



Comparative Chloroplast Genome Analyses of Streptophyte Green Algae Uncover Major Structural Alterations in the Klebsormidiophyceae, Coleochaetophyceae and Zygnematophyceae

Claude Lemieux*, Christian Otis and Monique Turmel

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

OPEN ACCESS

Edited by:

David Domozych,
Skidmore College, USA

Reviewed by:

David Roy Smith,
University of Western Ontario, Canada
Burkhard Becker,
University of Cologne, Germany
Robert K. Jansen,
The University of Texas at Austin, USA

*Correspondence:

Claude Lemieux
claude.lemieux@bcm.ulaval.ca

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 25 February 2016

Accepted: 06 May 2016

Published: 24 May 2016

Citation:

Lemieux C, Otis C and Turmel M
(2016) Comparative Chloroplast
Genome Analyses of Streptophyte
Green Algae Uncover Major Structural
Alterations
in the Klebsormidiophyceae,
Coleochaetophyceae
and Zygnematophyceae.
Front. Plant Sci. 7:697.
doi: 10.3389/fpls.2016.00697

The Streptophyta comprises all land plants and six main lineages of freshwater green algae: Mesostigmatophyceae, Chlorokybophyceae, Klebsormidiophyceae, Charophyceae, Coleochaetophyceae and Zygnematophyceae. Previous comparisons of the chloroplast genome from nine streptophyte algae (including four zygnematophyceans) revealed that, although land plant chloroplast DNAs (cpDNAs) inherited most of their highly conserved structural features from green algal ancestors, considerable cpDNA changes took place during the evolution of the Zygnematophyceae, the sister group of land plants. To gain deeper insights into the evolutionary dynamics of the chloroplast genome in streptophyte algae, we sequenced the cpDNAs of nine additional taxa: two klebsormidiophyceans (*Entransia fimbriata* and *Klebsormidium* sp. SAG 51.86), one coleochaetophycean (*Coleochaete scutata*) and six zygnematophyceans (*Cylindrocystis brebissonii*, *Netrium digitus*, *Roya obtusa*, *Spirogyra maxima*, *Cosmarium botrytis* and *Closterium baillyanum*). Our comparative analyses of these genomes with their streptophyte algal counterparts indicate that the large inverted repeat (IR) encoding the rDNA operon experienced loss or expansion/contraction in all three sampled classes and that genes were extensively shuffled in both the Klebsormidiophyceae and Zygnematophyceae. The klebsormidiophycean genomes boast greatly expanded IRs, with the *Entransia* 60,590-bp IR being the largest known among green algae. The 206,025-bp *Entransia* cpDNA, which is one of the largest genome among streptophytes, encodes 118 standard genes, i.e., four additional genes compared to its *Klebsormidium flaccidum* homolog. We inferred that seven of the 21 group II introns usually found in land plants were already present in the common ancestor of the Klebsormidiophyceae and its sister lineages. At 107,236 bp and with 117 standard genes, the *Coleochaete* IR-less genome is both the smallest and most compact among the streptophyte algal cpDNAs analyzed thus far; it

lacks eight genes relative to its *Chaetosphaeridium globosum* homolog, four of which represent unique events in the evolutionary scenario of gene losses we reconstructed for streptophyte algae. The 10 compared zygneumatophycean cpDNAs display tremendous variations at all levels, except gene content. During zygneumatophycean evolution, the IR disappeared a minimum of five times, the rDNA operon was broken at four distinct sites, group II introns were lost on at least 43 occasions, and putative foreign genes, mainly of phage/viral origin, were gained.

Keywords: Streptophyta, charophytes, phylogenomics, plastid DNA, large inverted repeat, chloroplast genome evolution, genome rearrangements, group II introns

INTRODUCTION

The green plants, also referred to as Viridiplantae or Chloroplastida, split early (1200-700 Mya) into two main clades, the Chlorophyta and Streptophyta (Lewis and McCourt, 2004; Leliaert et al., 2012; Becker, 2013), and about 470 Mya, green algae from the Streptophyta gave rise to all land plants (Gensel, 2008; Becker and Marin, 2009; Kenrick et al., 2012). The streptophyte green algae, also called charophyte algae, inhabit freshwater environments and display a variety of cellular organizations, ranging from unicellular (e.g., *Mesostigma viride*, the only species of the Mesostigmatophyceae and some species of the Zygneumatophyceae), to packets of cells (*Chlorokybus atmophyticus*) or filaments (Klebsormidiophyceae and Zygneumatophyceae), and to multicellular organization (Coleocheotophyceae and Charophyceae) (Graham et al., 2000; McCourt et al., 2004; Umen, 2014). Identifying which of the six main lineages of streptophyte algae is the closest to land plants proved to be challenging (Karol et al., 2001; Turmel et al., 2006, 2007, 2013; Laurin-Lemay et al., 2012; Zhong et al., 2013); but there is now solid evidence based on both nuclear (Wodniok et al., 2011; Laurin-Lemay et al., 2012; Timme et al., 2012; Zhong et al., 2013; Wickett et al., 2014) and chloroplast phylogenomic studies that the Zygneumatophyceae is sister to all land plants (Turmel et al., 2006; Turmel et al., 2007; Civan et al., 2014; Zhong et al., 2014). This morphologically diverse group comprises all green algae that reproduce sexually by conjugation and is the only streptophyte algal class that displays substantial diversity (at least 4000 species) (Gerrath, 2003), with several major lineages identified on the basis of the nuclear-encoded small subunit (SSU) rDNA and chloroplast-encoded *rbcL* gene sequences (Gontcharov et al., 2003, 2004), or on the basis of the chloroplast and mitochondrial *psaA*, *rbcL* and *cox3* genes (Hall et al., 2008). Based on the structure of the cell wall, the Zygneumatophyceae has been divided into two orders: the Zygneumatales feature a smooth cell wall (the ancestral trait) and the Desmidiiales an ornamented and segmented cell wall (McCourt et al., 2000; Gerrath, 2003). Cell division in the three earliest-diverging lineages of the Streptophyta, the Mesostigmatophyceae, Chlorokybophyceae and Klebsormidiophyceae, occurs by furrowing, but as in land plants, the morphologically more complex Charophyceae and Coleocheotophyceae use a mechanism of cell division involving a phragmoplast and possess cell walls with plasmodesmata (Graham et al., 2000; McCourt et al., 2004).

To date, only nine chloroplast genome sequences of streptophyte algae are available in public databases: four for the Zygneumatophyceae and a single genome for each of the remaining streptophyte algal lineages (Lemieux et al., 2000; Turmel et al., 2002, 2005, 2006; Lemieux et al., 2007; Civan et al., 2014). From their gene contents, which vary from 114 to 138 standard genes (i.e., genes whose orthologs are usually found in the chloroplast of photosynthetic eukaryotes), it was predicted that 144 unique standard genes were present in the common ancestor of all streptophytes (Turmel et al., 2006). Comparisons of streptophyte algal genomes with their land plant counterparts revealed that a large fraction of the structural features typically conserved in land plant chloroplast DNAs (cpDNAs) were inherited from streptophyte green algae occupying early diverging lineages (Turmel et al., 2006, 2007; Civan et al., 2014). For instance, 15 of the 21 group II introns found in land plants arose before the emergence of the Klebsormidiophyceae or during the evolutionary interval separating the latter lineage from the Charophyceae (Turmel et al., 2006, 2007; Civan et al., 2014). It was also inferred that the chloroplast genome remained unchanged or mostly unchanged in terms of gene content, gene order, and intron content during the transition from streptophyte algae to land plants (Turmel et al., 2006, 2007). In contrast, extensive gene shuffling and substantial structural alterations, including multiple intron losses, took place within the Zygneumatophyceae (Turmel et al., 2005; Civan et al., 2014). The highly conserved quadripartite structure of green plant cpDNAs — a structure characterized by two copies of a large inverted repeat (IR) sequence encoding the rRNA operon, which are separated from one another by small and large single-copy (SSC and LSC) regions — was found in only one of the analyzed zygneumatophyceans (*Roya anglica*); however, this IR contains a broken rRNA operon and two open reading frames (ORFs) not previously reported in other streptophyte IRs (Civan et al., 2014). This observation led Civan et al. (2014) to propose that either the IR was lost on three independent occurrences during the evolution of the Zygneumatophyceae or that it arose *de novo* from an ancestor lacking an IR. In *Klebsormidium flaccidum*, the IR is also unusual with regards to its exceptionally large size and its apparent lack of the 4.5S and 5S rRNA genes from the rRNA operon (Civan et al., 2014).

Sampling of additional streptophyte algal taxa is required to better understand the evolutionary history of the chloroplast genome in the Zygneumatophyceae and its evolutionary dynamics

in other classes. Toward these goals, we sequenced the chloroplast genomes of the klebsormidiophyceans *Entransia fimbriata* and *Klebsormidium* spec. SAG 51.86, the coleochaetophycean *Coleochaete scutata*, and six zygmatophyceans that were selected to represent distinct lineages in the SSU rDNA and *rbcL* phylogenies reported by Gontcharov et al. (2004): *Cylindrocystis brebissonii*, *Netrium digitus*, *Roya obtusa* and *Spirogyra maxima* belong to the Zygnematales, while *Cosmarium botrytis* and *Closterium baillyanum* represent the Desmidiaceae. Our comparative analyses of these genomes with their previously described streptophyte green algal counterparts indicate that the large IR was involved in major structural changes (IR losses or expansion/contraction) in all three sampled classes and that genes underwent extensive shuffling in both the Klebsormidiophyceae and the Zygnematophyceae.

MATERIALS AND METHODS

Strains and Culture Conditions

Strains of *Klebsormidium* sp. SAG 51.86, *Coleochaete scutata* SAG 110.80M, *Closterium baillyanum* SAG 50.89, *Cylindrocystis brebissonii* SAG 615-1, and *Roya obtusa* SAG 168.80 were obtained from the culture collection of algae at the University of Goettingen¹, whereas *Entransia fimbriata* UTEX LB 2353, *Cosmarium botrytis* UTEX 175, *Netrium digitus* UTEX LB 561, and *Spirogyra maxima* UTEX LB 2495 originated from the culture collection of algae at the University of Texas in Austin². All strains were grown in medium C (Andersen, 2005) at 18°C under alternating 12-h light/dark periods.

Genome Assemblies and Sequence Analyses

For all strains, except *Cosmarium* and *Cylindrocystis*, total cellular DNA was extracted as described (Turmel et al., 1999a) and A + T-rich organellar DNA was separated from nuclear DNA by CsCl-bisbenzimidazole isopycnic centrifugation (Lemieux et al., 2014). Total cellular DNA from *Cosmarium* and *Cylindrocystis* was isolated using the EZNA HP Plant Mini Kit of Omega Bio-Tek (Norcross, GA, USA).

For Illumina sequencing of the *Closterium*, *Cosmarium*, and *Cylindrocystis* chloroplast genomes, libraries of 700-bp fragments were constructed using the TrueSeq DNA Sample Prep Kit (Illumina, San Diego, CA, USA) and paired-end reads were generated on the Illumina HiSeq 2000 (100-bp reads) or the MiSeq (300-bp reads) sequencing platforms by the Innovation Centre of McGill University and Génome Québec³ and the “Plateforme d’Analyses Génomiques de l’Université Laval⁴,” respectively. Reads were assembled using Ray v2.3.1 (Boisvert et al., 2010) and contigs were visualized, linked and edited using CONSED v22 (Gordon et al., 1998). Contigs of chloroplast origin were identified by BLAST searches against a local database of

organelle genomes. Regions spanning gaps in cpDNA assemblies were amplified by polymerase chain reaction (PCR) with primers specific to the flanking sequences. Purified PCR products were sequenced using Sanger chemistry with the PRISM BigDye Terminator Ready Reaction Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) on ABI model 373 or 377 DNA sequencers (Applied Biosystems).

For 454 sequencing of the *Entransia*, *Netrium*, *Roya* and *Spirogyra* chloroplast genomes, shotgun libraries (700-bp fragments) of A + T-rich DNA fractions were constructed using the GS-FLX Titanium Rapid Library Preparation Kit of Roche 454 Life Sciences (Branford, CT, USA). Library construction and 454 GS-FLX DNA Titanium pyrosequencing were carried out by the “Plateforme d’Analyses Génomiques de l’Université Laval⁵.” Reads were assembled using Newbler v2.5 (Margulies et al., 2005) with default parameters, and contigs were visualized, linked and edited using CONSED v22 (Gordon et al., 1998). Identification of cpDNA contigs and gap filling were performed as described above for Illumina sequence assemblies.

For Sanger sequencing of the *Klebsormidium* and *Coleochaete* chloroplast genomes, random clone libraries were prepared from 1500 to 2000-bp fragments derived from A + T rich DNA fractions using the pSMART-HCKan (Lucigen Corporation, Middleton, WI, USA) 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) on ABI model 373 or 377 DNA sequencers (Applied Biosystems), using SR2 and SL1 primers as well as oligonucleotides complementary to internal regions of the plasmid DNA inserts. The resulting sequences were edited and assembled using Sequencher v5.1 (Gene Codes Corporation, Ann Arbor, MI, USA). Genomic regions not represented in the sequence assemblies or plasmid clones were directly sequenced from PCR-amplified fragments using primers specific to the flanking contigs.

Genes and ORFs were identified on the final assemblies using a custom-built suite of bioinformatics tools as described previously (Turmel et al., 2006). tRNA genes were localized using tRNAscan-SE v1.3.1 (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. Circular genome maps were drawn with OGDRAW v1.2 (Lohse et al., 2007). Genome-scale sequence comparisons of the pairs of *Roya* and *Klebsormidium* species were carried out with LAST v7.1.4 (Frith et al., 2010). For all compared genomes, G + C contents of a set of 88 protein-coding genes were determined at the three codon positions using DAMBE v5 (Xia, 2013).

To estimate the proportion of small repeated sequences, repeats ≥ 30 bp were retrieved using REPFIND of the REPuter v2.74 program (Kurtz et al., 2001) with the options $-f -p -l -allmax$ and were then masked on the genome sequences using

¹<http://www.uni-goettingen.de/de/184982.html>

²<http://utex.org/>

³<http://gqinnovationcenter.com/index.aspx>

⁴<http://pag.ibis.ulaval.ca/seq/en/>

⁵<http://pag.ibis.ulaval.ca/seq/en/>

RepeatMasker⁶ running under the Crossmatch search engine⁷. The G+C contents of the repeated and unique sequences were calculated from the outputs of RepeatMasker that were generated with the -xsmall option (under this option the repeat regions are returned in lower case and non-repetitive regions in capitals in the masked file).

