

Tracing the Evolution of Streptophyte Algae and Their Mitochondrial Genome

Monique Turmel*, Christian Otis, and Claude Lemieux

Institut de Biologie Intégrative et des Systèmes, Département de Biochimie, de Microbiologie et de Bio-Informatique, Université Laval, Québec, Canada

*Corresponding author: E-mail: monique.turmel@bcm.ulaval.ca.

Accepted: September 2, 2013

Data deposition: This project has been deposited at GenBank under the accession numbers KF060939, KF060940, KF060941, KF060942, and KF060943.

Abstract

Six monophyletic groups of charophycean green algae are recognized within the Streptophyta. Although incongruent with earlier studies based on genes from three cellular compartments, chloroplast and nuclear phylogenomic analyses have resolved identical relationships among these groups, placing the Zygnematales or the Zygnematales + Coleochaetales as sister to land plants. The present investigation aimed at determining whether this consensus view is supported by the mitochondrial genome and at gaining insight into mitochondrial DNA (mtDNA) evolution within and across streptophyte algal lineages and during the transition toward the first land plants. We present here the newly sequenced mtDNAs of representatives of the Klebsormidiales (*Entransia fimbriata* and *Klebsormidium* spec.) and Zygnematales (*Closterium baillyanum* and *Roya obtusa*) and compare them with their homologs in other charophycean lineages as well as in selected embryophyte and chlorophyte lineages. Our results indicate that important changes occurred at the levels of genome size, gene order, and intron content within the Zygnematales. Although the representatives of the Klebsormidiales display more similarity in genome size and intron content, gene order seems more fluid and gene losses more frequent than in other charophycean lineages. In contrast, the two members of the Charales display an extremely conservative pattern of mtDNA evolution. Collectively, our analyses of gene order and gene content and the phylogenies we inferred from 40 mtDNA-encoded proteins failed to resolve the relationships among the Zygnematales, Coleochaetales, and Charales; however, they are consistent with previous phylogenomic studies in favoring that the morphologically complex Charales are not sister to land plants.

Key words: Streptophyta, Charophyceae, charophytes, mitochondrial DNA, phylogenomics, genome evolution.

Introduction

Between 500 and 450 Ma, the first land plants emerged on Earth from green algal ancestors related to the contemporary freshwater green algae classified in the Charophyceae (Gensel 2008; Kenrick et al. 2012). Charophycean green algae and land plants form the Streptophyta (Bremer et al. 1987), whereas all other green algae belong to the sister lineage Chlorophyta (Lewis and McCourt 2004). Six monophyletic groups of charophycean green algae are currently recognized: the Mesostigmatales, Chlorokybales, Klebsormidiales, Zygnematales, Coleochaetales, and Charales, given here in order of increasing cellular complexity.

Which charophycean group gave rise to the land plants? As indicated by several phylogenetic studies of multigene data sets, this is a difficult question to address. Analyses of four genes from the nucleus (18S rRNA gene), chloroplast

(*atpB* and *rbcl*), and mitochondria (*nad5*) of 25 charophycean green algae, 8 land plants, and 5 chlorophytes identified with great support the Charales as the closest green algal relatives of land plants; however, the positions of the other charophycean groups received moderate support (Karol et al. 2001; McCourt et al. 2004). The best trees inferred in this four-gene analysis supported an evolutionary trend toward increasing cellular complexity; undeniably, this was an appealing result. The same topology was recovered in another analysis of combined genes from the three cellular compartments (Qiu et al. 2006). The latter two studies turned out to conflict with analyses of multiple chloroplast and nuclear genes. In agreement with trees inferred from the chloroplast small and large subunit rRNA genes (Turmel, Ehara et al. 2002), a whole-chloroplast genome study by our group resolved the Charales before the divergence of the Coleochaetales and

Zygnematales and showed that either the Zygnematales or a clade composed of the Zygnematales and Coleochaetales are the closest relatives of land plants (Turmel et al. 2006, 2007b). Recently, this conclusion received support from three independent nuclear phylogenomic analyses (Wodniok et al. 2011; Laurin-Lemay et al. 2012; Timme et al. 2012), thus reinforcing confidence in the basal position of the Charales. Furthermore, Laurin-Lemay et al. (2012) have provided strong evidence that the Coleochaetales + Zygnematales are sister to land plants. Why the analyses of combined genes yielded a different topology is unclear, but it is worth mentioning here that the misleading signal supporting the strong affiliation of the Charales with the land plants in the four-gene analysis was mainly contributed by *nad5*, the sole mitochondrial gene in the data set (Turmel et al. 2006). When using multigene sequence alignments for resolving deeply diverging relationships, misleading information due to violations of evolutionary models, bias in sequence composition, or sequence heterotachy may yield the wrong tree topology (Phillips et al. 2004; Jeffroy et al. 2006; Philippe et al. 2011b).

In the present investigation, we undertook a detailed analysis of streptophyte mitochondrial genomes in order to explore the evolutionary patterns of mitochondrial DNA (mtDNA) and also to determine whether a mitochondrial phylogeny congruent with those previously inferred from chloroplast and nuclear genes can be reconstructed. To date, the mitochondrial genomes of five charophyceans representing four distinct lineages, i.e., the genomes of *Mesostigma viride* (Mesostigmatales), *Chlorokybus atmophyticus* (Chlorokybales), *Chaetosphaeridium globosum* (Coleochaetales), and *Chara vulgaris* (Charales), have been described in the literature (Turmel et al. 2002a, 2002b, 2003, 2007a). These genomes show substantial variability in size, gene order, gene density, and intron content. The *Chara* mtDNA was found to be the most similar to its bryophyte counterparts at the levels of gene order and intron content, and it was also the closest to land plants mtDNAs in phylogenetic analyses of multiple genes (Turmel et al. 2003). More recently, the mtDNA sequence of another charalean green alga, *Nitella hyalina*, has become available (GenBank accession NC_017598).

Here, we report the mtDNA sequences of two charophyceans representing the Klebsormidiales, *Entransia fimbriata* and *Klebsormidium* spec., and of two others representing the Zygnematales, *Closterium baillyanum* and *Roya obtusa*. These genome sequences were compared with those previously reported for the abovementioned charophycean green algae and for selected embryophytes and chlorophytes. The results unveiled the dynamic nature of the mitochondrial genome in both the Klebsormidiales and Zygnematales. At the levels of gene order and gene content, we found that the Zygnematales and the Zygnematales + Coleochaetales uniquely share more structural genomic characters with bryophytes than the Charales. Our phylogenetic studies of 40

concatenated proteins were complicated by systematic errors causing the affiliation of the Charales with land plants, but when we minimized the impact of these errors, trees weakly supporting the notion that the Charales are not sister to land plants were recovered.

Materials and Methods

Strains and Culture Conditions

Strains of *Klebsormidium* spec. (SAG 51.86), *Closterium baillyanum* (SAG 50.89), and *Roya obtusa* (SAG 168.80) were obtained from the Sammlung von Algenkulturen Göttingen, whereas *Entransia fimbriata* (UTEX LB 2353) originated from the Culture Collection of Algae at the University of Texas at Austin. All four strains were grown in C medium (Andersen 2005) under alternating 12-h light/dark periods.

DNA Sequencing and Sequence Analyses

For each strain, an A + T rich organellar DNA fraction was obtained by CsCl-bisbenzimidazole isopycnic centrifugation of total cellular DNA as described earlier (Turmel et al. 1999). The *Entransia* and *Roya* organellar fractions were sequenced using 454 GS FLX Titanium technology at the Plate-forme d'Analyses Génomiques of Laval University (<http://pag.ibis.ulaval.ca/seq/en/index.php>, last accessed September 27, 2013). The 454 pyrosequencing reads were assembled using Newbler v2.5 (Margulies et al. 2005) with default parameters. For the *Klebsormidium* and *Closterium* mitochondrial genomes, random clone libraries were prepared from 1,500 to 2,000 bp fragments derived from the A + T rich DNA fractions using the pSMART-HCKan (Lucigen Corporation, Middleton, WI) plasmid. Positive clones were selected by hybridization of each plasmid library with the original DNA used for cloning. DNA templates were amplified using the Illustra TempliPhi Amplification Kit (GE Healthcare, Baie d'Urfé, Canada) and sequenced with the PRISM BigDye terminator cycle sequencing ready reaction kit (Applied Biosystems, Foster City, CA) on ABI model 373 or 377 DNA sequencers (Applied Biosystems), using T3 and T7 primers as well as oligonucleotides complementary to internal regions of the plasmid DNA inserts. The resulting sequences were edited and assembled using SEQUENCHER 4.8 (Gene Codes Corporation, Ann Arbor, MI). Genomic regions not represented in the sequence assemblies or plasmid clones were directly sequenced from polymerase chain reaction-amplified fragments using internal primers.

Genes and open reading frames (ORFs) were identified using a custom-built suite of bioinformatic tools as described previously (Pombert et al. 2005). tRNA genes were localized using tRNAscan-SE (Lowe and Eddy 1997). Intron boundaries were determined by modeling intron secondary structures (Michel et al. 1989; Michel and Westhof 1990) and by comparing intron-containing genes with intronless homologs using FRAMEALIGN of the Genetics Computer Group

software (version 10.3) package (Accelrys, San Diego, CA). To estimate the proportion of repeated sequences, repeats ≥ 30 bp were retrieved using REPFIND of the REPuter 2.74 program (Kurtz et al. 2001) with the options -f (forward) -p (palindromic) -l (minimum length = 30 bp) -allmax and then masked on the genome sequence using REPEATMASKER (<http://www.repeatmasker.org/>, last accessed September 27, 2013) running under the Crossmatch search engine (<http://www.phrap.org/>, last accessed September 27, 2013).

Analyses of Gene Order Data

We used a custom-built program to identify the regions that display the same gene order in selected pairs of streptophyte mtDNAs. This program was also employed to convert gene order in each of 16 streptophyte mtDNAs to all possible pairs of signed genes (i.e., taking into account gene polarity). The presence/absence of the signed gene pairs in two or more genomes was coded as binary Dollo characters using MacClade 4.08 (Maddison and Maddison 2000). The resulting gene order data set was subjected to maximum parsimony (MP) analysis using PAUP 4.0b10 (Swofford 2003). Confidence of branch points was estimated by 1,000 bootstrap replications.

Phylogenies based on inversion medians were inferred with GRAPPA 2.0 (<http://www.cs.unm.edu/~moret/GRAPPA/>, last accessed September 27, 2013) using the algorithm of Caprara (2003) and a data set of 55 gene/pseudogene positions. Although this software can use breakpoint medians to compute trees, we used inversion medians because this method has been shown to outperform breakpoint medians in phylogeny reconstructions (Moret et al. 2002).

Phylogenetic Reconstructions from Sequence Data

Mitochondrial genome sequences were retrieved from GenBank for the 25 green plant taxa listed in table 1. We selected for analysis the protein-coding genes that are shared by at least 14 of the 25 taxa. Forty genes met this criterion: *atp1*, 4, 6, 8, 9, *cob*, *cox1*, 2, 3, *mttB*, *nad1*, 2, 3, 4, 4L, 5, 6, 7, 9, *rpl2*, 5, 6, 10, 16, *rps1*, 2, 3, 4, 7, 10, 11, 12, 13, 14, 19, *sdh3*, 4, and *yejR*, U, V. An amino acid data set was prepared as follows. The deduced amino acid sequences from the 40 individual genes were aligned using MUSCLE 3.7 (Edgar 2004), and the ambiguously aligned regions in each alignment were removed using TRIMAL 1.3 (Capella-Gutierrez et al. 2009) with the following options: block = 7, gt = 0.7, st = 0.001, and sw = 3, and the protein alignments were concatenated. Missing characters represented 12.5% of the amino acid data set. This data set is available upon request to the corresponding author.

Phylogenies were inferred from the amino acid data set using the maximum likelihood (ML) and Bayesian inference (BI) methods. ML analyses were carried out using RAxML 7.2.8 (Stamatakis 2006) and the site-homogeneous LG + Γ 4 + F and GTR + Γ 4 models of sequence evolution.

