Mitochondrion-to-Chloroplast DNA Transfers and Intragenomic Proliferation of Chloroplast Group II Introns in *Gloeotilopsis* Green Algae (Ulotrichales, Ulvophyceae)

Monique Turmel, Christian Otis, and Claude Lemieux*

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

*Corresponding author: E-mail: claude.lemieux@bcm.ulaval.ca.

Accepted: July 31, 2016

GenBank accessions KX306821, KX306822, KX306823, and KX306824.

Abstract

To probe organelle genome evolution in the Ulvales/Ulotrichales clade, the newly sequenced chloroplast and mitochondrial genomes of *Gloeotilopsis planctonica* and *Gloeotilopsis sarcinoidea* (Ulotrichales) were compared with those of *Pseudendoclonium akinetum* (Ulotrichales) and of the few other green algae previously sampled in the Ulvophyceae. At 105,236 bp, the *G. planctonica* mitochondrial DNA (mtDNA) is the largest mitochondrial genome reported so far among chlorophytes, whereas the 221,431-bp *G. planctonica* and 262,888-bp *G. sarcinoidea* chloroplast DNAs (cpDNAs) are the largest chloroplast genomes analyzed among the Ulvophyceae. Gains of non-coding sequences largely account for the expansion of these genomes. Both *Gloeotilopsis* cpDNAs lack the inverted repeat (IR) typically found in green plants, indicating that two independent IR losses occurred in the Ulvales/Ulotrichales. Our comparison of the *Pseudendoclonium* and *Gloeotilopsis* cpDNAs offered clues regarding the mechanism of IR loss in the Ulotrichales, suggesting that internal sequences from the rDNA operon were differentially lost from the two original IR copies during this process. Our analyses also unveiled a number of genetic novelties. Short mtDNA fragments were discovered in two distinct regions of the *G. sarcinoidea* cpDNA, providing the first evidence for intracellular inter-organelle gene migration in green algae. We identified for the first time in green algal organelles, group II introns with LAGLIDADG ORFs as well as group II introns inserted into untranslated gene regions. We discovered many group II introns occupying sites not previously documented for the chloroplast genome and demonstrated that a number of them arose by intragenomic proliferation, most likely through retrohoming.

Key words: mitochondrial genome, plastid genome, inverted repeat, promiscuous DNA, retrohoming, repeated sequences.

Introduction

The Ulvophyceae is one of the three major classes of green algae belonging to the core Chlorophyta, a robustly supported assemblage that took roots from unicellular planktonic prasinophytes (Leliaert et al. 2012; Fucikova et al. 2014; Turmel et al. 2016). Although the class is best known for its macroscopic marine forms (the green seaweeds), several members live in freshwater or damp subaerial habitats (Friedl and Rybalka 2012; Leliaert et al. 2012). The Ulvophyceae far exceeds the other chlorophyte classes in diversity of morphological complexity (ranging from microscopic unicells to multicellular plants and giant-cells) and cellular sophistication (with four cytomorphological types). Numerous orders have been erected to reflect this diversity, but the relationships among them as well as the monophyly of the class remain ambiguous because the phylogenetic analyses that were inferred to elucidate these questions yielded conflicting results. A phylogenetic analysis based on ten genes (eight encoded in the nucleus and two in the chloroplast) recovered the Ulvophyceae as a well-supported monophyletic group composed of two main lineages: the Oltmannsiellopsidales–Ulvales–Ulotrichales (comprising most microscopic forms) and the Trentepohliales–Bryopsidales–Cladophorales–Dasycladales (Cocquyt et al. 2010). In contrast, recent chloroplast phylogenomic studies revealed that the class is not monophyletic (Fucikova et al. 2014; Leliaert and Lopez-Bautista 2015; Melton et al. 2015; Sun et al. 2016; Turmel et al. 2016): the Oltmannsiellopsidales–Ulvales–Ulotrichales was recovered in most studies but the affinity of this clade with the other lineages (Bryopsidales, Trentepohliales, and Dasycladales) received no support.

The genomic data currently available for the Ulvophyceae indicate that the chloroplast genome is highly variable in structure, gene density, gene order, and intron content. Complete chloroplast DNA (cpDNA) sequences have been reported for only seven ulvophycean taxa: the marine flagellate *Oltmannsiellopsis viridis* (Oltmannsiellopsidales) (Pombert et al. 2006), the freshwater filamentous and branched *Pseudendoclonium akinetum* (Ulotrichales) (Pombert et al. 2005), the marine macroalgae *Ulva* sp. UNA00071828 and *Ulva fasciata* (Ulvales) (Melton and Lopez-Bautista 2015; Melton et al. 2015), and three marine siphonous species from the *Bryopsis* and *Tydemania* genera (Bryopsidales) (Lu et al. 2011; Leliaert and Lopez-Bautista 2015). Their sizes range from 96 kb (in *Ulva fasciata*) to 196 kb (in *Pseudendoclonium*), while their gene repertoires contain 100 (in *Ulva* species) to 105 (in *Pseudendoclonium*) conserved genes. Only the chloroplast genomes sampled from the Oltmannsiellopsidales (*Oltmannsiellopsis*) and Ulotrichales (*Pseudendoclonium*) have retained the large rRNA operon-encoding inverted repeat (IR) that is present in most green plants, but as observed for the Chlorophyceae (de Cambiaire et al. 2006; Brouard et al. 2008), gene partitioning among the single-copy regions differs considerably between these two ulvophycean lineages and is also distinct from the patterns observed in other green algal groups. Another feature shared by ulvophycean and chlorophycean cpDNAs is their highly diverse and variable intron contents, which are generally characterized by a greater proportion of group I rather than group II introns. Twenty-seven introns, all belonging to the group I class, are present in *Pseudendoclonium* cpDNA, whereas *Oltmannsiellopsis* and the two *Ulva* species contain only five introns in their chloroplast. Moreover, a number of chloroplast group I introns are inserted at the same insertion sites in some members of the Ulvophyceae and Chlorophyceae. For instance, 11 of the introns found in *Pseudendoclonium* and *Oedogonium cardiacum* share the same positions (Brouard et al. 2008). This observation might reflect past events of intercellular horizontal DNA transfers involving these mobile elements, but more data on intron diversity in the Chlorophyceae and the Oltmannsiellopsidales–Ulvales–Ulotrichales clade are needed to provide support for this hypothesis.

Unexpectedly, the analysis of the *Pseudendoclonium* chloroplast and mitochondrial genomes provided indirect evidence for intracellular, inter-organellar DNA exchanges in the Ulotrichales (Pombert et al. 2005, 2004). Two observations support the occurrence of these genetic exchanges: first, the mitochondrial *atp1* gene contains a group I intron (site 522) that is highly similar in both sequence and secondary structure to a group I intron inserted at the identical site (site 489) in the chloroplast *atpA* gene, and second, the chloroplast and mitochondrial genomes contain an identical 15-bp repeat. The direction of transfer of these sequences, however, could not be elucidated. To date, inter-organellar DNA transfers have been documented solely for seed plants, with frequent transfers of cpDNA to the mitochondrion but only rare cases of mitochondrial DNA (mtDNA) migration to the chloroplast (Goremykin et al. 2009; Smith 2011, 2014; Straub et al. 2013). It is thought that the relative absence of mtDNA-derived sequences in the chloroplast might reflect the lack of a DNA uptake system in this organelle (Bock 2010).

In the present study, we have analyzed the chloroplast and mitochondrial genomes of two additional ulvophyceans belonging to the Ulotrichales, *Gloeotilopsis planctonica* and *Gloeotilopsis sarcinoidea*, and compared these genomes to their previously reported counterparts. The *Gloeotilopsis* lineage is sister to that comprising *Pseudendoclonium akinetum* and species from the genus *Hazenia* (Škaloud et al. 2013). The main goals of our study were to enhance our understanding of chloroplast and mitochondrial genome evolution in the Ulvales/Ulotrichales and to gain more information regarding inter-organellar DNA transfer. Our comparative analyses unveiled a number of novel genomic features not previously documented for green algal organelles, including the first case of intracellular transfer of mtDNA to the chloroplast.

Materials and Methods

Strains and Culture Conditions

Gloeotilopsis planctonica SAG 29.93 and *Gloeotilopsis sarcinoidea* UTEX 1710 were obtained from the culture collections of algae at the University of Gottingen (SAG) and the University of Texas at Austin (UTEX), respectively. Both strains were grown in C medium (Andersen 2005) at 18 °C under alternating 12 h light and 12 h dark periods.

Genome Assemblies, Annotations, and Sequence Analyses

The *G. planctonica* chloroplast and mitochondrial genomes were sequenced using the Roche 454 method. A shotgun library (700-bp fragments) of A + T-rich organelle DNA, which was obtained by CsCl-bisbenzimide isopycnic centrifugation of total cellular DNA (Turmel et al. 1999), was 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 Laval" (http://pag.ibis.ulaval.ca/seq/en/; last accessed 8 August 2016). Reads were assembled using Newbler v2.5 (Margulies et al. 2005) with default parameters, and contigs of chloroplast and mitochondrial origins were identified by BlastN and BlastX (Altschul et al. 1990) searches against a local database of organelle genomes. Following visualization and editing with the

CONSED 22 finishing package (Gordon et al. 1998), the 20 chloroplast contigs comprising a total of 47,577 reads (average coverage depth = 67) were ordered and linked by polymerase chain reaction (PCR) amplification of the regions spanning gaps using a set of 32 oligonucleotides (supplementary table S1, Supplementary Material online). Purified PCR products were sequenced using Sanger chemistry with the PRISM BigDye Terminator Ready Reaction Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). The five contigs that were recovered for the mitochondrial genome (7172 reads; average coverage depth = 24) were assembled using CONSED.

The *G. sarcinoidea* organelle genomes were sequenced using the Illumina method. Total cellular DNA was isolated using the EZNA HP Plant Mini Kit of Omega Bio-Tek (Norcross, GA, USA). A library of 500-bp fragments was constructed using the TrueSeq DNA Sample Prep Kit (Illumina, San Diego, CA, USA) and paired-end reads (300-bp) were generated on the MiSeq sequencer by the "Plateforme d'Analyses Génomiques de l'Université Laval" (http://pag.ibis.ulaval.ca/seq/en/index.php; last accessed 8 August 2016). These reads were trimmed to remove adapter and low-quality sequences with CUTADAPT (Martin 2011) and PRINTSEQ (Schmieder and Edwards 2011), respectively, and the paired-end sequences were merged using FLASH (Magoc and Salzberg 2011). Reads were then assembled using Ray 2.3.1 (Boisvert et al. 2010) with kmer values of 61 and 89. Identification, visualization and editing of organellar DNA contigs were performed as described above for the 454 sequence assemblies. A single contig with overlapping terminal sequences that allowed genome circularization was recovered for each organellar genome. The final chloroplast assembly contained a total of 281,367 reads with an average coverage depth of 319, while the final mitochondrial assembly contained 36,523 reads with an average coverage depth of 129.

