

Are Combined Analyses Better Than Single Gene Phylogenies? A Case Study Using SSU rDNA and *rbcL* Sequence Comparisons in the Zygnematophyceae (Streptophyta)

Andrey A. Gontcharov,¹ Birger Marin, and Michael Melkonian

Botanisches Institut, Lehrstuhl I, Universität zu Köln, Köln, Germany

Although the combination of different genes in phylogenetic analyses is a promising approach, the methodology is not well established and analyses often suffer from inadequate, noncongruent taxon sampling, long-branch attraction, or conflicting evolutionary models of the genes analyzed. Conflicts or congruence between multigene and single-gene phylogenies, as well as the assumed superiority of the multigene approach, are often difficult to assess solely because of incongruent taxon sampling. In the present study, a data set of 43 nuclear-encoded SSU rDNA and plastid-encoded *rbcL* gene sequences was generated from the same strains of conjugating green algae (Zygnematophyceae, Streptophyta). Phylogenetic analyses used the genes individually and in combination, either as concatenated sequences or with the log-likelihood summation method. Single-gene analyses, although mostly congruent, revealed some conflicting nodes and showed different patterns of statistical support. Combined analyses confidently resolved the conflicts between the single-gene analyses, enhanced phylogenetic resolution, and were better supported by morphological information. Long-branch taxa were not the same for the two genes analyzed, and, thus, their effect on phylogenetic resolution was minimized in the combined analyses.

Introduction

Recent years have seen combined analyses using two or more genes or even complete genomes become increasingly popular in molecular phylogenetic studies. Multigene phylogenies have been used to address the evolution of embryophyte land plants (Nickrent et al. 2000; Shaw and Allen 2000; Karol et al. 2001; Bowe, Coat, and dePamphilis 2002), animals (Mallat and Winchell 2002), various groups of algae (Nozaki et al. 2000; Hoef-Emden, Marin, and Melkonian 2002), and the radiation of eukaryotes (Baldauf et al. 2000; Baptiste et al. 2000). In most of these groups of organisms, single-gene analyses did not provide sufficient resolution or sometimes gave conflicting results, which is often ascribed to the limited number of alignable nucleotides or to differing rates of sequence evolution (Capesius and Bopp 1997; Nei, Kumar, and Takahashi 1998; Poe and Swofford 1999; Nickrent et al. 2000; Philippe 2000; Hoef-Emden, Marin, and Melkonian 2002). Thus, combined approaches are driven by the assumption that a larger number of characters improves phylogenetic accuracy and resolution (Hillis 1996). However, it is known that a strong bias in evolutionary rates (leading to long-branch attraction [LBA]) may persist and even increase when more and more characters are added (Sanderson and Shaffer 2002).

In practice, some multigene analyses are still leading to conflicting results, are sensitive to LBA, and do not significantly resolve all internal branches (e.g., Karol et al. 2001; Murphy et al. 2001; Baptiste et al. 2002). Insufficient taxon sampling because of limiting sequencing and computation capacities can be a major problem in multi-

gene approaches, whereas single-gene phylogenies may recover the correct topology because of a better taxon representation (Graybeal 1998; Baptiste et al. 2002). Moreover, multigene analyses often deal with data sets originating from incongruent taxon sampling (i.e., different genes representing the same taxon did not originate from the same clonal source) and are still affected by unresolved methodological problems, especially model misspecifications (Cao et al. 1998; Baptiste et al. 2002; Hoef-Emden, Marin, and Melkonian 2002; Pupko et al. 2002). Because the characteristics of sequence evolution are rarely identical in different genes, an “average” model for a multigene data set may sufficiently deviate from single-gene models to favor spurious relationships in the analysis. Ideally, the combined analysis should allow for different sets of model parameters to be used for the different genes (Yang 1996; Baptiste et al. 2002; Pupko et al. 2002).

These general questions of relationship between single-gene and multigene analyses have been rarely studied using real sequence data. An ideal two-gene data set to resolve such questions should have the following characteristics: congruent taxon sampling, comparable phylogenetic resolution in single-gene trees, and presence of some conflict between single-gene trees that could be tested with the multigene approach and compared with independent evidence derived from morphological information. For the present study, we have generated such a data set comprising nuclear-encoded SSU rDNA and plastid-encoded *rbcL* sequences of 43 taxa (clonal strains) of zygnematophycean green algae.

Our model taxon, the class Zygnematophyceae, is characterized by a unique mode of sexual reproduction (conjugation) and occupies a still unresolved position within the streptophyte green algae (Chapman et al. 1998; Karol et al. 2001; Turmel et al. 2002a). Absence of flagellate reproductive stages and any trace of centriolar centrosomes in the Zygnematophyceae are presumably unique among the streptophyte green algae. Previous single-gene phylogenies using SSU rDNA (Bhattacharya et al. 1994; Surek et al. 1994; Chapman et al. 1998;

¹ Present address: Institute of Biology and Soil Science, Vladivostok, Russia.

Key words: Combined analyses, phylogeny, *rbcL*, SSU rDNA, Streptophyta, Zygnematophyceae.

E-mail: gontcharov@ibss.dvo.ru.

Mol. Biol. Evol. 21(3):612–624. 2004

DOI: 10.1093/molbev/msh052

Advance Access publication January 22, 2004

Molecular Biology and Evolution vol. 21 no. 3

© Society for Molecular Biology and Evolution 2004; all rights reserved.

Besendahl and Bhattacharya 1999; Gontcharov, Marin, and Melkonian 2003) or *rbcL* (McCourt et al. 1995, 2000; Park et al. 1996) revealed some conflicting results at lower taxonomic levels (order, family, genus) but congruently resolved the Zygnematophyceae as a monophyletic lineage. All studies suggest an evolutionary trend from taxa with smooth nonornamented cell walls consisting of one piece (defining the order Zygnematales) toward taxa (order Desmidiatales) characterized by ornamented cell walls composed of more than one segment with pores, thus, implying that the more ancestral order Zygnematales is not monophyletic. One case of conflict concerns the zygnematalean genus *Roya*, which, according to the *rbcL* phylogeny, is embedded within the Desmidiatales (McCourt et al. 2000), whereas in the SSU rDNA phylogeny, *Roya* is sister to the whole Desmidiatales clade (Gontcharov, Marin and Melkonian 2003). However, direct comparison of these conflicting scenarios is impeded by the noncongruent usage of taxa and strains in published analyses.

In this study, we present evidence that combined analyses can be superior to single-gene analyses with respect to the resolution of internal branches as well as the position of taxa forming long branches in single-gene analyses. In addition, we critically compare statistical confidence measures obtained by Bayesian phylogenetics with those derived from traditional methods using the nonparametric bootstrap.

Materials and Methods

Cultures

The 43 strains of conjugating green algae used for this study were obtained from different sources (table 1) and grown in modified WARIS-H culture medium (McFadden and Melkonian 1986) at 20°C with a photon fluency rate of 40 $\mu\text{mol m}^2/\text{s}$ in a 14/10 h light/dark cycle.

DNA Extraction, Amplification, and Sequencing

After mild ultrasonication to remove mucilage (Surek and Sengbusch 1981), total genomic DNA was extracted using the QIAGEN DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). SSU rDNA and *rbcL* were amplified by polymerase chain reactions (PCR) using published protocols and 5'-biotinylated PCR primers (Marin, Klingberg, and Melkonian 1998). PCR and sequencing primers for SSU rDNA were described elsewhere (Marin, Klingberg, and Melkonian 1998; Gontcharov, Marin, and Melkonian 2003) (for newly designed *rbcL* primers, see table 2). PCR products were purified with the Dynabeads M-280 system (DynaL Biotech, Oslo, Norway) and used for bidirectional sequencing reactions (for protocols, see Hoef-Emden, Marin and Melkonian [2002]). Gels were run on a Li-Cor IR² DNA sequencer.

Sequence Alignments and Tree Reconstructions

Sequences were manually aligned using the Olson Multiple Sequence Alignment Editing Program (Olsen 1990). For coding regions of the SSU rDNA of the Zygnematophyceae, the alignment was guided by primary

and secondary structure conservation (Wuyts et al. 2000, 2001 [<http://oberon.rug.ac.be:8080/rRNA/>]). The alignments are available from the authors upon request. Phylogenetic trees were inferred with maximum-likelihood (ML), neighbor-joining (NJ), and maximum-parsimony (MP) criteria using PAUP version 4.0b10 (Swofford 1998) and Bayesian inference (BI) using MrBayes version 3.0b3 (Huelsenbeck and Ronquist 2001). SSU rDNA (1,722 unambiguously aligned nt) and *rbcL* data sets (1,353 nt) were analyzed separately and in combination (3,075 nt). Evolutionary models (for ML and NJ analyses) for the different data sets were selected via Modeltest version 3.04 (Posada and Crandall 1998). Distances used for NJ analyses were calculated by ML. ML and MP analyses used heuristic searches with a branch-swapping algorithm (tree bisection-reconnection). In BI, the Markov chains were run for one million generations, sampling every 100 generations for a total of 10,000 samples. The first 500 (*rbcL* set) or 1,000 (SSU rDNA and combined sets) samples were discarded as "burn-in." The remaining samples were combined into a single file and analyzed using the sumt command in MrBayes. The robustness of the trees was estimated by bootstrap percentages (BP [Felsenstein 1985]) using 1,000 (NJ and MP) or 100 (ML) replications and by posterior probabilities (PP) in BI. Nonsignificant BP less than 50% and PP less than 0.90 were not included in figures. In MP, the stepwise addition option (10 heuristic searches with random taxon input order) was used for each bootstrap replicate. ML-bootstrap used a single heuristic search (starting tree via stepwise addition) per replicate.

