaminant sequence. Properties of the ITS region sequences posed several analytical challenges. For the 5.8S sequences, these included (aside from the more broadly applicable problems of site- and lineage-specific evolutionary rate variation, nonstationary substitution patterns, and sampling limitations) the potentially inadequate number of variable sites and the possibility that the variable sites were either too functionally constrained and/or mutationally saturated. In the case of ITS2 sequences, diagnostic signal across the phylogenetic space considered here had never been demonstrated (cf. Baldwin et al. 1995). Besides resolving our original query, the exercise yielded a broader assessment of the evolutionary properties of 5.8S and ITS2 sequences and useful observations on sequence analysis in general.

5.8S Sequences

Ribosomal DNA 5.8S sequences yield phylogenetic trees qualitatively similar to those from 18S but less well supported. The 5.8S and 18S trees conflict similarly with

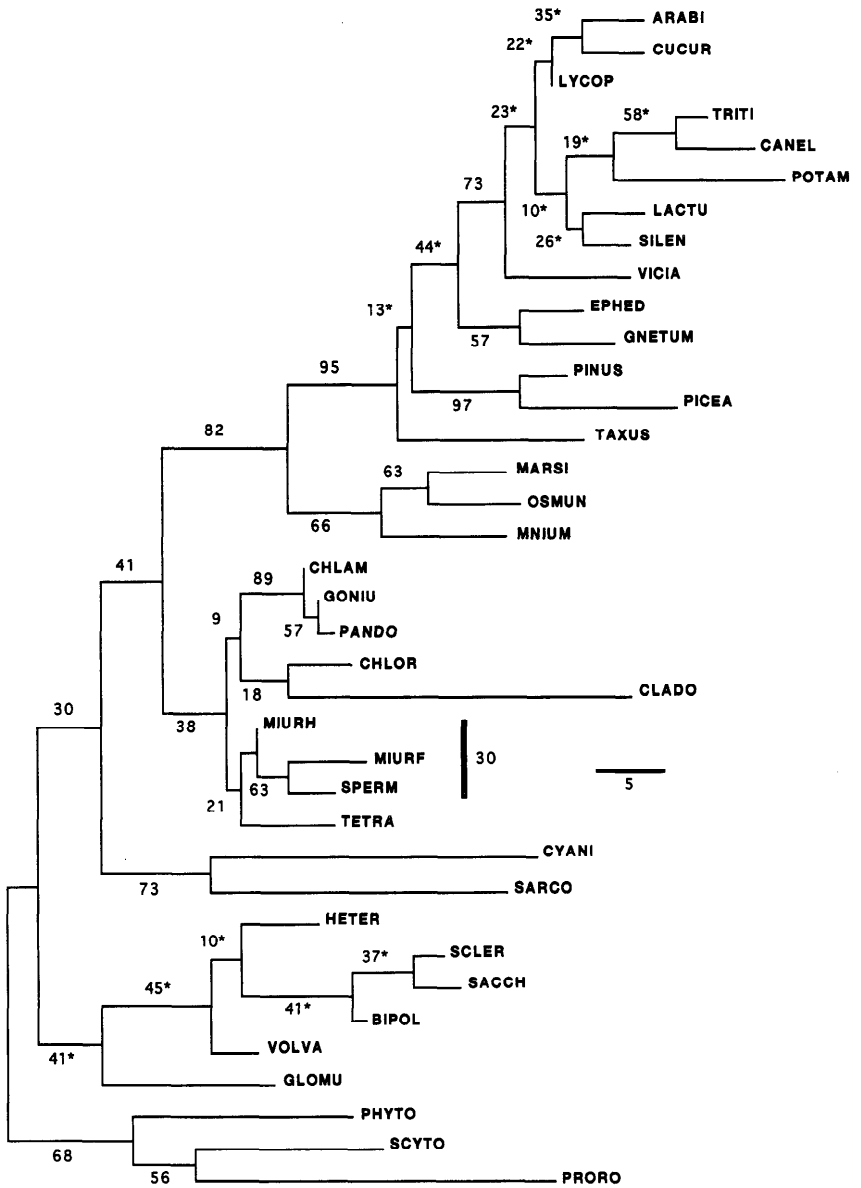

FIG. 3.—Maximum-parsimony (MP) tree for the 5.8S data set (appendix) modified from the 5.8S alignment (fig. 1). Taxon acronyms are listed in Table 1. Numbers above branches are percent bootstrap recovery. Asterisks denote branches collapsing in the strict consensus. Branch lengths are proportional to the number of changes assigned to that branch (see scale bar). The tree (length = 377, consistency index, CI, excluding uninformative characters = 0.48; retention index, RI = 0.69; rescaled consistency index, RC = 0.38, 79 parsimony-informative sites) was derived using the heuristic MP procedure (100 trees held at each step of stepwise taxon addition, tree bisection–reconnection branch swapping) in PAUP*.

independent evidence in the spurious grouping of ferns plus moss. The similarities in the trees probably reflect, in part, the genetic and functional linkage of these sequences. This linkage is perhaps best demonstrated by the high divergence of *Cladophora* in both the 5.8S and 18S data sets, evident in the figure 2, 3, and 5 trees. But linkage per se has no bearing on amount and/or quality of phylogenetic signal, as evidenced by the likewise linked internal spacers. The lower bootstrap recoveries for groups in the 5.8S tree is probably attributable mainly to the relatively low number of variable sites at a given phylogenetic level.

Sequence divergences are much greater in the 5.8S than 18S sequences, but, at any level of phylogenetic comparison, the proportion of variable sites is also much higher, and likelihood estimates of among-site rate heterogeneity are less extreme (table 3). This difference in the variation pattern offsets, to some degree, the >10-fold advantage of 18S in the number of analyzable sites. Divergences in 18S measured from angiosperms increase gradually with the inclusion of more distantly related chlorophytes (excluding the highly divergent *Cladophora*) and more with the inclusion of fungi and protists (table 2). Some angiosperm–fungi and angiosperm–protist divergences appear similar, but this discounts divergence in poorly alignable regions scored as missing. The 5.8S divergences from angiosperms to chlorophycean algae and to fungi are approximately the