Genes and open reading frames (ORFs) were identified on the final assemblies using a custom-built suite of bioinformatics tools allowing the automated execution of the following three steps: 1) ORFs were found using GETORF in EMBOSS (Rice et al. 2000), 2) their translated products were identified by BlastP (Altschul et al. 1990) searches against a local database of mtDNA- and cpDNA-encoded proteins or the nr database at the National Center for Biotechnology Information, and 3) consecutive 100 bp segments of the genome sequence were analyzed with BlastN and BlastX (Altschul et al. 1990) to identify gene sequences. Only the ORFs that revealed identities with genes of known functions or previously reported ORFs were annotated. Genes for rRNAs and tRNAs were independently identified and localized using RNAmmer (Lagesen et al. 2007) and tRNAscan-SE (Lowe and Eddy 1997), respectively. Intron boundaries were determined by comparing intron-containing genes with intronless homologs and by modeling intron secondary structures (Michel et al. 1989; Michel and Westhof 1990). Circular genome maps were drawn with OGDraw (Lohse et al. 2007). In the course of our study, we also used the methods mentioned above to verify the gene repertoires and intron contents of the organelle genomes previously reported for *Bryopsis* and *Ulva* species (Lu et al. 2011; Leliaert and Lopez-Bautista 2015; Melton and Lopez-Bautista 2015, 2016; Zhou et al. 2016a, 2016b).

Genome-scale sequence comparisons were carried out with LAST v7.1.4 (Frith et al. 2010) to map regions sharing similar sequences within a genome and to identify shared sequences between chloroplast and mitochondrial genomes. To estimate the proportion of small repeated sequences, repeats with a minimal size of 30 bp were retrieved using REPFIND of REPuter v2.74 (Kurtz et al. 2001) with the options -f -p -l -allmax and were then masked on the genome sequence using RepeatMasker (http://www.repeatmasker.org/; last accessed 8 August 2016) running under the Crossmatch search engine (http://www.phrap.org/; last accessed 8 August 2016). Tandem repeats were identified with TRF v4.09 (Benson 1999) and classification of the repeat sequences was performed using RECON (Bao and Eddy 2002).

Analyses of Gene Organization

The number of reversals separating the chloroplast or mitochondrial genomes of the two *Gloeotilopsis* species from one another and from their *Pseudendoclonium* counterparts was estimated with GRIMM v2.01 (Tesler 2002). We used a custom-built Perl script to identify the regions that display the same gene order in these genomes. This script employs a concatenated list of signed gene orders in the compared genomes as input file (i.e., taking into account gene polarity) and interacts with MySQL database tools (https://www.mysql.com; last accessed 8 August 2016) to identify syntenic regions.

Phylogenetic Analyses

To identify the relationships among *Gloeotilopsis* group II introns, nucleotides comprising the central wheel of the intron secondary structure as well as adjacent nucleotides from the conserved regions within the six domains radiating from this wheel were aligned based on secondary structure models and the resulting alignment was analyzed using RAxML v8.2.6 (Stamatakis 2014) and the GTR + Γ4 model. Confidence of branch points was estimated by fast-bootstrap analysis (f = a) with 1000 replicates.

To determine the timing of inter-organellar DNA transfers in the Ulotrichales, the *Gloeotilopsis* mitochondrial-like chloroplast sequences were aligned with the corresponding mitochondrial gene sequences of *Gloeotilopsis* and other green plants using MUSCLE v3.7 (Edgar 2004) and the resulting alignments were analyzed individually using RAxML v8.2.6 (Stamatakis 2014) and the GTR + Γ4 model. Confidence of branch points was estimated by fast-bootstrap analysis (f = a) with 500 replicates.

Results and Discussion

The *Gloeotilopsis* Mitochondrial Genomes Share Several Features with their *Pseudendoclonium* Counterpart

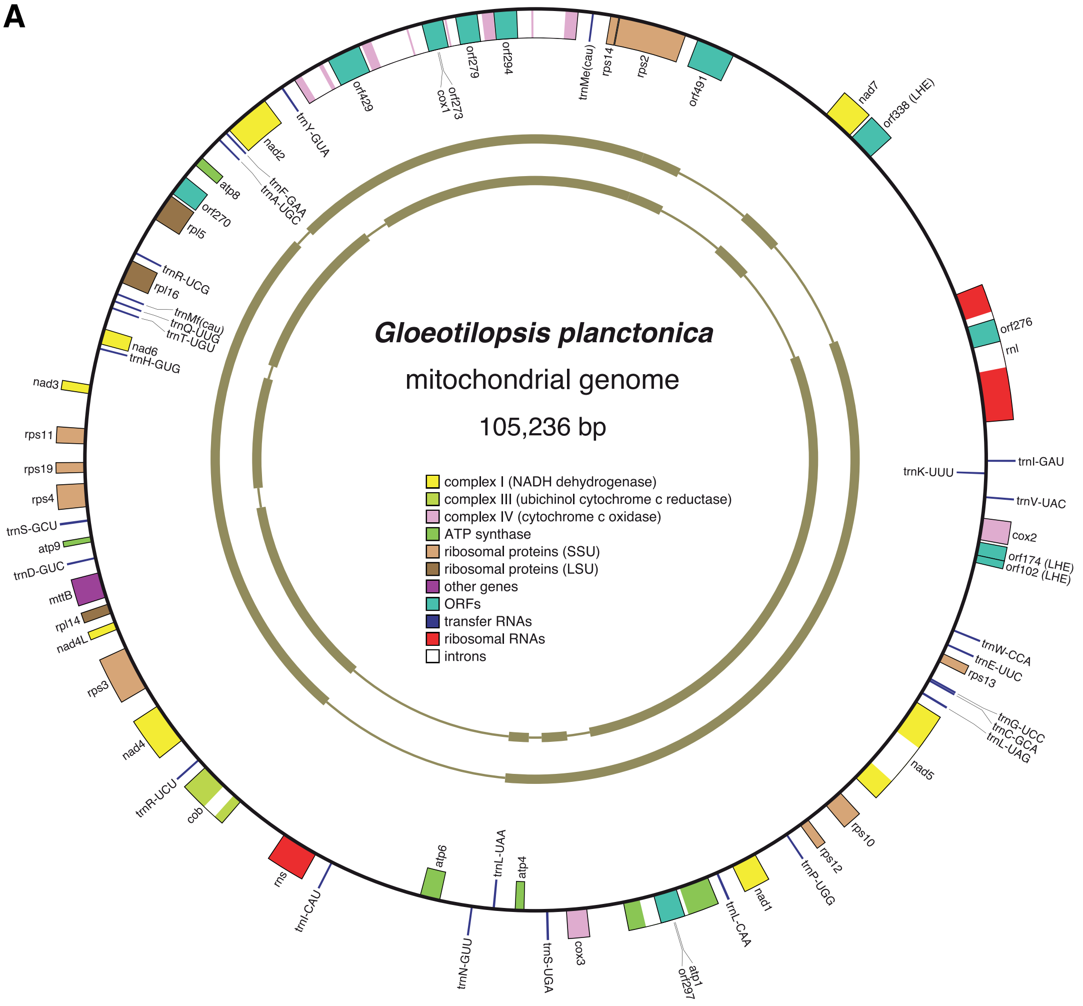
The *G. planctonica* and *G. sarcinoidea* mtDNAs were assembled as circular molecules of 105,236 bp and 85,108 bp, respectively (fig. 1A and supplementary fig. S1, Supplementary Material online). The *G. planctonica* genome is the largest reported so far among chlorophytes. Although the two newly sequenced ulotrichalean genomes and their *Pseudendoclonium* counterpart are larger than the mtDNAs currently available for the Olmannsiellopsidales and Ulvales, their repertoires of conserved genes are very similar to that of the *Ulva* genomes (table 1 and fig. 1B). The *Gloeotilopsis* mtDNAs are identical in gene content: their 58 conserved genes code for 30 proteins, two rRNAs, and 26 tRNAs. The only difference relative to the *Pseudendoclonium* gene repertoire is the presence of *trnL*(caa), a gene specific to the *Gloeotilopsis* lineage. The mitochondrial gene distribution currently available for ulvophycean green algae also suggests that *trnR*(ucg) is a lineage-specific tRNA gene in the Ulvales and that four genes present in the Oltmannsiellopsidales (*nad9*, *rrn5*, *trnG*(gcc) and *trnR*(acg)) were lost before the divergence of the Ulvales and Ulotrichales (fig. 1B). As reported for other algal lineages (Valach et al. 2014), it is possible that the gene coding for the 5S rRNA (*rrn5*) went undetected because it is too divergent in sequence. Aside from conserved genes, the two *Gloeotilopsis* genomes share with *Pseudendoclonium* mtDNA three free-standing ORFs that code for one LAGLIDADG homing endonuclease (LHE) and two proteins of unknown functions (fig. 1A and supplementary fig. S1, Supplementary Material online); the corresponding *Pseudendoclonium* ORFs are *orf307*, *orf325*, and *orf361*, respectively (Pombert et al. 2004). In addition, we found 5′ to *cox2*, one free-standing LAGLIDADG ORF in *G. sarcinoidea* (*orf193*) and two in *G. planctonica* (*orf102* and *orf174*) that are highly similar to the proteins encoded by group I introns in the mitochondrial *atp1* and chloroplast *atpA* genes of *Pseudendoclonium*.

Unlike the *Ulva* mtDNAs whose genes are all encoded on the same DNA strand (Melton et al. 2015; Melton and Lopez-Bautista 2016; Zhou et al. 2016a, 2016b), the *Gloeotilopsis* and *Pseudendoclonium* genomes have their genes distributed between the two strands. The four *Ulva* genomes exhibit the same gene order but are substantially rearranged compared to ulotrichalean genomes. The *Gloeotilopsis* mtDNAs show slight differences in gene organization: nearly 90% of their conserved genes and free-standing ORFs form four syntenic blocks (fig. 1) and as inferred by GRIMM analysis, only four reversals are required to convert gene order in one genome to that of the other. Our comparison of *G. planctonica* and *G. sarcinoidea* mtDNAs with their *Pseudendoclonium* counterpart revealed seven and eight syntenic blocks (fig. 1), respectively, with 11 and 12 reversals inferred using GRIMM.