Combined Analyses

For concatenated analyses, SSU rDNA and *rbcL* sequence data were fused as a "supergene" in one alignment and analyzed using a single "concatenated model" with averaged parameters. In addition, we performed a combined analysis via "log-likelihood summation" (LS [Yang 1996]) following the method described by Baptiste et al. (2002). The 1,000 "best" ML topologies of the SSU rDNA, *rbcL* and concatenated data sets (\sum 3,000 trees) were combined in a treefile and log-likelihood values were calculated separately for the SSU rDNA data set (with the SSU rDNA model) and the *rbcL* alignment (with the *rbcL* model). For each topology, the sum of both values was calculated and provided the tree optimality criterion. Similarly, the "LS" method was applied to the *rbcL* gene alone, and the data set was subdivided corresponding to the three codons. For each codon, as well as for the complete gene, the 1,000 best ML trees were determined (\sum 4,000 trees) with the appropriate model (table 3), and for each tree topology, the log-likelihood values were calculated separately for each codon (using the appropriate model). The sum of all three values again provided the tree optimality criterion.

Topology Tests

User-defined trees were generated by manually modifying the treefile of the "best tree" using TreeView version 1.6.2 (Page 1996). To compare user-defined

Table 1
Origin and Taxonomic Designation of Strains and EMBL/GenBank Accession Numbers of SSU and *rbcL* rDNA

Taxon	Strain	Accession Number	
		SSU rDNA	<i>rbcL</i>
<i>Bambusina borrieri</i> (<i>brebissonii</i>) (Ralfs) Cl.	CCAC 0045	AJ428118	AJ553935
<i>Closterium lunula</i> (Müll.) Nitzsch ex Ralfs	M 2156	AJ553916	AJ553936
<i>Cosmarium contractum</i> Kirchn.	SVCK 396	AJ428112	AJ553937
<i>Cylindrocystis brebissonii</i> Menegh. ex De Bary	ACOI 55	AJ549228	AJ553938
<i>C. brebissonii</i> var. <i>deserti</i> Flechtner et Johansen	UTEX B 2684	AJ553920	AJ553939
<i>C. crassa</i> De Bary	ACOI 788	AJ553918	AJ553940
<i>C. cf. cushleackae</i> Brook	M 2158	AJ553919	AJ553941
<i>Cylindrocystis</i> sp.	UTEX 1925	AJ553917	AJ553942
<i>Gonatozygon brebissonii</i> De Bary	SVCK 210	AJ428083	AJ553944
<i>G. kinahanii</i> (Arch.) Rabenh.	ACOI 350	AJ553921	AJ553945
<i>Gonatozygon</i> (<i>Genicularia</i>) <i>spirotaenium</i> De Bary	SVCK 329	X74753	AJ553946
<i>Groenbladia neglecta</i> (Racib.) Teil.	SVCK 478	AJ428119	AJ553943
<i>Haplotaenium</i> (<i>Pleurotaenium</i>) <i>minutum</i> (Ralfs) Bando	SVCK 302	AJ428090	AJ553947
<i>Heimansia</i> (<i>Cosmocladium</i>) <i>pusilla</i> (Hilse) Coes.	SVCK 428	AJ428125	AJ553948
<i>Mesotaenium caldariorum</i> (Lagerch.) Hansg.	ACOI 898	AJ549229	AJ553949
<i>M. cf. chlamydo sporum</i> De Bary	M 2155	AJ553923	AJ553950
[<i>Mesotaenium endlicherianum</i> Näg.]	SAG 12.97	AJ428078	AJ553951
<i>M. kramstai</i> Lemm.	UTEX 1025	AJ428079	AJ553952
<i>Mougeotia</i> sp.	SVCK 417	AJ428073	AJ553953
<i>Mougeotia</i> sp.	SVCK 240	AJ553924	AJ553954
<i>Netrium digitus</i> var. <i>latum</i> Hustedt.	SVCK 254	AJ553927	AJ553955
<i>N. interruptum</i> (Bréb.) Lütkem.	M 1021	AJ428071	AJ553956
<i>N. oblongum</i> (De Bary) Lütkem.	SVCK 255	AJ553925	AJ553957
<i>N. oblongum</i>	ASW 07201	AJ553926	AJ553958
<i>Penium cylindrus</i> (Ehr.) Bréb.	ACOI 780	AJ553930	AJ553959
<i>P. exiguum</i> W. West	M 2159	AJ553929	AJ553960
<i>P. spirostriolatum</i> Barker.	SVCK 189	AJ553928	AJ553961
<i>Phymatodocis nordstedtiana</i> Wolle	SVCK 327	AJ428122	AJ553962
<i>Roya anglica</i> G.S West	UTEX 934	AJ428081	AJ553963
<i>R. obtusa</i> (Bréb.) W. et G.S. West	SVCK 45	AJ428082	AJ553964
<i>Spinoclosterium cuspidatum</i> (Bayley) Hirano	NIES 325	AJ553931	AJ553965
<i>Spirogyra</i> (<i>Sirogonium</i>) <i>stictica</i> (J.E. Smith) Wille	UTEX 1985	AJ428076	AJ553966
<i>Spirogyra</i> sp.	M 2157	AJ549231	AJ549239
<i>Spirogyra</i> sp.	M 1810	AJ428074	AJ553967
<i>Spondylosium panduriforme</i> (Heimerl) Teil.	SAG 52.88	AJ428124	AJ553969
<i>S. pulchrum</i> (Bail.) Arch.	SVCK 331	AJ428129	AJ553970
<i>Staurastrum lunatum</i> Ralfs	SVCK 15	AJ428106	AJ553971
<i>S. tumidum</i> Bréb. ex Ralfs	SVCK 85	AJ428108	AJ553972
<i>Zygnema</i> sp.	ASW 07036	AJ428077	AJ553973
<i>Zygnema</i> sp.	M 1156	AJ553932	AJ553974
<i>Zygnemopsis minuta</i> Randhawa	ACOI 60	AJ553933	AJ553975
<i>Zygnemopsis</i> sp.	CCAP 699/1	AJ553934	AJ553976
<i>Zygonium tunetanum</i> Gauthier-Lievre	UTCC 136	AJ549230	AJ549235

NOTE.—Sequences determined for this study are indicated by accession numbers in bold face. ACOI = Coimbra Collection of Algae, University of Coimbra, Portugal (www.uc.pt/botanica/ACOI.htm); ASW = Culture Collection of Algae at the University of Vienna (www.univie.ac.at/IECB/hydrobotanik/); CCAC = Culture Collection of Algae at the University of Cologne, Germany (www.ccac.uni-koeln.de); M = Culture Collection Melkonian, Botanical Institute, University of Cologne, Germany; NIES = National Institute for Environmental Studies, Tsukuba, Japan (www.nies.go.jp/Biology/mcc/home.htm); SAG = Sammlung von Algenkulturen, University of Göttingen, Germany (www.gwdg.de/~epsag/phykologia/epsag.html); SVCK = Sammlung von Conjugaten-Kulturen, University of Hamburg, Germany (www.biologie.uni-hamburg.de/b-online/d44_1/44_1.htm); UTCC = University of Toronto Culture Collection of Algae and Cyanobacteria (www.botany.utoronto.ca/utcc/); UTEX = Culture Collection of Algae at the University of Texas at Austin, U.S.A. (www.bio.utexas.edu/research/utex/). Taxon names in (parentheses) correspond to those used in the culture collection catalogue. [*Mesotaenium endlicherianum*] presumably represents a wrong determination of strain SAG 12.97.

topologies with the “best tree,” site-wise log-likelihoods were calculated for each topology in PAUP and used as input for CONSEL (Shimodaira and Hasegawa 2001), which calculates the probability values according to the Kishino-Hasegawa test (KH [Kishino and Hasegawa 1989]), the Shimodaira-Hasegawa test (SH [Shimodaira

and Hasegawa 1999], both weighted [w] and unweighted), and the approximately unbiased test (AU) using the multi-scale bootstrap technique (Shimodaira 2002). CONSEL was also used to test incongruence between the three “best” ML trees (figs. 1–3) using the SSU rDNA, the *rbcL*, and the combined data set.

Table 2
Newly Designed Oligonucleotides Used for PCR and Sequencing (Seq) of *rbcL*

Primer	Sequence (5' to 3')
MaGo1F	ATGTCACCACAAAACNGAAAC
MaGo2F	ATGTCACCACAAAACNGAACTAAAGCWGG
MaGo3R	
(5'-biotinylated)	GTATCRATHGTWTCAAATTC
Seq-1F- <i>rbcL</i>	CACCACAAAACNGAACTAAAGCWGG
Seq-2F- <i>rbcL</i>	TRTTYACYTCYATTGTAGG
Seq-4R- <i>rbcL</i>	ATCRATHGTWTCAAATTC
Seq-5R- <i>rbcL</i>	TTWGGYTTAATDGTACAWCC

Results

Taxon Sampling

For this study, 18 new SSU rDNA and 41 new *rbcL* sequences were obtained from 43 strains of the Zygnematomyceae; the new sequences are available under GenBank accession numbers AJ553916 to AJ553976. For four nonmonophyletic genera (*Mesotaenium*, *Cylindrocystis*, *Netrium*, and *Penium*; [Gontcharov, Marin, and Melkonian 2003, and unpublished results]) three to five strains/species were included, whereas for the monophyla *Spirogyra*, *Mougeotia*, *Zygnema*, and *Gonatozygon*, two or three representative members (the most distantly related in SSU rDNA phylogenies) were selected. To cover the derived family Desmidiaceae (containing about 35 genera and approximately 2,000 species), 10 genera (one species/strain each) were analyzed, including its basal divergence in SSU rDNA phylogenies, *Phymato-*

docis nordstedtiana (Gontcharov, Marin, and Melkonian 2003).