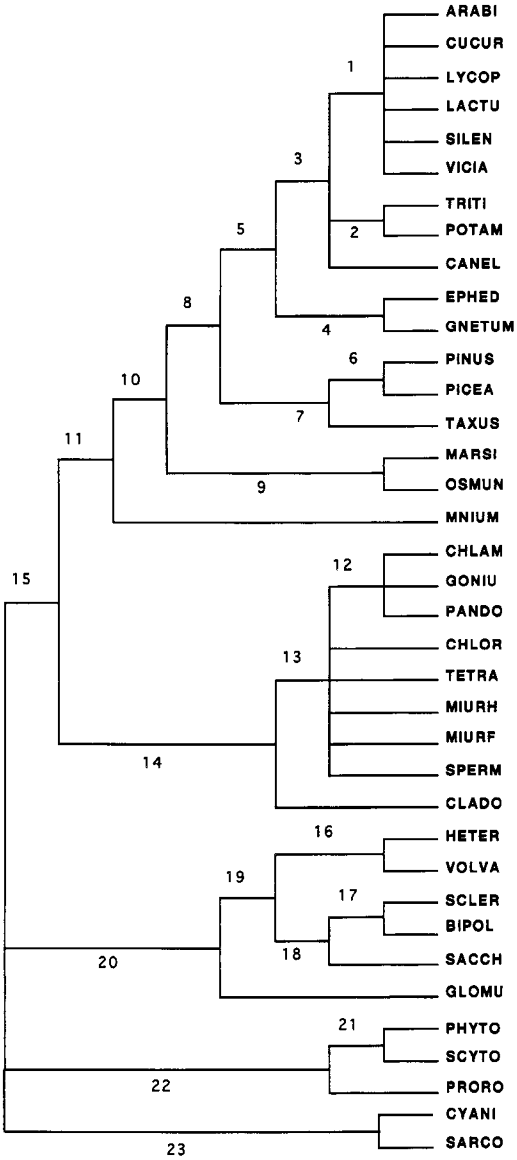

FIG. 4.—Consensus tree of relationships supported by diverse lines of evidence. Support for the numbered nodes is discussed and/or referenced in (1) eudicots, (2) monocots, and (3) angiosperms (Chase et al. 1993; Doyle, Donoghue, and Zimmer 1994) 4—8; (4) gnetophytes and (5) anthophytes: Doyle, Donoghue, and Zimmer 1994); (6) Pinaceae and (7) conifers: Page 1990; (8) seed plants: Doyle, Donoghue, and Zimmer 1994); (9) ferns, (10) vascular plants, and (11) embryophytes: Mishler et al. 1994; (12) Volvocales, (13) Chlorophyceae/Trebouxiophyceae/Micromonadophyceae and 14) Ulvophyceae plus node 13 (Mishler et al. 1994); (15) chlorophytes (Bhattacharya and Medlin 1995); (16) basidiomycetes, (17) filamentous ascomycetes, (18) ascomycetes, (19) septate fungi; and (20) fungi (Gargas et al. 1995); (21) heterokont stramenopiles and (22) alveolates plus node 21 (Leipe et al. 1994; Bhattacharya and Medlin 1995) and (23) rhodophytes (Bhattacharya and Medlin 1995).

same, suggesting mutational saturation at this divergence level, but the calculation ignores the length-variable region (fig. 1, positions 121–150, in which angiosperms are clearly more similar to green algae than to fungi. Moreover, the angiosperm–protist divergences are much higher than all other divergences from angiosperms.

Our estimates of α and ρ over taxon subsets of both the 18S and 5.8S data appear to reflect either variation in among-site rate heterogeneity throughout the phylogeny or the sensitivity of the estimate to divergence patterns among the sequences (i.e., sampling; cf. Yang 1994), or both. For the 18S data, the estimated variation possibly reflects, at least in part, our scoring as missing data the poorly alignable, hence hypervariable, regions in nonangiosperms. The ρ estimate, however, should be lowest rather than highest for the seed plant data subset, which includes a higher proportion of potentially variable sites. The α and ρ estimates for the tree with embryophytes pruned (see Results) fall in between those estimated over the entire taxon set and the set excluding protists. That the ρ estimate in this case (0.53) is still much less than those for the embryophyte and seed plant sets is counterintuitive, because the nonembryophyte sequences include all sites that align with embryophytes (i.e., the most conserved regions) minus those that do not. The 5.8S rate heterogeneity estimates vary more erratically over the taxon subsets than the 18S data, which probably reflects the small number of total sites, hence high variance. Considering variance, there is presumably a range of combinations of substitution and among-site heterogeneity parameters that could explain the data more or less equally well, but recovery and evaluation of all combinations is not readily feasible using current software. In any case, variation in estimates of rate heterogeneity are problematic, because the ML algorithm in PAUP* (and essentially all commonly applied phylogenetic algorithms) assumes that all rate parameters are constant.

Our results provide a practical illustration of problems that obfuscate phylogenetic reconstruction from DNA sequences. While the phylogenetic utility of 5.8S rDNA is obviously constrained by its short length, our comparison with 18S rDNA indicates that sequence length is not the sole determinant of tree accuracy, as demonstrated by relatively strong bootstrap support for the presumably inaccurate grouping of *Ephedra* with *Pinus* in our 18S tree. Large divergences, character nonindependence, and lineage-specific evolutionary rates, coupled with analytical methods that fail to correct for any or all of these conditions, can create an inverse correlation between data quantity and support for the true tree (Hillis, Huelsenbeck, and Swofford 1994; Lockhart et al. 1994).

ITS2 Sequences

The major factors limiting broader elucidation of ITS sequence conservation patterns within and among eukaryote groups are scattered and sparse sampling, limitations on secondary structural comparison imposed by sequence hypervariability, and an increased probability that hypervariable molecules in distantly related taxa will randomly include short stretches of sequence similarity. The present work demonstrates how a more taxonomically intensive sampling will bear upon understanding of eukaryotic ITS evolution. Previously, recognition of short conserved motifs in otherwise unalignable ITS sequences have been limited to examples

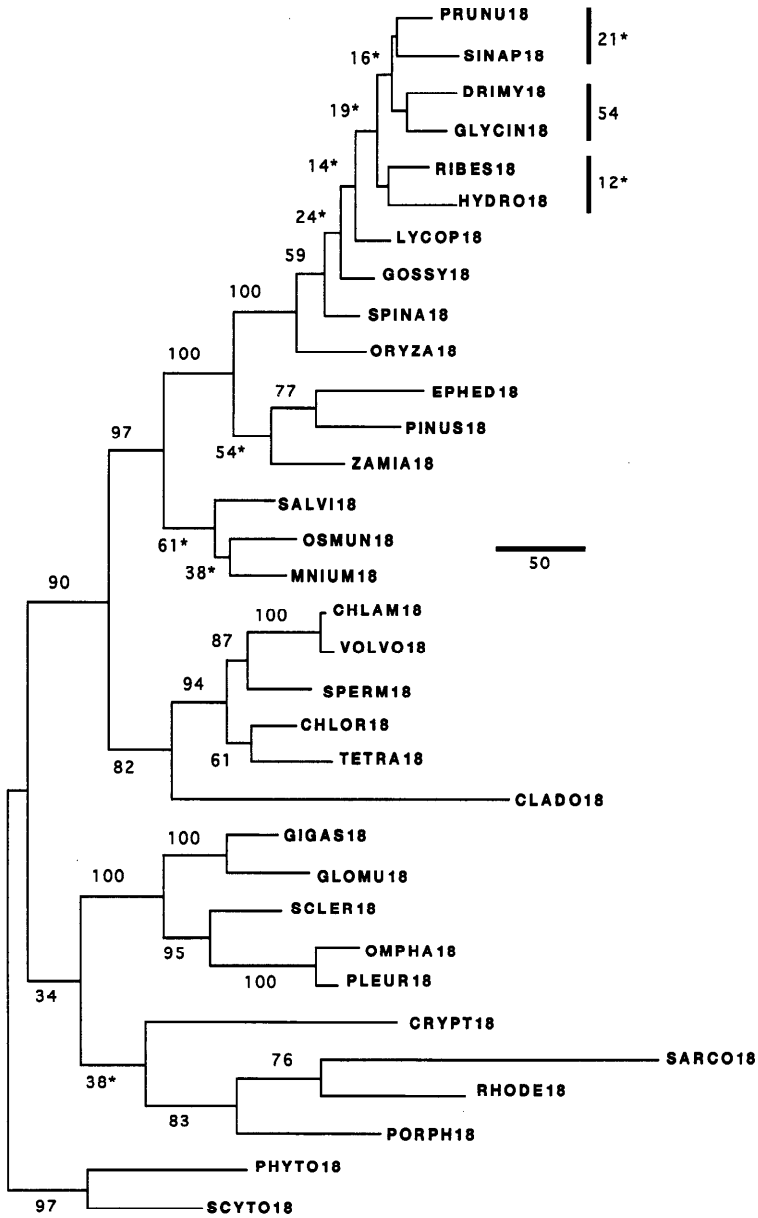

FIG. 5.—Maximum-parsimony (MP) tree for the 18S data set (available upon request). Taxon acronyms are listed in table 1. Numbers above branches are percent bootstrap recovery. Asterisks denote branches collapsing in the strict consensus. Branch lengths are proportional to the number of changes assigned to that branch (see scale bar). The tree (length = 2,315, CI excluding uninformative characters = 0.45; RI = 0.61; RC = 0.33, 535 parsimony-informative sites) was derived using the heuristic MP procedure (100 trees held at each step of stepwise taxon addition, tree bisection–reconnection branch swapping) in PAUP*.

in vertebrates (Crouch and Bachellerie 1986; cf. Schmickel et al. 1990) and angiosperms (Liu and Schardl 1994). Conserved motifs have been reported for fungi but these have been limited to ascomycetes, especially yeasts (Crouch and Bachellerie 1986; van Nues et al. 1995). Wesson, Porter, and Collins (1992) reported among insects, frog, and sea urchin, "a region [of ITS2] showing good homology" with sequence flanking ribosomal processing sites identified in rat and yeast. The specified region, however, is unconserved even among yeasts and, besides, is substantially dispensable for yeast rRNA processing (cf. van der Sande et al. 1992; van Nues et al. 1995).