Confidence of branch points was estimated by fast-bootstrap analysis (f option = a) with 100 pseudoreplicates. BI analyses were performed with PhyloBayes 3.3e (Lartillot et al. 2009) using the site-heterogeneous CAT + Γ 4 and CATGTR + Γ 4 models (Lartillot and Philippe 2004). To establish the appropriate conditions for these analyses, two independent chains were run for 10,000 cycles under the CAT + Γ 4 model and for 2,000 cycles under the CATGTR + Γ 4 model, and consensus topologies were calculated from the saved trees using the BPCOMP program of PhyloBayes after a burn-in of 2,000 and 500 cycles, respectively. Under these conditions, the largest discrepancy we observed across all bipartitions in the consensus topologies (maxdiff) was lower than 0.10, indicating that convergence between the two chains was achieved. For the bootstrap analyses, 100 pseudoreplicates were generated using the SEQBOOT program of the PHYLIP package (Felsenstein 1989), and BI chains were run on these pseudoreplicates using the conditions just described. In the CAT + Γ 4 analyses, one chain was thus run for 10,000 cycles (with each cycle sampled) for each pseudoreplicate and a consensus tree was computed with the READPB program of PhyloBayes after elimination of 2,000 burn-in trees. The same procedure was applied to each pseudoreplicate in the analysis done under the CATGTR + Γ 4 model, except that the chain was run for 2,000 cycles with a burn-in of 500 cycles. In both analyses, a bootstrap consensus tree was generated from the 100 resulting consensus trees using the CONSENSE program of the PHYLIP package.

Cross-validation tests were conducted to evaluate the fits of the four models of amino acid substitutions to the data set. They were carried out with PhyloBayes using ten randomly generated replicates. Cross-validation is a very general statistical method for comparing models. The procedure can be summarized as follows. The data set is randomly partitioned into two unequal subsets, the learning set (also called the training set) and the test set. The learning set serves to estimate the parameters of the model and these parameters are then used to compute the likelihood of the test set. To reduce variability, multiple rounds of cross-validation are performed using different partitions and the resulting log-likelihood scores (which measure how well the test sets were predicted by the model) are averaged over the rounds.

To analyze the amino acid composition of the data set, we first assembled a 20×25 matrix containing the frequency of each amino acid per species using the program Pepstats of the EMBOSS package (Rice et al. 2000). A correspondence analysis of this data set was then performed using the R package ca (Nenadic and Greenacre 2007).

Testing Robustness of Trees by Removal of Fast-Evolving Sites

The influence of removing increasing proportions of fast-evolving sites in the amino acid data set was investigated as

Table 1

General Features of the Streptophyte and Chlorophyte mtDNAs Compared in This Study

Taxon ^a	Name	Abbreviation	Accession	Size (bp)	A + T (%)	Genes	Introns		Repeats ^c (%)
							Group I	Group II ^b	
Charophyceae									
	<i>Mesostigma viride</i>	MESO	NC_008240	42,424	67.79	65	4	3 (2)	0.1
	<i>Chlorokybus atmophyticus</i>	CHLO	NC_009630	201,763	60.20	71	6	14	7.1
	<i>Entransia fimbriata</i>	ENTR	KF060941	61,645	57.30	66	3	2	0.0
	<i>Klebsormidium spec.</i> ^d	KLEB	KF060942	87,405	59.11	65	7	2	7.3
	<i>Chaetosphaeridium globosum</i>	CHAE	NC_004118	56,574	65.59	68	9	2	0.1
	<i>Chara vulgaris</i>	CHAR	NC_005255	67,737	59.10	69	14	13	2.6
	<i>Nitella hyalina</i>	NITE	NC_017598	80,183	59.17	69	7	13	1.1
	<i>Closterium baillyanum</i>	CLOS	KF060940	152,089	60.37	71	20	11	3.7
	<i>Roya obtusa</i>	ROYA	KF060943	69,465	73.17	69	0	2	3.0
Embryophyta									
	<i>Marchantia polymorpha</i>	MARC	NC_001660	186,609	57.59	70	7	25	6.9
	<i>Treubia lacunosa</i>	TREU	NC_016122	151,983	56.62	66	7	21	5.5
	<i>Pleurozia purpurea</i>	PLEU	NC_013444	168,526	54.63	68	7	21	5.0
	<i>Physcomitrella patens</i>	PHYS	NC_007945	105,340	59.43	66	3	24	1.0
	<i>Anomodon rugelli</i>	ANOM	NC_016121	104,239	58.80	66	3	24	1.0
	<i>Megaceros aenigmaticus</i> ^e	MEGA	NC_012651	184,908	53.99	42	0	28	2.0
	<i>Phaeceros laevis</i>	PHAE	NC_013765	209,482	55.40	43	0	32	2.5
	<i>Cycas taitungensis</i>	CYCA	NC_010303	414,903	53.08	64	0	25 (5)	24.9
	<i>Beta vulgaris</i>	BETA	NC_002511	368,801	56.14	50	0	20 (6)	22.1
	<i>Nicotiana tabacum</i>	NICO	NC_006581	430,597	55.04	57	0	23 (6)	17.6
	<i>Arabidopsis thaliana</i>	ARAB	NC_001284	366,924	55.23	49	0	23 (5)	11.6
	<i>Oryza sativa</i>	ORYZ	NC_007886	490,521	56.15	54	0	23 (6)	55.2
Chlorophyta									
	<i>Nephroselmis olivacea</i>	NEPH	NC_008239	45,223	67.20	66	4	0	0.1
	<i>Ostreococcus tauri</i>	OSTR	NC_008290	44,237	61.78	63	0	0	0.1
	<i>Monomastix sp.</i>	MONO	KF060939	60,883	59.87	65	8	0	4.8
	<i>Prototheca wickerhamii</i>	PROT	NC_001613	55,328	74.18	62	5	0	5.3

^aDifferent colors, also used in figures, denote the distinct charophycean and bryophyte lineages.

^bNumber of trans-spliced introns is given in parentheses.

^cNonoverlapping repeat elements were mapped on each genome with RepeatMasker using the repeats ≥ 30 bp identified with REPuter as input sequences.

^dThis species was formally called *Microspora stagnorum* (Mikhailiyuk et al. 2008).

^eThis species name has been changed to *Nothoceros* (Villarreal et al. 2010).

follows. Substitution rates among sites in the data set were estimated with CODEML in PAML 4.6 (Yang 2007) for the four alternative topologies observed for the streptophyte green algal lineages in bootstrap analyses with the CAT + $\Gamma 4$ and CATGTR + $\Gamma 4$ models. These rates were averaged for each site, and the fastest evolving sites were incrementally removed in 5% intervals using MESQUITE 2.75 (Maddison and Maddison 2011) in order to generate five subsets of data. RAXML analyses of these data subsets were performed as described above for the original amino acid data set. The five trimmed data sets were also analyzed with PhyloBayes under the CAT + $\Gamma 4$ and CATGTR + $\Gamma 4$ models. Two independent chains were run for 5,000 cycles (with each cycle sampled) with a burn-in of 1,000 cycles in each of the analyses under the CAT + $\Gamma 4$ model, whereas the two chains were run for 2,000 cycles with a burn-in of 500 cycles

in the analyses under the CATGTR + $\Gamma 4$ model. Consensus trees and posterior probability values were computed from saved trees after burn-in, using the BPCOMP program of PhyloBayes.

Bootstrap analyses of the trimmed data set lacking 20% of the fastest evolving sites were carried out with RAXML and PhyloBayes essentially as described for the original amino acid data set. For the BI analyses under the CAT + $\Gamma 4$ and CATGTR + $\Gamma 4$ models, we confirmed that the largest discrepancy across all bipartitions in the consensus trees obtained from two independent chains was lower than 0.10.

Saturation levels of the original amino acid data set and of the trimmed data set lacking 20% of the fastest evolving sites were estimated by computing the slopes of the regression plots of patristic distances versus observed distances using KaleidaGraph 4.1.3 (Synergy Software, Reading, PA). For

each data set, patristic distances were calculated using PATRISTIC 1.0 (Fourment and Gibbs 2006) from the branch lengths of the best tree inferred under the CATGTR + Γ 4 model, whereas the observed distances were derived from pair-wise comparisons between the raw sequences of the data set using MEGA 5 (Tamura et al. 2011).

Results

The members of the Klebsormidiales we selected for mtDNA sequencing, *Entransia* and *Klebsormidium*, represent distinct major lineages (Karol et al. 2001; Mikhailiyuk et al. 2008). In the Zygnematales, we also sampled two distantly related species, *Closterium* and *Roya*. Although *Closterium* is known to belong to the Desmidiaceae, it remains unclear whether *Roya* is part of the other major lineage recognized in the Zygnematales (McCourt et al. 2000; Gontcharov et al. 2004; Hall et al. 2008). Concerning the phylogenetic positions of the members of the Charales whose mtDNAs were compared here, it has been shown that the *Chara* and *Nitella* genera are part of separate major lineages (Karol et al. 2001).

Comparison of Structural Genomic Features

The gene maps of the klebsormidialean and zygnematalean mitochondrial genomes are shown in [supplementary figures S1 and S2, Supplementary Material](#) online, respectively. The structural features of these four newly sequenced genomes were compared with those previously observed for five charophyceans and seven bryophytes (table 1). The latter land plants represent all three known lineages of bryophytes and consist of three liverworts (*Marchantia polymorpha*, *Trebisia lacunosa*, and *Pleurozia purpurea*), two hornworts (*Megaceros aenigmaticus* and *Phaeceros laevis*), and two mosses (*Physcomitrella patens* and *Anomodon rugelii*). For certain analyses, we have also included representatives of chlorophytes and seed plants (table 1).

Genome Size and Gene Content

All compared charophycean mitochondrial genomes, except the mtDNAs of *Chlorokybus* and of the zygnematalean *Closterium*, are smaller in size than their bryophyte counterparts (table 1 and fig. 1). Important differences in mitochondrial genome size are found among distinct charophycean lineages as well as within individual lineages, in particular the *Mesostigma/Chlorokybus* clade and the Zygnematales. In this regard, it is noteworthy that the two representatives of the Charales, *Chara* and *Nitella*, carry very similar mitochondrial genomes; these mtDNAs, which differ by only 12,447 bp, contain the same gene complement and display exactly the same gene order. Overall, the total lengths of both the intergenic and intronic regions in mtDNA changed substantially during charophycean evolution, even though the total lengths of coding regions remained similar (fig. 1). Repeated

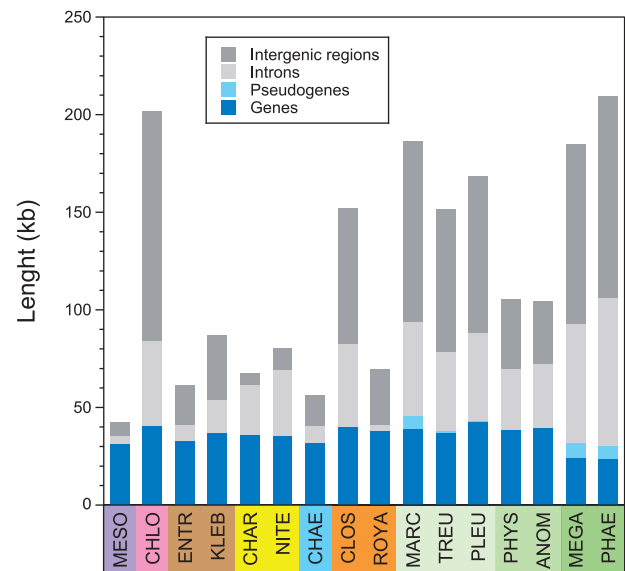


Fig. 1.—Total lengths of coding, intronic, and intergenic sequences in the streptophyte mtDNAs examined in this study. Note that intron-encoded genes were not considered as coding sequences but rather as intron sequences and that *trans*-spliced introns were not taken into account to estimate the total lengths of intron sequences because their sizes are usually not annotated in the GenBank accessions. Species names are abbreviated as in table 1.

sequences account for less than 7.5% of the total mitochondrial genome size in charophyceans, the observed variation being comparable to that found in bryophyte mtDNAs (table 1).