Although the monophyly of the Zygnematomyceae is usually recovered (McCourt et al. 2000; Gontcharov, Marin, and Melkonian 2003), the position of this class within the streptophytes remains unsettled. Therefore, only unrooted phylogenies are chosen in this study.

SSU rDNA Phylogeny

The selected model for the SSU rDNA data set was TrN with gamma shape (Γ) and proportion of invariable sites (I). The value of one substitution rate category, C \leftrightarrow T, was considerably elevated (6.9) compared to those of other categories (table 3).

In the ML phylogeny, the 43 sequences were arranged in two clusters corresponding to the orders Zygnematales and Desmidiaceae (fig. 1). This split was strongly supported by all methods except MP, and the longest internal branch (20 steps) separated these clusters. In the Desmidiaceae, two robust families/clades with long individual branches, Gonatozygaceae (GON) and Closteriaceae (CL), preceded a crown assemblage (designated DESM), which comprised the Desmidiaceae and two of three species of the Peniaceae analyzed (*Penium spirostriolatum* was a weakly supported sister taxon of DESM). The significance of DESM was only moderate in ML, NJ, and MP analyses (65% to 71%), but PP in BI analysis was 1.0 (fig. 1).

Within the Zygnematales, several genera obtained high support, namely *Mougeotia*, *Zygnema*, *Roya* (two identical sequences), and the long-branched *Spirogyra* (SPI). *Zygnema* and *Zygonium* formed the moderately supported ZYG clade. Other genera were not mono-

Table 3
Evolutionary Models, Log-Likelihood Values ($-\ln L$), and Settings Identified by Modeltest for Different Data Sets Used for Figure 1 to Figure 3 and for Special Analyses

Model parameter	SSU rDNA (Fig. 1)	<i>rbcL</i> (Fig. 2)	SSU rDNA + <i>rbcL</i> (Fig. 3)	SSU rDNA – SPI-clade	<i>rbcL</i> – 3 LB ^a	<i>rbcL</i> (1+2) ^b	<i>rbcL</i> (1) ^c	<i>rbcL</i> (2) ^c	<i>rbcL</i> (3) ^c
$-\ln L$	TrN+I+ Γ 9734.959	GTR+I+ Γ 16535.103	GTR+I+ Γ 27120.692	TrN+I+ Γ 7874.399	GTR+I+ Γ 14906.352	GTR+I+ Γ 4072.608	GTR+I+ Γ 2727.526	GTR+I+ Γ 1170.204	GTR+I+ Γ 11379.570
I	0.5097	0.5558	0.5052	0.6163	0.5718	0.7387	0.6393	0.8283	0.0445
Γ	0.5656	1.4549	0.6069	0.6158	1.7909	0.5028	0.7270	0.4056	1.9752
Base frequencies									
A	0.2647	0.2869	0.2601	0.2669	0.2844	0.2450	0.2071	0.2880	0.3312
C	0.2157	0.1529	0.1753	0.2057	0.1525	0.2489	0.2284	0.2560	0.1006
G	0.2684	0.1790	0.2331	0.2659	0.1784	0.2819	0.3687	0.2052	0.0563
T	0.2512	0.3812	0.3314	0.2615	0.3847	0.2242	0.1958	0.2507	0.5119
GC	0.4841	0.3319	0.4084	0.4716	0.3309	0.5308	0.5971	0.4612	0.1569
Rate matrix ([G \leftrightarrow T=1.00])									
[A \leftrightarrow C]	1.0000	1.5321	1.6666	1.0000	1.5572	2.6826	3.7679	1.0948	0.1674
[A \leftrightarrow G]	2.1110	7.4587	4.8221	2.3229	7.7026	1.2305	1.2985	1.6980	9.3712
[A \leftrightarrow T]	1.0000	3.1598	2.8498	1.0000	3.0591	0.6810	0.4504	1.1309	0.6730
[C \leftrightarrow G]	1.0000	1.6363	1.4417	1.0000	1.6368	1.2484	0.6887	5.4494	0.4214
[C \leftrightarrow T]	6.9202	14.2786	11.5814	7.4490	13.8966	9.7266	12.4698	5.2814	5.5969
Aligned nt	1722	1353	3075	1722	1353	902	451	451	451
Constant nt	1194	785	1979	1331	805	762	349	413	23
MP-inform. ^d	392	507	899	295	494	108	83	25	399
MP-uninf. ^d	136	61	197	96	54	32	19	13	29

^a *rbcL* minus three long branches (LB), that is, *Mesotaenium endlicherianum*, *M. caldariorum*, and *Netrium oblongum* SVCK 255.^b *rbcL* with first and second codon positions only.^c *rbcL* with first (1), second (2), or third (3) codon position only.^d Informative and uninformative (but not constant) characters in MP.

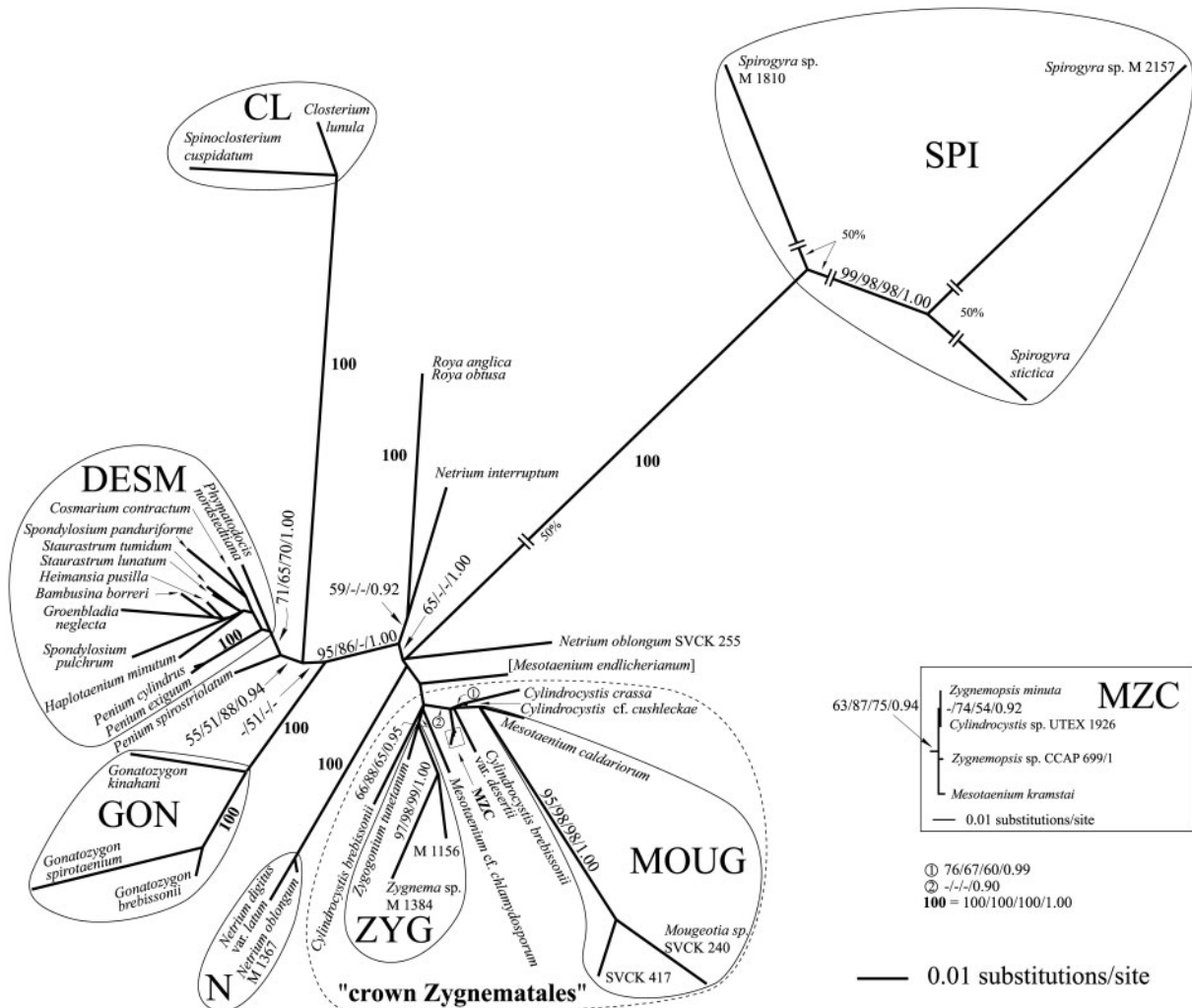


FIG. 1.—Unrooted ML (TrN+I+ Γ ; for parameters, see table 3) phylogeny of the Zygnematophyceae based on 43 SSU rDNA sequences (1,722 aligned positions); very long branches are graphically (||) reduced to 50%. Nodes are characterized by BP and PP values: ML/ NJ/ MP/ BI. [*Mesotaenium endlicherianum*] presumably represents a wrong determination. For clade abbreviations, see *Results*. The longest internal branch reflects the split between both zygnematophycean orders, Zygnematales and Desmidiaceae.

phyletic, specifically *Netrium*, *Mesotaenium*, and *Cylindrocystis*. A stable clade (MZC) combined four strains with almost identical sequences (1 to 4 nt difference only) belonging to three traditional genera: *Mesotaenium kramstai*, *Cylindrocystis* sp. strain UTEX 1926, *Zygnemopsis* sp. strain CCAP 699/1, and *Zygnemopsis minutum*. Within the Zygnematales, the branching order remained unresolved. *Roya* spp. and *Netrium interruptum* branched closer to the Desmidiaceae, followed by a polytomy comprising five branches (SPI, N-clade, *Netrium oblongum* SVCK 255, *Mesotaenium endlicherianum* and the “crown Zygnematales” [fig. 1]).