The *Gloeotilopsis* and *Ulva* Mitochondrial Genomes Boast Group II Introns with LAGLIDADG ORFs

Gloeotilopsis planctonica mtDNA contains ten group I introns, seven of which are present in *cox1*, as well as one group II intron, whereas *G. sarcinoidea* mtDNA contains five group I introns (fig. 2). The introns uncovered so far in ulvophycean mtDNAs represent 22 distinct insertion sites, half of which have not been identified in other clades of chlorophytes (see arrows in fig. 2). Note that a number of introns were not annotated or correctly annotated in the GenBank accessions of Ulvales; numerous changes related to intron types and insertion positions were introduced in the course of this study (fig. 2). All known ORF-containing group I introns in ulvophycean mitochondria encode putative LAGLIDADG homing endonucleases (LHEs), an observation that contrasts with the finding of intron-encoded homing endonucleases from three distinct families (LAGLIDADG, GIY-YIG, and H-N-H) in the ulvophycean chloroplast (Pombert et al. 2005).

Unexpectedly, group II introns containing LAGLIDADG ORFs were identified at three distinct mtDNA sites in the course of our study: site 916 within the *cox1* genes of *U. prolifera* (genomic positions 4994–6502) and *U. linza* (positions 7516–9024), site 1911 within *rnl* of *G. planctonica*, and site 2451 within *rnl* of *U. linza* (positions 37,293–38,685). LHEs are generally associated with group I introns, promoting their mobility by introducing double-strand breaks within intronless target sequences (Belfort et al. 2002). The great majority of ORF-containing group II introns encode proteins with reverse transcriptase (RT), intron maturase (X), and H-N-H endonuclease (En) domains that are required for intron splicing and mobility (Zimmerly and Semper 2015). The only few LHE-encoding group II introns that have been previously documented are confined to the giant sulfur bacterium *Thiomargarita namibiensis* (Salman et al. 2012) and the mitochondria of fungi belonging to the Ascomycota and Basidiomycota (Toor and Zimmerly 2002; Monteiro-Vitorello et al. 2009; Mullineux et al. 2010; Pfeifer et al. 2012). In fungal mitochondria, these unusual introns also occur in the *cox1* and rRNA genes but their insertion sites in these genes (*cox1* site 969, *rns* sites 785 and 952, and *rnl* site 2059) differ from those we uncovered in ulvophycean mtDNAs. Structural analyses of the four ulvophycean LHE-encoding introns revealed that the ORF is located in domain IV of the secondary structure, i.e., the same domain where RT-related ORFs are generally located in group II introns. The LHEs encoded by the introns inserted at distinct sites do not appear to be closely related, as BlastP searches against the non-redundant database of NCBI using these protein sequences identified different sets of sequences, which originated mostly from fungal mitochondria. Therefore, it is likely that the LAGLIDADG ORFs in ulvophycean group II introns represent three different events of integration which, as demonstrated for the LHEs encoded by the *Leptographium rns* intron (Mullineux et al. 2010) and the *Ustilago rnl* intron



B

	atp1	atp4	atp6	atp8	atp9	cob	cox1	cox2	cox3	mttB	nad1	nad2	nad3	nad4	nad4L	nad5	nad6	nad7	nad9	rpl5	rpl14	rpl16	rps2	rps3	rps4	rps10	rps11	rps12	rps13	rps14	rps19
Oltmannsiellopsis viridis NIES 360	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■			■	■	■			■	■	■	■	■
Ulva prolifera	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■
Ulva linza	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■
Ulva sp. UNA00071828	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■
Ulva fasciata	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■
Pseudendoclonium akinetum UTEX 1912	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■
Gloeotilopsis planctonica SAG 29.93	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■
Gloeotilopsis sarcinoidea UTEX 1710	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■

	rnl	rns	rrn5	trnA(ugc)	trnC(gca)	trnD(guc)	trnE(uuc)	trnF(gaa)	trnG(gcc)	trnG(ucc)	trnH(gug)	trnI(cau)	trnI(gau)	trnK(uuu)	trnL(caa)	trnL(uaa)	trnL(uag)	trnMe(cau)	trnMf(cau)	trnN(guu)	trnP(ugg)	trnQ(uug)	trnR(acg)	trnR(gcg)	trnR(ucg)	trnR(ucu)	trnS(gcu)	trnS(uga)	trnT(ugu)	trnV(uac)	trnW(cca)	trnY(gua)
Oltmannsiellopsis viridis NIES 360	■	■	■	■	■	■	■	■	■	■	■	■		■		■	■	■	■	■	■	■	■			■	■	■		■	■	■
Ulva prolifera	■	■		■	■	■	■	■		■	■	■	■	■		■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■
Ulva linza	■	■		■	■	■	■	■		■	■	■	■	■		■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■
Ulva sp. UNA00071828	■	■		■	■	■	■	■		■	■	■	■	■		■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■
Ulva fasciata	■	■		■	■	■	■	■		■	■	■	■	■		■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■
Pseudendoclonium akinetum UTEX 1912	■	■		■	■	■	■	■		■	■	■	■	■		■	■	■	■	■	■	■			■	■	■	■	■	■	■	■
Gloeotilopsis planctonica SAG 29.93	■	■		■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■			■	■	■	■	■	■	■	■
Gloeotilopsis sarcinoidea UTEX 1710	■	■		■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■			■	■	■	■	■	■	■	■

Fig. 1.—Mitochondrial gene content and organization in *Gloeotilopsis*. (*A*) Gene map of the *G. planctonica* mitochondrial genome. Filled boxes on the gene map represent genes, with colors denoting gene categories as indicated in the legend. Genes on the outside are transcribed counterclockwise, whereas those on the inside are transcribed clockwise. Thick lines in the innermost ring represent the gene clusters shared with *Pseudendoclonium*; those in the second outermost inner ring represent the gene clusters shared with *G. sarcinoidea*. (*B*) Comparison of gene content among ulvophycean mtDNAs. The presence of a gene is denoted by a blue box.

Table 1
General Features of *Gloeotilopsis* and Other Ulvophycean Mitochondrial Genomes

Taxon	Accession	Size	A+T	Genes[a]	Introns (no.)[b]		Repeats[c]
		(bp)	(%)	(no.)	GI	GII	(%)
Oltmannsiellopsidales							
Oltmannsiellopsis viridis NIES 360	NC_008256	56,761	66.6	54	2	1	8.9
Ulvales							
Ulva prolifera	NC_028538	63,845	66.0	58	7	2	3.0
Ulva linza	NC_029701	70,858	65.4	58	9	4	3.5
Ulva sp. UNA00071828	KP720617	73,493	67.8	58	8	2	2.3
Ulva fasciata	NC_028081	61,614	67.5	58	3	1	1.4
Ulotrichales							
Pseudendoclonium akinetum UTEX 1912	NC_005926	95,880	60.7	57	7	0	13.0
Gloeotilopsis planctonica SAG 29.93	KX306823	105,236	65.3	58	10	1	6.6
Gloeotilopsis sarcinoidea UTEX 1710	KX306822	85,108	66.2	58	5	0	3.9

[a] Intronic genes and freestanding ORFs not usually found in green plant mitochondrial genomes are not included in these values. Duplicated genes were counted only once.
[b] Number of group I (GI) and group II (GII) introns is given.
[c] Non-overlapping repeat elements were mapped on each genome with RepeatMasker using as input sequences the repeats of at least 30 bp identified with REPuter.

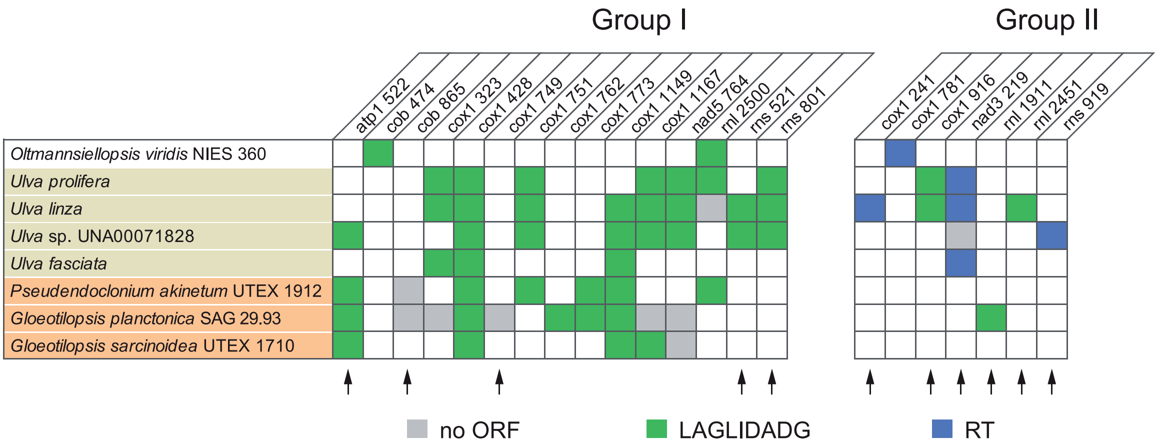


Fig. 2.—Group I and group II introns in *Gloeotilopsis* and other ulvophycean mtDNAs. A grey box represents an intron lacking an ORF; a green box indicates that the intron encodes a LAGLIDADG homing endonuclease, whereas a blue box indicates that the intron encodes a reverse transcriptase (RT) with or without H-N-H endonuclease and/or intron maturase domains. Arrows denote the intron insertion sites that have not been observed in other classes of the Chlorophyta. Intron insertion sites in protein-coding and tRNA genes are given relative to the corresponding genes in *Mesostigma viride* mtDNA (Turmel et al. 2002); insertion sites in *rrs* and *rrl* 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.

(Pfeifer et al. 2012), confer intron mobility by allowing the introduction of double-strand breaks at intronless target sites.