Among the 16 charophycean and bryophyte mtDNAs examined in this study, the total number of standard genes ranges from 42 (in the hornwort *Megaceros*) to 71 (in *Chlorokybus* and the zygnematalean *Closterium*), with most genomes exhibiting between 65 and 71 genes (table 1). As in other charophycean mtDNAs, all protein-coding genes in the newly sequenced genomes were annotated using the standard genetic code, thus providing no evidence for RNA editing. The few genes that were lost from the mitochondrial genome during charophycean evolution code mainly for tRNAs (*trn* genes), ribosomal proteins (*rps* and *rpl*), and subunits involved in cytochrome *c* biogenesis (*yej*, also designated as *ccm*) (fig. 2). In land plant mtDNAs, the same categories of genes are prone to extinction (see fig. 2) (Adams et al. 2002; Knoop 2013). Certain mitochondrial genes were lost in a lineage-specific fashion, whereas others were lost in two or more lineages. Lineage-specific gene losses occurred at highest frequency in the Klebsormidiales and in the hornworts. The genes whose absence was recorded in multiple lineages are of special interest as they can be phylogenetically informative. In this context, we note that *trnI(gau)* is uniformly absent from the three bryophyte lineages and the Zygnematales, whereas *rpl14* is missing from the bryophyte lineages, the Zygnematales and the Coleochaetales.

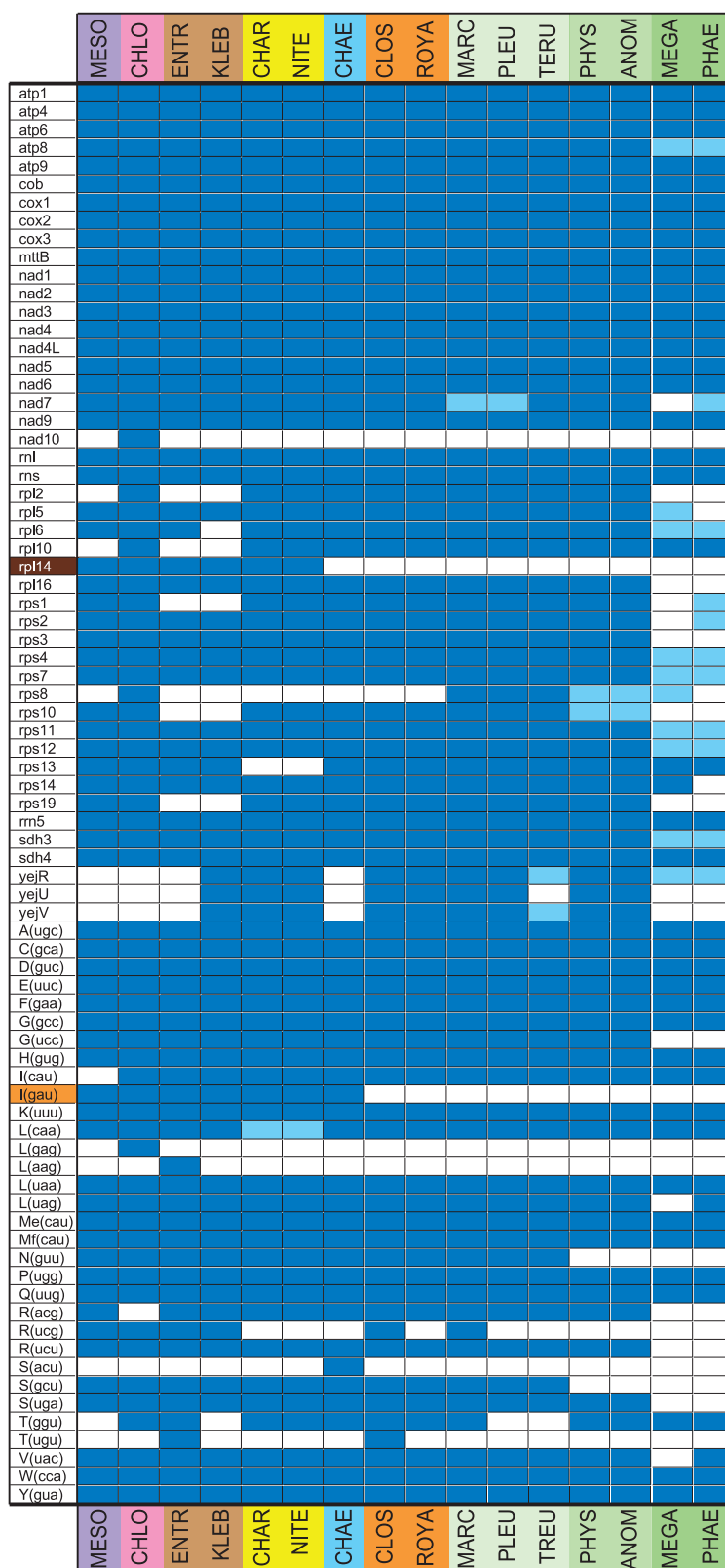


Fig. 2.—Gene repertoires of the streptophyte mitochondrial genomes examined in this study. The presence of a standard gene is indicated by a dark blue box and the presence of a pseudogene by a light blue box. The different colors on the left of the figure refer to gene distributions supporting distinct hypotheses concerning the sister group of land plants: orange, Zygnematales; brown, Coleochaetales + Zygnematales.

Two tRNA genes not previously identified in streptophyte mitochondria, *trnL(aag)* and *trnT(ugu)*, were annotated in the course of this study. The *trnT(ugu)* gene is present in the klebsormidialean *Entransia* and the zygneatalean *Closterium*, whereas *trnL(aag)* is found only in *Entransia*. Given that these mitochondrial genes have been described so far in only a few chlorophytes and are also rarely found or absent in other algal groups, they were almost certainly not inherited vertically from the common ancestor of all streptophytes. Our similarity searches against the nonredundant database of the National Center for Biotechnology Information using *Entransia trnL(aag)* as the query sequence failed to reveal any tRNA gene sharing substantial sequence identity. However, both the *Entransia* and *Closterium trnT(ugu)* genes were found to exhibit high sequence similarities with *trnT(ugu)* genes originating from the chloroplasts of red algae and glaucocystophytes. Two other tRNA genes showing a restricted distribution in streptophytes and other algal groups, *trnS(acu)* et *trnR(ucg)*, have presumably not been inherited through vertical inheritance from the last common ancestor of streptophytes. Our similarity searches suggest that *trnS(acu)*, documented uniquely in *Chaetosphaeridium* mtDNA, originates in part from the mitochondrial *trnT(ggu)*, as we detected substantial sequence similarity between the 3' portions of these two genes from *Chaetosphaeridium*. The evolutionary history of *trnR(ucg)* is more complex, given the evidence that it arose in chlorophyte and streptophyte mtDNAs through duplication and mutation of either *trnR(acg)* (in the *Mesostigma* and *Nephroselmis* lineages) or *trnR(ucu)* (in the *Chlorokybus* and *Marchantia* lineages) (Oda et al. 1992; Turmel et al. 2003; Wang et al. 2009). Our analyses showed that the mitochondrial *trnR(ucg)* genes of the two klebsormidialeans and the zygneatalean *Closterium* are highly similar to one another and also share high similarity with the mitochondrial *trnR(acg)* gene.

In the *Klebsormidium* and *Roya* mtDNAs, we discovered potential coding sequences that are not usually found in green plant mitochondrial genomes. The *orf209* of *Klebsormidium*, which is embedded within a 7.1-kb region containing only three tRNA genes (supplementary fig. S1, Supplementary Material online), shows strong sequence homology ($E = e^{-50}$) with bacterial and phycodnavirus genes encoding C-5 cytosine-specific DNA methyltransferases. The nonstandard gene (*orf235*) in the *Roya* mtDNA also resides in a relatively long intergenic region (supplementary fig. S2, Supplementary Material online) and its predicted product is related to phage integrases/recombinases ($E = 1e^{-09}$) and to putative proteins encoded by the mtDNAs of *Chaetosphaeridium* (*orf202*, $E = 9e^{-31}$) and the chlorophyte *Prototheca wickerhamii* (*orf304*, $E = 4e^{-24}$). In both the *Roya* and *Chaetosphaeridium* mtDNAs, the putative integrase/recombinase gene is located downstream of the large subunit rRNA gene (*rrl*).

Intron Content

In contrast to their land plant counterparts, charophycean mtDNAs are extremely variable in intron content (table 1 and figs. 3 and 4). All the introns identified so far in charophycean and bryophyte genomes, except for two group II introns in the tightly packed mtDNA of *Mesostigma viride*, are *cis*-spliced. Charophycean group I and group II introns lie at 37 and 44 distinct genomic sites (in 8 and 24 genes), respectively. Even within individual charophycean lineages, there are sharp differences in intron content. For instance, within the Zygnematales, the *Closterium* and *Roya* mtDNAs harbor as many as 31 introns and as few as two introns, respectively; the latter, which are both inserted into tRNA genes and belong to the group II family, are conserved between the two zygneatalean taxa and are also present in other charophycean or bryophyte lineages. Within the Klebsormidiales, the five *Entransia* and nine *Klebsormidium* introns share no homolog, although similar introns reside at identical insertion sites in other charophycean mtDNAs. Moreover, the mtDNAs of the charalean *Chara* (27 introns) and *Nitella* (20 introns), both of which contain numerous group I and group II introns, have in common a high proportion of group II introns but share only two group I introns at the same sites.

In a previous study (Turmel et al. 2003), we found that four group II introns within *cox2* (site 71), *nad3* (site 140), *nad4* (site 949), and *rps3* (site 77) are shared between *Chara* and some land plant lineages, including bryophytes (fig. 4). The expanded data reported here confirm that these sites are occupied by introns that are shared solely by members of the Charales (three of the four introns are present in both *Chara* and *Nitella*) and one or more land plant lineages. In addition, we identified an extra site in *cox2* (site 343) that is occupied by a group II intron shared between *Nitella*, mosses, hornworts, and some seed plants. Moreover, a group II intron at another site in *cox2* (site 220) as well as a group I intron in *nad5* (site 723) were found only in the zygneatalean *Closterium* and some bryophyte lineages (figs. 3 and 4).

Gene Order

To compare the gene organizations of the charophycean and bryophyte mtDNAs in our study group, we examined the gene pairs and gene clusters they share and also inferred a scenario of genome rearrangements. Figure 5 presents one of the two gene-pair analyses that we performed. The mtDNAs of four green algae representing early diverging lineages of the Chlorophyta (i.e., three members of the Prasinophyceae and *Prototheca wickerhamii*, a representative of the Trebouxiophyceae) were included in this analysis in an attempt to distinguish conserved gene clusters that predate the divergence of chlorophytes and streptophytes (i.e., ancestral clusters) from those that arose during charophycean evolution. The ancestral gene pairs appear to be more prevalent in the