Because lineages within the Zygnematophyceae differed profoundly in evolutionary rates of the SSU rRNA gene, long-branch attraction (LBA) may have affected the phylogenetic analyses. *Spirogyra* was the longest SSU rDNA branch, and an analysis without this taxon produced an almost identical tree topology with similar significance in ML (not shown). However, deletion of *Spirogyra* had a noticeable effect on NJ and MP analyses

in which the significance of many internal branches increased (not shown).

rbcL Phylogeny

The *rbcL* data set included the same strains as used for the SSU rDNA phylogeny and contained 1,353 aligned nt; the appropriate model was GTR+ Γ +I (table 3). Whereas the proportion of invariable sites (I) estimated for the *rbcL* alignment was comparable to that of the SSU rDNA data set, Γ was significantly higher (1.46 compared with 0.57), reflecting the more even distribution of substitutions in *rbcL*. In the substitution rate matrix, two categories ([A \leftrightarrow G], [C \leftrightarrow T]) attained much higher values in *rbcL* (7.5, 14.3) compared with SSU rDNA (2.1, 6.9 [table 3]).

In contrast to SSU rDNA analyses, the two zygnematophycean orders were not resolved as monophyla because in *rbcL* trees, two members of the Zygnematales (*Roya* and *Netrium oblongum* SVCK 255) diverged within the Desmidiaceae (fig. 2) (see below).

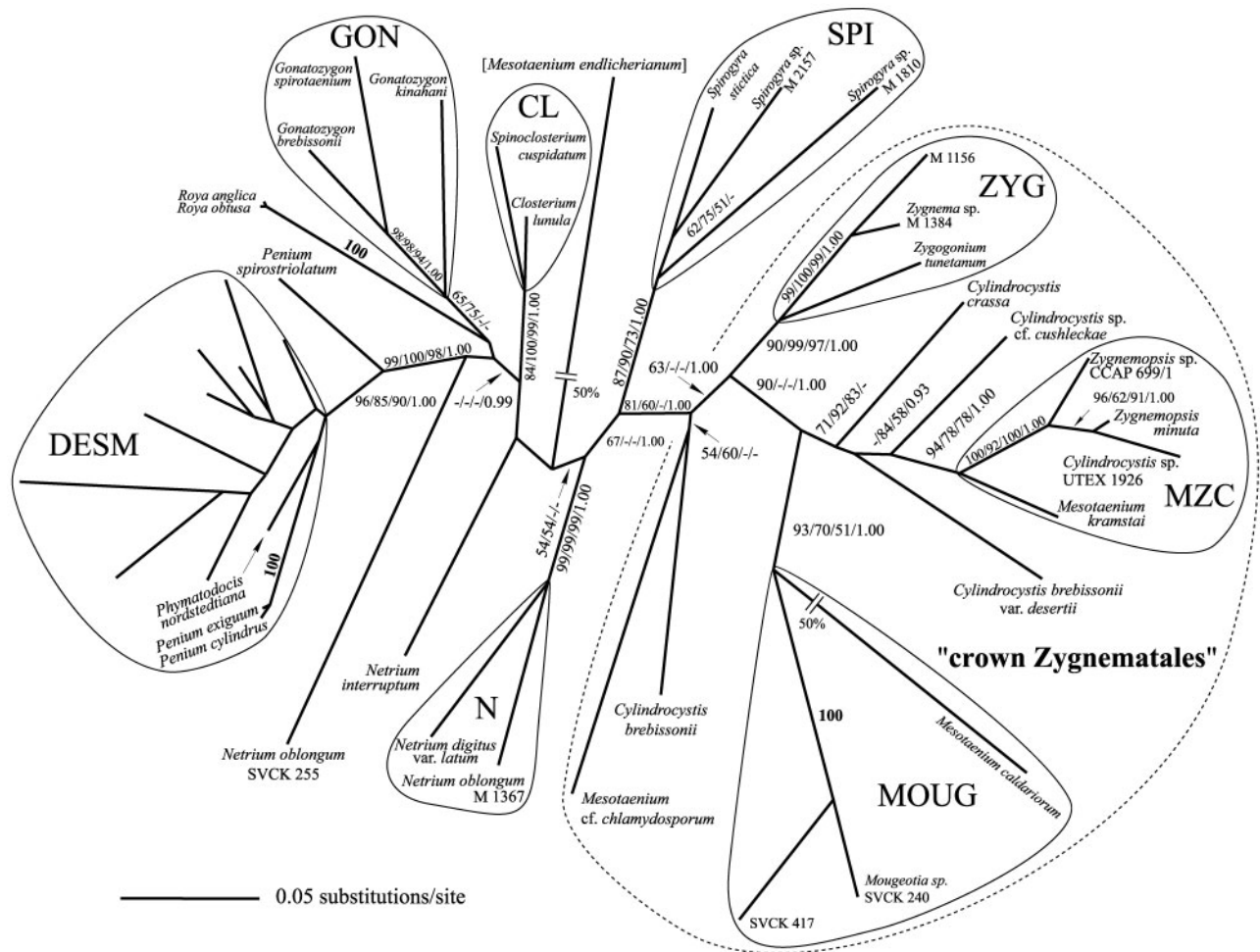


FIG. 2.—Unrooted ML (GTR+I+ Γ ; for parameters, see table 3) phylogeny of the Zygnematophyceae based on 43 *rbcL* sequences using all three codon positions (1,352 aligned nt). Most taxon designations within DESM are not shown; further details as in figure 1. Note the placement of the zygnematalean taxa *Roya* and *Netrium oblongum* SVCK 255 within the Desmidiaceae (conflicting with fig. 1), resulting in nonmonophyly of both orders.

Except for the orders Desmidiaceae and Zygnematales, *rbcL* analyses generally recovered the same clades as SSU rDNA phylogenies (figs 1 and 2). The lineages DESM, DESM/*Penium spirostriolatum*, ZYG, MCZ, and MOUG, which gained no (MOUG) or only moderate support in the SSU rDNA phylogeny, were robustly resolved by *rbcL* analyses ($\geq 90\%$ BP in ML [fig. 2]). Similarly, the “crown-Zygnematales” (only topological support in SSU rDNA sequence comparisons) were significantly supported in the *rbcL* phylogeny (in ML and BI [fig. 2]). Among those four branches, which in SSU rDNA analyses represented the closest relatives of the “crown-Zygnematales” without significance (see fig. 1), *rbcL* placed SPI as sister to it, separated from all other Zygnematophyceae with high PP (1.00) but no or low BP (branch separating N and SPI [fig. 2]). The position of *Roya* and the long-branch taxon *Netrium oblongum* SVCK 255 within the Desmidiaceae, conflicting with the SSU rDNA phylogeny, was supported by PP only (branch separating CL and *Roya*/GON [fig. 2]). The four taxa constituting the MCZ-clade showed considerably divergent *rbcL* sequences, in contrast to their almost identical SSU rRNA genes (see fig. 1). Their branching pattern, with *Mesotaenium kramstai* as basal

branch, was now significantly resolved (fig. 2). Moreover, *rbcL* revealed three *Cylindrocystis* species diverging as a paraphyletic assemblage grouped with MCZ (see fig. 2). The same *Cylindrocystis* species formed an unresolved polytomy in the SSU rDNA tree (see fig. 1).

In our *rbcL* analysis, *Roya* was only nonsignificantly associated with GON (fig. 2), in contrast to a previous *rbcL* study (McCourt et al. 2000), where *Roya anglica* (UTEX 934; accession number U38694) was sister to the *Gonatozygon*/*Genicularia* clade with 100% BP. Therefore, we resequenced the same strain (UTEX 934) and found 155 differences to the published sequence. When U38694 was included in our analysis (tree not shown), it was significantly positioned between *Gonatozygon kinahani* and the remaining *Gonatozygon* species, without affinity to *Roya*. We conclude that U38694 was derived from a *Gonatozygon* sp., presumably resulting from culture or DNA misplacement.