Conserved motifs in ITS2 (fig. 6) are responsible for the partitions between seed plants, green algae, and fungi in the guide tree (fig. 8). The motifs shared between conifers and angiosperms evidence conservation over at least 350 million years since these shared a common ancestor (Stewart and Rothwell 1993). Likewise, the distinctiveness of Caryophyllales ITS2 sequences among angiosperms evident in the alignment and figure 8 exemplifies suprafamilial phylogenetic signal in the spacer (see also Hershkovitz and Zimmer 1996). The degree of ITS2 sequence conservation evident among green algae (Fig. 7) appears greater than or equal to that among seed plants (cf. fig. 6). The divergence time

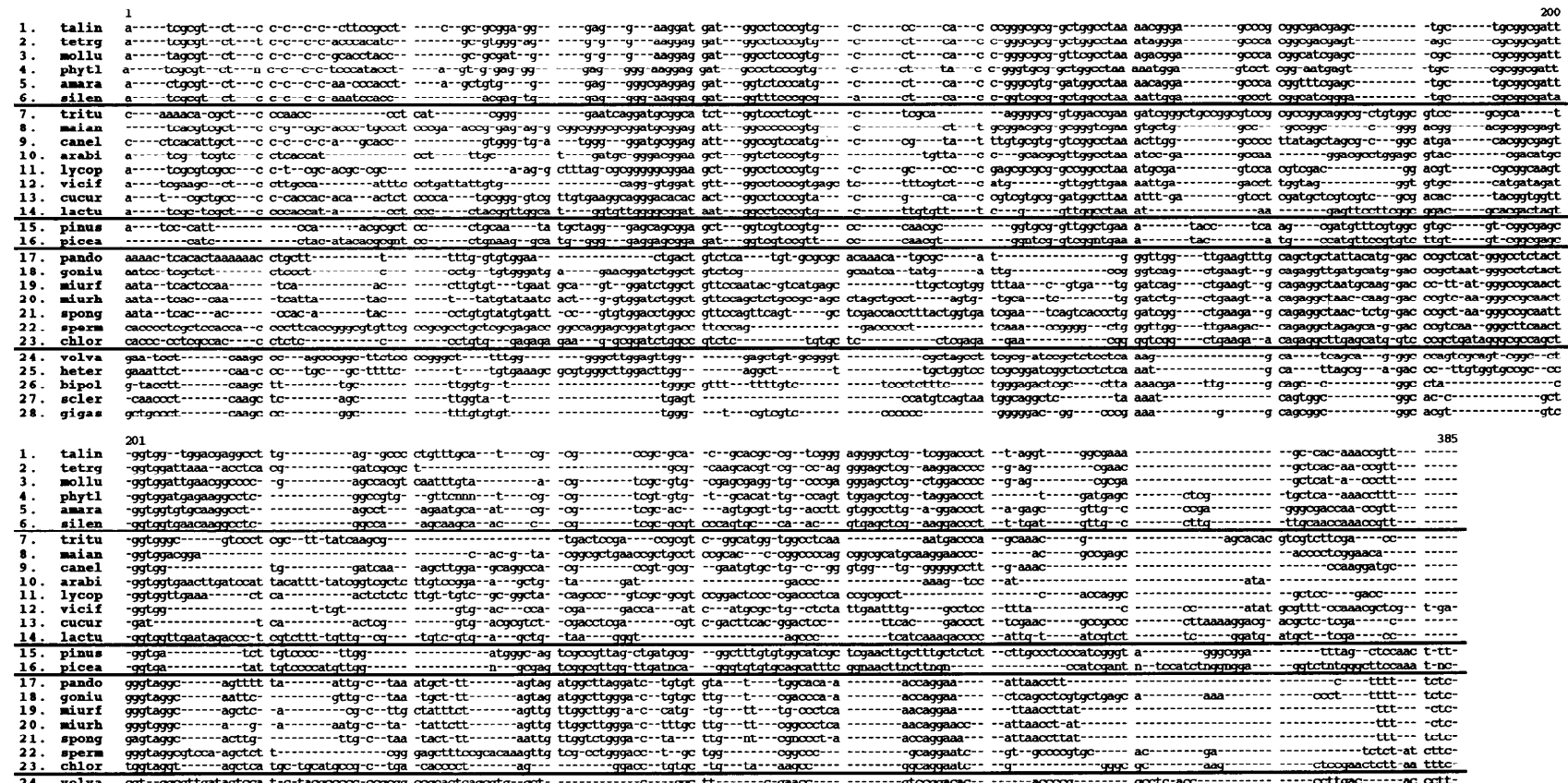

FIG. 6.—CLUSTALV alignment of ITS2 sequences of angiosperms, chlorophyceae algae, and fungi. Window size = 10, gap-opening penalty = 10, gap-extension penalty = 3. Taxon acronyms are listed in table 1. Taxa are organized by major group: caryophyllalean angiosperms: (1–6); other angiosperm (7–14); conifers (15–16); green algae (17–23); and fungi (24–28).

```
         1                                                                                                              100
SPERM    ttCACCCCTC GCTCCACCAC CCCTTCACCG GGCGTGTTCG CCG-CGCCTG CTCGCGAGAC CGGCCAGGAG CGGATGTGAC CTTCCCAGGA CCCCCTTCAA
CHLOR    ttCACCCC-C TCGCCAC--- CCTCTCCCCT GTGGAGAGAG AAGGCGGAT- CTGGCCGT-- ---CT----- ---CTGTG-- ---------- -CTCCT-CGA
PANDO    -tAAAA-CTC ACACTAAAAA ACCTGCTTTT TTGGTGTG-G AA--CTGA-- CTGTCTCAT- -------GTG CGCGCAC-A- ---------- ------AACA
GONIU    -tAATC-CTC GCTCTCT--- CCCTCCCTGT GTGGGATGAG AA--CGGAT- CTGGCTG-TC --------TC GGCAATC-A- ---------- ------TATG
SPONG    -tAATATCAC -ACCCACATA CCCTGTGTAT GTGAT-TCCG TG--TGGAC- CTGGCCGTTC CAGTTCAGTG CTCGACC-AC CTTTACTGGT GATC-GAATC
MIURM    -tAATATCAC TCTCCCTATA CCCTGTGTAT A--AC-TGAG ----TGGAC- CTGGTTGTTC C--------- ---------- -------GGT GATCCTT-TC
MIURH    -tAATATCAC CAATCATTAT ACTTATGTAT A--AT-CACT GG--TGGAT- CTGGCTGTTC CAGCTC--TG C-CG--C-AG CCTAGCTGC- ---C-TAGT-
MIURF    -tAATATCAC T-CCAATCAA CCTTGTGTTG A--ATGCA-- -G--TGGAT- CTGGCTGTTC CAATAC-GT- CATGAGC--- --TTGCTCGT GGTT-TAAC-

         101                                                                                                            200
SPERM    ACCGGGGCTG GGTTGGTTGA AGAC--CAGA GGCTAGAGCA -GGACCCG-T CAA-GGGCTT CAACTGGGTA GGCGTCCAAG CTCTTCGGGA GCTTTCCGCA
CHLOR    G--AGAACGG GGTCGGCTGA AGA--ACAGA GGCTTGAGCA TGGTCCCGCT GATAGGGCGC CAGCTTGGTA GG-----TAG CTCAT----- GCTGCATGCC
PANDO    TGCG--CATG GGTTGGTTGA AGTTTGCAGC TGCTATTACA TGGACCCGCT CAT-GGGCCT CTACTGGGTA GGC-----AG TTTTTAATT- GCTAAATGCT
GONIU    ATTG--CC-G GGTCAGCTGA AG--TGCAGA GGTTGATGCA TGGACCCGCT AAT-GGGCCT CTACTGGGTA GGC-----AA TTC---GTT- GCTAA-TGCT
SPONG    AGTCACCCTG GATCGGCTGA AGA--GCAGA GGCTAAC-TC TGGACCCGCT -AA-GGGCCG CAATTGAGTA GGC-----AC TT----GTT- GCTA-ATACT
MIURM    GGTGAACC-G GATCAGCTGA AGA--ACAGA GGTTAGCACA AGGACCCGTT -AA-GGGCCG-CAATTGGGTA GGC-----AG TTTTCT---- GCTC------
MIURH    -GTGCATCTG GATCTGCTGA AGT--ACAGA GGCTAAC-CA AGGACCCGTC -AA-GGGCCG CAACTGGGTG GGC-----AG AA----AT-- GCTATATTCT
MIURF    -GTGA---TG GATCAGCTGA AGT--GCAGA GGCTAATGCA AGGACCC-TT -AT-GGGCCG CAACTGGGTA GGC-----AG CTC---AC-- GCTTGCTATT

         201                                                                                        289
SPERM    CAAAGTTG-T CGCCTGGGAC C--TGC-TGG -CGGCCCG-- --CAGGAATC GTGCCCCGTG CAC-G----- ATCTC-T-AT -CTtcgacc
CHLOR    GCT--GACAC CCCTAGGGAC CTGTGC-TGT AAAGCCGG-- --CAGGAATC GGGGCGCAAG CTCCG----A ACTCT-TAAT -T-tcgacc
PANDO    T-TAGTAGAT GGCTTAGGAT CTGTGTGTAT -TGGCACA-A ACCAGGAAA- -TT------- ---------A ACCTTCTTTT -TCtcgacc
GONIU    T-TAGTAGAT GGCTTGGGA- CTGTGCTTGT -CGACCCA-A ACCAGGAAC- TCAGCCTCGT GCTGAGCAAA ACCCT-TTTT -TCtcgacc
SPONG    T-TAATTGTT GGTCTGGGA- CTAT---TGN TCGNCCCT-A ACCAGGAAA- ATT------- ---------A ACC-TTATTT TTCtcgacc
MIURM    C-TAATTGTT GGCTTGGGAC CATGAGC--- TCGACCCTCA ATCAGGAA-- -TT------- ---------A ACCCTTATTT -TCtcgacc
MIURH    T--AGTTGTT GGCTTGGGA- CTTTGCTTGT TCGGCCCTCA AACAGGAACC ATT------- ---------A ACC-T-ATTT -TCtcgacc
MIURF    TCTAGTTGTT GGCTTGG-A- CCATG--TGT TTG-CCCTCA AACAGGAA-- -TT------- ---------A ACC-TTATTT -TCtcgacc
```