Phylogenetic Analyses

The chloroplast genomes of 28 streptophyte taxa were used to generate the analyzed amino acid (PCG-AA) and nucleotide (PCG12) data sets. The latter were assembled from the following 88 protein-coding genes: *accD*, *atpA*, *B*, *E*, *F*, *H*, *I*, *ccsA*, *cemA*, *chlB*, *I*, *L*, *N*, *clpP*, *ftsH*, *infA*, *ndhA*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K*, *odpB*, *petA*, *B*, *D*, *G*, *L*, *N*, *psaA*, *B*, *C*, *I*, *J*, *M*, *psbA*, *B*, *C*, *D*, *E*, *F*, *H*, *I*, *J*, *K*, *L*, *M*, *N*, *T*, *Z*, *rbcL*, *rpl2*, *14*, *16*, *20*, *21*, *22*, *23*, *32*, *33*, *36*, *rpoA*, *B*, *C1*, *C2*, *rps2*, *3*, *4*, *7*, *8*, *11*, *12*, *14*, *15*, *16*, *18*, *19*, *ycf1*, *3*, *4*, *12*, *62*, *66*.

The PCG-AA data set was prepared as follows: the deduced amino acid sequences from the 88 individual genes were aligned using MUSCLE v3.7 (Edgar, 2004), the ambiguously aligned regions in each alignment were removed using TrimAl v1.3 (Capella-Gutierrez et al., 2009) with the options block = 6, gt = 0.7, st = 0.005 and sw = 3, and the protein alignments were concatenated using Phyutility v2.2.6 (Smith and Dunn, 2008). Phylogenies were inferred from the PCG-AA data set using the maximum likelihood (ML) and Bayesian methods. ML analyses were carried out using RAxML v8.2.3 (Stamatakis, 2014) and the GTR + Γ 4 model of sequence evolution; in these analyses, the data set was partitioned by gene, with the model applied to each partition. Confidence of branch points was estimated by fast-bootstrap analysis ($f = a$) with 100 replicates. Bayesian analyses were performed with PhyloBayes v4.1 (Lartillot et al., 2009) using the site-heterogeneous CATGTR + Γ 4 model (Lartillot and Philippe, 2004). Five independent chains were run for 2,000 cycles and consensus topologies were calculated from the saved trees using the BPCOMP program of PhyloBayes after a burn-in of 500 cycles. Under these conditions, the largest discrepancy observed across all bipartitions in the consensus topologies (maxdiff) was 0.0007, indicating that convergence between the chains was achieved.

The PCG12 nucleotide data set (first and second codon positions) was prepared as follows. The multiple sequence alignment of each protein was converted into a codon alignment, the poorly aligned and divergent regions in each codon alignment were excluded using Gblocks v0.91b (Castresana, 2000) with the $-t = c$, $-b3 = 5$, $-b4 = 5$ and $-b5 = \text{half}$ options, and the individual gene alignments were concatenated using Phyutility v2.2.6 (Smith and Dunn, 2008). The third codon positions of the resulting PCG123 alignment were then excluded using Mesquite v3.04 (Maddison and Maddison, 2015) to produce the PCG12 data set. ML analysis of the PCG12 data set was carried out using RAxML v8.2.3 (Stamatakis, 2014) and the GTR + Γ 4 model of sequence evolution. This data set was partitioned into gene groups, with the model applied to each partition. Confidence of

branch points was estimated by fast-bootstrap analysis ($f = a$) with 100 replicates.

dN, dS and dN/dS trees were inferred from a *tufA* codon alignment prepared as described above using PAML v4.8a (Yang, 2007) and the F3X4 codon frequencies model implemented in codeml. Positive selection was tested across the *tufA* sequences using the PARRIS module implemented in Datamonkey (Delpont et al., 2010).

Analyses of Gene Order and Reconstruction of Genomic Character Evolution

Syntenic regions in pairwise genome comparisons were identified using a custom-built program and the number of gene reversals between the compared genomes was estimated with GRIMM v2.01 (Tesler, 2002). The same custom-built program was employed to convert gene order to all possible pairs of signed genes (i.e., taking into account gene polarity); the gene pairs conserved in three or more genomes were visualized using Mesquite v3.04 (Maddison and Maddison, 2015). Gains and/or losses of genomic characters (standard genes, introns and gene pairs) were mapped on the streptophyte tree topology inferred in this study using MacClade v4.08 (Maddison and Maddison, 2000) and the Dollo principle of parsimony.

A ML tree based on gene adjacency was inferred using the phylogeny reconstruction option of the MLGO web server (Hu et al., 2014) and a gene order matrix containing all standard genes (including all copies of duplicated genes). Confidence of branch points was estimated by 1000 bootstrap replications. A gene reversal tree with the same topology as the MLGO tree was also computed; branch lengths were estimated using the $-t$ option of MGR v2.03 (Bourque and Pevzner, 2002) and a gene order matrix of the 89 genes shared by all compared genomes; because MGR cannot handle duplicated genes, only one copy of the IR and of each duplicated gene was included in this analysis.

Availability of Supporting Data

The chloroplast genome sequences generated in this study are available in the GenBank database under the accession numbers KU646489-KU646497. The data sets supporting the results of this article are available in the TreeBASE repository (Study ID 19085).

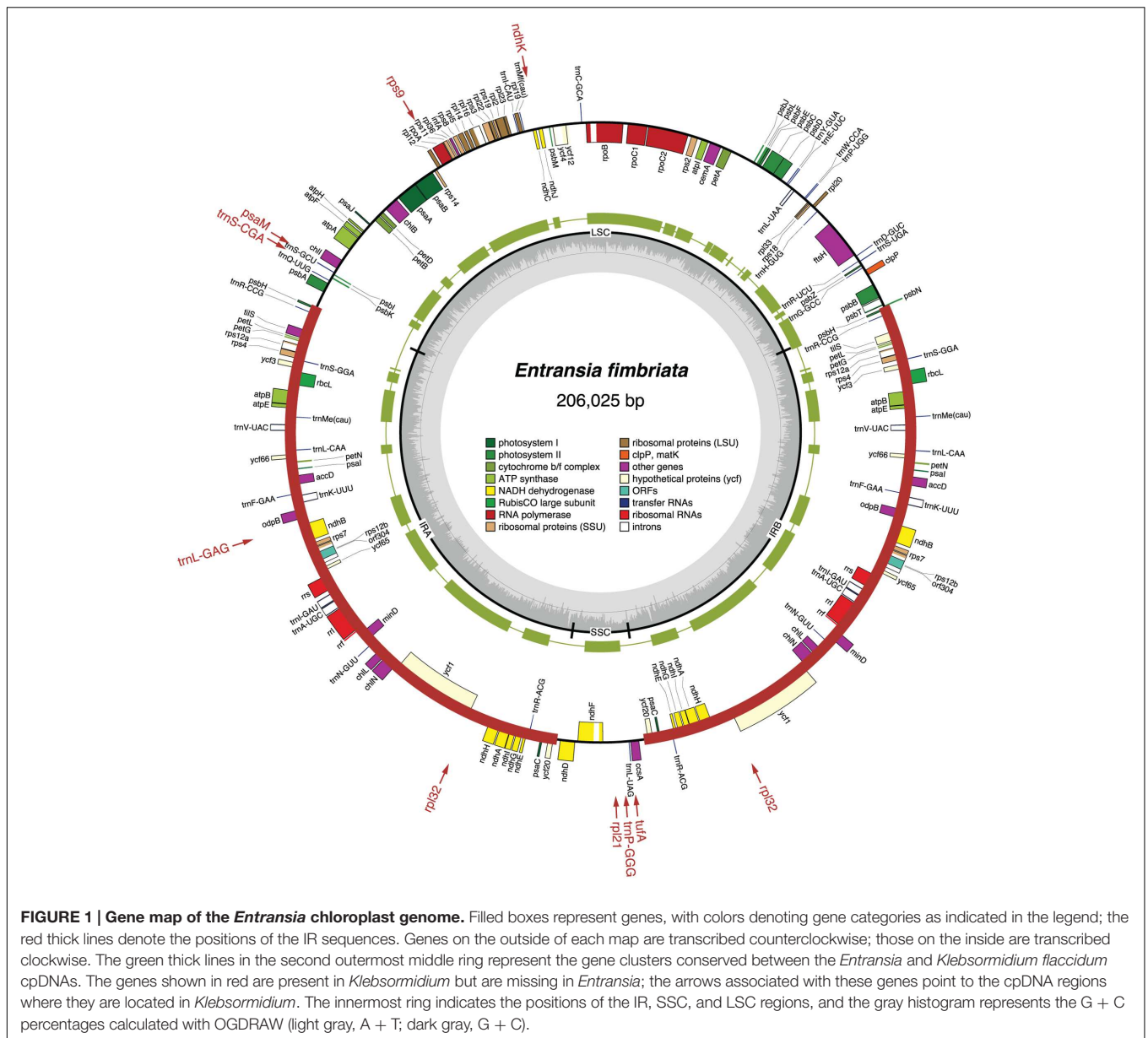
RESULTS

General Features

The chloroplast genomes of the nine sampled taxa, except that of *Klebsormidium* sp. SAG 51.86, were entirely sequenced and assembled as circular-mapping molecules (Figures 1–4). Although these genomes assembled as circles, it is possible that they exist *in vivo* as multi-genomic, linear-branched structures, as reported for land plant cpDNAs (Bendich, 2004). Their sizes range from 107 (for *Coleochaete*) to 208 kb (for *Cosmarium*) (Figures 1–4). Genome size variation is also important among lineages of the same class, in particular within the Zygnematophyceae where the variation is 1.6-fold (from

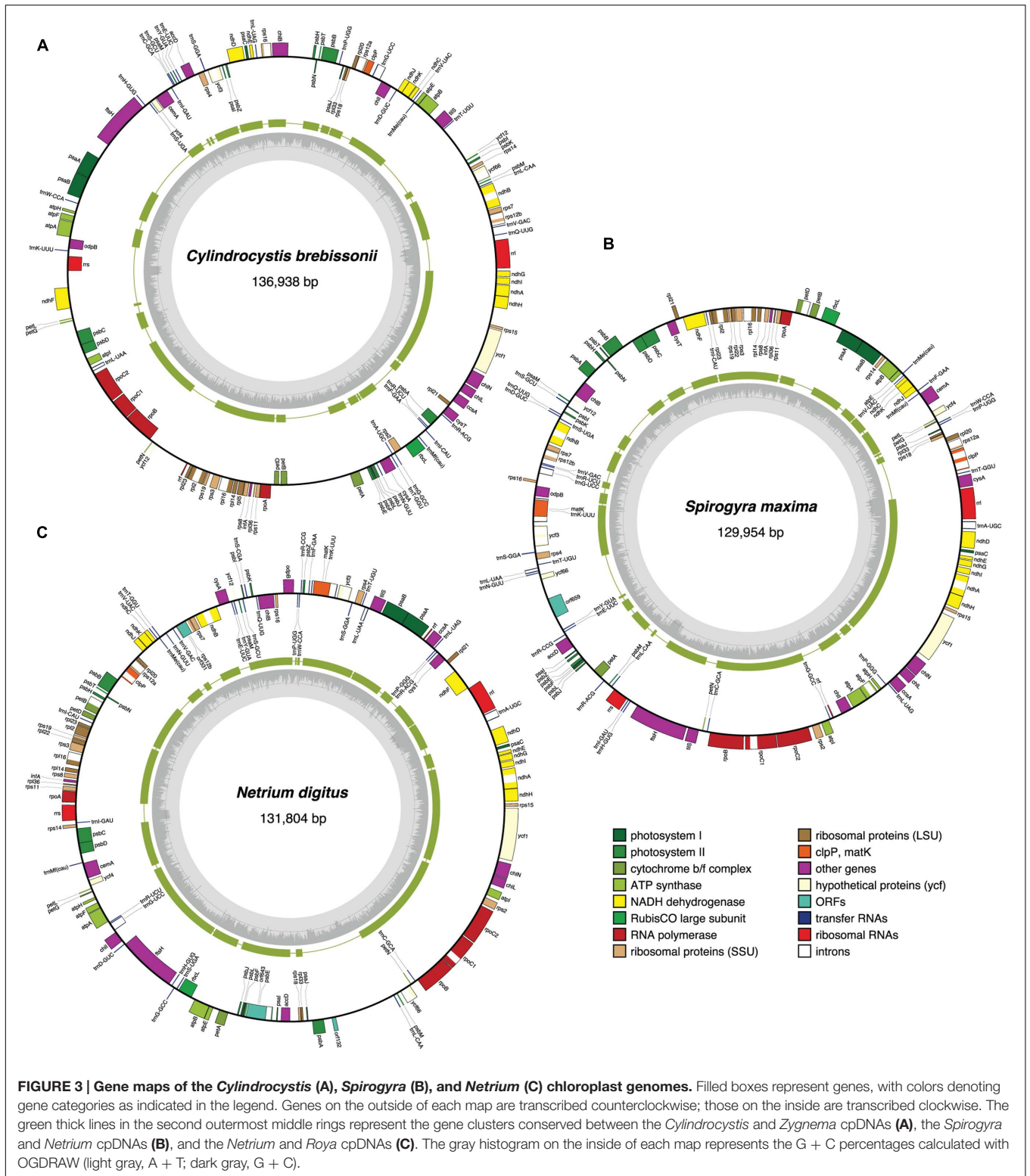
⁶<http://www.repeatmasker.org/>

⁷<http://www.phrap.org/>



130 kb in *Spirogyra* to 208 kb in *Cosmarium*). Within the Klebsormidiophyceae, differences in IR size, intron content, and lengths of intergenic regions essentially account for the increased size of the *Entransia* genome (206 kb) compared to that of *Klebsormidium flaccidum* (177 kb) (Table 1; Figure 5). Note that we detected an IR in the partially sequenced *Klebsormidium* sp. SAG 51.86 genome but were unable to identify the IR/SSC junction; 91% of this 130,962-kb genome sequence could be aligned with the *Klebsormidium flaccidum* genome in a strictly colinear fashion, highlighting the absence of just two genes [*trnN(guu)* and *ccsA*] and a total of 7,551 divergent positions, i.e., 6.3% of the aligned sequence. Within the Coleochaetophyceae, the 24-kb difference between the *Chaetosphaeridium* and *Coleochaete* genomes is largely due to the absence of the IR and of eight genes, as well as to shorter intergenic regions in

Coleochaete (Table 1; Figure 5). For the Zygnematophyceae, the observed genome size variation is attributable to combinations of all the abovementioned factors (Table 1; Figure 5). Only half of the six newly sequenced zygnematophycean genomes display an IR: the 12.6-kb IR of *Roya obtusa* is about two-fold smaller than those of *Closterium* and *Cosmarium*, but is about the same size as that found in the coleochaetophycean *Chaetosphaeridium*. The newly sequenced zygnematophycean genomes also differ greatly in intron content. For instance, the *Closterium* and *Cosmarium* genomes, which are the largest among the examined zygnematophyceans, display the lowest and highest numbers of introns, respectively; moreover, they exhibit the highest proportion of repeats ≥ 30 bp (Table 1). Note that the alignment of the *Roya obtusa* and *Roya anglica* cpDNAs revealed that these genomes are colinear over their entire length and that



Gene Content Standard Genes

All compared genomes share a set of 90 genes coding for three rRNAs (*rrs*, *rml* and *rfl*), 24 tRNAs, and 63 proteins (see legend of