The *Gloeotilopsis* Chloroplast Genomes Are Large, Lack the IR, and Differ in Gene Order

The *G. planctonica* and *G. sarcinoidea* cpDNAs are the largest ulvophycean chloroplast genomes analyzed so far (table 2); they were assembled as circular molecules of 221,431 and 262,888 bp, respectively (fig. 3A and supplementary fig. S2, Supplementary Material online). Like their counterparts in the Bryopsidales and Ulvales (Lu et al. 2011; Leliaert and Lopez-Bautista 2015; Melton et al. 2015), they lack a large IR encoding the rRNA genes. Their gene repertoire, which only differs from that of *Pseudendoclonium* by the absence of *trnR*(ccu) (fig. 3B), comprises 104 different conserved genes coding for 73 proteins, three rRNAs and 28 tRNAs. As revealed by GRIMM analyses, seven reversals account for the alterations in gene order between the two *Gloeotilopsis* cpDNAs; 101 of their 104 shared conserved genes form six syntenic blocks, with the longest containing 69 genes (fig. 3A). The G. *sarcinoidea* and *G. planctonica* genomes differ from their

Table 2
General Features of *Gloeotilopsis* and Other Ulvophycean Chloroplast Genomes

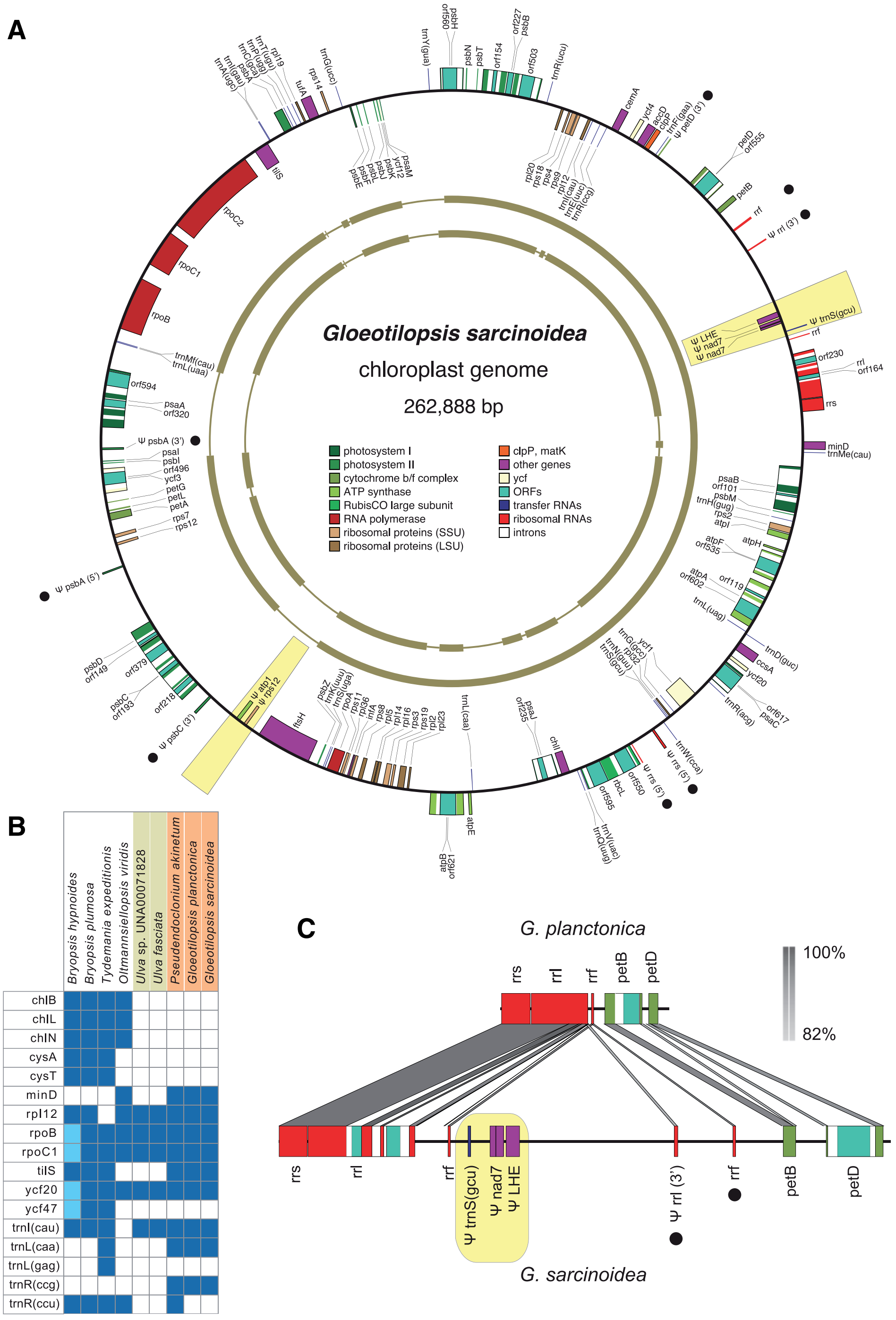
Taxon	Accession	Size (bp)		A+T	Genes[a]	Introns (no.)[b]		Repeats[c]
		Genome	IR	(%)	(no.)	GI	GII	(%)
Bryopsidales								
Bryopsis hypnoides	NC_013359	153,429	–	66.9	108[d]	6	6	9.9
Bryopsis plumosa West4718	NC_026795	106,859	–	69.2	108	7	6	2.4
Tydemania expeditionis FL1151	NC_026796	105,200	–	67.2	109	8	3	0.4
Oltmannsiellopsidales								
Oltmannsiellopsis viridis NIES 360	NC_008099	151,933	18,510	59.5	104	5	0	11.1
Ulvales								
Ulva sp. UNA00071828	KP720616	99,983	–	74.7	100	4	1	0.5
Ulva fasciata	NC_029040	96,005	–	75.1	100	4	1	0.5
Ulotrichales								
Pseudendoclonium akinetum UTEX 1912	NC_008114	195,867	6,039	68.5	105	27	0	5.3
Gloeotilopsis planctonica SAG 29.93	KX306824	221,431	–	68.5	104	14	17	3.7
Gloeotilopsis sarcinoidea UTEX 1710	KX306821	262,888	–	68.5	104	15	12	11.6

[a] Intronic genes and freestanding ORFs not usually found in green plant chloroplast genomes are not included in these values. Duplicated genes were counted only once.
[b] Number of group I (GI) and group II (GII) introns is given.
[c] Non-overlapping repeat elements were mapped on each genome with RepeatMasker using as input sequences the repeats of at least 30 bp identified with REPuter.
[d] This value includes four pseudogenes (*rpoB, rpoC1, ycf20* and *ycf47*).

Pseudendoclonium counterpart by 21 and 23 reversals, respectively; the conserved genes shared by the *Gloeotilopsis*/ *Pseudendoclonium* cpDNAs form 14 syntenic blocks (fig. 3A). In addition to the ancestral gene clusters that were reported to be fragmented in *Pseudendoclonium* (Pombert et al. 2005), the *Gloeotilopsis* genomes display a disrupted rRNA operon.

The 41-kb size difference between the *G. planctonica* and *G. sarcinoidea* cpDNAs is essentially accounted for by longer intergenic regions in the latter genome (133 kb versus 92 kb). Both intergenic regions and introns, however, contribute to the larger sizes of the *Gloeotilopsis* cpDNAs relative to previously sequenced ulvophycean genomes. For instance, in the *Pseudendoclonium* genome, the intergenic regions are 81 kb in length, and the intron sequences span 30 kb as compared to 50 kb in the *Gloeotilopsis* genomes, even though the total number of introns is comparable in both ulvophycean lineages (table 2). In contrast to the *Oltmannsiellopsis* and *Pseudendoclonium* cpDNAs, whose introns all belong to the group I family, the *Gloeotilopsis* genomes contain many group II introns (representing 14.1% and 11.8% of the *G. planctonica* and *G. sarcinoidea* cpDNAs, respectively); the latter introns, in particular those containing ORFs, account for the increased proportion of intron sequences in the *Gloeotilopsis* lineage.

Considering our current understanding of the phylogenetic relationships among major ulvophycean lineages (Škaloud et al. 2013) and also the fact that there is no solid evidence for *de novo* gain of the IR by chloroplast genomes, our results indicate that the IR was lost at least once in the Ulotrichales, bringing to three the number of independent IR losses known in the Ulvophyceae. Although multiple IR losses have also been reported in the Trebouxiophyceae (Turmel et al. 2015), Prasinophyceae (Lemieux et al. 2014), and streptophytes (Ruhlman and Jansen 2014; Blazier et al. 2016; Lemieux et al. 2016), the molecular mechanism(s) underlying these events remain(s) unclear. One might envision that multiple rounds of IR contraction by a mechanism involving illegitimate recombination (Goulding et al. 1996; Wang et al. 2008) ultimately lead to IR loss. This hypothesis predicts that intermediate IR forms housing partial rRNA operons would be created; however, such observations are extremely rare (Guisinger et al. 2011) and recent reports suggest that the rRNA operon may be disrupted before IR loss (Turmel et al. 2015; Lemieux et al. 2016). Our comparison of the *Gloeotilopsis* cpDNA regions corresponding to the IR and adjacent sequences in the *Pseudendoclonium* genome provides hints into the process of IR loss in the Ulotrichales (fig. 4). The rRNA operon, which exists in its ancestral form in *Pseudendoclonium*, has been broken into two pieces in the *Gloeotilopsis* lineage: *trnI*(gau) and *trnA*(ugc) are still linked but are far apart from the *rrs*, *rrl* and *rrf* genes, which are also clustered and in the same order as in the ancestral operon (fig. 4A). Importantly, if we exclude a small inversion involving the *minD*/*trnMe*(cau) genes, the rRNA gene cluster is bordered on each side by a long segment of conserved genes that is perfectly colinear with the corresponding *Pseudendoclonium* sequence neighboring the IR. As a possible interpretation of these results, we suggest that the *trnI*(gau) and *trnA*(ugc) genes were deleted from the rRNA operon ancestrally located in the large syntenic region and that these two genes were the only genetic elements that were retained from the other copy of the rRNA operon during the process of



	Bryopsis hypnoides	*Bryopsis plumosa*	*Tydemania expeditionis*	*Oltmannsiellopsis viridis*	*Ulva* sp. UNA00071828	*Ulva fasciata*	*Pseudendoclonium akinetum*	*Gloeotilopsis planctonica*	*Gloeotilopsis sarcinoidea*
chlB	■	■	■	■					
chlL	■	■	■	■					
chlN	■	■	■	■					
cysA	■	■	■						
cysT	■	■	■						
minD				■			■	■	■
rpl12	■	■		■	■	■	■	■	■
rpoB	▨	■	■	■	■	■	■	■	■
rpoC1	▨	■	■	■	■	■	■	■	■
tilS	■	■	■				■	■	■
ycf20	▨	■	■	■	■	■	■	■	■
ycf47	▨	■	■						
trnI(cau)	■	■	■		■	■	■	■	■
trnL(caa)			■				■	■	■
trnL(gag)			■						
trnR(ccg)							■	■	■
trnR(ccu)	■	■	■	■			■		