FIG. 7.—Manually optimized alignment of chlorophycean ITS2 sequences. Taxon acronyms are listed in table 1. Lowercase letters signify putative 5′ and 3′ adjacent coding regions. Underlined bases indicate the distalmost bases of ITS2 according to Genbank documentation.

among these groups is unclear, in part because their relationships remain poorly resolved, but also because fossilized coccoid green algae cannot be confidently identified and might belong to many different lineages. Volvocales are known from the upper Jurassic (ca. 200 million years ago; Bold and Wynne 1985). The three fungal classes are distinct before the Permian, with zygomycetes and "probable" ascomycetes known as early as

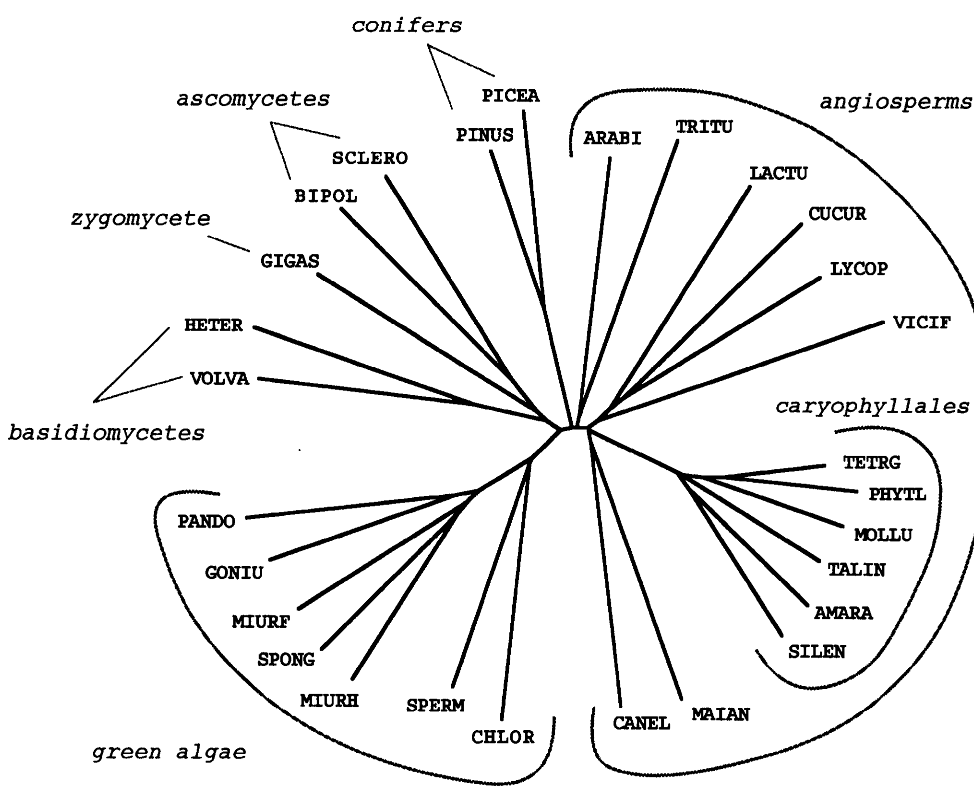

FIG. 8.—Guide tree for ITS2 sequences derived using CLUSTALW. Taxon acronyms are listed in table 1. The tree is derived using neighbor-joining on scores for independent pairwise "optimal" alignments using a gap-opening penalty of 25 and gap-extension penalty of 0.3. The tree is a phenogram, rather than phylogram or cladogram: the apparent divergences do not reflect identically aligned base positions across all taxa.

the Silurian (Stewart and Rothwell 1993). Collectively, these results reveal that ITS2 has retained detectably conserved sequence/structure through 350 or more million years of evolution.

ITS hypervariability and poor alignability across major taxonomic groups had previously yielded the opinion that "most of ITS [is] devoid of sequence-specific functions" (Crouch and Bachellerie 1986), i.e., is free to vary randomly. In vivo analyses in yeasts, however, have demonstrated that heterologous ITS2 sequences from only "closely" and not more distantly related yeast species were functional and that mutations/deletions of conserved ITS2 motifs reduced or eliminated 26S accumulation (van der Sande et al. 1992; van Nues et al. 1994, 1995). Deleting certain combinations of unconserved regions reduced function, and even mutations in unconserved regions reduce cell growth and viability. Evidence for functional constraints on ITS sequence elements in other taxa is lacking, but the yeast experiments suggest that patterns of ITS sequence variation are not random but rather reflect rapid coevolution between sequence and ribosomal processing factors (ribonucleolar proteins and RNAs; van Nues et al. 1995). If this is the case, ITS variation might be useful for guiding taxonomic sampling of the processing factors themselves, which might provide useful phylogenetic evidence, even at levels where ITS itself does not.

Conclusions

The purpose of this study was to provide an updated comparison of 5.8S and ITS2 sequences in diverse eukaryotes in order to determine whether the organismic origin of such sequences could be confidently estimated. Our analyses yield several conclusions:

1) Although our results do not challenge the notion that these sequences are not ideal for deep-level phylogenetic analysis, they do suggest that the deep-level phylogenetic signal contained in 5.8S and ITS2, however small, is generally consistent with that from other molecular and nonmolecular data. Because the ITS region contains at least some diagnostic signal spanning interkingdom- to interspecific-level divergences, the organismal origins, as well as phylogenetic relationships, of any ITS region sequence can be accurately approximated. This result is significant given the increasing popularity of ITS region sequences for lower-level phylogenetic analyses and concomitant increase in the number of taxa for which only ITS sequences are now or ever likely to become available.