Figure 7). Therefore, of the 144 standard genes predicted to have been present in the common ancestor of all streptophytes, 54 (42 protein-coding genes, 11 tRNA genes and the *ssrA* gene coding for tmRNA) experienced losses from the chloroplast during

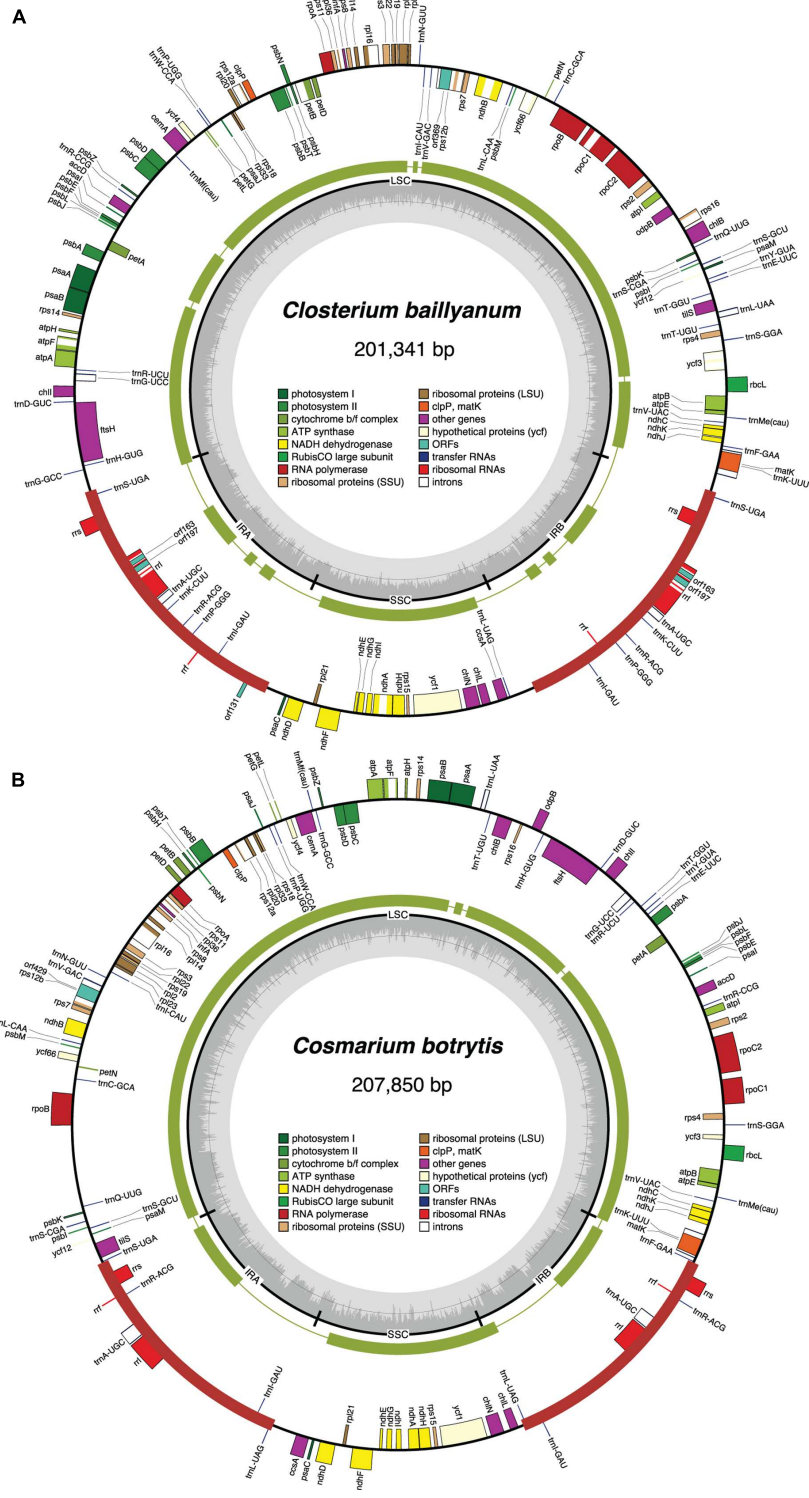


FIGURE 4 | Gene maps of the *Closterium* (A) and *Cosmarium* (B) chloroplast genomes. Filled boxes represent genes, with colors denoting gene categories as indicated in the legend; the red thick lines denote the positions of the IR sequences. Genes on the outside of each map are transcribed counterclockwise; those on the inside are transcribed clockwise. The green thick lines in the second outermost middle rings represent the gene clusters conserved between the *Closterium* and *Roya* cpDNAs (A) and the *Cosmarium* and *Staurastrum* cpDNAs (B). The innermost ring indicates the positions of the IR, SSC and LSC regions. The innermost ring indicates the positions of the IR, SSC and LSC regions, and the gray histogram represents the G + C percentages calculated with OGDRAW (light gray, A + T; dark gray, G + C).

TABLE 1 | General features of the streptophyte green algal chloroplast genomes compared in this study.

Taxon	Accession no. ^a	Size (bp)			G+C (%)		Genes		Introns (no.)		Repeats ^b (%)
		Total	IR	LSC	SSC	No. ^c	% genome ^d	Group I	Group II ^e		
Mesostigmatoephyceae											
<i>Mesostigma viride</i> NIES 296	NC_002186	118,360	6,057	83,627	22,619	30.1	137	73.2			0.2
Chlorokybophyceae											
<i>Chlorokybus atmophyticus</i> SAG 48.80	NC_008822	152,254	7,640	109,098	27,876	38.2	138	58.8	1		0.5
Klebsormidiophyceae											
<i>Klebsormidium</i> sp. SAG 51.86 ^f	KU646497*	> 130,962 ^g				40.6	112	67.7		12 (2)	0
<i>Klebsormidium flaccidum</i> SAG 121.8	NC_024167	176,832	51,118	72,779	1,817	42.0	114	76.3		12 (2)	0.2
<i>Entransia fibriliata</i> UTEX LB 2353	KU646490*	206,025	60,590	75,629	9,216	33.0	118	64.4	1	10 (1)	0.4
Charophyceae											
<i>Chara vulgaris</i>	NC_008097	184,933	10,919	135,815	27,280	26.2	127	60.9	2	16 (1)	1.6
Coleochaetophyceae											
<i>Chaetosphaeridium globosum</i> M1311	NC_004115	131,183	12,431	88,682	17,639	29.6	125	76.9	1	17 (1)	0.7
<i>Coleochaete scutata</i> SAG 110.80M	KU646493*	107,236				27.6	117	79.9	1	14 (0)	0.6
Zygnematophyceae											
<i>Mesotaenium endlicherianum</i> SAG 12.97	NC_024169	142,017				42.2	124	68.4	1	16 (1)	0.6
<i>Zygnema circumcarinatum</i> SAG 698-1a	NC_008117	165,372				31.1	125	57.8	1	12 (1)	1.2
<i>Cylindrocapsa brevissonii</i> SAG 615-1	KU646495*	136,938				29.8	122	63.8	1	9 (1)	1.5
<i>Spirogyra maxima</i> UTEX LB 2495	KU646489*	129,954				30.1	124	73.3	1	16 (1)	1.0
<i>Netrium digitus</i> UTEX LB 561	KU646491*	131,804				31.4	125	74.5	1	13 (1)	0.5
<i>Roya anglica</i> ACCI 799	NC_024168	138,275	12,568	92,926	20,213	33.1	122	69.7	1	9 (1)	0.1
<i>Roya obtusa</i> SAG 168.80	KU646496*	138,272	12,568	92,924	20,212	33.1	122	69.9	1	9 (1)	0.1
<i>Closterium baillyanum</i> SAG 50.89	KU646494*	201,341	26,784	120,746	27,027	39.3	124	53.1	4	15 (1)	4.9
<i>Cosmarium botrytis</i> UTEX 175	KU646492*	207,850	24,465	132,863	26,057	39.8	122	47.9	1	6 (1)	6.0
<i>Staurastrum punctulatum</i> SAG 679-1	NC_008116	157,089				32.5	122 ^h	58.4	1	7 (1)	0.3

^aThe asterisks denote the nine genomes described here for the first time. ^bNon-overlapping repeat elements were mapped on each genome with RepeatMasker using the repeats ≥ 30 bp identified with REPuter as input sequences. ^cOnly standard genes conserved in green plant chloroplast genomes are included in these values. Duplicated genes were counted only once. ^dThese values include the sequences of standard genes and ORFs coding for proteins of known functions. ^eThe first value provides the total number of group I introns; the number of trans-spliced introns is given in parentheses. ^fThis species was formally called *Microspora stagnorum* (Mikhailiyuk et al., 2008). ^gThe exact size of the *Klebsormidium* sp. SAG 51.86 genome could not be determined because the IR/SSC junction was not identified. The features of this genome were estimated from the partial sequence. ^hThis value includes the *trnS(cgg)* pseudogene found in the *Staurastrum* genome (Turner et al., 2005).

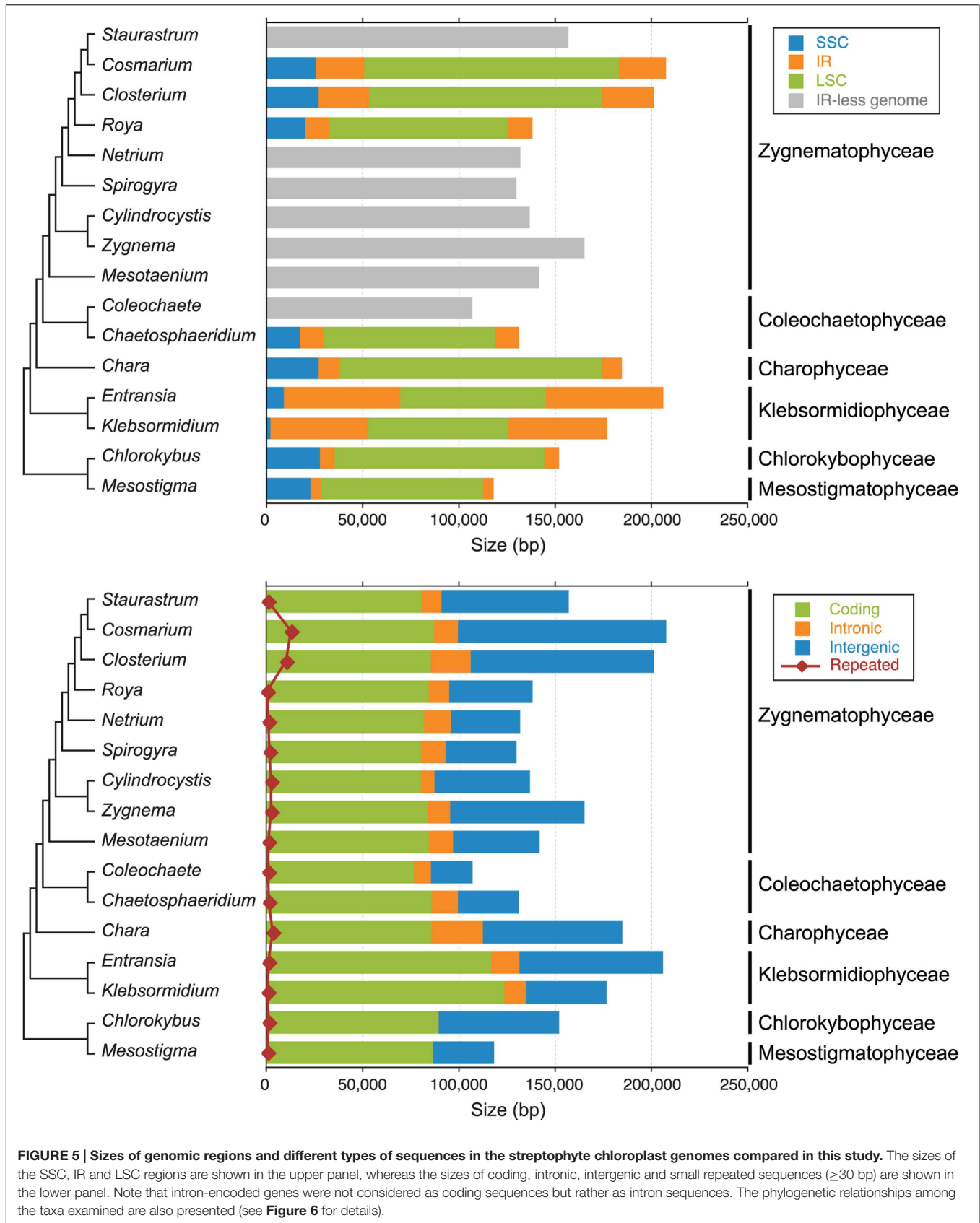


FIGURE 5 | Sizes of genomic regions and different types of sequences in the streptophyte chloroplast genomes compared in this study. The sizes of the SSC, IR and LSC regions are shown in the upper panel, whereas the sizes of coding, intronic, intergenic and small repeated sequences (≥ 30 bp) are shown in the lower panel. Note that intron-encoded genes were not considered as coding sequences but rather as intron sequences. The phylogenetic relationships among the taxa examined are also presented (see **Figure 6** for details).