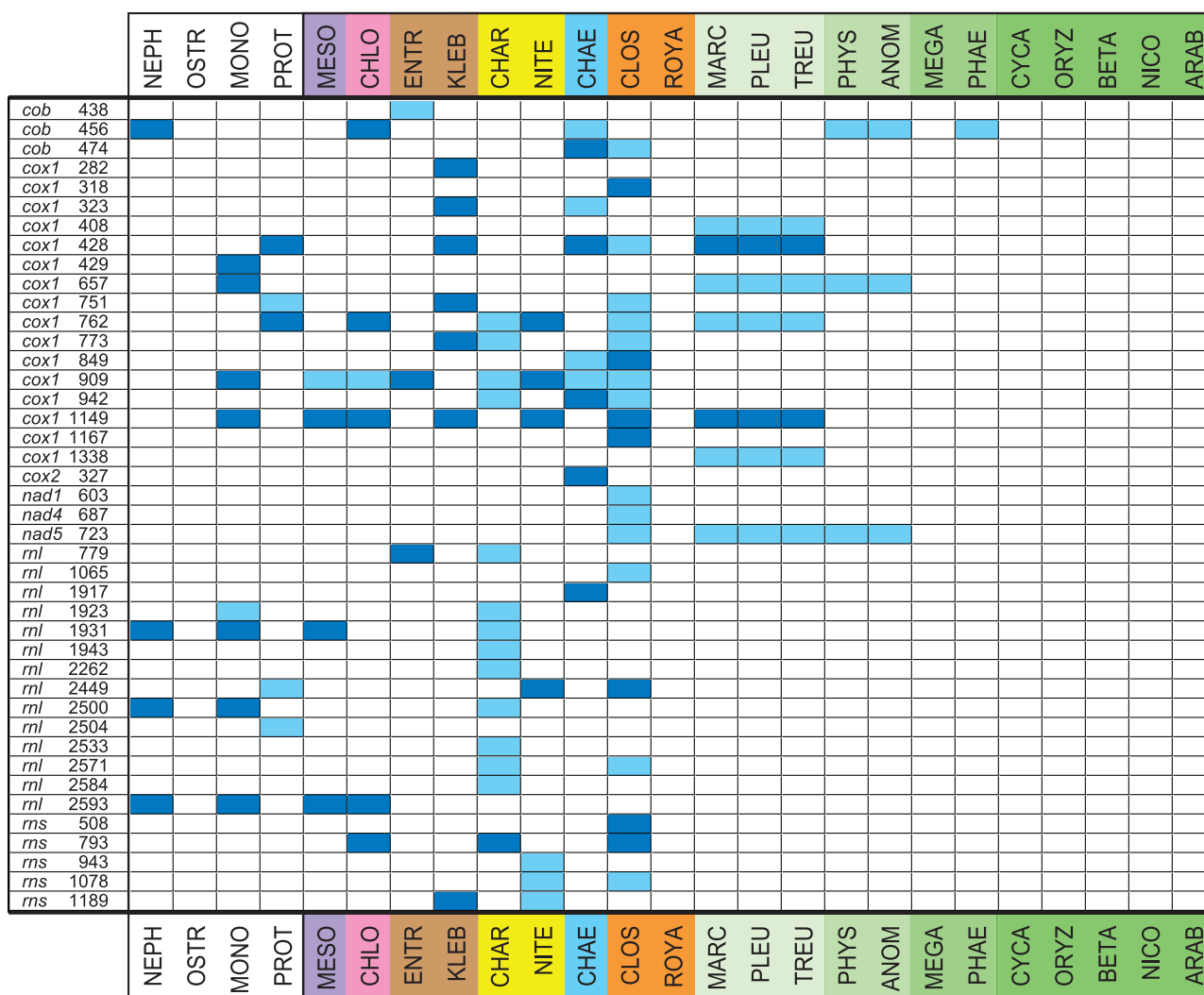


Fig. 3.—Distribution of group I introns among the chlorophyte and streptophyte mtDNAs examined in this study. The presence of an intron containing an ORF is denoted by a dark blue box, whereas the presence of an intron lacking an ORF is denoted by a lighter blue box. Intron insertion sites in protein-coding and tRNA genes are given relative to the corresponding genes in *Mesostigma* mtDNA; insertion sites in *rns* and *rnl* are given relative to the *Escherichia coli* 16S and 23S rRNAs, respectively. For each insertion site, the position corresponding to the nucleotide immediately preceding the intron is reported.

Klebsormidiales and Charales than in the other streptophyte lineages. With regard to the gene pairs that are streptophyte-specific, they are most abundant in the Zygnematales and Charales and their proportions in these lineages are comparable to those observed in liverworts and mosses. We identified the streptophyte-specific gene pairs (synapomorphies) that are shared exclusively between the Charales and bryophytes; between the Coleochaetales, Zygnematales, and bryophytes; and also between the Zygnematales and bryophytes in order to evaluate which of the three current hypotheses concerning the identity of the charophycean lineage(s) being sister to land plants is best supported (fig. 5, compare the gene pairs shown in different colors). The hypothesis that the Coleochaetales + Zygnematales are sister to the land plant lineage received the strongest support. In this context, it is worth

pointing out that several gene pairs are uniquely shared between the Charales, Zygnematales, and land plants (fig. 5), thus providing support for a basal position of the Coleochaetales relative to these lineages. In a separate analysis, we assembled a data set of 134 gene pairs from 16 streptophytes (128 phylogenetically informative characters), in which the gene pairs were coded as binary Dollo characters (presence/absence), and used this data set to infer a phylogeny using MP (fig. 6). The majority-rule consensus tree of the bootstrap replicates resolved with weak bootstrap (BP) support the Coleochaetales and Zygnematales as sister to land plants and the Charales as sister to the Coleochaetales + Zygnematales + land plants.

Conserved gene pairs in streptophyte mtDNAs form larger conserved clusters whose numbers, sizes, and arrangements

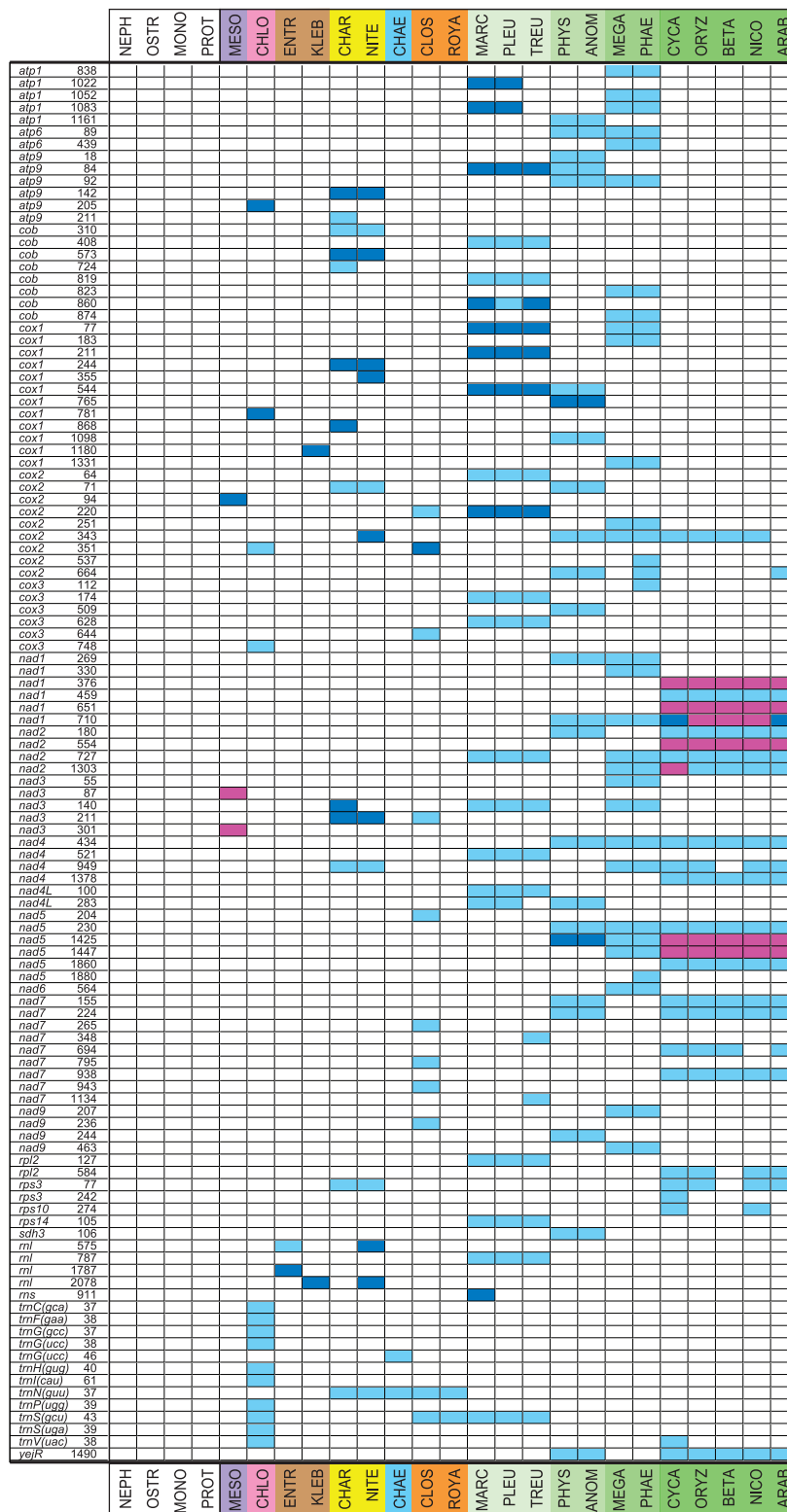


Fig. 4.—Distribution of group II introns among the chlorophyte and streptophyte mtDNAs examined in this study. The presence of a *cis*-spliced intron containing an ORF is denoted by a dark blue box, whereas the presence of a *cis*-spliced intron lacking an ORF is denoted by a lighter blue box. *Trans*-spliced introns are represented by purple boxes. Intron insertion sites in protein-coding and tRNA genes are given relative to the corresponding genes in *Mesostigma* mtDNA; insertion sites in *ms* and *rnl* are given relative to the *Escherichia coli* 16S and 23S rRNAs, respectively. For each insertion site, the position corresponding to the nucleotide immediately preceding the intron is reported.

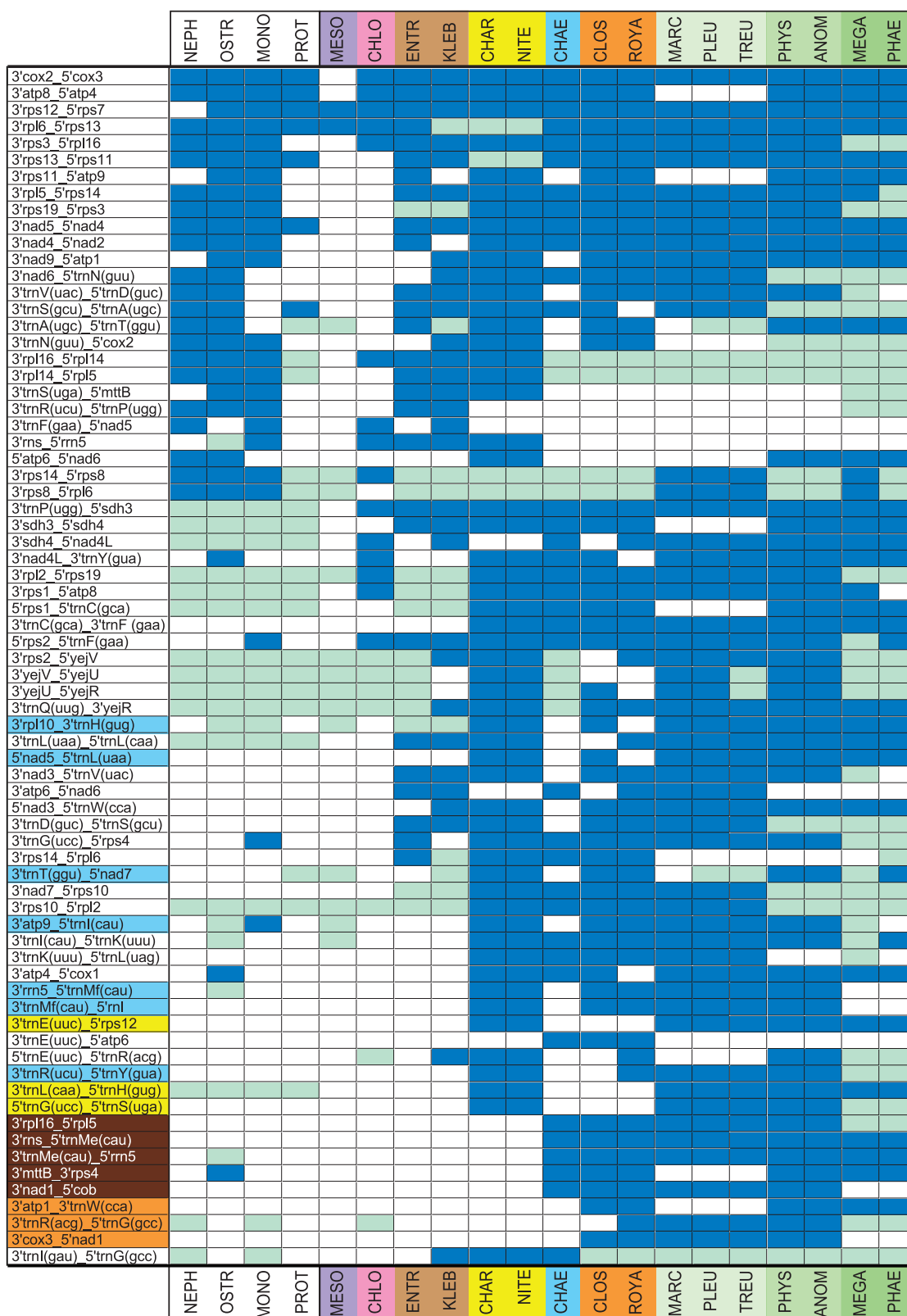


Fig. 5.—Distribution of mitochondrial gene/pseudogene pairs that are shared by at least two charophycean and two chlorophyte taxa, by at least three charophycean taxa, or by at least one charophycean and two bryophyte lineages. The presence of a gene pair is denoted by a dark blue box. A light green box refers to a gene pair in which at least one gene is missing in some lineages due to gene loss. Ancestral gene pairs, that is, gene pairs common to