2) Our results suggest that the diagnostic limitations of 5.8S sequences do not strictly reflect short sequence length but are also influenced by factors affecting sequences of any length.

3) Because the ITS region comprises, across a relatively short sequence, three structurally and functionally distinct molecules, each of which is the focus of ongoing experimental analysis in yeast and other model organisms, ITS region comparisons among both closely and distantly related organisms should provide a readily exploitable tool for deriving unique insights into DNA sequence evolution in general.

4) Our results demonstrate that both 5.8S and spacer sequence variation at relatively deep phylogenetic levels is nonrandom and thus probably reflects the evolution of underlying elements of the translational apparatus. Thus, ITS region variation should prove useful for experimental analyses of translational apparatus evolution, which in turn might yield phylogenetic data not directly derivable from the ITS region itself.

5) Finally, the importance of exercising caution at all stages of DNA-based research is underscored by the extent of apparent errors and ambiguities in DNA sequence database documentation that we uncovered in the process of analyzing suspected PCR contaminant sequences.

Acknowledgments

We are indebted to Dave Swofford for access to and permission to publish results using prerelease versions of PAUP* 4.0,. We thank Dave Swofford, Paul Lewis, and Jack Sullivan for discussions and insights into inference methods; Liz Zimmer, Matt Kane, Tod Reeder, Robb Brumfield, Bruce Baldwin, and an anonymous reviewer for critically reading the manuscript; and participants of the Laboratory of Molecular Systematics journal club for critical comments. M.A.H. was supported by a Smithsonian Postdoctoral Fellowship in Molecular Evolution. L.A.L. was supported by a Postdoctoral Fellowship from a Mellon Foundation grant awarded to the Botany Department, Duke University.

LITERATURE CITED

BAKKER, F. T., J. L. OLSEN, and W. T. STAM. 1995. Evolution of nuclear rDNA ITS sequences in the *Cladophora albida/sericea* clade (Chlorophyta). J. Mol. Evol. **40**:640–651.

BALDWIN, B. G., M. J. SANDERSON, J. M. PORTER, M. F. WOJCIECHOWSKI, C. S. CAMPBELL, and M. J. DONOGHUE. 1995. The ITS regions of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Mo. Bot. Gard. **82**:247–277.

BHATTACHARYA, D., and L. MEDLIN. 1995. The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. J. Phycol. **31**:489–498.

BOLD, H. C., and M. J. WYNNE. 1985. Introduction to the algae. Prentice-Hall, Englewood Cliffs, N.J.

BRUNS, T. D., R. VILGALYS, S. M. BARNS, D. GONZALEZ, D. S. HIBBETT, D. J. LANE, L. SIMON, S. STICKEL, T. M. SZARO, W. G. WEISBURG, and M. L. SOGIN. 1992. Evolutionary relationships within the Fungi: analysis of nuclear small subunit rRNA sequences. Mol. Phyl. Evol. **1**:231–241.

CHASE, M. W., D. E. SOLTIS, R. G. OLMSTEAD et al. (39 co-authors). 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Mo. Bot. Gard. **80**:528–580.

COLEMAN, A. W., A. SUAREZ, and L. J. GOFF. 1994. Molecular delineation of species and syngens in volvocacean green algae (Chlorophyta). J. Mol. Evol. **30**:808–90.

CRONQUIST, A. 1981. An integrated system of classification of flowering plants. Columbia Univ. Press, New York, N.Y.

CROUCH, R. J., and J.-P. BACHELLERIE. 1986. Ribosomal RNA processing sites. Pp. 47–80 *in* S. K. DUTTA, ed. DNA systematics, Vol. **1**. Evolution. CRC Press, Boca Raton, Fla.

DELWICHE, C. F., M. KUHSEL, and J. D. PALMER. 1995. Phylogenetic analysis of tufA sequences indicates a cyanobacterial origin of all plastids. Mol. Phyl. Evol. **4**:110–128.

DOYLE, J. A., M. J. DONOGHUE, and E. A. ZIMMER. 1994. Integration of morphological and ribosomal RNA data on the origin of angiosperms. Ann. Missouri Bot. Gard. **81**: 419–450.

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

FRIEDL, T. 1995. Inferring taxonomic positions and testing genus level assignments in coccoid green lichen algae: a phylogenetic analysis of 18S ribosomal RNA sequences from *Dictyochloropsis reticulata* and from members of the genus *Myrmecia* (Chlorophyta, Trebouxiophyceae cl. nov.). J. Phycol. **31**:632–639.

FRITZ, G. N., J. CONN, A. COCKBURN, and J. SEAWRIGHT. 1994. Sequence analysis of the ribosomal internal transcribed spacer 2 from populations of *Anopheles nuneztovari* (Diptera: Culidae). Mol. Biol. Evol. **11**:406–416.

GARGAS, A., P. T. DEPRIEST, M. GRUBE, and A. TEHLER. 1995. Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. Science **268**:1492–1495.

GARBARY, D. J., and P. W. GABRIELSEN. 1990. Taxonomy and evolution. Pp. 477–498 *in* K. M. COLE and R. G. SHEATH, eds. Biology of the red algae. Cambridge University Press, Cambridge.

GARBARY, D. J., K. S. RENZAGLIA, and J. G. DUCKETT. 1992. The phylogeny of land plants: a cladistic analysis based on male gametogenesis. Plant Syst. Evol. **188**:237–269.

GENETICS COMPUTER GROUP. 1994. Program Manual for the Wisconsin Package, version 8. Genetics Computer Group, Madison, Wis.

HENRY, Y., H. WOOD, J. MORRISEY, E. PETFALSKI, S. KEARSEY, and D. TOLLERVEY. 1994. The 5′ end of yeast 5.8S rRNA is generated by exonuclease from an upstream cleavage site. EMBO J. **13**:2452–2463.

HERSHKOVITZ, M. A., and E. A. ZIMMER. 1996. Conservation patterns in angiosperm rDNA-ITS2 sequences. Nucl. Acids. Res. **24**:2857–2867.

HIGGINS, D. G. A., J. BLEASBY, and R. FUCHS. 1992. Clustal V: improved software for multiple sequence alignment. CABIOS **8**:189–191.

HILLIS, D. M., J. P. HUELSENBECK, and D. L. SWOFFORD. 1994. Hobgoblin of phylogenetics? Nature **369**:363–364.

HOLMBERG, L., Y. MELANDER, and O. NYGÅRD. 1994. Probing the conformational changes in 5.8S, 18S, and 28S rRNA upon association of derived subunits into complete 80S ribosomes. Nucleic Acids Res. **22**:2776–2783.

HUSS, V. A. R., and M. L. SOGIN. 1990. Phylogenetic position of some *Chlorella* species within the Chlorococcales based upon complete small-subunit ribosomal RNA sequences. J. Mol. Evol. **31**:432–442.

KRANZ, H. D., D. MIKS, M.-L. SIEGLER, I. CAPESIUS, and C. W. SENSEN. 1995. The origin of land plants: phylogenetic relationships among chlorophytes, bryophytes, and vascular plants inferred from complete small-subunit ribosomal RNA gene sequences. J. Mol. Evol. **41**:74–84.

LEIPE, D. D., P. O. WAINWRIGHT, J. H. GUNDERSON, D. PORTER, D. J. PATTERSON, F. VALOIS, S. HIMMERICH, and M. L. SOGIN. 1994. The stramenopiles from a molecular perspective: 16S-like sequences from *Labyrinthuloides* and *Cafeteria roenbergensis*. Phycologia **33**: 369–377.

LINDAHL, L., R. H. ARCHER, and J. M. ZENGEL. 1994. Alternate pathways for processing the internal transcribed spacer 1 in pre-rRNA of *Saccharomyces cerevisiae*. Nucleic Acids Res. **22**:5399–5407.