Fig. 3.—Chloroplast gene content and organization in *Gloeotilopsis*. (*A*) Gene map of the *G. sarcinoidea* chloroplast genome. Filled boxes on the gene map represent genes, with colors denoting gene categories as indicated in the legend. Duplicated gene sequences are denoted with filled circles and sequences of mitochondrial origin are highlighted in yellow. Genes on the outside are transcribed counterclockwise, whereas those on the inside are transcribed clockwise. Thick lines in the innermost ring represent the gene clusters shared with *Pseudendoclonium*; those in the second outermost inner ring represent the gene clusters shared with *G. sarcinoidea*. (*B*) Comparison of gene content in ulvophycean cpDNAs. Genes and pseudogenes are shown in dark and light blue, respectively. Only the conserved genes showing a variable distribution are indicated. All compared genomes share the following set of 95 genes: *accD, atpA, B, E, F, H, I, ccsA, cemA, chlI, clpP, ftsH, infA, petA, B, D, G, L, psaA, B, C, I, J, M, psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z, rbcL, rpl2, 5, 14, 16, 19, 20, 23, 32, 36, rpoA, C2, rps2, 3,4, 7, 8, 9, 11, 12, 14, 18, 19, rrf, rrl, rrs, tufA, ycf1, 3, 4, 12, trnA*(ugc), *C*(gca), *D*(guc), *E*(uuc), *F*(gaa), *G*(gcc), *G*(ucc), *H*(gug), *I*(gau), *K*(uuu), *L*(uaa), *L*(uag), *Me*(cau), *Mf*(cau), *N*(guu), *P*(ugg), *Q*(uug), *R*(acg), *R*(ucu), *S*(gcu), *S*(uga), *T*(ugu), *V*(uac), *W*(cca), *Y*(gua). (*C*) Alignment of the *G. sarcinoidea* and *G. planctonica* cpDNA regions extending from *rrs* to *petD*. Sequence identity is denoted by the grey scale.

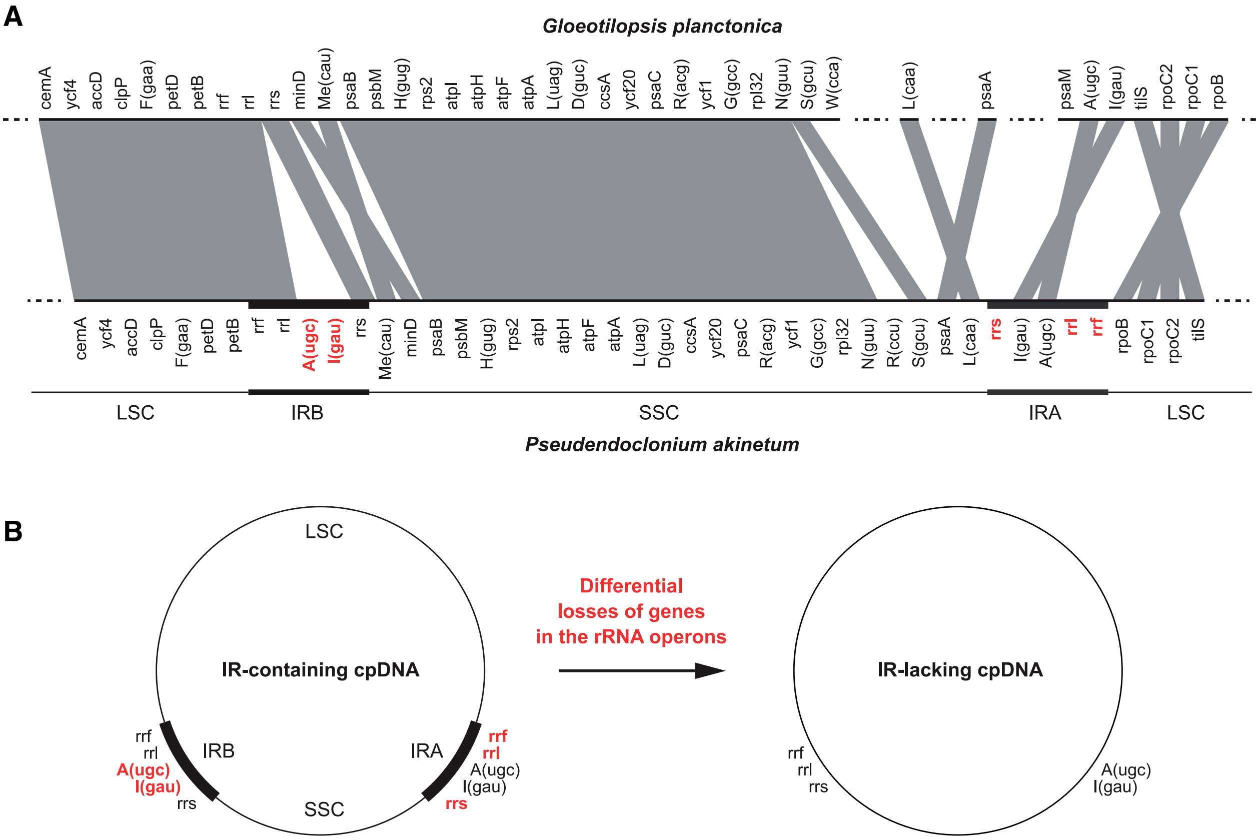


Fig. 4.—Model of IR loss from the chloroplast genome based on the comparison of the *Pseudendoclonium* and *G. planctonica* cpDNAs. (*A*) Comparison of gene order and gene content in the regions surrounding the genes making up the rRNA operon. Shown in red are the genes from the rRNA operon that are missing in *Gloeotilopsis*. (*B*) Model of IR loss proposed for the Ulotrichales. Genes differentially lost from the two copies of the rRNA operon are shown in red.

IR loss in the *Gloeotilopsis* lineage (fig. 4B). The alternative explanation that the *trnI*(gau) and *trnA*(ugc) gene pair in this large syntenic block was translocated to another locus is unlikely, given that this type of gene rearrangements have not been documented for chloroplast genomes. Analyses of additional ulotrichalean chloroplast genomes will be necessary to further support or reject the notion that genes making up the rRNA operon can be differentially lost from the IR copies, thereby leading to disruption of the rRNA operon, during the elimination of the quadripartite structure.

Intergenic Regions of the *G. sarcinoidea* cpDNA Include Sequences of Mitochondrial Origin, Gene Fragments, and Small Dispersed Repeats

Remarkably, short sequences of mitochondrial origin were detected at two distinct loci of the *G. sarcinoidea* cpDNA. The *psbC*/*ftsH* spacer includes a 530-bp sequence corresponding to the 5′ part of *atp1* and a 408-bp sequence spanning the entire coding region of *rps12*, whereas the *petB*/*rrf* spacer contains four coding sequences with a total length of 1529 bp that originate from *nad7* and the adjacent LHE-encoding gene present in ulotrichalean mtDNAs, as well as from the separate locus containing *trnS*(gcu) (fig. 3A). Although the chloroplast mitochondrial-like *rps12* and *trnS*(gcu) sequences cover the complete coding regions of these genes, they have clearly undergone pseudogenization: the *rps12* sequence displays frameshifts at two distinct locations resulting from 4-bp and 5-bp insertions and includes a UGA stop codon, whereas the *trnS*(gcu) sequence contains a 4-bp deletion within the stem of the T-arm and a C→T substitution in the highly conserved TΨC sequence of the T-loop.

The mitochondrial sequences we uncovered in the *G. sarcinoidea* cpDNA are unlikely to be the result of a chimeric assembly for the following reasons. First, the assembly of this genome yielded a single contig with overlapping sequences at both ends that allowed circularization of the genome. Second, coverage depth was high (average depth = 319) and uniform throughout the genome, and crucially, several hundreds of reads spanned the junctions between the mtDNA insertions and adjacent cpDNA sequences. Third, the sequences of the mtDNA insertions differed with regards to their homologs in the mitochondrial genome (see below).

To identify the potential donor(s) of the chloroplast mitochondrial-like *atp1*, *rps12*, *nad7*, and *trnS*(gcu) sequences, phylogenetic trees were inferred with RAxML using individual gene datasets comprising these sequences and the corresponding mitochondrial sequences available for *G. sarcinoidea*, *G. planctonica* and other green plants (fig. 5). All chloroplast mitochondrial-like sequences were found to be nested within the clade containing the *G. sarcinoidea* and *G. planctonica* mitochondrial genes, occupying a position sister to one of the *Gloeotilopsis* sequences or to both algal sequences. The *Gloeotilopsis* clade received strong bootstrap support in the *nad7* and *atp1* trees (bootstrap values of 92% and 94%, respectively). These results support the notion that mtDNA fragments were transferred to the chloroplast during the evolution of the *Gloeotilopsis* genus. Only the *nad7* tree is sufficiently resolved to distinguish unambiguously that the chloroplast captured this gene in the lineage leading to *G. sarcinoidea*. Although no mitochondrial-like sequences were identified with confidence in the *G. planctonica* cpDNA (*E*-value threshold of 1e−20), it is possible that mitochondrion-to-chloroplast DNA transfers occurred in a common ancestor of the two *Gloeotilopsis* species or in the lineage leading to *G. planctonica* and that the transferred mitochondrial sequences diverged rapidly following their insertion. In agreement with this hypothesis, we identified a short sequence showing limited similarity to *cox1* in the *G. planctonica* cpDNA (genomic coordinates 48,919–48,980, *E*-value of 2e−12).

To the best of our knowledge, our study provides the first evidence for the transfer of gene sequences from the mitochondria to the chloroplast in green algae. In flowering plants, only two cases of mitochondrion-to-plastid DNA transfer have been documented (Smith 2014); the first in a carrot ancestor (Goremykin et al. 2009; Iorizzo et al. 2012a, 2012b) and the second in the common milkweed (Straub et al. 2013). The scarcity of inter-organelle DNA exchanges that occurred in this direction in land plants has been suggested to be due to the lack of a specific DNA-uptake mechanism in the chloroplast (Bock 2010); however, it has been shown that under certain conditions that might be encountered during environmental stress, DNA can move across the chloroplast envelope due to transient alteration of its permeability (Cerutti and Jagendorf 1995). The DNA transfer reported for the carrot ancestor was proposed to result from transposition of a non-LTR retrotransposon, while that reported for the milkweed was attributed to homologous recombination events involving sequences of chloroplast origin that were identified near the transferred sequence in milkweed mtDNA. Although we have not detected any sequences of obvious chloroplast origin in *Gloeotilopsis* mitochondrial genomes, mtDNA fragments might have entered the *Gloeotilopsis* chloroplast genome through homologous recombination using small repeats shared by the two organelle genomes (see below).