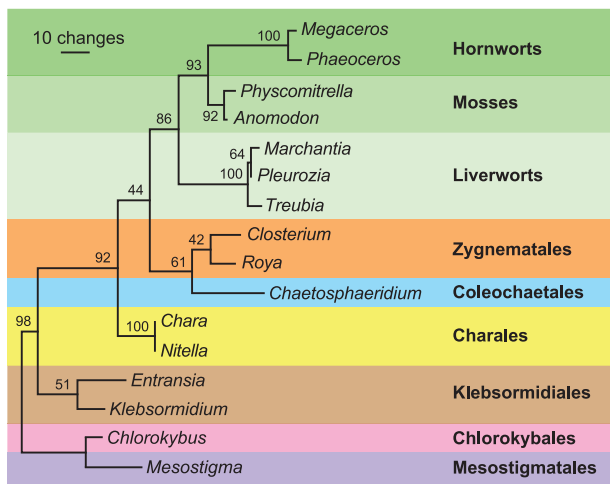


Fig. 6.—Majority-rule consensus tree inferred by MP analysis using a data set of 134 gene pairs. BP support values are indicated on the nodes.

differ according to the lineage examined. The gene clusters in the *Chara/Nitella*, *Roya*, *Closterium*, and *Chaetosphaeridium* mtDNAs that are conserved with the mtDNA of the moss *Physcomitrella* are highlighted in [supplementary figure S3, Supplementary Material](#) online. It can be seen that the mtDNAs of *Chara/Nitella* show the highest level of synteny with the *Physcomitrella* genome, as they feature both the minimal number of gene clusters and the maximal number of genes in these clusters. A phylogeny based on inversion medians was inferred from a data set of 55 gene positions from eight streptophytes using GRAPPA and the algorithm of Caprara (which minimizes the number of reversals) ([supplementary fig. S4, Supplementary Material](#) online). Note here that inversions might not be the predominant mode of rearrangements for the mtDNAs we examined, as models involving duplication and deletion have been proposed to explain structural variation for the complex mitochondrial genomes of two angiosperms, maize (Darracq et al. 2010) and *Brassica* (Chang et al. 2011). No representative of the hornworts was included in our analysis because too many mitochondrial genes were lost in this bryophyte lineage (see [fig. 2](#)) and only the genes common to all taxa can be analyzed using GRAPPA. We recovered a single best tree with 109 inversion medians. This tree shows that the two zygnematalean genomes are the most similar to their bryophyte counterparts; however, they

do not form a clade. The branch corresponding to the mtDNAs of the Charales is basal relative to those corresponding to the mtDNAs of the Coleochaetales and Zygnematales.

Phylogenetic Analyses of Sequence Data

We assembled an amino acid data set of 9,013 unambiguously aligned positions from 40 mitochondrial protein-coding genes of 21 streptophytes and four chlorophytes. The chlorophyte sequences served as outgroup to root the trees. The data set was analyzed by ML and BI methods using the site-homogeneous LG + Γ 4 and GTR + Γ 4 models of amino acid substitutions (ML) and the site-heterogeneous CAT + Γ 4 and CATGTR + Γ 4 models (BI). The fits of these models of sequence evolution to the data set of 9,013 positions were assessed using cross-validation ([supplementary table S1, Supplementary Material](#) online). As expected, the site-heterogeneous models proved to be superior to the site-homogeneous models. The CATGTR + Γ 4 model is the model that best fits the data set, whereas the CAT + Γ 4 model has a better fit than the GTR + Γ 4 and LG + Γ 4 models. Considering that the GTR + Γ 4 model has a better fit than the LG + Γ 4 model, it appears that the size of the data set is sufficiently large to estimate a GTR amino acid substitution matrix that models more accurately our data than the empirical LG matrix. We analyzed the amino acid composition of the data set of 9,013 positions because heterogeneity in amino acid composition may violate the stationarity assumption made by most models of sequence evolution (Lockhart et al. 1994; Gowri-Shankar and Rattray 2006). No large deviations in amino acid composition of mitochondrial proteins were found among the taxa examined ([supplementary fig. S5, Supplementary Material](#) online).

Regardless of the method of analysis and the model of sequence evolution used, we observed with high BP support that the *Mesostigma* + *Chlorokybus* clade occupies the first branch of the Streptophyta and that the clade formed by the two klebsormidialean green algae represents the second deepest divergence ([fig. 7](#)). In contrast, the relationships among the Coleochaetales, Zygnematales, and Charales were influenced by the model used. The ML analyses under the LG + Γ 4 and GTR + Γ 4 models as well as the BI phylogeny under the CAT + Γ 4 model favored the placements of the Charales as sister to the land plants and of the Zygnematales as sister to the Charales + land plants clade (T1 topology). On the other hand, the BI analysis done

Fig. 5.—Continued

chlorophytes and streptophytes, are presented first. The yellow, orange, and brown colors on the left of the figure refer to derived gene pairs (synapomorphies) supporting the three current hypotheses concerning the sister group of land plants: yellow, Charales; orange, Zygnematales; brown, Coleochaetales + Zygnematales. The blue color on the left of the figure refers to the gene pairs supporting the alternative hypothesis that the Coleochaetales are sister to the Charales, Zygnematales, and land plants. Note that the 3'yejV–5'yejU pair was considered to be an ancestral linkage because it is found in the few protist lineages that have retained these genes, more specifically in the bacteria-like and gene-rich mitochondrial genomes of jakobids (Burger et al. 2013).

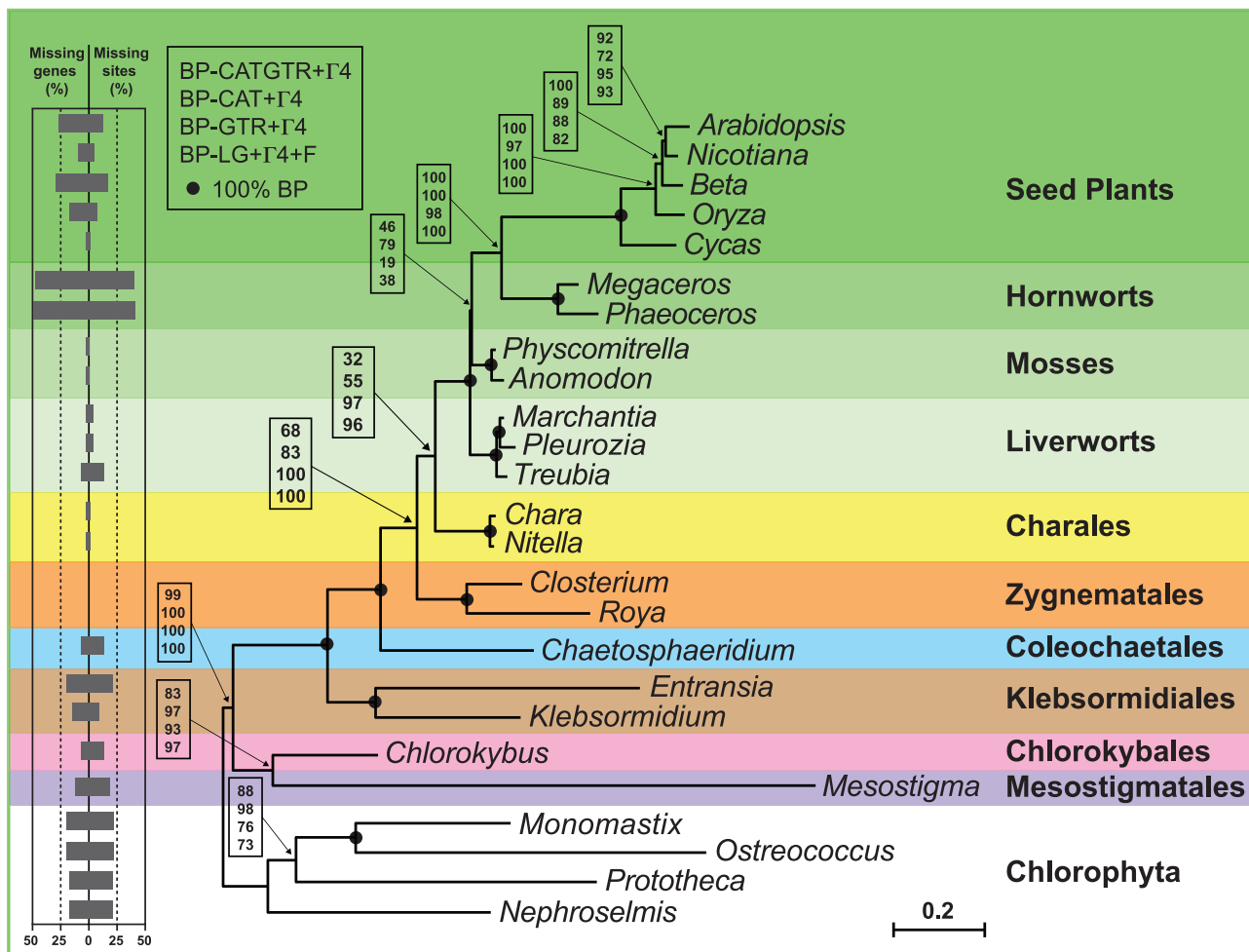


Fig. 7.—Majority-rule consensus tree inferred by PhyloBayes under the CAT + Γ 4 model using a data set of 9,013 positions assembled from 40 mtDNA-encoded proteins. BP support values are reported on various nodes: from top to bottom are shown the values for the PhyloBayes CATGTR + Γ 4 and CAT + Γ 4 analyses and for the RAxML GTR + Γ 4 and LG + Γ 4 analyses. Black dots indicate that the corresponding branches received 100% BP support in all four analyses. The histograms on the left indicate the proportion of missing data for each taxon. The scale bar denotes the estimated number of amino acid substitutions per site.

under the CATGTR + Γ 4 model favored a sister group relationship between the Zygnematales and land plants and placed the Charales just before the divergence of the Zygnematales (T2 topology), albeit BP support for these positions was weak. Use of the best-fitting model not only lowered the frequency with which the T1 topology was detected among the bootstrap pseudoreplicates but also increased the frequencies of the alternative topologies T3 and T4 in which the Charales diverged earlier than the Coleochaetales and Zygnematales (table 2). The sister group to all land plants is the Coleochaetales + Zygnematales clade in T3, whereas it is the Zygnematales in T4. Regarding the relationships among land plants, the method of analysis used was found to influence the branching order of the three major bryophyte lineages. Consistent with previously reported phylogenies inferred from multiple genes (Qiu et al. 2006), the liverworts are sister to all

other land plants in the mitochondrial trees constructed with the BI method (CATGTR + Γ 4 and CAT + Γ 4 models); however, both the liverworts and the mosses are united in the same clade in the ML trees (GTR + Γ 4 and LG + Γ 4 models).