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

LIU, J.-S., and C. L. SCHARDL. 1994. A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. Plant Mol. Biol. **26**:775–778.

LYGEROU, Z., P. MITCHELL, E. PETFALSKI, B. SÉRAPHIN, and D. TOLLERVEY. 1994. The POP1 gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins. Genes & Dev. **8**:1423–1433.

MATTOX, K. R., and K. D. STEWART. 1984. Classification of the green algae: a concept based on comparative cytology. Pp. 29–72 *in* D. E. G. IRVINE and D. M. JOHN, eds. Systematics of the green algae. Academic Press, London.

MISHLER, B. M., L. A. LEWIS, M. A. BUCHHEIM, K. S. RENZAGLIA, D. J. GARBARY, C. F. DELWICHE, F. W. ZECHMAN, T. S. KANTZ, and R. L. CHAPMAN. 1994. Phylogenetic relationships of the "green algae" and "bryophytes". Ann. Mo. Bot. Gard. **81**:451–483.

PAGE, C. N. 1990. Interrelationships between families of conifers. Pp. 292–293 *in* K. U. KRAMER and P. S. GREEN, eds. The families and genera of vascular plants. Springer-Verlag, Berlin.

RITLAND, C., K. RITLAND, and N. A. STRAUS. 1993. Variation in the internal transcribed spacers (ITS1 and ITS2) among eight taxa of the *Mimulus guttatus* species complex. Mol. Biol. Evol. **10**:1273–1278.

RITLAND, C., and N. A. STRAUS. 1993. High evolutionary divergence of the 5.8S ribosomal DNA in *Mimulus glaucescens* (Scrophulariaceae). Plant Mol. Biol. **22**:691–696.

SCHMICKEL, R. D., J. SYLVESTER, D. STAMBOLIAN, and I. L. GONZALES. 1990. The ribosomal DNA sequence for the study of the evolution of human, chimpanzee, and gorilla. Pp. 11–32 *in* S. K. DUTTA and W. P. WINTER, eds. DNA systematics, Vol. **3**. Human and higher primates. CRC Press, Boca Raton, Fla.

SMITH, S. 1994. GDE version 2.2. University of Illinois, Urbana, Ill.

STEELE, K. P., K. E. HOLSINGER, R. K. JANSEN, and D. W. TAYLOR. 1991. Assessing the reliability of 5S rRNA sequence data for phylogenetic analysis. Mol. Biol. Evol. **8**: 240–248.

STEINKÖTTER, J., D. BHATTACHARYA, I. SEMMELROTH, C. BIBEAU, and M. MELKONIAN. 1994. Prasinophytes form independent lineages within the Chlorophyta: evidence from ribosomal RNA sequence comparisons. J. Phycol. **30**:340–345.

STEWART, W. N., and G. W. ROTHWELL. 1993. Paleobotany and the evolution of plants, 2nd edition. Cambridge University Press, Cambridge, England.

SUH, Y., L. B. THIEN, and E. A. ZIMMER. 1992. Nucleotide sequences of the internal transcribed spacers and 5.8S rRNA gene in *Canella winterana* (Magnoliales; Canellaceae). Nucleic Acids Res. **20**:6101–6102.

SWOFFORD, D. L. 1996. PAUP* version 4.0. Sinauer Assoc., Inc., Sunderland, Mass. (in press).

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic Inference. Pp. 407–514 *in* D. M. HILLIS, C. MORITZ, and B. K. MABLE, eds. Molecular systematics, 2nd edition. Sinauer Associates, Inc. Sunderland, Mass.

THOMPSON, A. J., and D. L. HERRIN. 1994. A chloroplast group I intron undergoes the first step of reverse splicing into host cytoplasmic 5.8S rRNA. J. Mol. Biol. **236**:455–468.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, po-

sition specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

TROITSKY, A. V., and V. K. BOBROVA. 1986. 23s rRNA-derived small ribosomal RNAs: their structure and evolution with references to plant phylogeny. Pp. 137–170 *in* S. K. DUTTA, ed. DNA systematics, Vol. **2**. Plants. CRC Press, Boca Raton, Fla.

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

VAN DER SANDE, C. A. F. M., M. KWA, R. W. VAN NUES, H. VAN HEERIKHUIZEN, H. A. RAUÉ, and R. J. PLANTA. 1992. Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA. J. Mol. Biol. **223**: 899–910.

VAN NUES, R. W., J. M. J. RIENTES, S. A. MORRÉ, E. MOLLEE, R. J. PLANTA, J. VENEMA, and H. A. RAUÉ. 1995. Evolutionarily conserved structural elements are critical for processing of internal transcribed spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA. J. Mol. Biol. **250**: 24–36.

VAN NUES, R. W., J. M. J. RIENTES, C. A. F. M. VAN DER SANDE, S. F. ZERP, C. SLUITER, J. VENEMA, R. J. PLANTA, and H. A. RAUÉ. 1994. Separate structural elements within internal transcribed spacer 1 of *Saccharomyces cerevisiae* precursor ribosomal RNA direct the formation of 17S and 26S rRNA. Nucleic Acids Res. **22**:912–919.

WESSON, D. M., C. H. PORTER, and F. H. COLLINS. 1992. Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera: Culicidae). Mol. Phylogenet. Evol. **1**:253–269.

WHEELER, W. C., and R. L. HONEYCUTT. 1988. Paired sequence differences in ribosomal RNAs: evolution and phylogenetic implications. Mol. Biol. Evol. **5**:90–96.

WILCOX, L. W., L. A. LEWIS, P. A. FUERST, and G. L. FLOYD. 1992. Assessing the relationships of autosporic and zoosporic chlorococcalean green algae with 18S rDNA sequence data. J. Phycol. **28**:381–386.

YANG, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. **39**:306–314.

YEH, L.-C. C., and J. C. LEE. 1991. Higher-order structure of the 5.8S rRNA sequence within the yeast precursor ribosomal RNA synthesized *in vitro*. J. Mol. Biol. **217**:649–659.

BARBARA A. SCHAAL, reviewing editor

Accepted July 24, 1996

APPENDIX

Alignment of 5.8S Sequences Used for Phylogenetic Analysis in This Study, Including Gap and Missing Data Designations[a]