In *rbcL* phylogenies, *Mesotaenium endlicherianum*, *M. caldariorum*, and *Netrium oblongum* SVCK 255 were characterized by fast evolutionary rates and had long individual branches (>125 apomorphies), in contrast to the SSU rDNA analysis. Reversely, the SPI clade

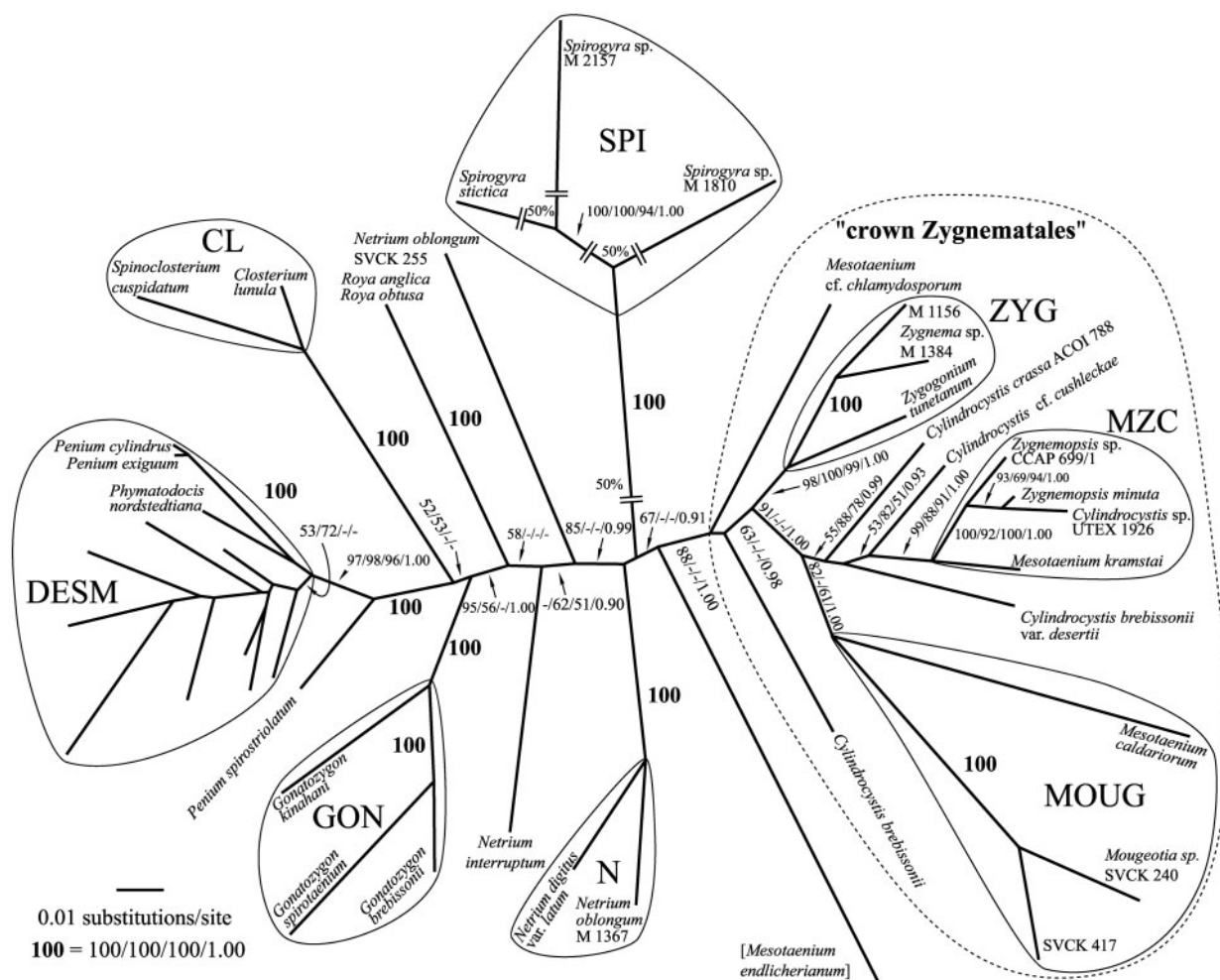


FIG. 3.—Combined (concatenated) analysis of 43 Zygnematomyceae inferred from SSU rDNA and *rbcL* sequences (3,075 aligned nt) using ML (GTR+I+ Γ ; for parameters, see table 3). Details as in figure 1. As in the SSU rDNA phylogeny (fig. 1), the Zygnematales and Desmidiaceae are monophyletic, revealing the placement of *Roya* and *Netrium oblongum* SVCK 255 in the *rbcL* tree (fig. 2) as an artifact. Compared with both single-gene analyses (figs. 1 and 2), the combined tree reveals much better topological resolution, higher statistical support for internal branches, and a more regular significance distribution.

(extremely long branched in SSU rDNA) revealed an average evolutionary rate in *rbcL* (figs. 1 and 2). Whereas in *Mesotaenium caldariorum*, the first and second codon positions contributed only 10 of 128 autapomorphic characters, this number was higher (34 of 166 and 24 of 154) in *Mesotaenium endlicherianum* and *Netrium oblongum*, respectively, reflecting their deviating amino acid composition. Exclusion of the long-branch species from the *rbcL* data set led to very similar tree topologies (not shown) and model parameters except for a higher gamma-shaped parameter (table 3).

To investigate the possible impact of homoplasy and saturation in third codon positions (Nickrent et al. 2000; Nozaki et al. 2000), the first two (902 nt) and the third position (451 nt) were analyzed separately. Analyses using first and second codon positions (trees not shown) revealed the same basic topology as for the complete *rbcL* alignment (fig. 2). However, resolution and significance decreased considerably and previously supported clades were either not recovered at all (GON, SPI, and ZYG) or received no statistical support (DESM and the DESM/*Penium spiro-*

striolatum clade). In contrast, analyses using third codon positions recovered the same clades as the complete data set with similar statistical confidence (trees not shown; for model parameters see table 3). The log-likelihood summation (LS) method using separate models for each codon (table 3) identified a tree with the lowest sum of log-likelihoods (not shown), which was identical to the “concatenated” tree shown in figure 3 with two exceptions: (1) the branches of *Phymatodocis* and *Penium exiguum*/*P. cylindricum* were separate, and (2) *Netrium oblongum* SVCK 255 was sister to *Roya* spp. (still positioned as sister to GON within the Desmidiaceae). Neither of these differences refers to significantly supported clades in figure 2.

Combined Analysis (SSU + *rbcL*)

The analysis of combined SSU rDNA and *rbcL* data sets (3,075 nt [table 3]) with a concatenated model resolved almost all internal branches separating zygnematomycean lineages (fig. 3). The tree topology was similar to the SSU rDNA tree (fig. 1), but internal branches were

Table 4
Comparison Between the Best ML Trees (Figure 1 to Figure 3) Using the SSU rDNA, *rbcL*, and Combined Data Sets and User-Defined Trees by Kishino-Hasegawa and Shimodaira-Hasegawa Tests

Tree compared	P^a															
	Diff-inL ^b			AU			KH			SH			wSH			
	SSUr DNA	<i>rbcL</i>	SSU + <i>rbcL</i>	SSUr DNA	<i>rbcL</i>	SSU + <i>rbcL</i>	SSUr DNA	<i>rbcL</i>	SSU + <i>rbcL</i>	SSUr DNA	<i>rbcL</i>	SSU + <i>rbcL</i>	SSUr DNA	<i>rbcL</i>	SSU + <i>rbcL</i>	
Best ML trees																
SSUr DNA	Best	152.2	118.8	0.965	**	**	0.961	**	0	0.976	**	**	0.988	**	0	0
<i>rbcL</i>	83.8	Best	39.4	**	0.982	0.031	**	0.966	0.038	**	0.974	0.138	**	0.989	0.064	0.064
SSU + <i>rbcL</i>	24.1	31.0	Best	0.035	0.018	0.970	0.039	0.034	0.962	0.113	0.140	0.996	0.066	0.058	0.998	0.998
2) UD trees																
1		8.2			0.061			0.057			0.510			0.152		
2		12.4			0.162			0.144			0.342			0.267		
3	14.5	14.4	25.7	0.041	0.039	0.048	0.058	0.050	0.038	0.182	0.278	0.163	0.111	0.124	0.087	0.087
4	9.3	34.8	27.8	0.216	0.003	0.010	0.155	0.004	0.017	0.312	0.025	0.104	0.345	0.010	0.056	0.056

NOTE.—AU = P -value of the approximately unbiased test calculated from the multiscale bootstrap. KH = Kishino-Hasegawa test. SH = Shimodaira-Hasegawa test. wSH = weighted Shimodaira-Hasegawa test. The tree topologies tested: 1. The orders Desmidiatales and Zygnematales are monophyletic (*Netrium oblongum* SVCK 255 and *Roya* are sisters to the Desmidiatales); 2. *Mesotaenium endlicherianum* is a sister to the “crown Zygnematales”; 3. *Penium cylindrus* and *P. spirostriolatum* are sisters; 4. The genus *Netrium* is monophyletic. ** $0 < P < 0.001$.

^a Probability of obtaining a more extreme T -value under the null hypothesis of no difference between the two trees.

^b Difference in $-\log$ -likelihood.

longer and received better support. Especially, the branch separating the Zygnematales and Desmidiatales was again recovered, in contrast to *rbcL* analyses, and received significant support by ML and BI (branch separating *Roya* and GON [fig. 3]). Notably, the combined analysis resolved the position of *Netrium oblongum* SVCK 255 and *Mesotaenium endlicherianum*, unlike the *rbcL* tree (the latter presumably because of their long branches). *Roya* and two of three (paraphyletic) *Netrium* lineages were resolved as sister taxa to the Desmidiatales with high significance by ML and BI (branch between N and *Netrium oblongum* SVCK 255 [fig. 3]).

In the combined analysis, the SPI clade again comprised the longest branch (>180 apomorphic characters of 3,075), although significantly shortened compared with the SSU rDNA tree (159 apomorphies of 1,722 [fig. 1]). In combined analyses without SPI sequences, the significance of almost every internal branch increased (not shown).