We tested whether eliminating various proportions of the fastest evolving sites (i.e., the sites prone to yield a non-phylogenetic signal) from the amino acid data set has any effect on the robustness of the consensus tree and on the frequencies of the alternative topologies recovered. Five trimmed data sets, in which the sites with the highest rates of sequence evolution were deleted by 5% increment, were analyzed using the ML (LG- Γ 4 model) and BI (CAT + Γ 4 and CATGTR + Γ 4 models) methods. As illustrated in figure 8, removal of the fastest evolving sites (5–25%) markedly lowered support for the placement of the Charales as sister to all land plants (i.e., the T1 topology) and concomitantly increased support for

Table 2

Alternative Topologies Recovered among 100 Bootstrap Pseudoreplicates in BI Analyses of the Amino Acid Data Sets of 9,013 and 7,210 Positions under the CAT + Γ 4 and CATGTR + Γ 4 Models

Topology		Data Set of 9,013 Positions		Data Set of 7,210 Positions	
		CAT + Γ 4	CATGTR + Γ 4	CAT + Γ 4	CATGTR + Γ 4
T1	((Meso,Chlo),(Kleb,(Cole,(Zygn,(Char,LP))))))	55	32	33	21
T2	((Meso,Chlo),(Kleb,(Cole,(Char,(Zygn,LP))))))	25	34	64	59
T3	((Meso,Chlo),(Kleb,(Char,((Coleo,Zygn),LP))))	17	24	3	12
T4	((Meso,Chlo),(Kleb,(Char,(Coleo,(Zygn,LP))))))	2	8	0	7
Unresolved topologies		1	2	0	1

NOTE.—Meso, Mesostigmatales; Chlo, Chlorokybales; Kleb, Klebsormidiales; Cole, Coleochaetales; Zygn, Zygnematales; Char, Charales; LP, land plants.

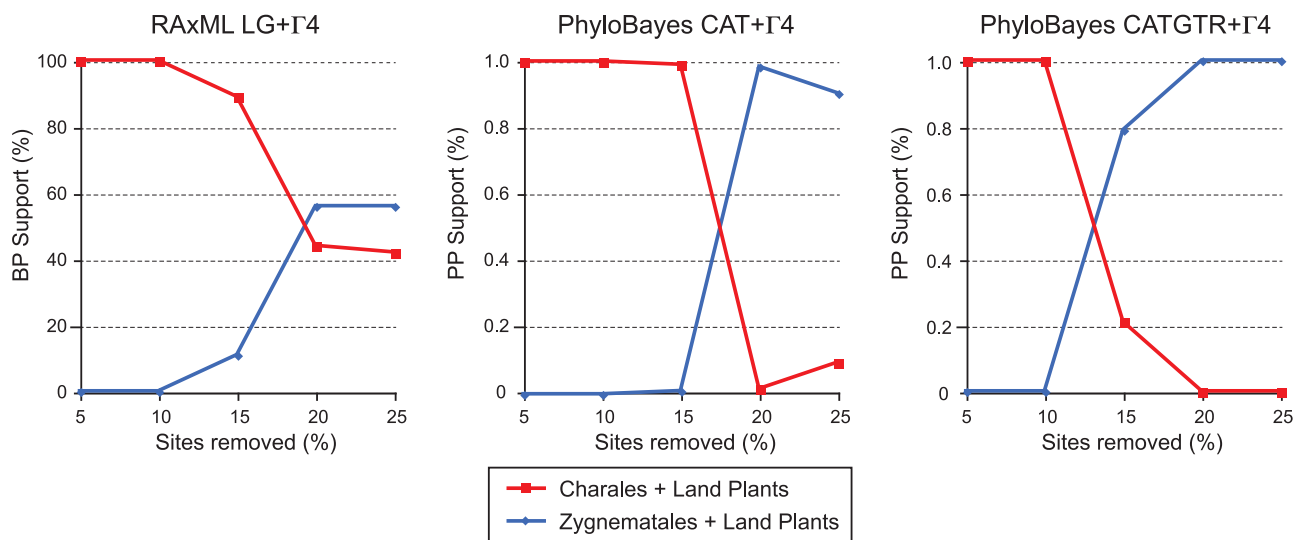


Fig. 8.—Changes in statistical support for the sister group of land plants in ML (RAxML LG + Γ 4) and BI (PhyloBayes CAT + Γ 4 and CATGTR + Γ 4) analyses when increasing proportions of the fastest evolving sites were removed from the data set of 9,013 amino acid positions.

the sister group relationship of the Zygnematales to the land plants (i.e., T2 topology). For the data set lacking 20% of the fastest evolving sites, we also showed that it is much less saturated with substitutions than the original data set (supplementary fig. S6, Supplementary Material online). This reduced data set of 7,210 positions was further analyzed by bootstrapping using the ML and BI methods (fig. 9). The consensus trees derived from all four analyses uncovered identical relationships among the charophycean lineages, that is, the relationships corresponding to the T2 topology. The sister group relationship of the Zygnematales to the land plants received weak BP support (56–66%); however, the better fitting was the model of sequence evolution, the higher was the observed support level for this grouping. It is also worth noting that, as the fit of the model improved, the clade uniting the Charales, Zygnematales, and land plants received lowered support due to the increased frequencies of the T3 and T4 topologies (table 2).

Discussion

Prior to this study, the mtDNA of *Chara* was known to bear more resemblance to its bryophyte counterparts than the mtDNAs of the Mesostigmatales/Chlorokybales and Coleochaetales lineages (Turmel et al. 2003, 2007a). Although this observation is consistent with the view that the Charales are sister to land plants, it conflicts with a chloroplast phylogeny based on the small and large subunit rRNA genes (Turmel, Ehara et al. 2002) and with chloroplast and nuclear phylogenomic analyses placing the Charales before the divergence of the Zygnematales and Coleochaetales (Turmel et al. 2006, 2007b; Wodniok et al. 2011; Laurin-Lemay et al. 2012; Timme et al. 2012). In the present investigation, we have examined, for the first time, mtDNAs from representatives of the Klebsormidiales and Zygnematales to determine whether the branching order recovered in the latter studies is supported by the mitochondrial genome and also to gain insights into the dynamics of mtDNA evolution

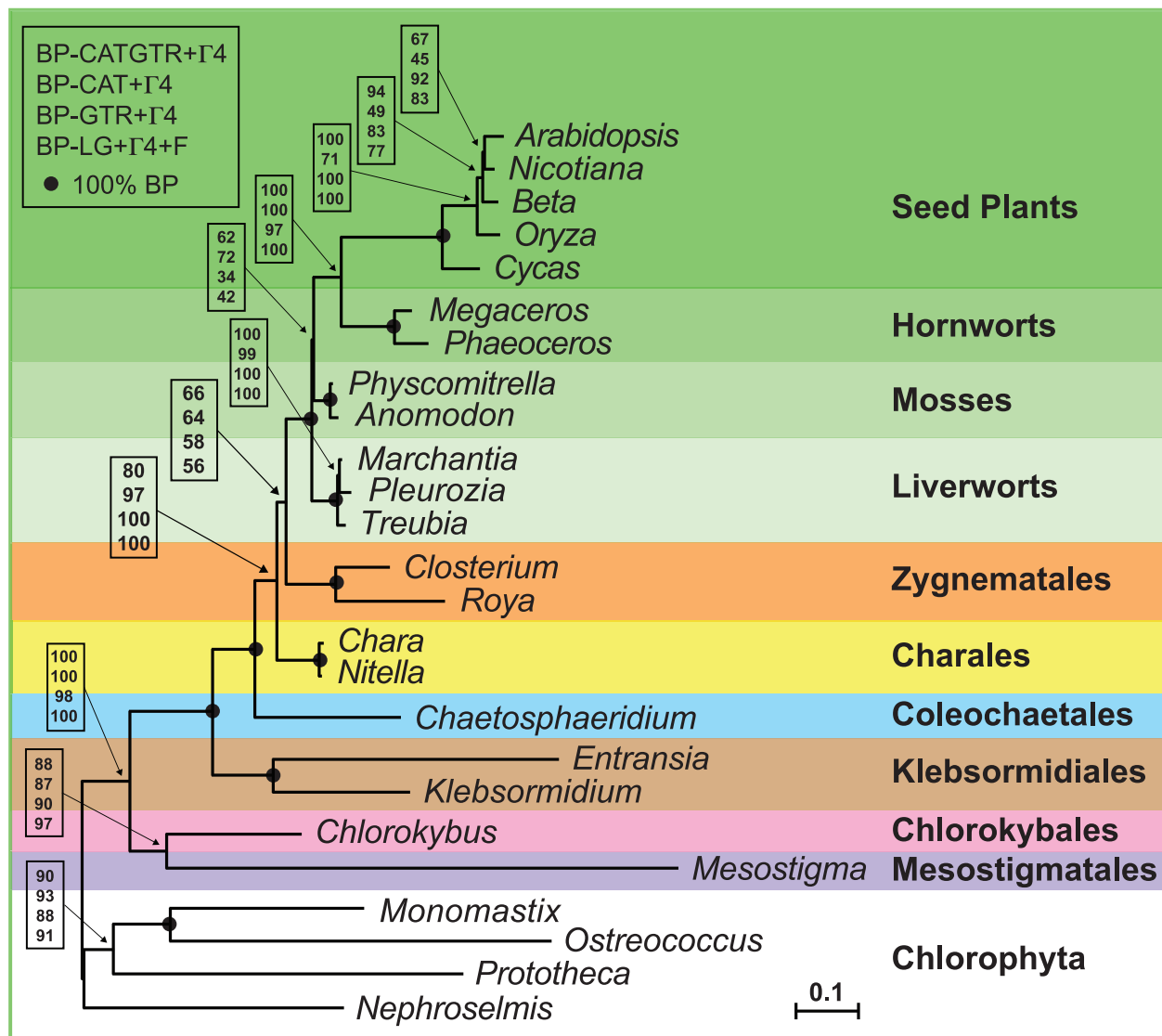


FIG. 9.—Majority-rule consensus tree inferred by PhyloBayes under the CATGTR + Γ 4 model using a data set of 7,210 positions assembled from 40 mtDNA-encoded proteins. This data set was generated from the data set of 9,013 positions by removing 20% of the fastest evolving sites. BP support values are reported on various nodes: from top to bottom are shown the values for the PhyloBayes CATGTR + Γ 4 and CAT + Γ 4 analyses and for the RAxML GTR + Γ 4 and LG + Γ 4 analyses. Black dots indicate that the corresponding branches received 100% BP support in all four analyses. The scale bar denotes the estimated number of amino acid substitutions per site.

during the diversification of charophyceans and the transition from green algae to land plants.

Phylogenetic Relationships among Charophyceans and Interrelationships with Land Plants

At the gene content and gene order levels, the mtDNAs of the zygnemataleans *Closterium* and *Roya* closely resemble both their charalean and bryophyte homologs, making it difficult to identify at a glance which of these two green algal groups has the most affinity with land plants. Even though the *Chara* and *Nitella* mitochondrial genomes appear to be the least

rearranged relative to bryophyte mtDNAs in comparisons of shared gene clusters (supplementary fig. S3, Supplementary Material online), the distribution patterns of *rp14* and *trn(gau)* (fig. 2), both analyses of gene pairs (figs. 5 and 6) as well as the best scenario of gene rearrangements inferred with GRAPPA (supplementary fig. S4, Supplementary Material online) lend support to the hypothesis that the Zygnematales or a clade consisting of the Zygnematales and Coleochaetales are sister to land plants. Taken altogether, the gene content and gene order data reported here offer ample evidence that there exist close relationships among the Charales, Coleochaetales, and Zygnematales.