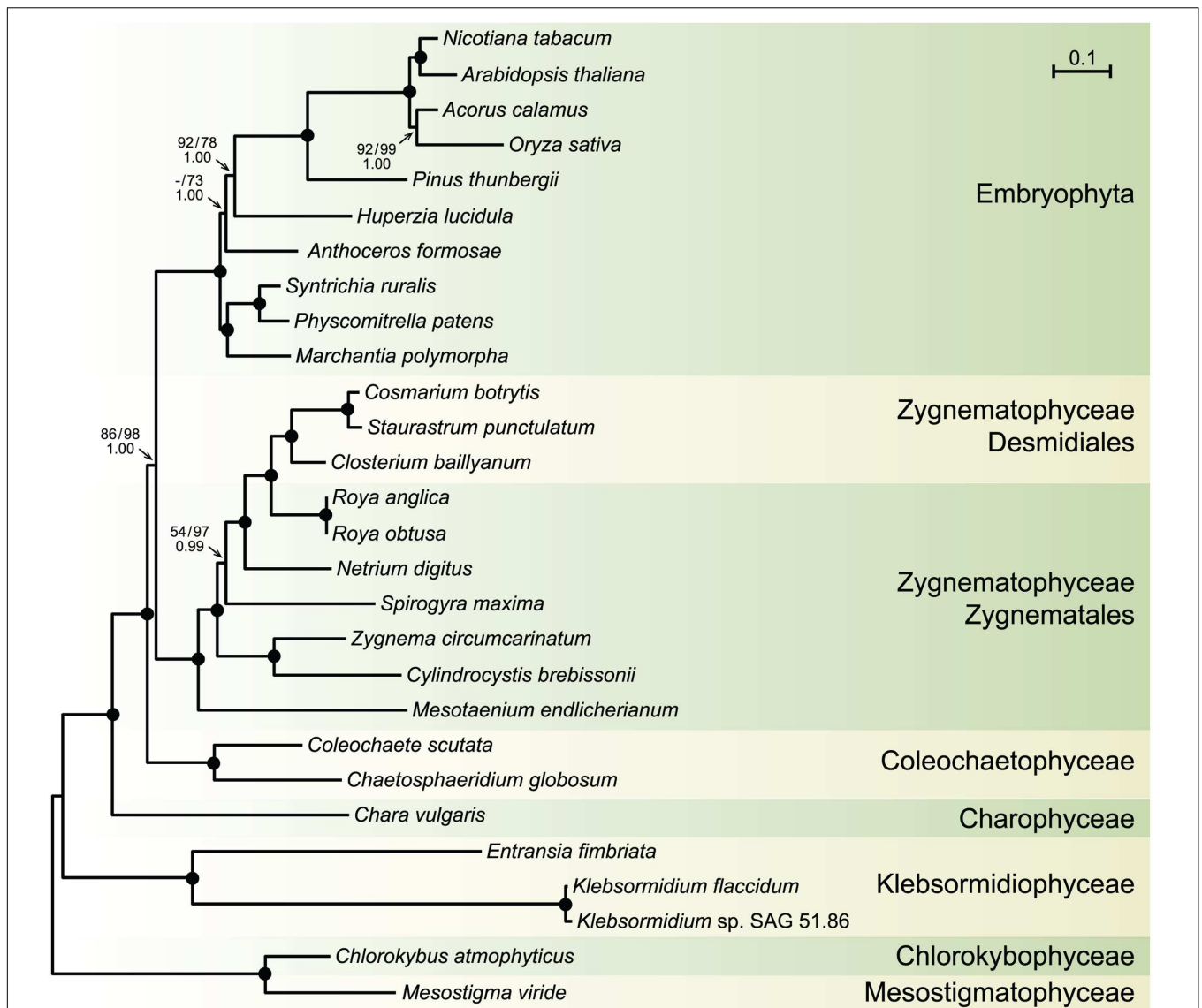


FIGURE 6 | Phylogeny of streptophytes inferred using amino acid and nucleotide data sets assembled from 88 chloroplast protein-coding genes. The best-scoring tree inferred with Phylobayes from the PCG-AA data set under the CATGTR + Γ 4 model is presented. *Mesostigma* and *Chlorokybus* were used as outgroup. Bootstrap support (BS) and posterior probability (PP) values are reported on the nodes: on top from left to right, are shown the BS values for the RAxML analyses of the PCG-AA and PCG12 data sets, respectively; the PP values of the Phylobayes CATGTR + Γ 4 analysis are presented under the BS values. A black dot indicates that the corresponding branch received maximum support in all three analyses; a dash represents a BS value <50%. The scale bar denotes the estimated number of amino acid substitutions per site. The GenBank accession numbers of the streptophyte algal chloroplast genomes included in the PCG-AA and PCG12 data sets are provided in **Table 1**. The GenBank accession numbers of the land plant chloroplast genomes are as follows: *Marchantia polymorpha* NC_001319, *Physcomitrella patens* NC_005087, *Syntrichia ruralis* NC_012052, *Anthoceros formosae* NC_004543, *Huperzia lucidula* NC_006861, *Pinus thunbergii* NC_001631, *Oryza sativa* NC_001320, *Acorus calamus* NC_007407, *Arabidopsis thaliana* NC_000932, *Nicotiana tabacum* NC_001879.

diversification of streptophyte algae (**Figure 7**). A total of 103 gene losses was inferred by mapping these genes on the phylogeny shown in **Figure 6**. Only 26 of these events occurred once, many of which represent signatures uniting different streptophyte classes or divergent lineages belonging to the same class.

Subsequent to the losses of three genes during the evolutionary period preceding the emergence of the common ancestor of *Chaetosphaeridium* and *Coleochaete*, eight genes disappeared in the lineage leading to *Coleochaete*. The reduced number of genes

in klebsormidiophyceans is the result of eight gene losses that occurred during the interval leading to the common ancestor of the *Klebsormidium* and *Entransia* genera and of 22 gene losses that occurred in the lineages leading to *Klebsormidium* and *Entransia*. Note that the *rrf* gene encoding the 5S rRNA is not included in the list of *Klebsormidium* losses; Civan et al. (2014) failed to identify this gene in [GenBank:NC_024167] but we discovered its highly divergent sequence at coordinate positions 92980–93102 and 156632–156510 during the course of

The gene loss scenario discussed above does not include *trnK*(cuu), a tRNA gene found exclusively in the zygneataleans *Mesotaenium* and *Closterium*. BlastN similarity searches against the non-redundant database of NCBI indicated that this gene arose from duplication and subsequent sequence divergence of *trnN*(guu). Prior to our study, the trebouxiophycean *Stichococcus bacillaris* was the only known green alga carrying *trnK*(cuu) in its chloroplast; however, it was found to originate from duplication of *trnK*(uuu) (Turmel et al., 2015).

Increased Mutation Rates of *tufA* in the Coleochaetophyceae

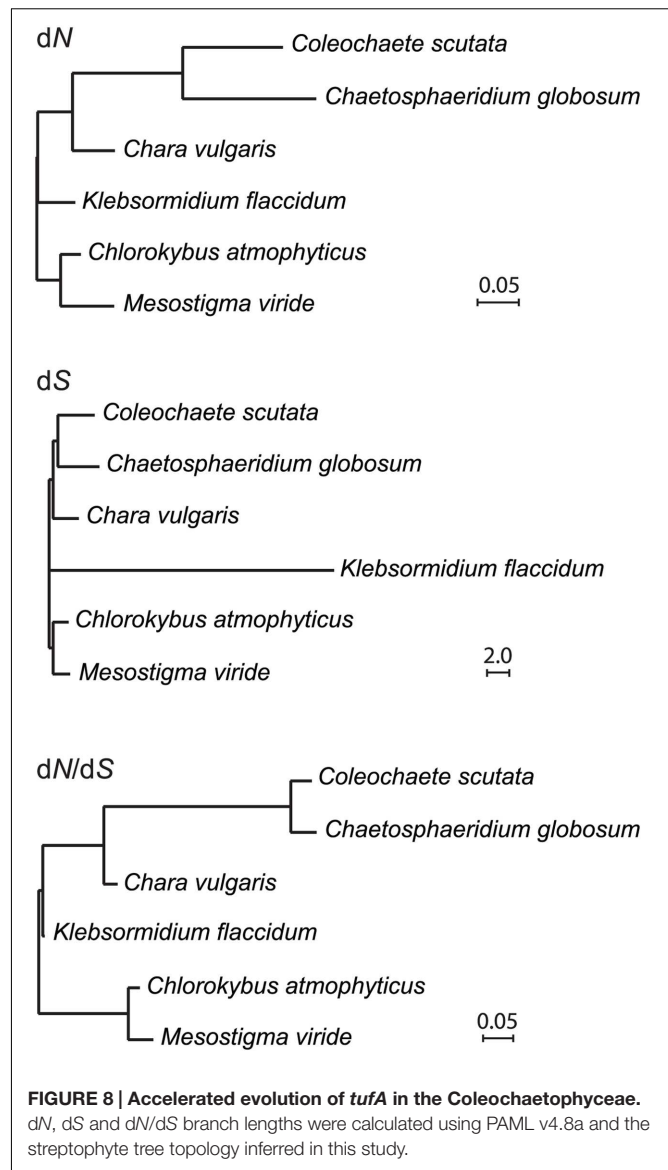
The *Coleochaete scutata* *tufA* gene encoding the protein synthesis factor Tu has retained an intact open reading frame but like its *Chaetosphaeridium* and *Coleochaete orbicularis* counterparts, it is highly divergent in sequence compared to the corresponding genes found in *Mesostigma*, *Chlorokybus* and *Chara*. Our alignment of streptophyte *tufA* sequences also revealed a marked G + C bias at the third codon positions of the *Klebsormidium* sequence. We calculated dN, dS and dN/dS branch lengths for *tufA* based on the relationships we inferred in **Figure 6** and found that the dN branches subtending and within the *Coleochaete* + *Chaetosphaeridium* clade are very long compared to all other examined streptophyte algae (**Figure 8**). The long branch subtending the *Coleochaete* + *Chaetosphaeridium* clade in the dN/dS tree might suggest that *tufA* experienced positive selection during early evolution of coleochaetophyceans. However, no evidence for positive selection across the *tufA* sequence ($P < 0.1$) was obtained in the likelihood ratio test implemented in the PARRIS module of Datamonkey (Delport et al., 2010).

Non-standard Genes

Freestanding ORFs showing similarities (*E*-value threshold of $1e-08$) with recognized protein domains were identified in zygneatophycean chloroplast genomes (**Table 2**). They encode type II intron maturases (*MatK*), a phage DNA primase and DNA breaking-rejoining enzymes (recombinase/integrase).

The *Roya obtusa* genome contains an ORF (*orf230*, coordinates 102713-103405 in GenBank:KU646496) that is similar to chloroplast ORFs present in *Roya anglica* and two ferns of the Ophioglossaceae, *Mankyua chejuensis* and *Ophioglossum californicum*. In addition, the *psbJ-petA* intergenic regions of *Netrium*, *Roya obtusa*, *Closterium*, and *Cosmarium* display an ORF encoding a hypothetical protein resembling those encoded at the same genomic locations in *Zygnema*, *Roya anglica*, and *Staurastrum*. The alignment of these hypothetical proteins indicates that the N-terminal portion is the most conserved (Supplementary Figure S2); however, given the small size and close proximity of this region to the 3' end of *petA*, the corresponding DNA sequence possibly represents a conserved regulatory sequence.

In *Klebsormidium* sp. SAG 51.86 cpDNA, *orf453* encodes a protein with reverse transcriptase (RT) and intron maturase domains that is highly similar to a freestanding ORF annotated as *matK* in the corresponding sequence of the *Klebsormidium flaccidum* genome. This coding sequence is more likely to be part



of the group II intron fragment linked with *psbA* exon 1, and consistent with this view, it shows less similarity to the *matK* genes encoded by *trnK*(uuu) introns than with the ORFs encoded by other group II introns (e.g., those in *Pyramimonas parkeae atpB*, *Tydemania expeditionis psbC* and *Jenufa minuta psbB*).

Intron Distribution

Introns in streptophyte algal cpDNAs are inserted at 38 distinct sites, 22 of which are shared with most land plants [21 group II intron sites and the *trnL*(uaa)₃₅ group I intron site]. Only four sites hold group I introns (*rrl*₂₄₄₉, *rrl*₂₅₀₀, *rrl*₂₅₉₃, and *trnL*(uaa)₃₅). **Figure 9** shows the scenario of chloroplast intron gains and losses that we reconstructed by mapping the presence/absence of introns on the phylogeny inferred in this study. All streptophyte algal genomes, except *Mesostigma* and *Klebsormidium* cpDNAs, contain the *trnL*(uaa)₃₅ intron. Group II introns shared with land plants were acquired during four

TABLE 2 | Non-standard genes identified as freestanding ORFs in the chloroplast genomes compared in this study.

Taxon	ORF ^a	Coordinates ^b	Conserved domain
<i>Mesotaenium endlicherianum</i>	519	108900–110459	Type II intron maturase MatK (CHL00002)
<i>Zygnema circumcarinatum</i>	531	66206–67801	Type II intron maturase MatK (CHL00002)
<i>Spirogyra maxima</i>	659	72632–74611	Phage integrase family (pfam00589) and integrase core domain (pfam00665)
<i>Netrium digitus</i>	132	102944–103342	Phage integrase family (pfam00589)
<i>Roya obtusa / Roya anglica</i>	268	96776–97582	Phage integrase family (pfam00589)
<i>Staurostrum punctulatum</i>	108	142226–142552	Phage DNA primase, D5 N terminal like domain (pfam08706)

^aReported here are the freestanding ORFs larger than 100 codons that revealed similarity (*E*-value threshold of $1e-08$) with recognized protein domains in BlastP searches. Each ORF is identified by the number of amino acid residues in the encoded protein. ^bGenomic coordinates of the ORFs in the GenBank accessions provided in **Table 1**.

distinct evolutionary intervals (these branches are denoted by roman numerals in **Figure 9**), but most of these introns were later lost on one or more occasions. Within the Zygnematophyceae, 43 events of intron losses involving 18 sites were recorded; only the *trans*-spliced *rps12_114* and the *cis*-spliced *rpl16_9* and *trnG(ucc)_23* introns were spared from losses. The *trnI(gau)_39* and *trnV(uac)_37* introns were lost before the emergence of this clade.

The only ORF-containing group II introns that streptophyte algae share with land plants are the *trnK(uuu)_37* and *trans*-spliced *rps12_114* introns, but the ORF is not conserved in all algal taxa carrying these introns. The ORF encoding the intron maturase MatK is missing from the *trnK(uuu)_37* introns of *Coleochaete* and the Klebsormidiophyceae, and there is no *trnK(uuu)_37* intron in the early diverging zygnematophyceans *Mesotaenium*, *Cylindrocystis* and *Zygnema*; however, the *matK* ORF is freestanding in both *Mesotaenium* and *Zygnema* (**Table 2**). An ORF is also absent from the *trans*-spliced *rps12_114* intron in *Klebsormidium*, *Chara*, *Spirogyra*, and the three zygnematophyceans just mentioned.

The group II introns that are not shared with land plants occur in *Klebsormidium* (sites 3, 4, 6, 10, 12, 15, and 20), *Entransia* (sites 11, 17, 21, and 33), *Chara* (site 1) and *Netrium* (site 16) (**Figure 9**). Only two of these lineage-specific introns (the *Klebsormidium psbA* and *Netrium psbE* introns) harbor an ORF coding for a putative protein with RT and/or maturase domains. The group I introns that are not shared with land plants are found in *Chara* (site 36) and *Closterium* (sites 35, 36, and 37).

Gene Organization

Comparative analyses of gene organization were carried out using four complementary approaches. First, syntenic regions were identified in pairwise genome comparisons (**Figures 1–4**). Second, we compared the gene partitioning patterns of the IR-containing genomes and examined whether the genes found in the IR, SSC and LSC are also clustered in IR-lacking genomes (**Figure 10**). Third, phylogenetic trees based on gene order were inferred using the MLGO web server (Hu et al., 2014) and MGR v2.03 (Bourque and Pevzner, 2002); MLGO reconstructs ML phylogenies based on gene adjacency, whereas MGR estimates the number of reversals required to interconvert gene order in pairs of genomes and construct tree topologies based on rearrangement distance (**Figure 11**). The data set analyzed with MLGO contained all standard genes in each genome (including

both copies of duplicated genes), while the data set analyzed with MGR was restricted to the 89 genes common to all compared genomes. Finally, the presence/absence of signed gene pairs in three or more genomes (**Figure 12**) were coded as binary Dollo characters and the gene pairs representing synapomorphic gains and losses were mapped on the streptophyte phylogeny reported in this study using MacClade v4.08 (**Figure 13**). The Dollo principle assumes that characters can be lost independently in several evolutionary lineages but cannot be regained. Because inversion endpoints have been shown to be reutilized multiple times in land plant chloroplast genomes experiencing frequent rearrangements (Jansen and Ruhlman, 2012), coding of gene pairs as Dollo characters might not be fully justified. For this reason, we have also coded gene pairs as unordered (Fitch parsimony) or ordered (Wagner parsimony) characters and found no difference in the evolutionary scenarios we inferred.