The log-likelihood summation method (LS) identified a tree with the lowest sum of log-likelihoods (designated LS tree, not shown), which was nearly identical to the “concatenated” tree shown in figure 3. This holds for tree topologies—in the LS-tree, the branches of *Mesotaenium endlicherianum* and SPI were interchanged with respect to figure 3—as well as their similar log-likelihood values (irrespective of using the concatenated or LS-method; data not shown). Among the “best” 100 topologies identified by LS, no SSU rDNA tree and only two *rbcL*-trees were found, whereas the remaining 98 trees originated from the concatenated analysis. A 95% majority-rule consensus of these 100 trees was again almost identical to figure 3—only the branch between *Mesotaenium endlicherianum* and SPI collapsed.

User-Defined Trees and Comparison of Best Trees

The first user-defined tree (UD-tree) addressed the nonmonophyly of the orders Zygnematales and Desmidiatales in the *rbcL*-analysis. To restore the monophyly of both orders, *Netrium oblongum* SVCK255 and *Roya* were positioned as sisters to the remaining Zygnematales in UD-tree 1, which was not rejected (table 4). Testing another conflicting case, the deviant position of *Mesotaenium endlicherianum* in *rbcL* trees, by moving it to the base of the “crown Zygnematales,” UD-tree 2 was also not significantly different from the best tree (fig. 2). In UD-tree 3, the monophyly of the genus *Penium* (polyphyletic in all phylogenies) was analyzed. However, this UD-tree was rejected by AU and KH but not by more relaxed SH and wSH tests. Similarly, enforcing a monophyletic *Netrium* (a polytomy in the analyses) was rejected by most tests in the *rbcL* and combined data sets but not in the SSU rDNA data set (UD-tree 4 [table 4]).

Comparison of the “best” ML trees from figures 1 and 3 using the SSU rDNA, the *rbcL* and the combined data sets revealed that the SSU rDNA tree was rejected by the *rbcL* and combined data sets with all tests performed ($P < 0.001$ [table 4]). The best *rbcL* tree was also rejected by the SSU rDNA data set with $P < 0.001$. However, the combined data set rejected this topology only in the AU

and KH tests ($0.03 < P < 0.05$). The combined ML tree was the least rejected (table 4). In the SH and wSH tests, this topology was not significantly rejected ($P > 0.05$).

Discussion

To exclude incongruent taxon sampling and likely artifacts associated with it, the two genes (*rbcL* and SSU rDNA) were sequenced exclusively from the same strain. The importance of this strict approach is shown by two examples: (1) two strains designated *Netrium oblongum* (M 1367 and SVCK 255; in morphology both corresponding to the species description) are in fact unrelated to each other, and (2) a database sequence (U38694) of strain UTEX 934 (*Roya anglica*) apparently was not derived from this strain, but refers to another genus, as shown in this study. Combining SSU rDNA and *rbcL* sequences of different strains may thus introduce taxon sampling artifacts, which could result in chimeric sequences and in single-gene trees in conflicting topologies. Our approach ensures that conflicts between single-gene topologies are derived from different patterns of sequence evolution between the genes.

SSU rDNA and *rbcL* sequence comparisons of 43 strains of the Zygnematophyceae were used to analyze the relation between single-gene versus combined analyses, gene concatenation versus log-likelihood summation, and bootstrap percentages (ML, NJ, and MP) versus posterior probabilities (BI), as an evolutionary case study based on empirical data. Previously published phylogenetic analyses using SSU rDNA sequence comparisons in the Zygnematophyceae suffered mostly from lack of resolution (Besendahl and Bhattacharya 1999; Denboh, Hendrayanti, and Ichimura 2001; Gontcharov, Marin, and Melkonian 2003), whereas in *rbcL* studies, taxon sampling was limited, with only one species per genus included (McCourt et al. 2000). In phylogenetic analyses of streptophyte green algae, different molecular markers favored conflicting tree topologies; for example, concerning the position of the genera *Mesostigma*, *Klebsormidium*, or the group studied here, the Zygnematophyceae (Marin and Melkonian 1999; Lemieux, Otis, and Turmel 2000; McCourt et al. 2000; Karol et al. 2001; Cimino and Delwiche 2002; Delwiche et al. 2002; Martin et al. 2002; Turmel et al. 2002a).

Single-Gene Data and Analyses

In the Zygnematophyceae, ribosomal (SSU rDNA) and chloroplast (*rbcL*) genes studied revealed considerable differences in their evolutionary dynamics as reflected by model parameters (patterns of nucleotide substitutions, Γ , base composition). In both genes, the C \leftrightarrow T substitution category is conspicuously high in comparison with the remaining frequencies (table 3), a situation, which for SSU rDNA may be related to pairing constraints at the transcript (rRNA) level (G-C \leftrightarrow G-U) and for *rbcL* (an even higher C \leftrightarrow T value) is caused by asymmetrical codon usage (Morton 1994). The gamma-shaped parameter estimated for the SSU rDNA data set is nearly three-fold lower than that for *rbcL*, in which the variability distribution reflects the regular codon structure. Moreover,

the length of the Rubisco large subunit is conserved in the Viridiplantae (476 amino acids), and the variability at the amino acid level is rather moderate (reviewed by Kellogg and Juliano [1997]). In the Zygnematophyceae, less than 80% of the sequence variability of *rbcL* refers to third codon positions. Although third codon positions are sometimes down-weighted or excluded from the phylogenetic analyses because of codon degeneration and homoplasy (Nickrent et al. 2000; Nozaki et al. 2000), restriction of our analyses to first and second codon positions resulted in greatly diminished resolution (see also McCourt et al. [2000]). However, using only the third codon position, the topology and resolution reflected that obtained for all positions, thus, demonstrating that most of the phylogenetic signal resides in the third codon.

Not surprisingly, SSU rDNA and *rbcL* data led to selection of different models of evolution and model parameters (table 3), but these models were not in conflict (TrN is nested within GTR) and do not prevent a combined analysis using a concatenated (averaged) model (here: GTR).

In general, both single-gene trees show a high degree of congruence and largely recover the same taxa and clades with comparable BP and PP support. However, the analyses also reveal some conflicts that are not caused by incongruence in taxon sampling. The most obvious discrepancy between the two single-gene trees relates to the split between the orders Zygnematales and Desmidiiales, which is resolved in the SSU rDNA analysis but not in *rbcL* phylogenies (McCourt et al. 2000; this study); in our *rbcL* study, the Desmidiiales are mixed with two zygnematalean branches (*Roya* and *Netrium oblongum* SVCK 255). However, even this incongruence refers to internal branches without significant bootstrap support in the *rbcL* analysis (in contrast to PP [see below]), and moving *Roya* and *N. oblongum* SVCK 255 to the Zygnematales is not rejected in user-defined topology tests (table 4). As an example for congruence, the polyphyly of *Mesotaenium*, *Cylindrocystis*, and *Netrium* is clearly revealed in both single-gene phylogenies.

Combined Analyses

When both data sets were combined as a concatenated "supergene" and analyzed with a single average model, the resulting phylogeny was superior to both single-gene analyses when the statistical support of internal branches is considered. Specifically, concatenated data significantly (BP and PP) resolved the major conflict between *rbcL* and SSU rDNA trees (Desmidiiales-Zygnematales divergence) in favor of the SSU rDNA analysis (i.e., monophyly of both groups in the unrooted phylogeny). Several other internal branches (especially basal branches of the Zygnematales), which in single-gene phylogenies were not or weakly supported, obtained higher significance in the combined analysis (see *Results*). In general, the combined analysis was dominated by the phylogenetic signal of the SSU rDNA data set, whereas *rbcL* contributed sufficient sequence diversity to improve resolution within clades (see MZC clade where SSU rDNAs are almost

identical) but also to increase the overall significance of the branches.

It is somewhat illegitimate (a logical circle) to use higher statistical support for branches as the only criterion to regard combined analysis as superior (i.e., closer to the “true tree” than the *rbcL* analysis). Of course, there is no a priori knowledge of the “true tree.” However, cell-wall characters provide some independent evidence for comparing evolutionary hypotheses in the Zygnematales. The SSU rDNA tree reveals a single evolutionary transition from simple cell walls (Zygnematales) towards complex cell walls in the monophyletic Desmidiaceae, without homoplasious character changes. The conflicting *rbcL* scenario, if correct, would imply two additional changes (i.e., reversals: complex→simple wall) for those Zygnematales that are rooted within the Desmidiaceae (*Roya* and *Netrium oblongum* SVCK 255). Thus, the parsimony criterion applied to cell-wall structures (one versus three character changes), as well as the combined SSU rDNA+*rbcL* analyses, both favor the SSU rDNA topology and emphasize the value of complex cell walls as a phylogenetic marker.

In a concatenated analysis, it is possible that the longer or the more variable data set dominates the “average” model of sequence evolution and, thus, the resulting topology. However, the LS method applied here (which avoids the use of an average model) agreed with the results obtained by the concatenated analysis. There are, of course, conditions under which concatenated phylogenetic analyses will fail. Whenever different genes evolve under deviating rules (i.e., models), a concatenated analysis may be significantly worse than phylogenies that allow the application of separate models, as shown by Pupko et al. (2002). In the Zygnematales, the two single-gene models were not in conflict (see above), but some model parameters, especially the gamma-shaped and the C↔T substitution category, differed considerably between SSU rDNA and *rbcL* (see *Results*). The concatenated model (averaged parameters) apparently did not violate the evolutionary characteristics of both genes.