		1									100
1	**ARABI**	AACGACTCTC	GGCAACGGAT	ATCTCGGCTC	TCGCATCGAT	GAAGAACGTA	GCGAAATGCG	ATACTTGGTG	TGAATTGCAG	AATCCCGTGA	ACCATCGAGT
2	**LYCOP**	AACGACTCTC	GGCAACGGAT	ATCTCGGCTC	TCGCATCGAT	GAAGAACGTA	GCGAAATGCG	ATACTTGGTG	TGAATTGCAG	AATCCCGTGA	ACCATCGAGT
3	**CUCUR**	AACGACTCTC	GGCAACGGAT	ATCTCGGCTC	TCGCATCGAT	GAAGAACGTA	GCGAAATGCG	ATACTTGGTG	TGAATTGCAG	GATCCCGCGA	ACCACCGAGT
4	**triti**	?ACGACTCTC	GGCAACGGAT	ATCTCGGCTC	TCGCATCGAT	GAAGAACGTA	GCGAAATGCG	ATACCTGGTG	TGAATTGCAG	AATCCCGCGA	ACCATCGAGT
5	**LACTU**	AACGACTCTC	GGCAACGGAT	ATTTCGGCTC	ACGCATCGAT	GAAGAACGTA	GCAAAATGCG	ATACTTGGTG	TGAATTGCAG	AATCCCGTGA	ACCATCGAGT
6	**SILEN**	AACGACTCTC	GGCAACGGAT	ATCTCGGCTC	TCGCATCGAT	GAAGAACGTA	GCGAAATGCG	ATACTT-GTG	TGAATTGCAG	AATCCCGTGA	ACCATCGAGT
7	**vicia**	AATGACTCTC	GGCAACGGAT	ATCTAGGCTC	TTGCATCGAT	GAAGAACGTA	GCGAAATGCG	ATACTTGGTG	TGAATTGCAG	AATCCCGTGA	ACCATCGAGT
8	**CANEL**	GATGACTCTC	GGCAACGGAT	ATCTCGGCTC	TCGCATCGAT	GAAGAACGTA	GCAAAATGCG	ATACTTGGTG	TGAATTGCAG	AATCCCGCGA	ACCATCGAGT
9	**potam**	CATGACTCTC	GGCAACGGAC	ATCTTGGCTC	TCGCATCGAT	GAAGAACGTA	GCGAAATGCG	ATACTTGGTG	TGAATGGCAG	AATCCCGTGA	ACCATCGAGT
10	**ephed**	TACGACTCTC	GGCAATGGAT	ATCTCGGCTC	TCGCATCGAT	GAAGAACGTA	GCGAAATGCG	ATACTTAGTG	TGAATTGCAG	AATCCCGTGA	ATCATCGAGT
11	**gnetu**	CACGACTCT-	GACAATGGAT	ATCTCGGCTC	TCGT-TCGAT	GAAGAACGTA	GCGAAATGCG	ATACTTGGTG	TGAATTGCAG	AATCCCGTGA	ATCATCGAAT
12	**PINUS**	AACGACTCTC	GTCAACGGAT	ATCTCGGCTC	TTGTTACGAT	GAAGAACGTA	-CGAAATGCG	ATACTTAGTG	TGAATTGCAG	AATCCCGTGA	ATCATCGAGT
13	**PICEA**	?ATGACTCTC	GGC-A-GGA-	ATCTCGGCTC	TTGTCAACA-	GAAGAACGTA	GCGA-ATGCG	ATACTTAGTG	TGAATTGCAG	AATCCCGTGA	ATCATCGAGT
14	**TAXUS**	TGGCACTCTC	GGCAACGGAT	ATCTCGGCTC	TCGC-ACGAT	GAAGAACGTA	GCGAAATGCG	ATACTTAGTG	TGAATTGCAG	AATCCCGTGA	ATGATCGAGT
15	**marsi**	?AGGACTCTC	AGCAACGGAT	ATCTTGGCTC	TTGCAACGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAATG	TGAATTGCAG	AATTCCGCGA	ATCATCGAAT
16	**osmun**	?ACGACTCTC	AGCAACGGAT	ATCTTGGCTC	TTGCAACGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAGTG	TGAACTGCAG	AATTCCGCGA	ATAATCGAGT
17	**mnium**	?ATAACCCTC	AGCAACGGAT	ATCTTGGCTC	TTGCAACGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAGTG	TGAATTGCAG	AATTCCGCGA	ATCATCGAGT
18	**CHLAM**	???AACTCTC	AACAACGGAT	ATCTTGGCTC	TCGGATCGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAGTG	TGAATTGCAG	AAATACGTGA	ATCATCGAAT
19	**GONIU**	GACAACTCTC	AACAACGGAT	ATCTTGGCTC	TCAGATCGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAGTG	TGAATTGCAG	AAATACGTGA	ATCATCGAAT
20	**PANDO**	GACAACTCTC	AACAACGGAT	ATCTTGGCTC	TCAGATCGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAGTG	TGAATTGCAG	AAATACGTGA	ATCATCGAAT
21	**MIURH**	GACAACTCTC	AACAACGGAT	ATCTTGGCTC	TCGCAACGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAGTG	TGAATTGCAG	AATTCCGTGA	ACCATCGAAT
22	**MIURF**	GACAACTGTG	TACAACGGAT	ATCTTGGCTC	TCGTAACGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAGTG	TGAATTGCAG	AATTCCGTGA	ACCATCAAAT
23	**SPERM**	GACAACTCTC	AACAACGGAT	ATCTTGGCTC	TCGTAACGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAGTG	TGAATTGCAG	AATTCCGTGA	ACCATCGAAT
24	**TETRA**	GACAACTCTC	AACAACGGAT	ATCTTGGCTC	TTACAACGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAGTG	TGAATTGCAG	AATTCCGTGA	ACCATCGAAT
25	**CHLOR**	GACAACTCTC	AACAACGGAT	ATCTTGGCTC	CCGCATCGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAGTG	TGAATTGCAG	AATTCCGTGA	ATCATCGAAT
26	**CLADO**	AGTAACACTG	TACAATGGAT	TTCTTGGCTC	CCACATCGAT	GAAGAACGCA	GCAAAGCGCG	ATAGGTAGTG	TGAATTGCAG	AATTCCGTGA	-TCATCGAAT
27	**HETER**	TACAACTTTC	AACAATGGAT	CTCTCGGCTC	TCGCATCGAT	GAAGAACGCA	GCGAAATGCG	ATAAGTAATG	TGAATTGCAG	AATTCAGTGA	ATCATCGAAT
28	**VOLVA**	TACAACTTTC	AACAACGGAT	CTCTTGGCTC	TCGCATCGAT	GAAGAACGCA	GCGAAATGCG	ATAAGTAATG	TGAATTGCAG	AATTCAGTGA	ATCATCGAAT
29	**SCLER**	TAAAACTTTC	AACAACGGAT	CTCTTGGTTC	TGGCATCGAT	GAAGAACGCA	GCGAAATGCG	ATAAGTAATG	TGAATTGCAG	AATTCAGTGA	ATCATCGAAT
30	**sacch**	??AAACTTTC	AACAACGGAT	CTCTTGGTTC	TCGCATCGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAATG	TGAATTGCAG	AATTCCGTGA	ATCATCGAAT
31	**BIPOL**	TACAACTTTC	AACAACGGAT	CTCTTGGTTC	TGGCATCGAT	GAAGAACGCA	GCGAAATGCG	ATACGTAGTG	TGAATTGCAG	AATTCAGTGA	ATCATCGAAT
32	**GLOMU**	GATCACTTTC	AACAACGGAT	CTCTTGGCTC	TCGCATCGAT	GAAGAACGCA	GCGAAATGCG	ATAAGTAGTG	TGAATTGCAT	AATTTTGTGA	ATCATCGAAT
33	**CYANI**	AGAAGCTCTC	AACGGTGGAT	ATCTCGGCTC	TCTTGACGAT	GAAGAACGCA	-CAAAAAGCG	AAATGTAGTG	TGAACTGCAA	GTTTTCGTGA	ATCATTGAAT
34	**SARCO**	TACAACTCAT	GACGGTGGAT	GTCTTGGCTC	CAGCAACGAT	GACGAACGCA	GCGAATTGCG	AAA-GTAGTA	CGAACTGCAG	AA-CCCGTGA	ATCATCGAAT
35	**PHYTO**	AGCAACTTTC	AGCAGTGGAT	GTCTAGGCTC	GCACATCGAT	GAAGAACGCT	GCGAACTGCG	ATACGTAATG	CGAATTGCAG	GATTCAGTGA	GTCATCGAAA
36	**SCYTO**	TAAAACTTTC	AGCGACGGAT	GTCTTGGCTC	CCACAACGAT	GAAGAACGCA	GCGAAATGCG	ATACGTCTTG	CGACTTGCAG	AATCCAGTGA	ATCATCAAAA
37	**crypt**	?ACAACTTTC	AGCAGTTGAT	TCCTTGGTTC	AGACCTCGAT	GAAGGGCACT	GCGAAA-GTG	A-ATGGCATG	TGAA-TGCAG	GCATCCGGGA	ATTGAGAGCT
38	**PRORO**	TACAACTTTC	AGCGACGGAT	GTCTCGGCTC	GAACAACGAT	GAAGGGCGCA	GCGAAATGTG	ATAAGCAATG	TGAATTGCAG	AATTCCGTGA	ACCAATAGGG