Although gene order is highly variable among streptophyte algal genomes, the branching order inferred with MLGO is entirely congruent with the phylogeny based on gene and protein sequences (**Figures 6 and 11**). The reversal distances estimated by MGR reveal that the chloroplast genomes of the bryophytes, Coleochaetophyceae and especially *Chara* are the least rearranged relative to those of *Mesostigma* and *Chlorokybus*, and that they are followed closely by the klebsormidiophycean genomes. By comparison, the Zygnematophyceae display very long branches indicative of massive gene shuffling. Of the examined zygnematophyceans, the representative of the earliest-diverging lineage (*Mesotaenium*) boasts the least rearranged genome.

The *Coleochaete* and *Chaetosphaeridium* cpDNAs share nine blocks of sequences containing 116 of their 117 and 125 conserved genes, respectively (**Figure 2**). Only eight reversals are required to interconvert the gene order of these two genomes. Following the loss of the IR in the *Coleochaete* lineage, the partitioning of genes between the SSC and LSC regions has been barely affected (**Figure 10**); in other words, the genes corresponding to those located 5' and 3' of the rRNA operon in the *Chaetosphaeridium* genome have largely remained clustered in *Coleochaete*. Three of the eight genes missing in *Coleochaete* (*rps14*, *tilS* and *ccsA*) correspond to breakpoints between syntenic blocks while the remaining, which are all part of ribosomal protein operons, have been cleanly excised from internal regions of syntenic blocks (**Figures 2 and 12**).

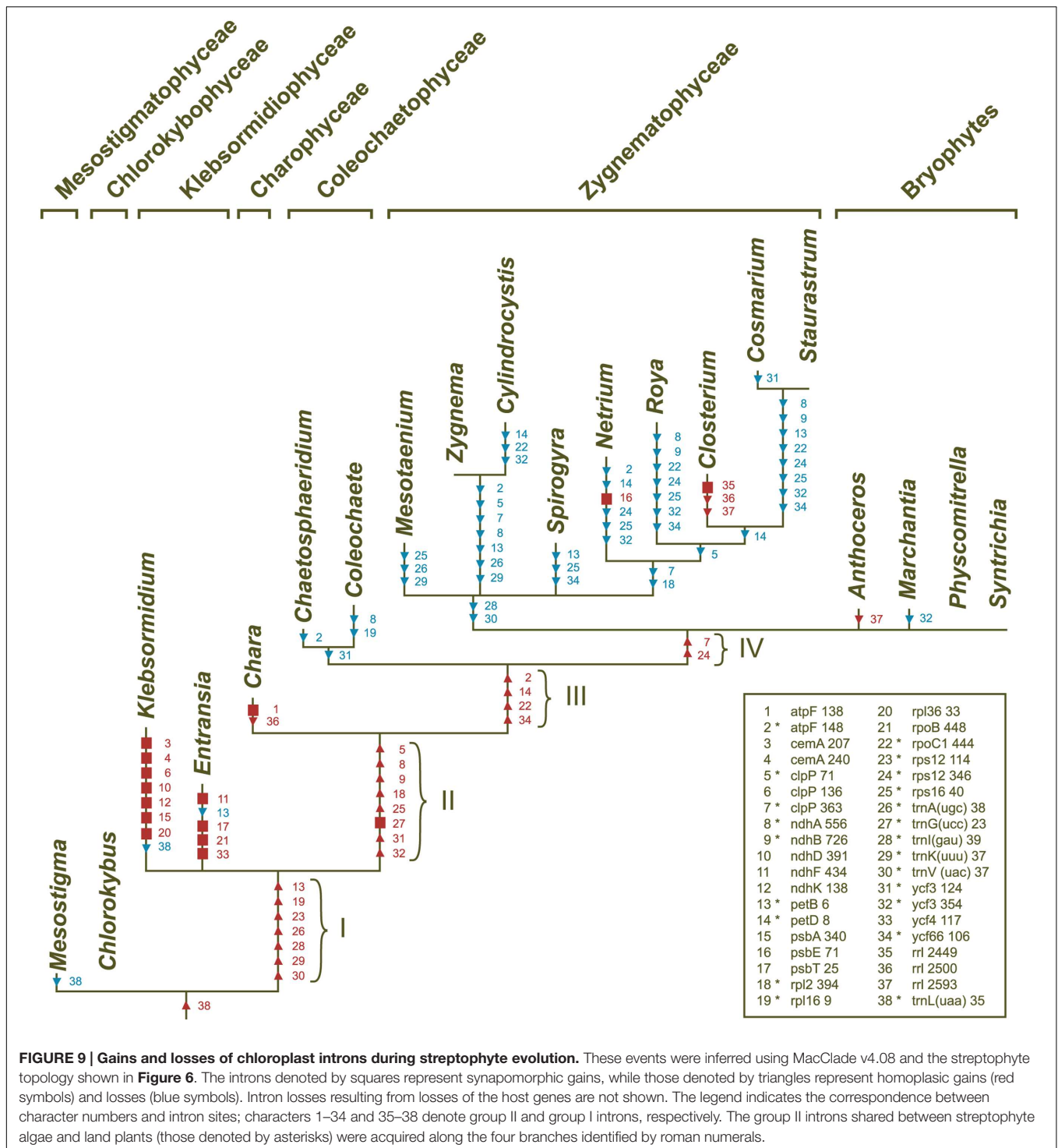


FIGURE 9 | Gains and losses of chloroplast introns during streptophyte evolution. These events were inferred using MacClade v4.08 and the streptophyte topology shown in **Figure 6**. The introns denoted by squares represent synapomorphic gains, while those denoted by triangles represent homoplastic gains (red symbols) and losses (blue symbols). Intron losses resulting from losses of the host genes are not shown. The legend indicates the correspondence between character numbers and intron sites; characters 1–34 and 35–38 denote group II and group I introns, respectively. The group II introns shared between streptophyte algae and land plants (those denoted by asterisks) were acquired along the four branches identified by roman numerals.

The *Klebsormidium flaccidum* and *Entransia* genomes display 26 syntenic blocks containing 85 of their 114 and 118 conserved genes, respectively (**Figure 1**). The vastly expanded IRs of these algae differ considerably from one another and from streptophyte IRs displaying an ancestral organization (i.e., those of *Mesostigma*, *Chlorokybus*, *Chara*, coleochaetophyceans and bryophytes) with respect to gene content (**Figure 10**). Of the 32

and 44 genes present in the *Klebsormidium* and *Entransia* IRs, respectively, only 14 are shared besides the five genes making up the rRNA operon; however, all genes typically found 5' and 3' of the rRNA operon in ancestral IR-containing genomes, except for *tufA* in *Klebsormidium*, have remained ancestrally segregated. Eight of the 22 genes lost in the lineages leading to *Klebsormidium* and *Entransia* [*accD*, *minD*, *psaM*, *rpl12*, *rpl32*, *tufA*, *trnV(uac)*

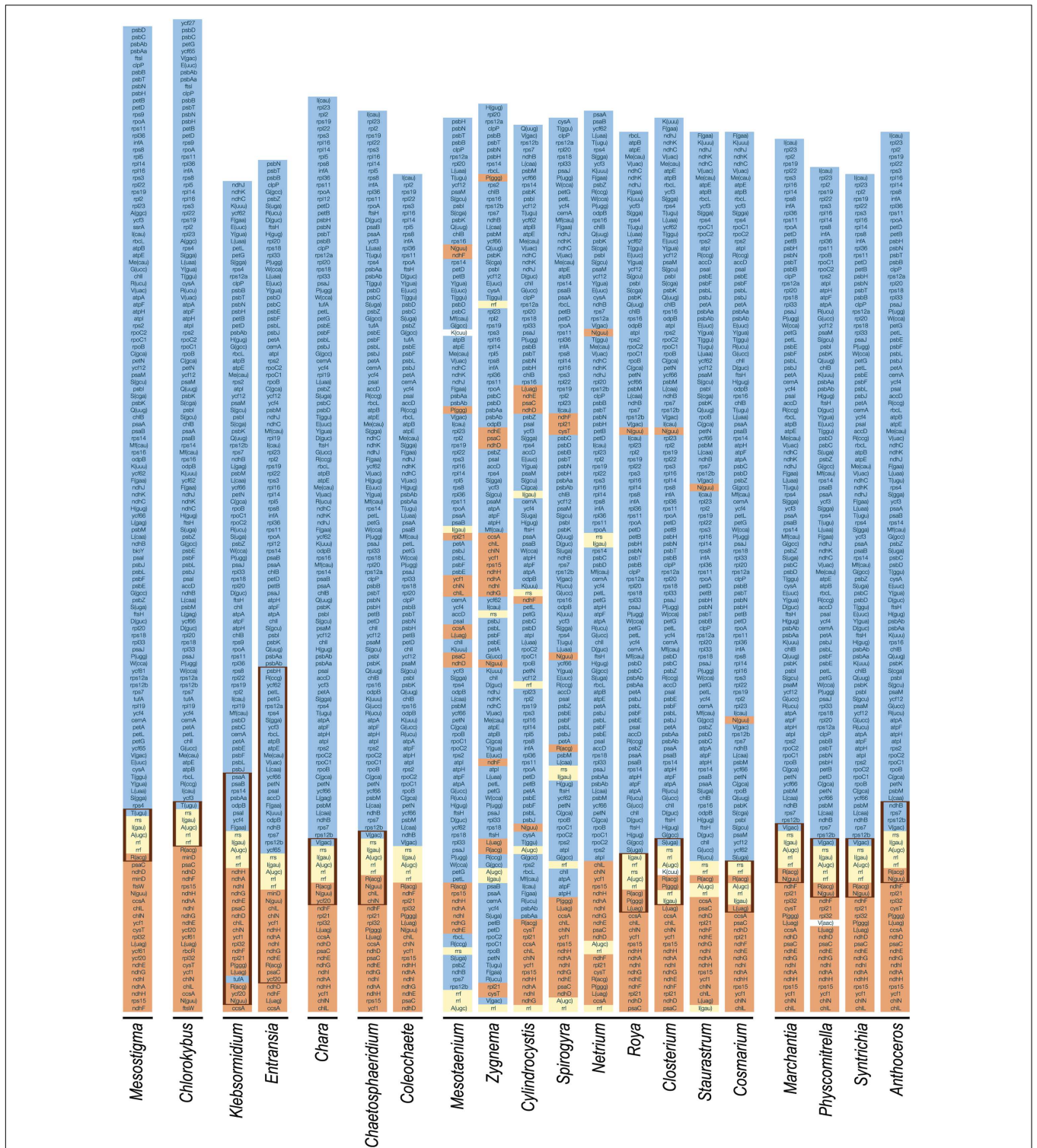
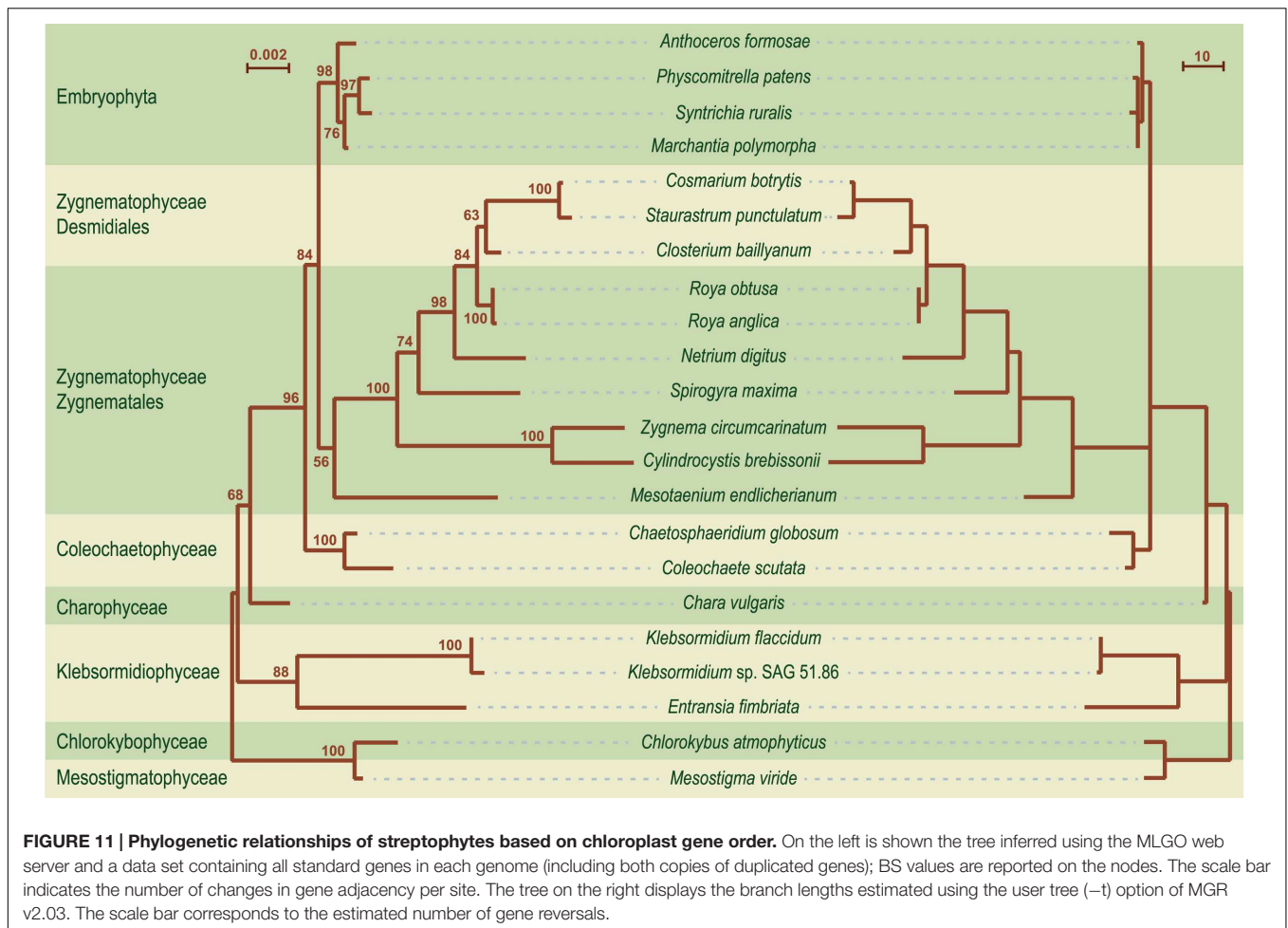


FIGURE 10 | Gene partitioning patterns of streptophyte chloroplast genomes. For each genome, one copy of the IR (thick vertical lines) and the entire SSC and LSC regions are represented. The five genes composing the rDNA operon are highlighted in yellow. The color assigned to each of the remaining genes is dependent upon the position of the corresponding gene relative to the rDNA operon in the cpDNA of the streptophyte alga *Mesostigma viride*, a genome displaying an ancestral gene partitioning pattern. The genes highlighted in orange are found within or near the SSC region in this streptophyte genome (downstream of the rDNA operon), whereas those highlighted in blue are found within or near the LSC region (upstream of the rDNA operon). To simplify the comparison of gene order, some genomes are represented in their alternative isomeric forms relative to the genome sequences deposited in GenBank. Note that the *trnK*(*cuu*) genes of *Mesotaenium* and *Closterium*, and the *trnV*(*acc*) gene of *Physcomitrella* were not color-coded because their evolutionary origins are unclear.



and *ycf65*] correspond to breakpoints between synteny blocks, while the others have been cleanly excised from internal regions of synteny blocks (Figures 1 and 12). Gene rearrangements have disrupted some ancestral gene clusters, e.g., *rpoC2* is no longer adjacent to *rps2* and *atpI* is no longer beside *atpH* (Figure 12).