Another unusual feature of the *G. sarcinoidea* cpDNA is the presence of duplicated sequences for *psbA*, *psbC, petD* and the three rRNA genes. These sequences, which are denoted by large dots in figure 3A, map to six distinct intergenic regions, including the two harboring mtDNA insertions, and except for *rrf*, which is entirely duplicated, they display either the 5′ or 3′ coding region. While repeats consisting of gene fragments (i.e., pseudogenes) are often found in highly rearranged chloroplast genomes of angiosperms (Guisinger et al. 2011), they have been rarely observed in green algal cpDNAs (Smith et al. 2010). The *G. sarcinoidea psbC*, *petD* and *rrl* pseudogenes correspond to the 3′ coding region and are located immediately downstream and on the same DNA strand as the functional gene copies. Similar arrangements, presumably resulting from the insertion of a mobile element, have been documented for the chloroplast of the chlorophycean green alga *Dunaniella salina* (Smith et al. 2010) and the mitochondria of the fungus *Allomyces macrogynus* (Paquin et al. 1994), and in both cases, an ORF encoding a homing endonuclease of the GIY-YIG family separates the functional protein-coding gene from the pseudogene. In this context, it is worth noting that for the *G. sarcinoidea* and *G. planctonica rrf*/*petB* intergenic regions, which differ by an addition/deletion 15.6 kb, the extra DNA segment of *G. sarcinoidea* contains at one junction a mtDNA insertion that includes part of a LAGLIDADG ORF and at the other junction the 3′-coding region of *rrl* and the second *rrf* copy (fig. 3C). In contrast to the *G. sarcinoidea* genome, the *G. planctonica* cpDNA exhibits only repeats of 5′ coding regions (*rns*, *psbA* and *rpoC2*) and the four detected pseudogenes lie within intergenic regions that coincide with endpoints of gene rearrangements between the two *Gloeotilopsis* genomes (supplementary fig. S2, Supplementary Material online).

The proportion of small repeats in the *G. sarcinoidea* genome (11.6%) is three-fold higher than in the *G. planctonica* cpDNA but is comparable to the values observed for *Oltmannsiellopsis* and *Bryopsis hypnoides* (table 2). Most of the *G. sarcinoidea* repeats reside in intergenic regions where they represent a diversified collection of sequences; a small proportion of repeats (spanning 2.0% of the genome) is accounted for by the presence of gene fragments and highly conserved group II intron sequences (see below). We identified many tandem repeats in *G. sarcinoidea*, several of which are dispersed throughout the genome. Notably, repeat units of 15 and 13 bp were found to be shared with the mtDNAs of *G. sarcinoidea* (TGGGAAAACTTTTCC, 11 identical copies in cpDNA and 8 in mtDNA) and *G. planctonica* (TTTATCAAA AAAG, 112 identical copies in cpDNA and 12 in mtDNA), respectively. The 13-bp unit is part of a 20-bp sequence (TTT ATCAAAAAAGTTTTTTG) that is repeated in tandem at 19 different sites in the *G. sarcinoidea* chloroplast genome.

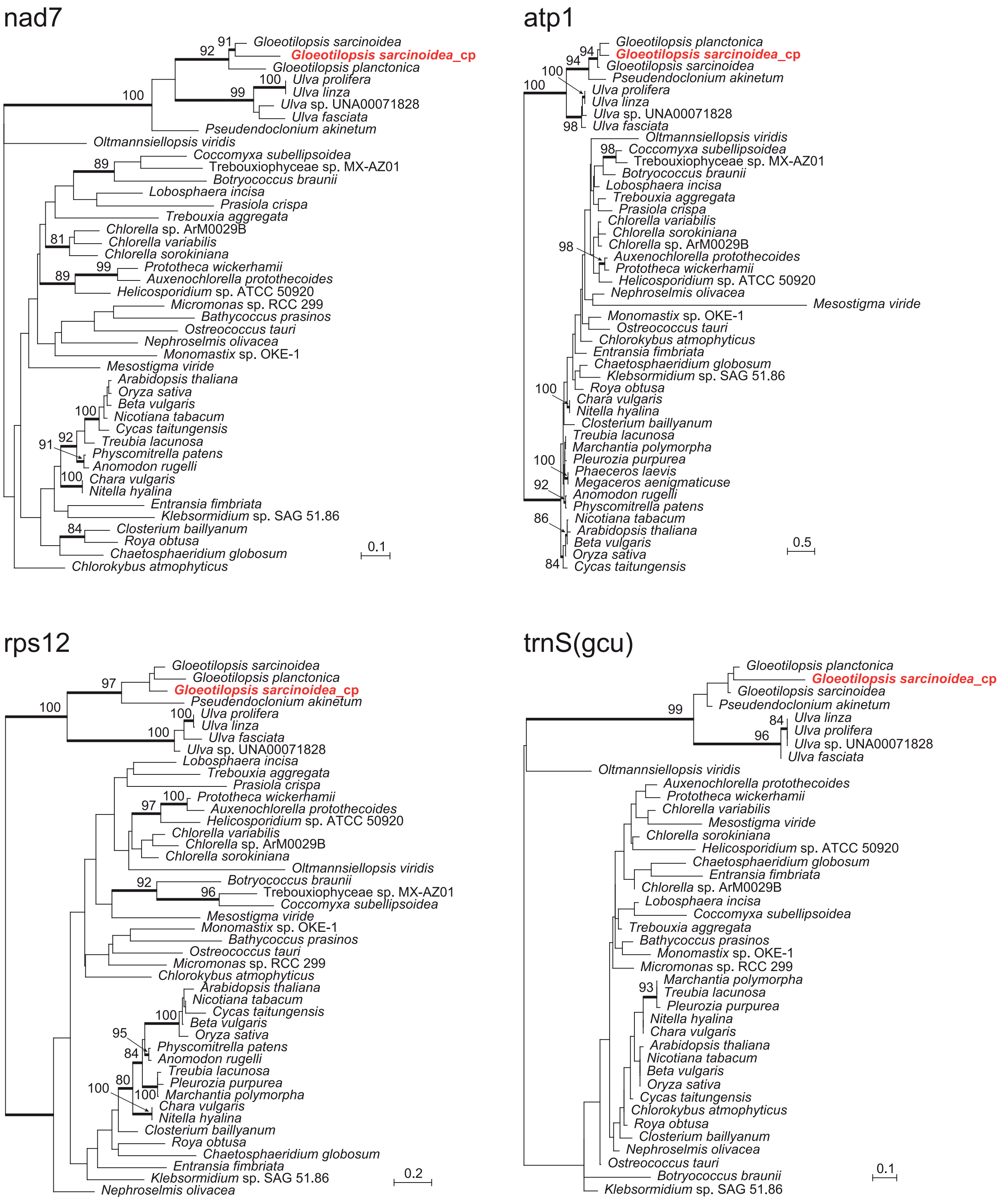
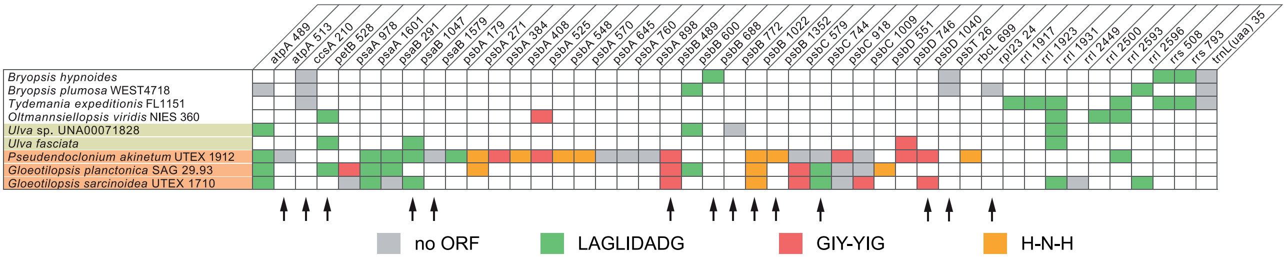


Fig. 5.—Phylogenetic relationships between the mitochondrial-like sequences in *G. sarcinoidea* cpDNA and the corresponding genes of green plant mtDNAs. The RAxML trees were inferred using the GTR + Γ4 model. Branches that received ≥ 80% bootstrap support are denoted by thick lines. The names of the *G. sarcinoidea* mitochondrial-like cpDNA sequences are highlighted in red.

All *Gloeotilopsis* Chloroplast Group I introns Occur at Previously Identified Insertion Sites

The group I introns identified so far in ulvophycean cpDNAs are distributed among 15 genes and represent a total of 45 insertion sites, most of which also occur in other groups of chlorophytes (fig. 6). The 14 *G. planctonica* and 15 *G. sarcinoidea* introns are inserted at 19 distinct sites. Considering that most of these sites are also found in *Pseudendoclonium*



	atpA 489	atpA 513	ccsA 210	petB 528	psaA 978	psaA 1601	psaB 291	psaB 1047	psaB 1579	psbA 179	psbA 271	psbA 384	psbA 408	psbA 525	psbA 548	psbA 570	psbA 645	psbA 760	psbA 898	psbB 489	psbB 600	psbB 688	psbB 772
Bryopsis hypnoides			▨																			🟩	
Bryopsis plumosa WEST4718	▨		▨																		🟩		
Tydemania expeditionis FL1151			▨																				
Oltmannsiellopsis viridis NIES 360				🟩										🟥									
Ulva sp. UNA00071828	🟩																				🟩		▨
Ulva fasciata				🟩				🟩															
Pseudendoclonium akinetum UTEX 1912	🟩	▨				🟩	🟩	🟩	▨	🟩	🟧	🟥	🟧	🟥	🟧	🟧	▨	▨	▨	🟥			
Gloeotilopsis planctonica SAG 29.93	🟩			🟩	🟥	🟩	🟩				🟧									🟥	🟩		
Gloeotilopsis sarcinoidea UTEX 1710	🟩				▨	🟩	▨	🟩												🟥			
		↑	↑	↑				↑	↑											↑		↑	↑

	psbB 1022	psbB 1352	psbC 579	psbC 744	psbC 918	psbC 1009	psbD 551	psbD 746	psbD 1040	psbT 26	rbcL 699	rpl23 24	rrl 1917	rrl 1923	rrl 1931	rrl 2449	rrl 2500	rrl 2593	rrl 2596	rrs 508	rrs 793	trnL(uaa) 35
Bryopsis hypnoides										▨										🟩	🟩	▨
Bryopsis plumosa WEST4718										▨		▨							🟩			▨
Tydemania expeditionis FL1151													🟩	🟩	🟩			🟩		🟩	🟩	▨
Oltmannsiellopsis viridis NIES 360															🟩		🟩	🟩				
Ulva sp. UNA00071828															🟩							
Ulva fasciata								🟥							🟩							
Pseudendoclonium akinetum UTEX 1912	🟧	🟧	▨	▨	🟥	▨		🟥	🟥		🟧							🟩				
Gloeotilopsis planctonica SAG 29.93	🟧		🟥	🟩	▨	▨	🟧															
Gloeotilopsis sarcinoidea UTEX 1710	🟧		🟥	🟩	▨	🟥			🟥						🟩	▨			🟩			
	↑	↑		↑					↑	↑		↑										