Overall, our phylogenetic analyses of two mitochondrial data sets derived from the same alignments of 40 proteins using two inference methods and four alternative evolutionary models failed to resolve unambiguously the relationships among the Charales, Coleochaetales, and Zygnematales (figs. 7 and 9). The Charales were robustly recovered as sister to land plants in the analyses of the data set of 9,013 positions under the site-homogeneous models (fig. 7), but under the most complex and best-fitting model (the site-heterogeneous CATGTR + Γ 4 model), the Zygnematales or the Zygnematales + Coleochaetales clade were recovered as sister to land plants with weak BP support (42% and 24%, respectively; see table 2). Systematic errors stemming from the difficulties to detect and interpret multiple substitutions most likely caused the Charales to affiliate with land plants when suboptimal models of amino acid substitutions were used. Systematic errors are common in phylogenomic studies with sparse taxon sampling and also when some of the taxa produce long branches (Philippe et al. 2011b); in trees inferred with ML, these long branches tend to associate erroneously with those of other taxa showing high sequence divergence, yielding the well-known, long-branch attraction artifact (Felsenstein 1978). Our analyses may have been hampered by this long-branch attraction phenomenon, as the branches leading to the Zygnematales, *Chaetosphaeridium*, the Klebsormidiales, and *Mesostigma* are longer than those leading to the Charales, liverworts, and mosses. As reported in a number of studies (e.g., Philippe et al. 2011a, 2011b), our results demonstrate that the impact of non-phylogenetic signal can be substantially reduced by using the most realistic model of sequence evolution. Another approach for reducing the misleading effect of nonphylogenetic signal is to remove the fastest evolving sites from the alignment (Delsuc et al. 2005; Rodriguez-Ezpeleta et al. 2007). This approach proved effective in our study: when 20% of the fastest evolving sites were deleted from the data set of 9,013 positions, we observed that, irrespective of the model of sequence evolution used, one of the topologies displaying the sister group relationship between the Zygnematales and land plants (T2) became predominant among the bootstrap pseudoreplicates, whereas the frequency observed for the topology showing the Charales as sister to land plants (T1) decreased concomitantly (table 2). In light of chloroplast and nuclear phylogenomic studies (Turmel et al. 2006, 2007b; Wodniok et al. 2011; Laurin-Lemay et al. 2012; Timme et al. 2012), the Charales would be expected to be basally positioned relative to the Coleochaetales, Zygnematales, and land plants (T3 or T4); however, in all analyses of the reduced data set, this placement received less support than that showing the Charales as sister to the Zygnematales and land plants (table 2), suggesting that *Chaetosphaeridium* is still attracted to the other taxa at the base of the tree. Improving taxon sampling by including additional representatives of the Klebsormidiales, Coleochaetales, and Zygnematales should be helpful to infer

more precisely multiple substitutions and hence resolve the correct relationships among the Charales, Coleochaetales, and Zygnematales, provided that sufficient phylogenetic signal has been retained in the mitochondrial genome during the radiation of these lineages.

Therefore, taken altogether, our analyses of mitochondrial gene content, gene order, and amino acid sequences concur with previously reported chloroplast and nuclear phylogenomic studies (Turmel et al. 2006, 2007b; Wodniok et al. 2011; Laurin-Lemay et al. 2012; Timme et al. 2012) in favoring the notion that the Charales are not the closest relatives of land plants. The only mtDNA data reported here that do favor the Charales as sister to land plants are the distributions we observed for the introns uniquely shared between charophytes and land plants (fig. 4). However, introns cannot be regarded as reliable phylogenetic indicators because they are often mobile and can be lost frequently (see next section).

Evolutionary Trends of mtDNA During Charophycean Evolution and the Transition to Land Plants

We concluded in our report on the comparison of the *Mesostigma* and *Chlorokybus* mtDNAs—two genomes differing extensively in size, gene order, and intron content—that these DNAs “exemplify the marked fluidity of the mitochondrial genome in streptophyte green algae” (Turmel et al. 2007a). It was then thought that the mitochondrial genome remained compact during charophycean evolution and expanded near the time the first land plants appeared and later when the seed plants evolved (Li et al. 2009; Liu et al. 2011). The four newly sequenced genomes we report here for representatives of the Klebsormidiales and of the Zygnematales as well as our comparison of the *Nitella* and *Chara* mtDNAs now provide us with a better understanding of the dynamics of mitochondrial genome evolution in other charophycean lineages, throwing also further light into the events that occurred during the radiation of charophytes and the origin of land plants. Our comparative analyses indicate that the charophycean mitochondrial genome experienced structural changes mainly at the levels of gene order, intron content, and sizes of intergenic regions, with the rate of changes depending on the lineage.

As in the deepest branch of the Charophyceae, the mitochondrial genome evolved in a rather dynamic fashion in the Zygnematales. Relative to the mtDNA of *Roya*, the 152.1-kb *Closterium* mtDNA is 2.2 times larger in size and 13% richer in A + T (boasting the highest A + T content among the streptophyte mtDNAs examined in our study) (table 1), contains 15 times as many introns (being the most intron-rich charophycean mtDNA identified to date), and is separated from the latter genome by as many rearrangements as found between the liverwort and moss mtDNAs (supplementary fig. S4, Supplementary Material online). The only two extra genes it contains are *trnR*(ucg) and *trnT*(ugu), two genes missing in

many charophyceans and land plants (fig. 2). In this context, it is interesting to mention that, although the mtDNAs of *Closterium* and the liverwort *Marchantia* both encode *trnR(ucg)*, this gene has different origins in the two streptophytes. Indeed, contrary to the *Chlorokybus* and *Marchantia trnR(ucg)* genes, which have been hypothesized to be independently derived from *trnR(ucu)* (Oda et al. 1992; Turmel et al. 2007a; Wang et al. 2009), the *Closterium trnR(ucg)* resembles its *Mesostigma* and klebsormidialean counterparts in being closely related to *trnR(acg)*. Another interesting difference between the two zygneatalean mtDNAs concerns the structure of *yejR*. The *Roya yejR* gene resembles its land plant homologs in being split into two ORFs; however, here again, this attribute must have originated from convergent events as the sites of break-up in the zygneatalean and land plant genes are different. In light of the wide diversity of the Zygneatales—at least 4,000 species divided into two major clades, Zygneatales and Desmidiales (Gerrath 2003)—and the remarkable differences in architecture uncovered for the *Closterium* and *Roya* mtDNAs, it is almost certain that we have sampled only part of the spectrum of mtDNA variations displayed by the Zygneatales. Examination of more zygneatalean taxa will be required to determine what was the ancestral condition of the mitochondrial genome in this lineage, an information that is also needed to better understand the mtDNA changes that occurred during the transition from green algae to land plants. In this regard, our results concerning the mtDNAs of both the Zygneatales and Coleochaetales support the hypotheses that important alterations took place at the level of intron content and that RNA editing was acquired during the emergence of the first land plants (Liu et al. 2011). On the other hand, our finding that the size of the *Closterium* mtDNA is comparable or even higher relative to those of their liverwort and moss counterparts calls into question the notion that the mitochondrial genome increased significantly in size during this period (Li et al. 2009; Liu et al. 2011). It is entirely possible, however, that the large size of the *Closterium* genome was not inherited from the most recent common ancestor of land plants and charophyceans but resulted rather from genome expansion events that took place in a specific zygneatalean lineage.

Although the mtDNAs of the representatives of the Klebsormidiales exhibit less important differences in genome size and intron content than their zygneatalean homologs (table 1), they appear to be more fluid at the level of gene order as evidenced by MP analysis of gene pairs (fig. 6) and the inferred scenario of genome rearrangements (supplementary fig. S4, Supplementary Material online). Moreover, they tend to be more prone to gene losses (fig. 2) than other charophycean mtDNAs. Indeed, a distinctive feature of the *Entransia* and *Klebsormidium* mtDNAs is the absence of several standard genes in one or both of these genomes. Most of these genes are missing in both *Entransia* and *Klebsormidium*,

indicating that they were lost from the common ancestor of the two algae. Included in this category are *rps8*, *rps10*, *rpl2*, and *rps19*, which are found in the cluster of ribosomal protein genes inherited from the prokaryotic progenitor of mitochondria, that is, the stretch delimited by *rps10* and *rps11* in other charophycean genomes. Because *rps10*, *rpl2*, and *rps19* are contiguous, they were probably transferred together to the nucleus in the common ancestor of *Entransia* and *Klebsormidium*. Mitochondrial gene losses also characterize mitochondrial genome evolution in the hornworts; but in contrast to what we observed in the Klebsormidiales, these events were more frequent and vestiges of many of the transferred genes, that is, pseudogenes, are still detectable in the mitochondria of extant hornworts (Li et al. 2009; Xue et al. 2010).

In sharp contrast to the charophycean lineages discussed above, the two representatives of the Charales display an extremely conservative pattern of mitochondrial genome evolution, sharing exactly the same gene content (fig. 2) and gene order (fig. 5) as well as a high proportion of group II introns (fig. 4). This gene organization includes ancestral gene linkages that have been retained in liverworts and/or mosses but have been disrupted in the Zygneatales and Coleochaetales (e.g., the clusters containing the *yej* genes, *ms* and *rrn5*, and *trnR(ucu)* and *trnY(gua)*). The strong conservation of structural features between the *Chara* and *Nitella* mtDNAs is correlated with the very high level of similarity found at the sequence level, which is illustrated by the short distance separating *Nitella* and *Chara* in mitochondrial trees (figs. 7 and 9). This observation suggests that the mitochondrial genome evolves extremely slowly within the Charales, both at the structural and sequence levels. Obviously, sampling of mtDNAs from earliest diverging lineages will be needed to determine whether *Nitella* and *Chara* are representative of all extant members of the Charales. With regard to land plants, mitochondrial genome evolution has been reported to be exceptionally conservative in liverworts and mosses since their origins (Terasawa et al. 2007; Wang et al. 2009; Liu et al. 2011) as well as in the magnoliid *Liriodendron tulipifera* (Richardson et al. 2013); however, the mitochondrial genomes of later-diverging angiosperms are known to evolve more rapidly. Considering that the differences between the evolutionary pattern of the magnoliid mtDNA and those of other angiosperms are thought to be related to differences in growth and mode of reproduction, it is possible that such factors could explain the seemingly slower pattern observed for the morphologically complex Charales compared with other charophycean lineages.

Our comparative analysis of mitochondrial gene order highlights the less conservative evolutionary trajectory followed by the mitochondrial genome in the major lineage of the Coleochaetales occupied by *Chaetosphaeridium*. We have inferred that the emergence of the Charales gave rise to a number of gene linkages that have been preserved in the Zygneatales, liverworts, and mosses but not in the

representative of the Coleochaetales (fig. 5). Given that the Coleochaetales represent a branch that emerged after the Charales or is sister to the Zygnematales in chloroplast and nuclear phylogenomic trees (Turmel et al. 2006, 2007b; Wodniok et al. 2011; Laurin-Lemay et al. 2012; Timme et al. 2012), extensive genome rearrangements likely led to the disappearance of these derived clusters in *Chaetosphaeridium*. This is consistent with our MP analysis of gene pairs (fig. 6) and scenario of genome rearrangements (supplementary fig. S4, Supplementary Material online), which reveal that the *Chaetosphaeridium* mtDNA forms a longer branch than its charalean and zygnematalean homologs. Whether a more conservative evolutionary pattern characterizes the other recognized main lineage known in the Coleochaetales will await sampling of *Coleochaete* species.

Introns are highly volatile components of the charophycean mitochondrial genome (figs. 3 and 4), making it difficult to trace their origins and use them as phylogenetic indicators. Regarding the group II introns, only six sites of insertion are conserved between two or more charophycean groups (*cox2_351*, *nad3_211*, *rnl_575*, *rnl_2078*, *trnN(guu)_37*, and *trnS(gcu)_43*) (fig. 4), implying that they may have been transmitted by vertical inheritance from a common ancestor. The putatively ancestral intron insertion site in *trnS(gcu)* has also been identified in liverworts (this gene is missing in other land plant lineages); four separate sites (in *cox2*, *nad3*, *nad4*, and *rps3*) are shared by the Charales and some land plant lineages (mainly bryophytes); and a single site in *cox2* is shared specifically between the Zygnematales and the liverworts. It is intriguing that the Charales have more group II introns in common with land plants than the Zygnematales; this may reflect a higher retention of ancestral characters in land plant mtDNAs compared with the investigated coleochaetalean and zygnematalean genomes. Considering that most of the group II introns found in the Mesostigmatales/Chlorokybales, Klebsormidiales, and Zygnematales are lineage-specific and are not shared by the two members sampled in these individual lineages, we hypothesize that they arose by horizontal transfer or transposition of existing introns, although we cannot exclude the possibility that some are the result of differential losses from a common ancestor. In contrast, the group I introns reveal a broader distribution of insertion sites across charophycean lineages, with sites 909 and 1149 in *cox1* being presumably the most ancient (fig. 3). Our data suggest that five of the eight insertion sites found in bryophyte mtDNAs (one in *cob*, three in *cox1*, and the *nad5* site) trace back to a charophycean ancestor. The site in *nad5* is the only one that is shared specifically between a single charophycean lineage (i.e., the *Closterium* lineage) and land plants. Given the broad distribution of group I introns across charophycean lineages, the total absence of members of this family from the compact mtDNA of *Roya* makes credible the hypothesis that many events of intron losses took place in the lineage leading to this alga. In this context, it is

interesting to note that the chloroplast genome also lost numerous group II introns during the evolution of the Zygnematales (Turmel et al. 2005). Because homologous recombination of intronless cDNA copies generated by reverse transcriptase is thought to lead to intron losses (Dujon 1989), the frequency of homologous recombination events or the level of reverse transcriptase activity might be higher in both the mitochondria and chloroplasts of some zygnematalean green algae.