Combined Analyses and Long Branches

It has been proposed that fast-evolving taxa with long undivided branches should be better analyzed by using another, slow-evolving, gene to avoid LBA (Philippe 2000). Fortunately, all long-branch taxa in our analyses are either fast-evolving in *rbcL* (*M. endlicherianum* and *N. oblongum* SVCK 255) or in the SSU rDNA gene (*Spirogyra*) but not in both data sets. Our results show that the combination of slow-evolving and fast-evolving genes can also resolve the phylogenetic position of a taxon with a fast-evolving gene, in particular when improved taxon sampling does not help to subdivide the long branch (e.g., SSU rDNA in *Spirogyra*). It appears that the combined approach can successfully extract the phylogenetic signal from the fast-evolving gene and nevertheless reduce LBA (see also Hoef-Emden, Marin, and Melkonian [2002]). The latter is illustrated by comparison of the positions of *Mesotaenium endlicherianum* and *Netrium oblongum* SVCK 255 in single-gene and combined analyses. Both taxa have long branches in the *rbcL* phylogeny but not in the

SSU rDNA phylogeny. Their position in the *rbcL* tree (fig. 2) most likely reflects a LBA artifact because in combined analyses, they comprise shorter (albeit still relatively long) branches (no LBA), and their placement confirms the SSU rDNA topology with better significance. The most conspicuous long branch in our analyses refers to the genus *Spirogyra* in SSU rDNA trees, in which *Spirogyra* was previously positioned as a sister to all other Zygnematophyceae (Besendahl and Bhattacharya 1999) or as one of the basal divergences of the class (Gontcharov, Marin, and Melkonian 2003). In the *rbcL* phylogeny, *Spirogyra* is not a long-branch taxon, but interestingly, the position of *Spirogyra* does not contradict rooted (Gontcharov, Marin, and Melkonian 2003) or unrooted (this study) SSU rDNA trees. We conclude that the position of *Spirogyra* in the SSU rDNA trees was likely not the result of an LBA artifact.

The examples discussed above may provide some confidence that the deep-level phylogeny of the Zygnematophyceae based on the conservative SSU rRNA gene agrees better with organismal data and is less sensitive to various artifacts than analyses using the homoplasious and largely saturated gene *rbcL*. However, the latter gene resolves shallow evolutionary relationships much better where the conserved SSU rDNA lacks variability.

Phylogeny of the Zygnematophyceae

Although the Zygnematales, characterized by a simple cell wall (consisting of only one piece, no pores [Mix 1972]; plesiomorphic character state), form a clade in the unrooted trees, this taxon is not monophyletic because the root of the Zygnematophyceae falls within the Zygnematales, and, thus, reveals this order as paraphyletic (McCourt et al. 2000; Gontcharov, Marin, and Melkonian 2003). Chloroplast shape (three types) and level of organization (unicellular or filamentous) vary in the Zygnematales and have previously been used for classification (e.g., Palla 1894; Randhawa 1959; Yamagishi 1963). Generally, none of these proposals gains support by our molecular phylogenetic analyses. We have tentatively identified seven lineages within the traditional Zygnematales, namely *Roya*, N, SPI, the “crown-Zygnematales,” and three individual taxa (*Mesotaenium endlicherianum*, *Netrium interruptum*, and *N. oblongum* SVCK 255). Three filamentous genera (*Spirogyra*, *Mougeotia*, *Zygnema*) are resolved as monophyletic, whereas other genera are not monophyletic. Obviously, in the Zygnematales, the genetic diversity at the genus level was severely underestimated by traditional taxonomists because the importance of the organizational level (unicellular versus filamentous) and chloroplast shape (axial platelike, stellate, or helical) has been overestimated. These characters may have originated or have been lost several times independently in the group. Our phylogenies place unicellular organisms as basal divergences of some zygnematalean clades (“crown-Zygnematales,” MZC, and MOUG) tentatively suggesting that unicells could perhaps be ancestral to these lineages.

Probably the most interesting genus resolved here as nonmonophyletic is *Netrium*—species analyzed form three independent branches, each characterized by a different number of chloroplasts per cell (1, 2, or 4), differing

positions of the nucleus in the cell, and varying nuclear behavior during cytokinesis (Pickett-Heaps 1975; Jarman and Pickett-Heaps 1990; unpublished observations). The three *Netrium* branches occupy a key position between the other Zygnematales and the Desmidiaceae, supporting previous rooted analyses containing only one *Netrium* species (as sister of the Desmidiaceae [McCourt et al. 2000; Gontcharov, Marin, and Melkonian 2003]). In the latter publication, we erroneously reported two species of *Netrium* as having identical SSU rDNA sequences; however, these sequences actually originated from the same culture, *Netrium interruptum* strain M 1021. Our expanded taxon sampling in this study reveals the Desmidiaceae as originating from a paraphyletic stock of derived unicellular Zygnematales (i.e., *Netrium* and *Roya* branches).

The Desmidiaceae, a clade defined by derived cell-wall characters, is well supported (but not in the *rbcL* phylogeny [see above]). The molecular phylogeny within the Desmidiaceae as revealed here and by previous studies (McCourt et al. 2000; Gontcharov, Marin, and Melkonian 2003) reflects the increasing complexity of the cell-wall ultrastructure. Among the four families described, the Gonatozygaceae and Closteriaceae are confirmed, whereas the concept of the Peniaceae (cell wall consisting of several segments separated by shallow groove(s); simple pores perforating only the outer cell wall layer) and the Desmidiaceae (constricted cells composed of two semicells with complex [simple in *Phymatodocis* {Engels and Lorch 1981}] cell wall pores) are in need of revision. Two of three species of *Penium* analyzed form a robust clade with the Desmidiaceae (DESM). The third *Penium* species (*P. spirostriolatum*) forms a clade with DESM, although no morphological synapomorphy is presently known. Because of its simple cell-wall structure, *Penium* is usually not regarded as closely related to the Desmidiaceae.

Bayesian Phylogenetics

Bayesian inference, a recently introduced method for inferring molecular phylogenies (Huelsenbeck and Ronquist 2001; Rannala and Yang 1996), provides a statistical confidence measure (PP) for branches and is much faster than ML bootstrap analysis. However, PP values are often much higher than ML BP and thus, the reliability of this method has recently been controversially discussed (Huelsenbeck et al. 2002; Suzuki, Glazko, and Nei 2002; Alfaro, Zoller, and Lutzoni 2003; Douady et al. 2003). Based on simulation studies or using real sequence data, some authors considered ML BP as too conservative (Hillis and Bull 1993; Murphy et al. 2001; Wilcox et al. 2002; Alfaro, Zoller, and Lutzoni 2003), whereas others concluded that PP is too optimistic (Suzuki, Glazko, and Nei 2002). In our empirical study, the level of support for branches by PP or BP is roughly similar, although we also found several branches with significant PP support (≥ 0.95), which were not substantiated by significant BP values. One branch in the *rbcL* phylogeny, namely the branch separating CL and *Roya*/GON (fig. 2), defines an artificial divergence that does not exist in the SSU rDNA topology and in the combined analysis (see *Results*). This artificial branch receives no BP

values in the *rbcL* analysis, but considerable support by Bayesian inference (PP = 0.99 [fig. 2]). It is known that Bayesian analysis is sensitive to small-model misspecifications (Waddell, Kishino, and Ota 2001; Buckley 2002; Buckley et al. 2002), here probably related to the individual long-branch taxa *Mesotaenium endlicherianum* and *Netrium oblongum* SVCK255 and the high level of homoplasy in the *rbcL* gene. We conclude that Bayesian inference can be positively misleading as exemplified in our case study and suggest that PP support should always be confirmed by traditional bootstrap analyses.

Acknowledgments

We thank Kerstin Hoef-Emden for help with the CONSEL program and two anonymous reviewers for helpful comments. This study was supported by a grant from the Alexander von Humboldt-Stiftung to A.G.

Literature Cited

- Alfaro, M. E., S. Zoller, and F. Lutzoni. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol. Biol. Evol.* **20**:255–266.
- Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**:972–977.
- Bapteste, E., H. Brinkmann, and J. A. Lee et al. (11 co-authors) 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl. Acad. Sci. USA* **99**:1414–1419.
- Besendahl, A., and D. Bhattacharya. 1999. Evolutionary analyses of small-subunit rDNA coding regions and the 1506 group I introns of Zygnematales (Charophyceae, Streptophyta). *J. Phycol.* **35**:560–569.
- Bhattacharya, D., B. Surek, M. Rüsing, S. Damberger, and M. Melkonian. 1994. Group I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Chlorophyta). *Proc. Natl. Acad. Sci. USA* **91**:9916–9920.
- Bowe, L. M., G. Coat, and C. W. dePamphilis. 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc. Natl. Acad. Sci. USA* **97**:4092–4097.
- Buckley, T. R. 2002. Model misspecification and probabilistic tests of topology: Evidence from empirical data sets. *Syst. Biol.* **51**:509–523.
- Buckley, T. R., P. Arensburger, C. Simon, and G. K. Chambers. 2002. Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. *Syst. Biol.* **51**:4–18.
- Cao, Y., M. Fujiwara, M. Nikaido, N. Okada, and M. Hasegawa. 1998. Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. *Gene* **259**:149–158.
- Capesius I., and M. Bopp. 1997. New classification of liverworts based on molecular and morphological data. *Plant Syst. Evol.* **207**:87–97.
- Chapman, R. L., M. A. Buchheim, C. F. Delwiche et al. (11 co-authors). 1998. Molecular Systematics of the Green Algae. Pp. 508–540 in D. E. Soltis, P. S. Soltis, and J. J. Doyle, eds.