▨ no ORF 🟩 LAGLIDADG 🟥 GIY-YIG 🟧 H-N-H

Fig. 6.—Group I introns in *Gloeotilopsis* and other ulvophycean cpDNAs. A grey box represents an intron lacking an ORF, whereas a colored box represents an intron containing an ORF (see the color code for the type of intron-encoded protein). Arrows denote the intron insertion sites that have not been observed in other groups of chlorophytes. Intron insertion sites in protein-coding and tRNA genes are given relative to the corresponding genes in *M. viride* cpDNA (Lemieux et al. 2000); insertion sites in *rrs* and *rrl* are given relative to *E. coli* 16S and 23S rRNAs, respectively.

and/or *Ulva* species, many group I introns were likely inherited vertically in ulotrichalean and ulvalean lineages. The majority of the *Gloeotilopis* introns (those at 17 sites) display an ORF, with the LAGLIDADG family of homing endonucleases being the most represented (nine sites), followed by the GIY-YIG family (five sites). Based on the observation that the *Pseudendoclonium* chloroplast *atpA* intron at site 489 and the mitochondrial *atp1* intron at site 522 share not only the same insertion site but also highly similar secondary structures and LAGLIDADG ORFs, it has been hypothesized earlier that an inter-organellar, lateral DNA transfer event involving a group I intron took place specifically in the Ulvophyceae (Pombert et al. 2005). The direction of this inter-organellar DNA transfer still remains unclear at this point owing to the limited information currently available for the distributions of the *atpA* 489 and *atp1* 522 introns among chlorophytes.

Many Group II introns in the *Gloeotilopsis* Chloroplast Genomes Arose by Intragenomic Proliferation

Group II introns are scattered among 26 protein-coding genes where they occupy 38 distinct sites, and in contrast to the situation prevailing for group I introns, the great majority of the insertion sites appear to be unique to ulvophycean green algae, with 24 sites specific to *Gloeotilopsis* species, just one site specific to the Ulvales, and nine to the Bryopsidales (fig. 7A). *G. planctonica* and *G. sarcinoidea* share only two insertion sites, one of which is also present in the Bryopsidales (*atpA* 753). All 12 *G. sarcinoidea* group II introns encode proteins with RT, intron maturase and/or H-N-H endonuclease domains, whereas most of the 17 *G. planctonica* introns are lacking ORFs and thus depend on proteins encoded by other group II introns for splicing and mobility. As reported for the genome of the thermophilic cyanobacterium *Thermosynechoccus elongatus* in which 20 of the 28 group II introns do not encode any proteins (Mohr et al. 2010), the high proportion of ORF-less introns in *G. planctonica* may reflect an evolutionary pressure for a smaller and more compact intron structure enabling increased efficiency of splicing and mobility.

Remarkably, the first intron in *G. planctonica petA* as well as the *psbZ* intron of the same alga are located in 5′ untranslated gene sequences. These observations represent the first cases of introns in non-coding regions of green plant organelle genes. The *petA* and *psbZ* introns were detected owing to their high sequence identity with several other group II introns of the *G. planctonica* and *G. sarcinoidea* cpDNAs. Indeed, in the course of identifying repeats using LAST and REPUTER, we found that several chloroplast group II introns at distinct insertion sites share long regions with high nucleotide identities. For instance, the *G. sarcinoidea ycf3* and *psbH* introns are 85.6% identical over an alignment of 2563 bp. Note that LAST analysis also revealed in the 3′ untranslated sequence of the *G. planctonica psbZ* the remnant sequence of a group II intron (positions 151,853–152,004) that is similar (*E*-value of 1.3e−31) to the 5′ part of domain I of the *ycf12* intron in the same alga.

To delineate the relationships among the 29 *Gloeotilopsis* group II introns, a global alignment of 286 nucleotides corresponding to the core secondary structures of these introns was submitted to phylogenetic analysis using RAxML under the GTR + G4 model (fig. 7B). Four clades (I–IV) with a total of 23 introns received bootstrap support ≥79%. The introns in clades I, II, and III occupy distinct insertion sites and form three families of very closely related sequences as judged by the very short lengths of the observed branches. Clade III is unique in comprising only ORF-less introns from *G. planctonica*; deletion of the ORF probably occurred prior to the divergence of the five introns, with both splicing and dispersal to novel (ectopic) sites being promoted by the protein encoded by a closely related, but not yet identified group II intron. In addition to the *rbcL* intron unique to *G. sarcinoidea*, clade IV comprises the *Gloeotilopsis atpA* and *atpB* introns sharing common insertion sites; the distinct clades formed by the latter two pairs of introns support the idea that the *G. planctonica* and *G. sarcinoidea* introns in each clade are related through vertical descent. The fifth clade that we identified displays the most divergent introns. These findings, which highlight the first case of intragenomic proliferation of organellar group II introns in the Viridiplantae, suggest that several independent waves of

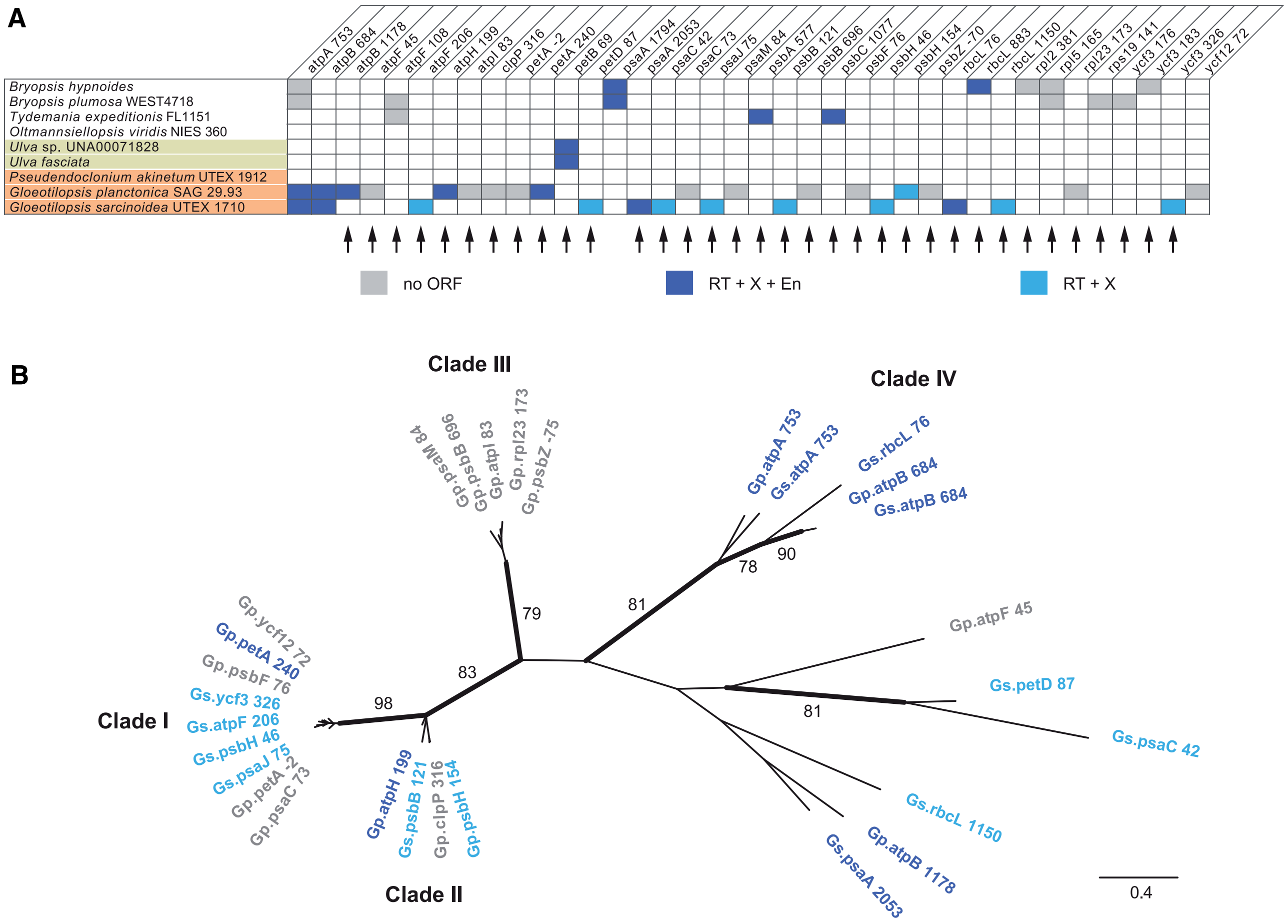
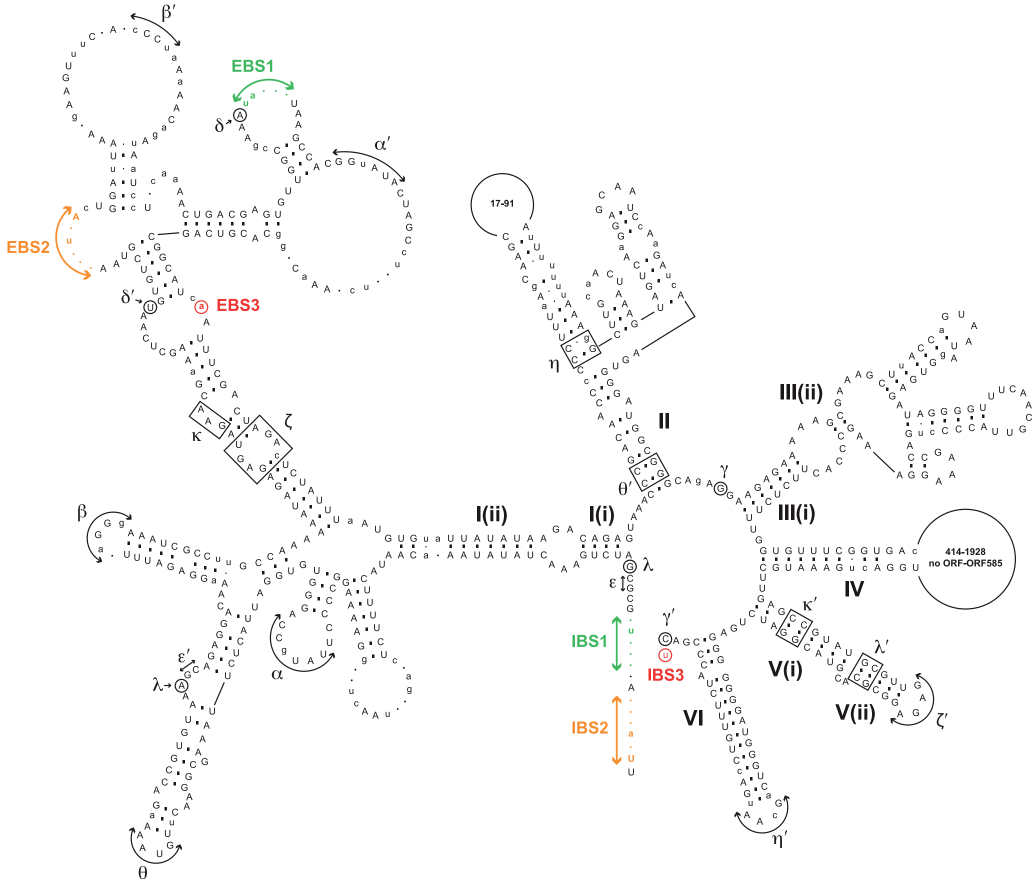