The chloroplast genomes of the Zygnematophyceae exhibit a wide range of divergence at the gene order level (Figure 11). The *Cosmarium* and *Staurastrum* genomes are the most similar, with six syntenic blocks containing all 122 encoded genes except *trnS(uga)* (Figure 4). All investigated zygnematophycean genomes feature a disrupted rRNA operon; two, three or four breakage sites are observed depending on the species (Figure 12). Like the IRs of the *Roya* species, those of *Cosmarium* and *Closterium* harbor two to four tRNA genes in addition to the genes composing the rRNA operon (Figure 10). Of these tRNA genes, only *trnR(acg)* is shared with the IRs of streptophyte algal cpDNAs having an ancestral partitioning pattern. All the genes ancestrally located in the SSC and LSC regions, with the single exception of *trnN(guu)*, still form ancestrally segregated groups in the *Roya*, *Cosmarium*, and *Closterium* genomes (Figure 10). In contrast, reshuffling of gene order in the *Mesotaenium*, *Zygnum* + *Cylindrocystis* and *Spirogyra* lineages led to extensive dispersal, throughout the genome, of the IR-encoded genes and of

the genes typically found in the SSC and LSC regions of ancestral streptophyte IR-containing cpDNAs (Figure 10).

Several gene pairs representing synapomorphic signatures of distinct lineages were lost or acquired before the Klebsormidiophyceae, Coleochaetophyceae and bryophyte lineage each arose. But just a single synapomorphy, corresponding to loss of linkage between the *trnI(gau)* and *trnA(ugc)* genes in the rDNA operon, unites the Zygnematophyceae (Figure 13). Following the emergence of the *Mesotaenium* lineage, two additional gene rearrangements occurred in the common ancestor of the remaining zygnematophyceans: the rDNA operon was broken at a second site (between *rri* and *rri'*) and the ancestral pair *5'psbZ-5'trnS(uga)* was lost. Reversal of the *ycf1* gene in the ancestral *3'chlN-3'ycf1* pair is a unique gene rearrangement shared by all Coleochaetophyceae, Zygnematophyceae and bryophytes. In contrast, no synapomorphic loss of gene pairs accompanied the emergence of the Charophyceae.

DISCUSSION

The comparative analyses presented in this study, which include a sampling of nine additional chloroplast genomes

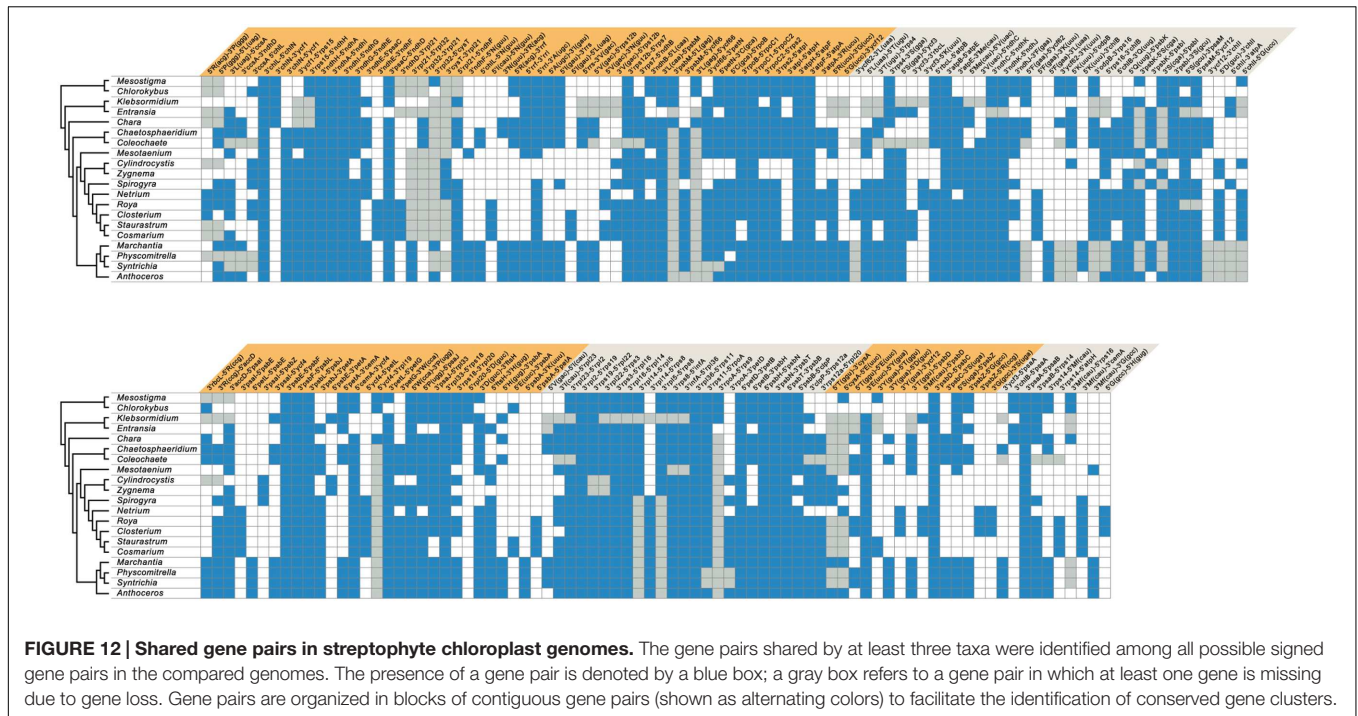


FIGURE 12 | Shared gene pairs in streptophyte chloroplast genomes. The gene pairs shared by at least three taxa were identified among all possible signed gene pairs in the compared genomes. The presence of a gene pair is denoted by a blue box; a gray box refers to a gene pair in which at least one gene is missing due to gene loss. Gene pairs are organized in blocks of contiguous gene pairs (shown as alternating colors) to facilitate the identification of conserved gene clusters.

from the Klebsormidiophyceae, Coleochaetophyceae and Zygnematophyceae, alter markedly our view of chloroplast genome evolution in streptophytes. Our results clearly indicate that the chloroplast genome is evolving in a dynamic fashion not only in the Zygnematophyceae but also in the Klebsormidiophyceae and Coleochaetophyceae. The diversity of genomic structures and organizations found in these three classes is reminiscent of the diversity that has recently been reported for various classes of the Chlorophyta (Brouard et al., 2010; Lemieux et al., 2014; Leliaert and Lopez-Bautista, 2015; Turmel et al., 2015), and contrasts with the extremely conservative evolutionary trend observed in most land plants (Wicke et al., 2011; Jansen and Ruhlman, 2012; Ruhlman and Jansen, 2014). In the following sections, we highlight the evolutionary trends observed in each streptophyte algal class.

The phylogenies we inferred from 88 chloroplast protein-coding genes and proteins are congruent with recent phylogenomic studies indicating that the Zygnematophyceae is sister to land plants (Wodniok et al., 2011; Laurin-Lemay et al., 2012; Timme et al., 2012; Zhong et al., 2013; Wickett et al., 2014). The relationships among the members of the Zygnematophyceae are consistent with the phylogeny reported by Gontcharov et al. (2004). Moreover, the differences in branching order of the bryophyte lineages between the protein and gene trees are in agreement with previously reported phylogenomic studies (Cox et al., 2014; Wickett et al., 2014).

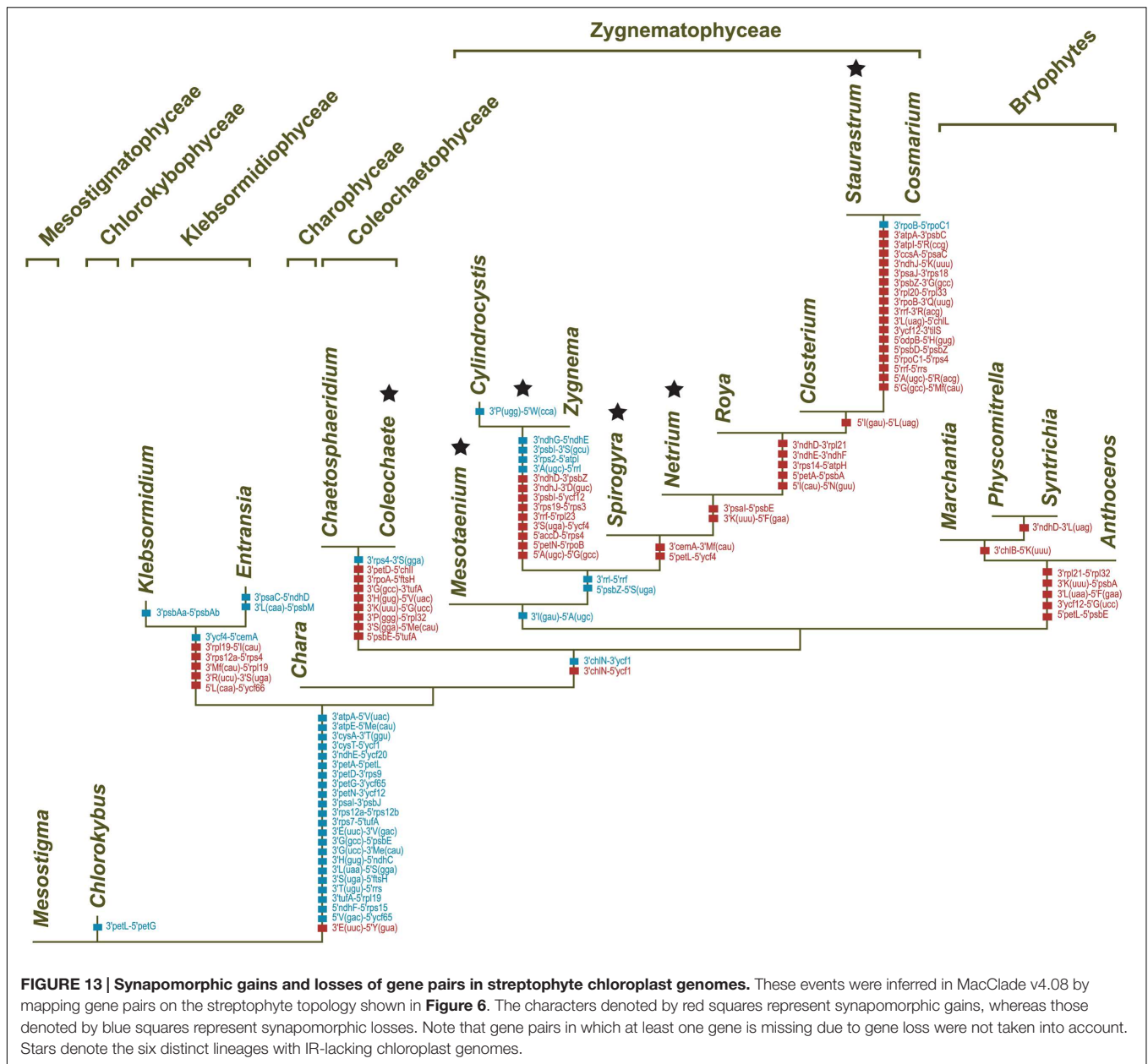
Mesostigmatophyceae and Chlorokybophyceae

The *Mesostigma* and *Chlorokybus* chloroplast genomes are the most rich in ancestral traits among the Viridiplantae (Lemieux

et al., 2000, 2007). They feature the most extensive gene content, are almost devoid of introns, contain many operons typically found in cyanobacteria, and have retained a quadripartite architecture with a pattern of gene partitioning that closely resembles those found in early diverging members of the Chlorophyta (e.g., *Nephroselmis*; Turmel et al., 1999b; Lemieux et al., 2014 and *Pyramimonas*; Turmel et al., 2009) (Table 1; Figure 10). These ancestral features mirror the deep-branching positions of *Mesostigma* and *Chlorokybus* and illustrate the great structural stability of the chloroplast genome in the clade uniting these algae.

Klebsormidiophyceae

The *Entansia* and *Klebsormidium flaccidum* chloroplast genomes have retained a quadripartite architecture, but their IRs are greatly enlarged and include many genes typically found in the SSC and LSC regions (Figures 1, 5, and 10). Reconstruction of the ancestral genome reveals that the IR underwent considerable expansion toward the SSC region before the split of the two klebsormidiophycean lineages and that it expanded predominantly toward the LSC region following this divergence. At 61 kb, the *Entansia* IR is the largest known among the green algae examined so far. Compared to the *Klebsormidium flaccidum* IR, it contains twice as many genes of LSC origin but fewer genes of SSC origin, suggesting that shifts of both the IR/SSC and IR/LSC junctions are on-going events in the Klebsormidiophyceae. Sampling of additional taxa from this class should provide more information on the directionality and extent of these shifts in various lineages. Like klebsormidiophycean cpDNAs, the chloroplast genomes of the chlorodendrophycean green algae *Scherffelia* and *Tetraselmis* feature enlarged IRs with a rich gene content, but the ancestral



partitioning pattern has not been maintained (Turmel et al., 2016).

Thirty standard chloroplast genes were lost during the evolution of the Klebsormidiophyceae; losses of *ndhK*, *rpl14*, *rpl16*, *rps3*, *trnG(ucc)*, *trnT(ggu)*, *trnT(ugu)*, *trnV(gac)* and *trnV(uac)*, in particular, are unique among streptophyte algae (Figure 7). As a consequence of the substantial losses of *trn* gene, the complement of tRNAs encoded in klebsormidiophycean chloroplast genomes is not sufficient to decode the entire set of codons found in these genomes. For example, there is no chloroplast-encoded tRNA^{Thr} in *Klebsormidium* and *Entransia*, and no chloroplast-encoded tRNA^{Val} in *Klebsormidium*. It thus appears that these missing tRNAs are imported from the cytosol into the chloroplast. Import of nuclear-encoded tRNAs into

plastids has previously been suggested for non-photosynthetic land plants (Wicke et al., 2011).