Our finding of unusual free-standing genes in the genomes newly described here reinforces the notion that horizontal transfer of foreign sequences played a role in shaping the mitochondrial genome during the evolution of charophyceans. We previously reported the presence of an integrase/recombinase gene, presumably of phage origin, in the *Chaetosphaeridium* mtDNA (Turmel et al. 2002a). This gene is found in the same gene context in the *Roya* mtDNA, implying that it might have arisen in a common ancestor of the Zygnematales and Coleochaetales. In the *Klebsormidium* mtDNA, we have uncovered a potential coding sequence related to bacterial or phycodnavirus C-5 cytosine-specific DNA methyltransferase genes, which are often associated with restriction–modification systems. To our knowledge, this is the first DNA methyltransferase gene reported in any organelle genome. As viruses belonging to the *Phycodnaviridae* infect marine or freshwater green algae (for chloroviruses, it has been shown that the restriction–modification system helps them to degrade host DNA early in infection [Agarkova et al. 2006]), it is possible that the *Klebsormidium* DNA methyltransferase gene is a vestige of viral infection. Finally, the *Entransia* and *Closterium* mtDNA sequences disclosed *trnT(ugu)*, a gene missing in all other streptophyte mtDNAs examined so far and presumably originating from the chloroplasts of red algae or glaucocystophytes.

Conclusions

The structural analyses reported here enhance considerably our understanding of the dynamics of mitochondrial genome evolution within and across streptophyte algal lineages and during the transition from green algae to land plants. Collectively, our analyses of gene order and gene distribution and the phylogenies we inferred from 40 mtDNA-encoded proteins failed to resolve unambiguously the relationships among the Zygnematales, Coleochaetales, and Charales; however, they are consistent with previous phylogenomic studies in favoring that the morphologically complex Charales are not the closest relatives of land plants. Our investigation not only provides another example of how caution must be exercised when large multigene sequence alignments with a limited taxon sampling are analyzed but it also exemplifies the utility of using structural genomic features to validate phylogenetic hypotheses.

Supplementary Material

Supplementary table S1 and figures S1–S6 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

The authors thank Antony Vincent for his help in comparing the amino acid compositions of the concatenated proteins used in our phylogenetic analyses. This work was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada (to C.L. and M.T.).

Literature Cited

- Adams KL, Qiu YL, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc Natl Acad Sci U S A*. 99:9905–9912.
- Agarkova IV, Dunigan DD, Van Etten JL. 2006. Virion-associated restriction endonucleases of chloroviruses. *J Virol*. 80:8114–8123.
- Andersen RA. 2005. *Algal culturing techniques*. Boston (MA): Elsevier/Academic Press.
- Bremer K, Humphries CJ, Mishler BD, Churchill SP. 1987. On cladistic relationships in green plants. *Taxon* 36:339–349.
- Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome Biol Evol*. 5:418–438.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- Caprara A. 2003. The reversal median problem. *INFORMS J Comput*. 15: 93–113.
- Chang SX, et al. 2011. Mitochondrial genome sequencing helps show the evolutionary mechanism of mitochondrial genome formation in *Brassica*. *BMC Genomics* 12:497.
- Darracq A, Varre JS, Touzet P. 2010. A scenario of mitochondrial genome evolution in maize based on rearrangement events. *BMC Genomics* 11:233.
- Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet*. 6:361–375.
- Dujon B. 1989. Group-I introns as mobile genetic elements: facts and mechanistic speculations—a review. *Gene* 82:91–114.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.
- Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool*. 27:401–410.
- Felsenstein J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* 5:164–166.
- Fourment M, Gibbs MJ. 2006. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evol Biol*. 6:1.
- Gensel PG. 2008. The earliest land plants. *Annu Rev Ecol Evol Syst*. 39: 459–477.
- Gerrath JF. 2003. Conjugating green algae and desmids. In: Wehr JD, Sheath RG, editors. *Freshwater algae of North America: ecology and classification*. San Diego (CA): Academic Press. p. 353–381.
- Gontcharov AA, Marin B, Melkonian M. 2004. Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and *rbcl* sequence comparisons in the Zygnematomyceae (Streptophyta). *Mol Biol Evol*. 21:612–624.
- Gowri-Shankar V, Rattray M. 2006. On the correlation between composition and site-specific evolutionary rate: implications for phylogenetic inference. *Mol Biol Evol*. 23:352–364.
- Hall JD, Karol KG, McCourt RM, Delwiche CF. 2008. Phylogeny of the conjugating green algae based on chloroplast and mitochondrial nucleotide sequence data. *J Phycol*. 44:467–477.
- Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? *Trends Genet*. 22:225–231.
- Karol KG, McCourt RM, Cimino MT, Delwiche CF. 2001. The closest living relatives of land plants. *Science* 294:2351–2353.
- Kenrick P, Wellman CH, Schneider H, Edgecombe GD. 2012. A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. *Philos Trans R Soc Lond B Biol Sci*. 367:519–536.
- Knoop V. 2013. Plant mitochondrial genome peculiarities evolving in the earliest vascular plant lineages. *J Syst Evol*. 51:1–12.
- Kurtz S, et al. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 29:4633–4642.
- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.
- Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol*. 21:1095–1109.
- Laurin-Lemay S, Brinkmann H, Philippe H. 2012. Origin of land plants revisited in the light of sequence contamination and missing data. *Curr Biol*. 22:R593–R594.
- Lewis LA, McCourt RM. 2004. Green algae and the origin of land plants. *Am J Bot*. 91:1535–1556.
- Li LB, Wang B, Liu Y, Qiu YL. 2009. The complete mitochondrial genome sequence of the hornwort *Megaceros aenigmaticus* shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. *J Mol Evol*. 68:665–678.
- Liu Y, Xue JY, Wang B, Li LB, Qiu YL. 2011. The mitochondrial genomes of the early land plants *Treubia lacunosa* and *Anomodon rugelii*: dynamic and conservative evolution. *PLoS One* 6:e25836.
- Lockhart PJ, Steel MA, Hendy MD, Penny D. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol Biol Evol*. 11:605–612.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 25: 955–964.
- Maddison DR, Maddison WP. 2000. *MacClade 4: analysis of phylogeny and character evolution*. Sunderland (MA): Sinauer Associates.
- Maddison WP, Maddison DR. 2011. *Mesquite: a modular system for evolutionary analysis*. Version 2.75. [cited 2013 Sep 27]. Available from: <http://mesquiteproject.org>.
- Margulies M, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.
- McCourt RM, Delwiche CF, Karol KG. 2004. Charophyte algae and land plant origins. *Trends Ecol Evol*. 19:661–666.
- McCourt RM, et al. 2000. Phylogeny of the conjugating green algae (Zygnemophyceae) based on *rbcl* sequences. *J Phycol*. 36: 747–758.
- Michel F, Umeson K, Ozeki H. 1989. Comparative and functional anatomy of group II catalytic introns—a review. *Gene* 82:5–30.
- Michel F, Westhof E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol*. 216:585–610.
- Mikhailiyuk TI, et al. 2008. New streptophyte green algae from terrestrial habitats and an assessment of the genus *Interfilum* (Klebsormidiophyceae, Streptophyta). *J Phycol*. 44:1586–1603.
- Moret BME, Siepel AC, Tang J, Liu T. 2002. Inversion medians outperform breakpoint medians in phylogeny reconstruction from gene-order data. in *Proceedings of the Second International Workshop on Algorithms in*

- Bioinformatics; 2002 September 17–21; Vol. 2452. Rome, Italy: Springer. p. 521–536.
- Nenadic O, Greenacre M. 2007. Correspondence analysis in R, with two- and three-dimensional graphics: the ca package. *J Stat Software* 20: 1–13.
- Oda K, et al. 1992. Transfer-RNA genes in the mitochondrial genome from a liverwort, *Marchantia Polymorpha*—the absence of chloroplast-like transfer-RNAs. *Nucleic Acids Res.* 20:3773–3777.
- Philippe H, et al. 2011a. Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. *Nature* 470:255–260.
- Philippe H, et al. 2011b. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 9:e1000602.
- Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol Biol Evol.* 21:1455–1458.
- Pombert JF, Otis C, Lemieux C, Turmel M. 2005. Chloroplast genome sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. *Mol Biol Evol.* 22: 1903–1918.
- Qiu YL, et al. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci U S A.* 103:15511–15516.
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16:276–277.
- Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 2013. The “fossilized” mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* 11:29.
- Rodriguez-Ezpeleta N, et al. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst Biol.* 56:389–399.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Swofford DL. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.
- Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
- Terasawa K, et al. 2007. The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. *Mol Biol Evol.* 24:699–709.
- Timme RE, Bachvaroff TR, Delwiche CF. 2012. Broad phylogenomic sampling and the sister lineage of land plants. *PLoS One* 7:e29696.
- Turmel M, Ehara M, Otis C, Lemieux C. 2002. Phylogenetic relationships among streptophytes as inferred from chloroplast small and large subunit rRNA gene sequences. *J Phycol.* 38:364–375.
- Turmel M, Otis C, Lemieux C. 2002a. The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. *Proc Natl Acad Sci U S A.* 99:11275–11280.
- Turmel M, Otis C, Lemieux C. 2002b. The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Mol Biol Evol.* 19: 24–38.
- Turmel M, Otis C, Lemieux C. 2003. The mitochondrial genome of *Chara vulgaris*: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. *Plant Cell* 15: 1888–1903.
- Turmel M, Otis C, Lemieux C. 2005. The complete chloroplast DNA sequences of the charophycean green algae *Staurastrum* and *Zygnema* reveal that the chloroplast genome underwent extensive changes during the evolution of the Zygnematales. *BMC Biol.* 3:22.
- Turmel M, Otis C, Lemieux C. 2006. The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Mol Biol Evol.* 23:1324–1338.
- Turmel M, Otis C, Lemieux C. 2007a. An unexpectedly large and loosely packed mitochondrial genome in the charophycean green alga *Chlorokybus atmophyticus*. *BMC Genomics* 8:137.
- Turmel M, Pombert JF, Charlebois P, Otis C, Lemieux C. 2007b. The green algal ancestry of land plants as revealed by the chloroplast genome. *Int J Plant Sci.* 168:679–689.
- Turmel M, et al. 1999. The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*. Two radically different evolutionary patterns within green algae. *Plant Cell* 11: 1717–1730.
- Villarreal JC, Goffinet B, Duff RJ, Cargill DC. 2010. Phylogenetic delineation of *Nothoceros* and *Megaceros* (Dendrocerotaceae). *Bryologist* 113:106–113.
- Wang B, Xue JY, Li LB, Liu Y, Qiu YL. 2009. The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea* reveals extremely conservative mitochondrial genome evolution in liverworts. *Curr Genet.* 55:601–609.
- Wodniok S, et al. 2011. Origin of land plants: do conjugating green algae hold the key? *BMC Evol Biol.* 11:104.
- Xue JY, Liu Y, Li LB, Wang B, Qiu YL. 2010. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Curr Genet.* 56:53–61.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Associate editor: Shu-Miaw Chaw