- Molecular systematics of plants II. Kluwer Academic Publishers, Boston.
- Cimino, M. T., and C. F. Delwiche. 2002. Molecular and morphological data identify a cryptic species complex in endophytic members of the genus *Coleochaete* Breb. (Charophyta: Coleochaetaceae). *J. Phycol.* **38**:1213–1221.
- Delwiche, C. F., K. G. Karol, M. T. Cimino, and K. J. Sytsma. 2002. Phylogeny of the genus *Coleochaete* (Coleochaetales, Charophyta) and related taxa inferred by analysis of the chloroplast gene *rbcL*. *J. Phycol.* **38**:394–403.
- Denboh, T., D. Hendrayanti, and T. Ichimura. 2001. Monophyly of the genus *Closterium* and the order Desmidiaceae (Charophyceae, Chlorophyta) inferred from nuclear small subunit rDNA data. *J. Phycol.* **37**:1063–1072.
- Douady, C. J., F. Delsuc, Y. Boucher, W. F. Doolittle, and E. J. P. Douzery. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.* **20**:248–254.
- Engels, M., and D. W. Lorch. 1981. Some observations on cell wall structure and taxonomy of *Phymatodocis nordstediana* (Conjugatophyceae, Chlorophyta). *Plant Syst. Evol.* **138**:217–225.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- Gontcharov, A. A., B. Marin, and M. Melkonian. 2003. Molecular phylogeny of conjugating green algae (Zygnemophyceae, Streptophyta) inferred from SSU rDNA sequence comparisons. *J. Mol. Evol.* **56**:89–104.
- Graybeal, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47**:9–17.
- Hillis, D. M. 1996. Inferring complex phylogenies. *Nature* **383**:130–131.
- Hillis, D. M., and J. J. Bull. 1993. An empirical-test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42**:182–192.
- Hoef-Emden, K., B. Marin, and M. Melkonian. 2002. Nuclear and nucleomorph SSU rDNA phylogeny in the cryptophyta and the evolution of cryptophyte diversity. *J. Mol. Evol.* **55**:161–179.
- Huelsenbeck, J. P., B. Larget, R. E. Miller, and F. Ronquist. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* **51**:673–688.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754–755.
- Jarman, J., and J. Pickett-Heaps. 1990. Cell division and nuclear movement in the saccoderm desmid *Netrium interruptus*. *Protoplasma* **157**:136–143.
- Karol, K. G., R. M. McCourt, M. T. Cimino, and C. F. Delwiche. 2001. The closest living relatives of land plants. *Science* **294**:2351–2353.
- Kellogg, E. A., and N. D. Juliano. 1997. The structure and function of RuBisCO and their implications for systematic studies. *Am. J. Bot.* **84**:413–428.
- Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of the Hominoidea. *J. Mol. Evol.* **29**:170–179.
- Lemieux C., C. Otis, and M. Turmel. 2000. Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature* **403**:649–652.
- Mallatt, J., and C. J. Winchell. 2002. Testing the new animal phylogeny: First use of combined large-subunit and small-subunit rRNA gene sequences to classify the protosomes. *Mol. Biol. Evol.* **19**:289–301.
- Marin, B., M. Klingberg, and M. Melkonian. 1998. Phylogenetic relationships among the Cryptophyta: analysis of nuclear-encoded SSU rRNA sequences support the monophyly of extant plastid-containing lineages. *Protist* **149**:265–276.
- Marin, B., and M. Melkonian. 1999. Mesostigmatophyceae, a new class of streptophyte green algae revealed by SSU rRNA sequence comparisons. *Protist* **150**:399–417.
- Martin, W., T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B., Stoebe, M. Hasegawa, and D. Penny. 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. USA* **99**:12246–12251.
- McCourt, R. M., K. G. Karol, J. Bell, K. M. Helm-Bychowski, A. Grajewska, M. F. Wojciechowski, and R. W. Hoshaw. 2000. Phylogeny of the conjugating green algae (Zygnemophyceae) based on *rbcL* sequences. *J. Phycol.* **36**:747–758.
- McCourt, R. M., K. G. Karol, S. Kaplan, and R. W. Hoshaw. 1995. Using *rbcL* sequences to test hypotheses of chloroplast and thallus evolution in conjugating green algae (Zygnematales, Charophyceae). *J. Phycol.* **31**:989–995.
- McFadden, G. I., and M. Melkonian. 1986. Use of Hepes buffer for microalgal culture media and fixation for electron microscopy. *Phycologia* **25**:551–557.
- Mix, M. 1972. Die Feinstruktur der Zellwände bei Mesotaeniaceae und Gonatozygaceae mit einer vergleichenden Betrachtung der verschiedenen Wandtypen der Conjugatophyceae und über deren systematischen Wert. *Arch. Mikrobiol.* **81**:197–220.
- Morton, B. R. 1994. Codon use and the rate of divergence of land plant chloroplast gene. *Mol. Biol. Evol.* **11**:231–238.
- Murphy, W. J., E. Eizirik, S. J. O'Brien et al. (11 co-authors). 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**:2348–2351.
- Nei, M., S. Kumar, and K. Takahashi. 1998. The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. *Proc. Natl. Acad. Sci. USA* **95**:12390–12397.
- Nickrent, D. L., C. L. Parkinson, J. D. Palmer, and R. J. Duff. 2000. Multi-gene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol. Biol. Evol.* **17**:1885–1895.
- Nozaki, H., K. Misawa, T. Kajita, M. Kato, S. Nohara, and M. M. Watanabe. 2000. Origin and evolution of the colonial Volvocales (Chlorophyceae) as inferred from multiple, chloroplast gene sequences. *Mol. Phylogenet. Evol.* **17**:256–268.
- Olsen, G. J. 1990. Sequence editor and analysis program. University of Illinois, Urbana.
- Page, R. D. M. 1996. TreeView: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**:357–358.
- Palla, E. 1894. Über eine neue, pyrenoidlose Art und Gattung der Conjugaten. *Ber. Dt. Bot. Ges.* **12**:228–236.
- Park, N. E., K. G. Karol, R. W. Hoshaw, and R. M. McCourt. 1996. Phylogeny of *Gonatozygon* and *Genicularia* (Gonatozygaceae, Desmidiaceae) based on *rbcL* sequences. *Eur. J. Phycol.* **31**:309–313.
- Philippe, H. 2000. Opinion: long branch attraction and protist phylogeny. *Protist* **151**:307–316.
- Pickett-Heaps, J. 1975. Green algae: structure, function and evolution in selected genera. Sinauer, Sunderland, Mass.
- Poe, S., and D. L. Swofford. 1999. Taxon sampling revisited. *Nature* **398**:299–300.
- Posada, D., and K. A. Crandal. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- Pupko, T., D. Huchon, Y. Cao, N. Okada, and M. Hasegawa. 2002. Combining multiple data sets in a likelihood analysis: Which models are the best? *Mol. Biol. Evol.* **19**:2294–2307.

- Randhawa, M. S. 1959. Zygnemaceae. Indian Council of Agriculture Research, New Dehli.
- Rannala, B., and Z. Yang. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol.* **43**:304–311.
- Sanderson, M. J., and H. B. Shaffer. 2002. Troubleshooting molecular phylogenetic analyses. *Annu. Rev. Ecol. Syst.* **33**:49–72.
- Shaw, A. J., and B. Allen. 2000. Phylogenetic relationships, morphological incongruence, and geographic speciation in the Fontinalaceae (Bryophyta). *Mol. Phylogenet. Evol.* **16**:225–237.
- Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**:492–508.
- Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**:1114–1116.
- . 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**:1246–1247.
- Surek, B., U. Beemelmans, M. Melkonian, and D. Bhattacharya. 1994. Ribosomal RNA sequence comparisons demonstrate an evolutionary relationship between Zygnematales and charophytes. *Plant Syst. Evol.* **191**:171–181.
- Surek, B., and P. Sengbusch. 1981. The localization of galactosyl residues and lectin receptors in the mucilage and the cell walls of *Cosmocladium saxonicum* (Desmidiaceae) by means of fluorescent probes. *Protoplasma* **108**:149–161.
- Suzuki, Y., G. V. Glazko, and M. Nei. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci. USA* **99**:16138–16143.
- Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Mass.
- Turmel, M., M. Ehara, C. Otis, and C. Lemieux. 2002a. Phylogenetic relationships among streptophytes as inferred from chloroplast small and large subunit rRNA gene sequences. *J. Phycol.* **38**:364–375.
- Waddell, P. J., H. Kishino, and R. Ota. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Informat. Ser.* **12**:141–155.
- Wilcox, T. P., D. J. Zwickl, T. A. Heath, and D. M. Hillis. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol. Phylogenet. Evol.* **25**:361–371.
- Wuyts, J., P. De Rijk, Y. Van de Peer, G. Pison, P. Rousseeuw, and R. De Wachter. 2000. Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA. *Nucleic Acids Res.* **28**:4698–4708.
- Wuyts, J., Y. Van de Peer, and R. De Wachter. 2001. Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. *Nucleic Acids Res.* **29**:5017–5028.
- Yamagishi, T. 1963. Classification of the Zygnemataceae. *Sci. Rep. Tokyo Kyoiku Daigaku B.* **11**:191–210.
- Yang, Z. 1996. Maximum-likelihood models for combined analyses of multiple sequence data. *J. Mol. Evol.* **42**:587–596.

Hervé Philippe, Associate Editor

Accepted November 4, 2003