Klebsormidiophycean chloroplast genomes contain an abundance of introns. Civan et al. (2014) previously inferred that the common ancestor of the Klebsormidiophyceae and its sister lineages (branch I in Figure 9) shared five group II introns with extant land plants. To this set of early acquired introns, we must now add the *rpl16_9* and *trnV(uac)_37* introns, which we identified in the *Entransia* genome. Although the *Klebsormidium flaccidum* and *Entransia* chloroplast genomes contain a large number of lineage-specific group II introns (seven are unique to *Klebsormidium* and four are unique to *Entransia*), they have no intron insertion sites in common other than those shared with land plants, thus suggesting that the lineage-specific introns were

mostly acquired through intragenomic proliferation of founding introns.

It is intriguing that the *tufA* gene resides in the SSC region rather than in the LSC region in the *Klebsormidium* chloroplast genome and that it is entirely missing in the *Entransia* genome. In *Mesostigma*, *Chlorokybus* and most chlorophyte cpDNAs, *tufA* is part of the *str* operon, which also comprises *rps12* and *rps7* (transcription order is 5'-*rps12-rps7-tufA*-3'). It appears that breakage of this operon through the acquisition of a *trans*-spliced group II intron at site 114 of *rps12* soon after the divergence of the *Mesostigma* + *Chlorokybus* clade led to relocation of *rps12* exon 1 outside the operon and ultimately to loss of linkage between *rps7* and *tufA*, which resulted in transfer of *tufA* to the SSC region in *Klebsormidium* and to the complete loss of this gene in *Entransia*. There is also no linkage between *rps7* and *tufA* in the *Chara* and *Chaetosphaeridium* genomes, and both *rps12* and *rps7* are completely missing in *Coleochaete*.

Charophyceae

Only the chloroplast genome sequence of *Chara vulgaris* is currently available for the Charophyceae (Turmel et al., 2006). Remarkably, this alga has retained the largest degree of ancestral traits among the streptophytes that diverged after the *Mesostigma* + *Chlorokybus* clade. This highly conservative evolutionary trend is apparent at all levels, including overall architecture, gene content, gene partitioning, and gene organization (Figure 10). The *Chara* genome is clearly the least rearranged relative to the *Mesostigma* and *Chlorokybus* cpDNAs, as revealed by the short branch length separating these taxa in the MGR tree shown in Figure 11. It will be interesting to see if sampling of additional taxa from the Charophyceae will support the notion that the chloroplast genome is evolving at a very slow rate in this class. In this context, it is noteworthy that the mitochondrial genomes of the distantly related charophycean algae *Chara vulgaris* and *Nitella hyalina* contain the same gene complement and display exactly the same gene order (Turmel et al., 2013).

Coleochaetophyceae

Our chloroplast genome analyses of a second representative of the Coleochaetophyceae unveiled a less conservative evolutionary history than previously thought for this class. Genome streamlining appears to be the main evolutionary force in the lineage leading to *Coleochaete*. The *Coleochaete* genome is both the smallest and most compact among the streptophyte algal cpDNAs investigated so far (Table 1). Unlike the *Chaetosphaeridium* cpDNA (Turmel et al., 2002), it lacks an IR and several standard genes. Notably, losses of four ribosomal protein-coding genes (*rps4*, *rps7*, *rps12*, and *rps14*) represent unique events in the evolutionary scenario we inferred for streptophyte algae (Figure 7). Otherwise, gene organization has been highly preserved in *Coleochaete* and *Chaetosphaeridium* and the genes usually present in the IR, SSC and LSC regions have retained an ancestral partitioning pattern in the *Coleochaete* genome (Figures 10 and 11).

The *tufA* gene is evolving at a fast pace in the Coleochaetophyceae (Figure 8). Considering that this gene

has completely disappeared from the chloroplast genome and is most probably nuclear-encoded in all streptophyte lineages that evolved after the divergence of the Coleochaetophyceae (Figure 7), it is possible that the chloroplast *tufA* sequences identified in the Coleochaetophyceae do not encode the functional elongation factor EF-Tu; instead, a nuclear gene product might play this role in protein synthesis. This hypothesis was proposed earlier by Baldauf et al. (1990) who reported that the chloroplast *tufA* sequence of *Coleochaete orbicularis* is unusually divergent and differs considerably at what was otherwise conserved amino acid positions. These authors speculated that, despite the presence of numerous mutations, long-term maintenance of an intact ORF at the *Coleochaete tufA* locus might be the result of selection to retain less constrained subsets of the original EF-Tu functions. This hypothesis is attractive, considering that EF-Tu has been shown to play an important role in cell shape maintenance in *Bacillus subtilis* through direct interaction with MreB (Defeu Soufo et al., 2010), a protein involved in septum synthesis and cell division (Fenton and Gerdes, 2013); however, it is not supported by our finding that there was no positive selection across the *tufA* sequence in the Coleochaetophyceae. Putative nuclear copies of *tufA* were detected in *Coleochaete orbicularis* by Southern blot analysis (Baldauf et al., 1990), and our BLASP searches against the 1000 Plants (oneKP) database⁸ using the *Chara* chloroplast *tufA* sequence as query identified highly similar sequences (*E*-value threshold of 0.0) that contain all functional domains of EF-Tu in the RNA-seq assemblies of *Coleochaete scutata* (VQBJ-2010477), *Coleochaete irregularis* (QPDY-2029449) and *Chaetosphaeridium globosum* (DRGY-2007378). Taken together, these observations support the notion that the nucleus houses the functional coding sequence for the chloroplast EF-Tu and that the divergent *tufA* sequence in the chloroplast genome is undergoing pseudogenization.

Zygnematophyceae

Previous studies revealed that the zygnematophycean chloroplast genome is highly variable in overall structure, gene order and intron content (Turmel et al., 2005, 2007; Civan et al., 2014). The comparative genome analyses reported here, which include six additional taxa sampled from the Zygnematales and Desmidiaceae, underscore the exceptionally dynamic evolution of this genome.

To account for the presence of an IR in one of the four zygnematophycean taxa they examined (*Roya*), Civan et al. (2014) proposed that the IR was either lost three times or gained once *de novo* during the diversification of the Zygnematophyceae. Our finding that the cpDNAs of the desmidiales *Closterium* and *Cosmarium* also have an IR suggests that the IR was lost a minimum of five times (Figures 5 and 13). It is unlikely that the IR was acquired *de novo* on one or more independent occasions because the ancestral gene partitioning pattern has been retained in late-diverging zygnematophyceans with IR-less chloroplast genomes (*Staurastrum* and *Netrium*). Indeed,

⁸<https://www.bioinfodata.org/Blast4OneKP/home>

considering that the genes originally present in the IR and SSC regions were dispersed throughout the genome as a result of high frequency gene rearrangements in early diverging lineages of the Zygnematophyceae (Figure 10), it is difficult to envision that the region containing all the genes encoded by the rDNA operon became duplicated at exactly the same site as the ancestral LSC/SSC junction in the common ancestor of *Roya*, *Closterium*, *Staurastrum*, and *Cosmarium* and that the ancestral gene partitions were restored concomitantly (Figure 6). The notion that the quadripartite structure was eliminated multiple times in the Zygnematophyceae predicts that future studies with a broader taxon sampling of this class will uncover early diverging taxa harboring IR-containing chloroplast genomes. Independent losses of the IR also took place in other lineages of the Viridiplantae, including land plants (Wicke et al., 2011; Jansen and Ruhlman, 2012; Ruhlman and Jansen, 2014). The quadripartite structure was eliminated at least four times in prasinophytes (Turmel et al., 2009; Lemieux et al., 2014), seven times in the Trebouxiophyceae (de Cambiaire et al., 2007; Turmel et al., 2015), twice in the Ulvophyceae (Lu et al., 2011; Leliaert and Lopez-Bautista, 2015; Melton et al., 2015) and once in the Chlorophyceae (Bélanger et al., 2006; Brouard et al., 2010, 2011). In the case of the Chlorophyceae, the reported loss unites two major lineages of the OCC clade: the Chaetophorales and Chaetopeltidales (Brouard et al., 2010).

The long branches leading to the zygnematophycean taxa in the MGR tree (Figure 11) indicate that the chloroplast genome underwent more extensive gene scrambling in the Zygnematophyceae than in any other streptophyte algal classes. These gene rearrangements were accompanied by the disruption of several ancestral clusters, including the rDNA operon (Figure 13). The latter operon was broken at four distinct sites, three of which are associated with synapomorphic losses of gene pairs: (1) the *trnI(gau)*-*trnA(ugc)* pair in the common ancestor of all zygnematophyceans, (2) the *rrl-rrf* pair following the divergence of *Mesotaenium* and (3) the *rnl-rrf* pair in the common ancestor of *Zygnema* and *Cylindrocystis*. The fourth site, located between *rrs* and *trnI(gau)*, was disrupted along the branches leading to *Mesotaenium*, to the common ancestor of *Zygnema* and *Cylindrocystis*, and to the common ancestor of *Netrium* and its sister clade. Disruptions of the chloroplast rDNA operon are rare events among viridiplants but have been reported in some chlorophyte lineages featuring highly rearranged genomes (de Cambiaire et al., 2007; Turmel et al., 2009, 2015; Lu et al., 2011; Lemieux et al., 2014; Leliaert and Lopez-Bautista, 2015).

In land plants, IR loss and/or acquisition of short dispersed repeats have been associated with an increased rate of genome rearrangements (Palmer, 1991; Wicke et al., 2011; Jansen and Ruhlman, 2012; Weng et al., 2014), but these factors are unlikely to be the main force driving genome rearrangements in zygnematophyceans. Although gene order in the *Zygnema* and *Cylindrocystis* IR-less genomes was reconfigured at a faster rate compared to their IR-containing homologs, this is not the case for other zygnematophycean lineages (*Mesotaenium*, *Spirogyra*, and *Netrium*) that also feature IR-less genomes

but much shorter branches (Figure 11). Similarly, the extent of gene rearrangements and repeat contents are weakly correlated; zygnematophycean genomes are generally poor in short dispersed repeats and those having the highest proportion of these elements (*Cosmarium* and *Closterium* cpDNAs) show minor differences in gene rearrangements (Figure 11) compared to their closest relatives having much fewer repeats (*Staurastrum* and *Roya*).

In agreement with the study of Civan et al. (2014), we inferred that the common ancestor of zygnemataleans harbored all the 21 group II introns usually present in land plant genomes, except those in *trnI(gau)*, a gene that was part of the ancestral rDNA operon, and in *trnV(uac)*. Just three introns — *rpl16_9*, *rps12_114* and *trnG(ucc)_23* — were retained in all ten examined taxa. While Civan et al. (2014) predicted 20 independent events of intron loss in the Zygnematophyceae, we scored 43 losses in our study (Figure 9). Two introns were lost only once (those at sites 18 and 31 in Figure 9), six were lost twice (sites 2, 5, 7, 9, 26, 29), and the remaining introns on three, four or five occasions. The underlying cause of this intron instability remains unclear. It has been previously speculated that intron losses could be the result of retroposition events (reverse transcription of a spliced RNA copy, followed by recombination-dependent insertion into the genome) and that the protein encoded by the *trans*-spliced *rps12_114* intron in several zygnematophycean lineages could provide the RT activity required for these events (Turmel et al., 2005; Civan et al., 2014). Although RT-mediated intron loss is a mechanism that is very efficient in removing introns (Cohen et al., 2012), very few zygnematophycean genomes actually encode a protein with this activity. Our BLAST analyses of group II intron-encoded ORFs revealed that only the *orf643* in the *psbE_71* group II intron of *Netrium* (an intron unique to this alga) codes for a putative RT; all other intron-encoded ORFs were found to contain an intron maturase domain, including the *Staurastrum orf404* that we incorrectly annotated as a RT gene in a previous report (Turmel et al., 2005). These observations, however, do not necessarily invalidate the retroposition hypothesis for intron removal in the Zygnematophyceae, as RT genes might have been present in ancestral chloroplast genomes. Alternatively, RT activities of mitochondrial or nuclear origin might also be invoked to support this hypothesis. The finding of a group II intron-encoded RT in the mitochondrial *cox2* gene of *Closterium baillyanum* is consistent with this idea (Turmel et al., 2013).

Putative gains of foreign genes by the chloroplast genome have been documented for several green algal lineages: prasinophytes (Turmel et al., 2009; Lemieux et al., 2014), Trebouxiophyceae (Turmel et al., 2015), Ulvophyceae (Leliaert and Lopez-Bautista, 2015), Chlorophyceae (Brouard et al., 2008) and streptophyte algae (Civan et al., 2014). Here we report further evidence that genes encoding integrases/recombinases (Int) and DNA primases of phage/viral origin were transferred to the chloroplast during the evolution of the Zygnematophyceae (Table 2). Putative Int-encoding genes have been also identified in the mitochondrial genomes of various streptophytes: *Klebsormidium flaccidum* (*orf374* in GenBank:KP165386), *Chaetosphaeridium globosum*

(Turmel et al., 2002), *Roya obtusa* (Turmel et al., 2013) and the lycophyte *Phlegmariurus squarrosus* (*orf233* in GenBank JQ002659). Inter-organellar gene transfers might account for the presence of these genes in both the chloroplast and mitochondria.

Early insertions of viral genes in the IR of the zygmatophycean chloroplast genome might have contributed to the instability of the IR. Civan et al. (2014) proposed that both cpDNA-encoded RT and Int activities have shaped this genome. Given that these activities are essential components of the replicative machinery of retroelements, they speculated that invasion of retroviruses and/or retrotransposons in the chloroplasts of early diverging zygmatophyceans triggered massive genome rearrangements. However, this scenario is not consistent with the findings that the chloroplast-encoded Int sequences uncovered in the zygmatophyceans investigated so far more likely originate from phages/viruses than retroelements and that only *Netrium* carries a chloroplast-encoded RT. As proposed for other streptophyte lineages with unusually rearranged chloroplast genomes (Weng et al., 2014; Zhang et al., 2016), it appears more likely that nuclear-encoded, plastid-targeted genes involved in DNA replication, recombination, and repair (and also perhaps in reverse transcription) played a major important

role in reshuffling the zygmatophycean chloroplast genome.

AUTHOR CONTRIBUTIONS

CL and MT conceived and designed the research. CL and CO performed the research. CL, CO, and MT analyzed the data. MT and CL wrote the paper.