		101						170	[122-145		]
1	**ARABI**	CTTTGAACGC	AAGTTGCGCC	C???CAAGCC	TTCT?GGCCG	????AGGGCA	CGTCTGCCTG	GGTGTCACAA	?????????	??????????	??
2	**LYCOP**	CTTTGAACGC	AAGTTGCGCC	C???GAAGCC	ATTT?GGCCG	????AGGGCA	CGTCTGCCTG	GGCGTCACG?	?????????	??????????	??
3	**CUCUR**	CTTTGAACGC	AAGTTGCGCC	C???GGAGCC	TTCT?GGCCG	????AGGGCA	CGTCTGCCTG	GGCGTCACGC	?????????	??????????	??
4	**triti**	CTTTGAACGC	AAGTTGCGCC	C???GAGGCC	ACTC?GGCCG	????AGGGCA	CGCCTGCCTG	GGCGTCACGC	?????????	??????????	??
5	**LACTU**	TTTTGAACGC	AAGTTGCGCC	C???GAAGCC	ATCC?GGCTG	????AGGGCA	CGCCTGCCTG	GGCGTCACGC	?????????	??????????	??
6	**SILEN**	TTTTGAACGC	AAGTTGCGCC	C???GAAGC?	-TTC?GGCTG	????AGGGCA	CGTCTGCCTG	GGCGTCACGC	?????????	??????????	??
7	**vicia**	CTTTGAACGC	AAGTTGCGCC	C???GATGCC	ATTA?GGTTG	????AGGGCA	CGTCTGCCTG	GGTGTCACAT	?????????	??????????	??
8	**CANEL**	CTTTGAACGC	AAGTTGCGCC	C???GAGGCC	ACTA?GGCTG	????AGGGCT	CGCCTGCCT-	GGCGTCACGC	?????????	??????????	??
9	**potam**	TTTTGAACGC	AAGTTGCGCC	C???TAAGCT	TCCG?GGCCG	????AGGGCA	AGCCTGCGTG	GGCGTCAAGC	?????????	??????????	??
10	**ephed**	CTTTGAACGC	AAGTTG-GCC	C???GAAGC?	-?TC?CGCCA	????AGGGCA	CGTCTGCCTG	GGCGTCGCAA	?????????	??????????	??
11	**gnetu**	CTTTGAACGC	AAGTTG-GCC	CTCCG??GC?	-?TA?GGCCG	????AGGGCA	CGTCTGCCTG	GGCGTCGCAA	?????????	??????????	??
12	**PINUS**	TTTTGAACGC	AATTTGCGCC	C???GAGGC?	-?TC?GGTCG	????AGGGCA	CGTCTGTCTG	GGCGTCGCAT	?????????	??????????	??
13	**PICEA**	TTTTG-ACGC	AATTTGCGCC	C???GAG-C?	-?TC?GGTCG	????AGGGCA	CGTCTGTCTG	GGCGTCGCAT	?????????	??????????	??
14	**TAXUS**	CTTTGAACGC	AAGTTGGCCC	G???GAG-C?	-?TC?GGCCG	????AGGGCA	CGTCTGCTTG	GGCGT-GCAC	?????????	??????????	??
15	**marsi**	CTTTGAACGC	AA-TTG-GCC	C???GAGGC?	?-TC?GTCCG	????AGGGCA	CGTCTGCCTG	AGCGTCCAC?	?????????	??????????	??
16	**osmun**	CTTTGAACGC	AAGTTG-GCC	C???GCGGC?	?-TC?GTCCA	????AGGGCA	TGCCTGCCTG	AGCGTCCTC?	?????????	??????????	??
17	**mnium**	CTTTGAACGC	AAGTTGCGCC	C???GAGGC?	?-TC?GTCCG	????AGGGCA	TTTCCGTTAG	AGCGTCACC?	?????????	??????????	??
18	**CHLAM**	CTTTGAACGC	ATATTGCGCT	C???GAGGC?	-TTC?GGCCA	????AGAGCA	TGTCTGCCTC	AGCGTCGGGT	?????????	??????????	??
19	**GONIU**	CTTTGAACGC	ATATTGCGCT	C???GAGGC?	-TTC?GGCCA	????AGAGCA	TGTCTGCCTC	AGCGTCGGGT	?????????	??????????	??
20	**PANDO**	CTTTGAACGC	AAATTGCGCT	C???GAGGC?	-TTC?GGCCA	????AGAGCA	TGTCTGCCTC	AGCGTCGGGT	?????????	??????????	??
21	**MIURH**	CTTTGAACGC	ATATTGCGCT	C???GAGGC?	-TTC?GGCCA	????AGAGCA	TGTCTGCCTC	AGCGTCGGGT	?????????	??????????	??
22	**MIURF**	CTTTGAACGC	ATATTGCGCT	C???AAGGC?	-TTC?GGCCG	????AGAGCA	TGTCTGCCTC	AGCGTCGGGT	?????????	??????????	??
23	**SPERM**	CTTTGAACGC	ATCTTGCGCT	C???GAGGC?	-TTC?GGCCG	????AGAGCA	TGTTTGCCTC	AGCGTCGG-T	?????????	??????????	??
24	**TETRA**	CTTTGAACGC	ATATTGCGCT	C???GAGGC?	-CTC?GGCCA	????AGAGCA	CGCCTGCCTC	AGAGTCGGTT	?????????	??????????	??
25	**CHLOR**	CTTTGAACGC	ATATTGCGCT	C???GAGAC?	-CTC?GGTCA	????AGAGCA	TGTCTGGCTC	AGCGTCGGCT	?????????	??????????	??
26	**CLADO**	TTTTGAACGC	ACATTGCGCT	C???AAGTC?	-TAC?GGACT	????TGAGCA	TGTCTGCCTC	AGCGTCGTTT	?????????	??????????	??
27	**HETER**	CTTTGAACGC	ACCTTGCGCC	C?????????	??????????	????AGG-CA	CGCCTGTTTG	AGTGTCGT??	?TTT??GGT	-ATT????CC	GA
28	**VOLVA**	CTTTGAACGC	ACCTTGCGCT	C?????????	??????????	????AGAGCA	TGCCTGTTTG	AGTGTCATC?	?TTT??GGC	CATT????CC	GA
29	**SCLER**	CTTTGAACGC	ACATTGCGCC	C?????????	??????????	????GGGGCA	TGCCTGTTCG	AGCGTCATTT	?CTT??GGT	-ATT????CC	GG
30	**sacch**	CTTTGAACGC	ACATTGCGCC	C?????????	??????????	????GGGGCA	TGCCTGTTTG	AGCGTCATTT	?CTT??GGT	-ATT????CC	AG
31	**BIPOL**	CTTTGAACGC	ACATTGCGCC	C?????????	??????????	????AGGGCA	TGCCTGTTCG	AGCGTCATTT	?TTT??GGT	-ATT????CC	AA
32	**GLOMU**	CTTTGAACGC	AAATTGCACT	C?????????	??????????	????GGAGTA	TGCCTGTTTG	AGGGTCGTTA	?CCT??GGT	-ATT????CC	GG
33	**CYANI**	CTTTGAACGC	AAGTTGCGCT	T?????????	??????????	????GGAGCA	CGTCTGTTGG	AGTGTCTCTT	?????????	??????????	??
34	**SARCO**	TTTTGAACGC	AAG-?GCGCT	C?????????	??????????	????GCAG-?	TGTCTGTTTG	AGTGTCCGCA	?????????	??????????	??
35	**PHYTO**	TTTTGAACGC	ATATTGCACT	T?????????	??????????	????GGAGTA	TGCCTGTATC	AGTGTC????	?????????	??????????	??
36	**SCYTO**	CTTTGAACGC	AT-CTGCG-T	T?????????	??????????	????GGAGCA	TGCTTGTCGG	AGTGTCTGTT	?????????	??????????	??
37	**crypt**	TCTTCAATGC	ATACTG??-C	T?????????	??????????	??????????	TGTCAGTTGG	ATAATCTCCC	?????????	??????????	??
38	**PRORO**	ACTTGAACGT	ATACTGCGCT	T?????????	??????????	????AAGGCA	TGCCTGCTTC	AGTGTCTATT	?????????	??????????	??

[a] "?" Represents sites coded as missing; "–" are those positions coded as gaps.

[b] Abbreviations used and Genbank accession numbers are given in table 1.