FUNDING

This work was supported by the Natural Sciences and Engineering Research Council of Canada (http://www.nserc-crnsng.gc.ca/index_eng.asp) (Grant No. 2830-2007 to MT and CL).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00697>

REFERENCES

- Andersen, R. A. (2005). *Algal Culturing Techniques*. Boston, MA: Elsevier.
- Baldauf, S. L., Manhart, J. R., and Palmer, J. D. (1990). Different fates of the chloroplast *tufA* gene following its transfer to the nucleus in green algae. *Proc. Natl. Acad. Sci. U.S.A.* 87, 5317–5321. doi: 10.1073/pnas.87.14.5317
- Becker, B. (2013). Snow ball earth and the split of Streptophyta and Chlorophyta. *Trends Plant Sci.* 18, 180–183. doi: 10.1016/j.tplants.2012.09.010
- Becker, B., and Marin, B. (2009). Streptophyte algae and the origin of embryophytes. *Ann. Bot.* 103, 999–1004. doi: 10.1093/aob/mcp044
- Bélanger, A. S., Brouard, J. S., Charlebois, P., Otis, C., Lemieux, C., and Turmel, M. (2006). Distinctive architecture of the chloroplast genome in the chlorophycean green alga *Stigeoclonium helveticum*. *Mol. Genet. Genomics* 276, 464–477. doi: 10.1007/s00438-006-0156-2
- Bendich, A. J. (2004). Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 16, 1661–1666. doi: 10.1105/tpc.160771
- Boisvert, S., Lavolette, F., and Corbeil, J. (2010). Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol.* 17, 1519–1533. doi: 10.1089/cmb.2009.0238
- Bourque, G., and Pevzner, P. A. (2002). Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Res.* 12, 26–36.
- Brouard, J. S., Otis, C., Lemieux, C., and Turmel, M. (2008). Chloroplast DNA sequence of the green alga *Oedogonium cardiacum* (Chlorophyceae): unique genome architecture, derived characters shared with the Chaetophorales and novel genes acquired through horizontal transfer. *BMC Genomics* 9:290. doi: 10.1186/1471-2164-9-290
- Brouard, J. S., Otis, C., Lemieux, C., and Turmel, M. (2010). The exceptionally large chloroplast genome of the green alga *Floydiella terrestris* illuminates the evolutionary history of the Chlorophyceae. *Genome Biol. Evol.* 2, 240–256. doi: 10.1093/gbe/evq014
- Brouard, J. S., Otis, C., Lemieux, C., and Turmel, M. (2011). The chloroplast genome of the green alga *Schizomeris leibleinii* (Chlorophyceae) provides evidence for bidirectional DNA replication from a single origin in the Chaetophorales. *Genome Biol. Evol.* 3, 505–515. doi: 10.1093/gbe/evr037
- Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334
- Civan, P., Foster, P. G., Embley, M. T., Seneca, A., and Cox, C. J. (2014). Analyses of charophyte chloroplast genomes help characterize the ancestral chloroplast genome of land plants. *Genome Biol. Evol.* 6, 897–911. doi: 10.1093/gbe/evu061
- Cohen, N. E., Shen, R., and Carmel, L. (2012). The role of reverse transcriptase in intron gain and loss mechanisms. *Mol. Biol. Evol.* 29, 179–186. doi: 10.1093/molbev/msr192
- Cox, C. J., Li, B., Foster, P. G., Embley, T. M., and Civan, P. (2014). Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. *Syst. Biol.* 63, 272–279. doi: 10.1093/sysbio/syt109
- de Cambiaire, J. C., Otis, C., Turmel, M., and Lemieux, C. (2007). The chloroplast genome sequence of the green alga *Leptosira terrestris*: multiple losses of the inverted repeat and extensive genome rearrangements within the Trebouxiophyceae. *BMC Genomics* 8:213. doi: 10.1186/1471-2164-8-213
- Defeu Soufo, H. J., Reimold, C., Linne, U., Knust, T., Gescher, J., and Graumann, P. L. (2010). Bacterial translation elongation factor EF-Tu interacts and colocalizes with actin-like MreB protein. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3163–3168. doi: 10.1073/pnas.0911979107
- Delpont, W., Poon, A. F., Frost, S. D., and Kosakovsky Pond, S. L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26, 2455–2457. doi: 10.1093/bioinformatics/btq429
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Fenton, A. K., and Gerdes, K. (2013). Direct interaction of FtsZ and MreB is required for septum synthesis and cell division in *Escherichia coli*. *EMBO J.* 32, 1953–1965. doi: 10.1038/emboj.2013.129
- Frith, M. C., Hamada, M., and Horton, P. (2010). Parameters for accurate genome alignment. *BMC Bioinformatics* 11:80. doi: 10.1186/1471-2105-11-80
- Gensel, P. G. (2008). The earliest land plants. *Annu. Rev. Ecol. Syst.* 39, 459–477. doi: 10.1146/Annurev.Ecolsys.39.110707.173526
- Gerrath, J. F. (2003). “Conjugating green algae and desmids,” in *Freshwater Algae of North America: Ecology and Classification*, eds J. D. Wehr and R. G. Sheath (San Diego: Academic Press), 353–381.
- Gontcharov, A. A., Marin, B., and Melkonian, M. (2003). Molecular phylogeny of conjugating green algae (Zygnemophyceae, Streptophyta) inferred from SSU

- rDNA sequence comparisons. *J. Mol. Evol.* 56, 89–104. doi: 10.1007/s00239-002-2383-4
- Gontcharov, A. A., Marin, B., and Melkonian, M. (2004). Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematales (Streptophyta). *Mol. Biol. Evol.* 21, 612–624. doi: 10.1093/molbev/msh052
- Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res.* 8, 195–202. doi: 10.1101/gr.8.3.195
- Graham, L. E., Cook, M. E., and Busse, J. S. (2000). The origin of plants: body plan changes contributing to a major evolutionary radiation. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4535–4540. doi: 10.1073/pnas.97.9.4535
- Hall, J. D., Karol, K. G., McCourt, R. M., and Delwiche, C. F. (2008). Phylogeny of the conjugating green algae based on chloroplast and mitochondrial nucleotide sequence data. *J. Phycol.* 44, 467–477. doi: 10.1111/J.1529-8817.2008.00485.X
- Hu, F., Lin, Y., and Tang, J. (2014). MLGO: phylogeny reconstruction and ancestral inference from gene-order data. *BMC Bioinformatics* 15:354. doi: 10.1186/s12859-014-0354-6
- Jansen, R. K., and Ruhlman, T. A. (2012). “Plastid genomes of seed plants,” in *Genomics of Chloroplasts and Mitochondria*, eds R. Bock and V. Knoop (Berlin: Springer), 103–126.
- Karol, K. G., McCourt, R. M., Cimino, M. T., and Delwiche, C. F. (2001). The closest living relatives of land plants. *Science* 294, 2351–2353. doi: 10.1126/science.1065156
- Kenrick, P., Wellman, C. H., Schneider, H., and Edgecombe, G. D. (2012). A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 519–536. doi: 10.1098/rstb.2011.0271
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPputer: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633
- Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286–2288. doi: 10.1093/bioinformatics/btp368
- Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21, 1095–1109. doi: 10.1093/molbev/msh112
- Laurin-Lemay, S., Brinkmann, H., and Philippe, H. (2012). Origin of land plants revisited in the light of sequence contamination and missing data. *Curr. Biol.* 22, R593–R594. doi: 10.1016/J.Cub.2012.06.013
- Leliaert, F., and Lopez-Bautista, J. M. (2015). The chloroplast genomes of *Bryopsis plumosa* and *Tydemania expeditiones* (Bryopsidales, Chlorophyta): compact genomes and genes of bacterial origin. *BMC Genomics* 16:204. doi: 10.1186/s12864-015-1418-3
- Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F., et al. (2012). Phylogeny and molecular evolution of the green algae. *CRC Crit. Rev. Plant Sci.* 31, 1–46. doi: 10.1080/07352689.2011.615705
- Lemieux, C., Otis, C., and Turmel, M. (2000). Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature* 403, 649–652. doi: 10.1038/35001059
- Lemieux, C., Otis, C., and Turmel, M. (2007). A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. *BMC Biol.* 5:2. doi: 10.1186/1741-7007-5-2
- Lemieux, C., Otis, C., and Turmel, M. (2014). Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. *BMC Genomics* 15:857. doi: 10.1186/1471-2164-15-857
- Lewis, L. A., and McCourt, R. M. (2004). Green algae and the origin of land plants. *Am. J. Bot.* 91, 1535–1556. doi: 10.3732/ajb.91.10.1535
- Lohse, M., Drechsel, O., and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52, 267–274. doi: 10.1007/s00294-007-0161-y
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.0955
- Lu, F., Xu, W., Tian, C., Wang, G., Niu, J., Pan, G., et al. (2011). The *Bryopsis hypnoides* plastid genome: multimeric forms and complete nucleotide sequence. *PLoS ONE* 6:e14663. doi: 10.1371/journal.pone.0014663
- Maddison, D. R., and Maddison, W. P. (2000). *MacClade 4: Analysis of Phylogeny and Character Evolution*. Sunderland, MA: Sinauer Associates.
- Maddison, W. P., and Maddison, D. R. (2015). *Mesquite: a Modular System for Evolutionary Analysis*. Version 3.04. Available at: <http://mesquiteproject.org/>
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi: 10.1038/Nature03959
- McCourt, R. M., Delwiche, C. F., and Karol, K. G. (2004). Charophyte algae and land plant origins. *Trends Ecol. Evol.* 19, 661–666. doi: 10.1016/j.tree.2004.09.013
- McCourt, R. M., Karol, K. G., Bell, J., Helm-Bychowski, K. M., Grajewska, A., Wojciechowski, M. F., et al. (2000). Phylogeny of the conjugating green algae (Zygnematales) based on rbcL sequences. *J. Phycol.* 36, 747–758. doi: 10.1046/J.1529-8817.2000.99106.X
- Melton, J. T. III, Leliaert, F., Tronholm, A., and Lopez-Bautista, J. M. (2015). The complete chloroplast and mitochondrial genomes of the green macroalga *Ulva* sp. UN00071828 (Ulvophyceae, Chlorophyta). *PLoS ONE* 10:e0121020. doi: 10.1371/journal.pone.0121020
- Michel, F., Umesono, K., and Ozeki, H. (1989). Comparative and functional anatomy of group II catalytic introns - a review. *Gene* 82, 5–30. doi: 10.1016/0378-1119(89)90026-7
- Michel, F., and 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. doi: 10.1016/0022-2836(90)90386-Z
- Mikhailiyuk, T. I., Sluiman, H. J., Massalski, A., Mudimu, O., Demchenko, E. M., Kondratyuk, S. Y., et al. (2008). New streptophyte green algae from terrestrial habitats and an assessment of the genus *Interfilum* (Klebsormidiophyceae, Streptophyta). *J. Phycol.* 44, 1586–1603. doi: 10.1111/J.1529-8817.2008.00606.X
- Palmer, J. D. (1991). “Plastid chromosomes: structure and evolution,” in *The Molecular Biology of Plastids*, ed. I. K. Vasil (Cambridge: Academic Press), 5–53.
- Ruhlman, T. A., and Jansen, R. K. (2014). “The plastid genomes of flowering plants,” in *Chloroplast Biotechnology*, ed. P. Maliga (New York City, NY: Humana Press), 3–38.
- Smith, S. A., and Dunn, C. W. (2008). Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24, 715–716. doi: 10.1093/bioinformatics/btm619
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Tesler, G. (2002). GRIMM: genome rearrangements web server. *Bioinformatics* 18, 492–493. doi: 10.1093/bioinformatics/18.3.492
- Timme, R. E., Bachvaroff, T. R., and Delwiche, C. F. (2012). Broad phylogenomic sampling and the sister lineage of land plants. *PLoS ONE* 7:e29696. doi: 10.1371/journal.pone.0029696
- Turmel, M., de Cambiaire, J. C., Otis, C., and Lemieux, C. (2016). Distinctive architecture of the chloroplast genome in the chlorodendrophycean green algae *Scherffelia dubia* and *Tetraselmis* sp. CCMP 881. *PLoS ONE* 11:e0148934. doi: 10.1371/journal.pone.0148934
- Turmel, M., Gagnon, M. C., O’Kelly, C. J., Otis, C., and Lemieux, C. (2009). The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Mol. Biol. Evol.* 26, 631–648. doi: 10.1093/molbev/msn285
- Turmel, M., Lemieux, C., Burger, G., Lang, B. F., Otis, C., Plante, I., et al. (1999a). The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*. Two radically different evolutionary patterns within green algae. *Plant Cell* 11, 1717–1730. doi: 10.2307/3871049
- Turmel, M., Otis, C., and Lemieux, C. (1999b). The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10248–10253. doi: 10.1073/pnas.96.18.10248
- Turmel, M., Otis, C., and Lemieux, C. (2002). 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. doi: 10.1073/pnas.162203299
- Turmel, M., Otis, C., and 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. doi: 10.1186/1741-7007-3-22
- Turmel, M., Otis, C., and 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. doi: 10.1093/molbev/msk018
- Turmel, M., Otis, C., and Lemieux, C. (2013). Tracing the evolution of streptophyte algae and their mitochondrial genome. *Genome Biol. Evol.* 5, 1817–1835. doi: 10.1093/gbe/evt135
- Turmel, M., Otis, C., and Lemieux, C. (2015). Dynamic evolution of the chloroplast genome in the green algal classes *Pedinophyceae* and *Trebouxiophyceae*. *Genome Biol. Evol.* 7, 2062–2082. doi: 10.1093/gbe/evv130
- Turmel, M., Pombert, J. F., Charlebois, P., Otis, C., and Lemieux, C. (2007). The green algal ancestry of land plants as revealed by the chloroplast genome. *Int. J. Plant Sci.* 168, 679–689. doi: 10.1086/513470
- Umen, J. G. (2014). Green algae and the origins of multicellularity in the plant kingdom. *Cold Spring Harb. Perspect. Biol.* 6:a016170. doi: 10.1101/cshperspect.a016170
- Weng, M. L., Blazier, J. C., Govindu, M., and Jansen, R. K. (2014). Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* 31, 645–659. doi: 10.1093/molbev/mst257
- Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Muller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4859–E4868. doi: 10.1073/pnas.1323926111
- Wodniok, S., Brinkmann, H., Glockner, G., Heidel, A. J., Philippe, H., Melkonian, M., et al. (2011). Origin of land plants: do conjugating green algae hold the key? *BMC Evol. Biol.* 11:104. doi: 10.1186/1471-2148-11-104
- Xia, X. (2013). DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* 30, 1720–1728. doi: 10.1093/molbev/mst064
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Zhang, J., Ruhlman, T. A., Sabir, J. S., Blazier, J. C., Weng, M. L., Park, S., et al. (2016). Coevolution between nuclear-encoded DNA replication, recombination, and repair genes and plastid genome complexity. *Genome Biol. Evol.* 8, 622–634. doi: 10.1093/gbe/evw033
- Zhong, B., Liu, L., Yan, Z., and Penny, D. (2013). Origin of land plants using the multispecies coalescent model. *Trends Plant Sci.* 18, 492–495. doi: 10.1016/j.tplants.2013.04.009
- Zhong, B., Xi, Z., Goremykin, V. V., Fong, R., McLenachan, P. A., Novis, P. M., et al. (2014). Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes. *Mol. Biol. Evol.* 31, 177–183. doi: 10.1093/molbev/mst200

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Lemieux, Otis and Turmel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.