Fig. 7.—Analysis of group II introns in *Gloeotilopsis* cpDNAs. (*A*) Comparison of their insertion sites and ORFs with those of group II introns in other ulvophycean cpDNAs. A grey box represents an intron lacking an ORF, whereas a colored box represents an intron containing an ORF (see the color code for the domains present in the encoded protein: RT, reverse transcriptase; X, intron maturase; EN, H-N-H endonuclease). Arrows denote the intron insertion sites that have not been observed in other groups of chlorophytes. Intron insertion sites in protein-coding and tRNA genes are given relative to the corresponding genes in *M. viride* cpDNA (Lemieux et al. 2000); insertion sites in *rrs* and *rrl* are given relative to *E. coli* 16S and 23S rRNAs, respectively. For each insertion site, the position corresponding to the nucleotide immediately preceding the intron is reported. Note that a few introns were not correctly annotated in the GenBank accessions of Bryopsidales and Ulvales and that changes related to intron types and insertion positions were introduced during this study. (*B*) Phylogenetic relationships among *Gloeotilopsis* group II introns. This tree was inferred by RAxML analysis of an alignment of 286 nucleotides corresponding to the core secondary structures of the group II introns. Branches that received ≥ 78% bootstrap support are denoted by thick lines (the actual bootstrap values are shown). Gp, *G. planctonica*; Gs, *G. sarcinoidea*.

intron dispersal occurred in the *Gloeotilopsis* chloroplast. Evidence for group II intron proliferation to high copy number in chloroplast genomes has previously been reported only for euglenids (Hallick et al. 1993; Pombert et al. 2012; Bennett and Triemer 2015) and the red alga *Porphyridium purpureum* (Perrineau et al. 2015).

Different pathways of intron proliferation to ectopic sites have been elucidated for the *Lactococcus lactis* Ll.LtrB IIA intron in its native host and in *Escherichia coli*, and for the CL/IIB1 introns in the thermophilic cyanobacterium *Thermosynechoccus elongatus* (Lambowitz and Belfort 2015). Although these mechanisms have in common a target DNA-primed reverse transcription step in which the excised intron RNA reverse splices into one strand of a DNA target site and is then reverse transcribed by the intron-encoded protein to produce an intron cDNA that is integrated into the genome, they differ in a number of ways, including the recognition of the target site by the ribonucleoprotein complex formed by the group II intron RNA and intron-encoded protein. To gain insights into the pathway underlying the dissemination of group II introns in *Gloeotilopsis* cpDNAs, we compared the secondary structure models of the introns from clade I as well as the flanking 5′ and 3′ exon sequences that contain the intron-binding sites IBS1, IBS2, and IBS3. The consensus of these intron secondary structures is characteristic of the CL/IIB1 introns (Toor et al. 2001), with 84% of the represented positions (545/651) displaying identical residues in at least seven of the nine introns (fig. 8). Remarkably, the



Clade I

		5′ Exon	Intron			3′ Exon
		IBS2 IBS1	EBS2	EBS1	EBS3	IBS3
Gp.petA	-2	TCAACTTTAACTAAACTAAA	UAAUAGUAACUG	AAAUUUAGCAA	UCAAU	CATGAC
Gp.petA	240	CTGTTTTTGAAGCACTTGTT	UAAGCUUCAUUG	AAAAACAAUAA	UGUAU	AAAATC
Gp.psaC	73	ACGTGCTTGTCCAACAGATG	AAACGGACAA-G	AAACAUCUUAA	UCAAU	TTTTAG
Gp.psbF	76	TCATGCTTTAGCAGTTCCAA	UAAGGCUAACUG	AAAUUGGAUAA	UCGAU	CCGTTT
Gp.ycf12	72	AGGTCCATTAGTAATCGTTT	UAACACUAACUG	AAAAAACGUAC	UUAAU	TACTAG
Gs.atpF	206	AACAATTTACGAGAAGCAGA	UAACUCGUACUG	AAAUUUGCUAA	UCAAU	TAATCG
Gs.psaJ	5	TGACTTTTACAGCAGGTTTA	UAAGCUGUACCG	AAAUAAACUAA	UCAAU	TTAATT
Gs.psbH	46	GACACCTTTAGGTACTTTAT	UAAACCUAACUG	AAGUAAAUUAA	UCAAU	TACGTC
Gs.ycf3	326	CCTTCTTTACCTCAAGCTTT	UAAGAGGUAAUA	AAAAAAGCUAA	UUUAU	AAATAA

Clade II

		5′ Exon	Intron			3′ Exon
		IBS2 IBS1	EBS2	EBS1	EBS3	IBS3
Gp.atpH	199	GGAATCTTTAACAATTTATG	GUAAAGUUAAUG	UUACAUAAAAA	UCCAC	GATTAG
Gp.clpP	316	GGCTTCAATGGCTTCTTGCG	GUUAAGCCAAUA	UUACGCAAAAG	UCAAC	TTTTAG
Gp.psbH	154	ATTTTTAGTAATTATTTTAG	GUCAAAUUAACA	UCACUAAAACG	UCAAC	AAATTT
Gp.psbB	121	TGGTTCAATGGCTTTTTACG	GUUACGCCAAAU	UUACGUAAAAG	UCUAC	AACTTG

Clade III

		5′ Exon	Intron			3′ Exon
		IBS2 IBS1	EBS2	EBS1	EBS3	IBS3
Gp.psaM	84	TTCGTTTAGGAATAGCTTTA	CCUAAUUCCAAG	UUUUAAAGUGA	AAAUG	TATCGT
Gp.psbB	696	ACAATGCTTTACGAATGGGC	CUUAGUAAUAAG	UUAUUUCAUGA	AAUUG	AACGTT
Gp.atpI	83	GGAAAGCACTACTATTGGCA	CUUAAGUAACAG	UUUUGCCAAGA	AAUUG	AATTGG
Gp.rpl23	173	AATTCACATTTATTACCAAC	UUUAAUAAACAG	UUUGUUGAUGA	AACUG	CAAAAA
Gp.psbZ	-75	CGAATACAAATTTAACTTAA	CUUAAAAAACAG	UUUUUAAGUAA	AAGUG	CATTTT

Fig. 8.—Consensus secondary structure model of the nine clade-I group II introns and insertion sites of introns from clades I, II, and III. The consensus secondary structure model is displayed according to Toor et al. (2001). Highly conserved (in seven or more introns) and less conserved (in at least five introns) nucleotide positions are shown in uppercase and lowercase characters respectively; the other residues are denoted by dots. Highly conserved (in seven or more introns) and less conserved base pairings (in at least five introns) are denoted by thick and thin dashes respectively. Roman numbers specify the major structural domains, with long-range tertiary interactions being denoted by Greek letters. Variations in size of peripheral regions are indicated by numbers inside the loops. The 5′ and 3′ exon sequences flanking the introns from clades I, II, and III are aligned along with the intron sequences spanning the exon-binding sites EBS1, EBS2, and EBS3. Colors highlight EBS sequences and complementary nucleotide residues in the IBS exon sequences.

single-stranded loops of domain I containing the exon-binding sites EBS1, EBS2, and EBS3, which play a major role in exon recognition during intron splicing and retrohoming, feature highly variable sequences. But the EBS1 and EBS2 sequences of each intron show perfect or nearly perfect complementarity with the IBS sequences in the 5′ exon, and only the EBS3 motif of the *G. planctonica* intron located in the 5′ untranslated region of *petA* cannot base pair with IBS3 in the 3′ exon (fig. 8). Essentially the same observations were made regarding the EBS-IBS interactions of the clade-II and clade-III introns (fig. 8).

The above results suggest that the spread of group II introns to ectopic sites in the *Gloeotilopsis* chloroplast genome occurred through several independent waves of retrohoming—the mobility mechanism used to maintain group II introns at cognate sites—following mutations in the EBS sequences of founding introns that enabled insertions into different target sites. It has been previously shown that retrohoming and divergence of DNA target specificity also contributed to the proliferation of CL/IIB1 introns in the *Thermosynechoccus* genome (Mohr et al. 2010) and of a subset of group II introns in *Wolbachia* bacterial endosymbionts (Leclercq et al. 2011). In contrast, as reported for the yeast mitochondrial group II intron aI1 inserted in the *cox1* gene (Dickson et al. 2001), the Ll.LtrB intron spreads to ectopic sites in its native host by retrotransposition, a mechanism occurring at lower frequency than retrohoming and involving relaxed EBS-IBS sequence interactions for intron integration (Ichiyanagi et al. 2002).

Conclusion

Prior to this study, *Pseudendoclonium* was the sole representative of the Ulotrichales that had been sampled for mitochondrial and chloroplast genome sequencing. By incorporating the newly sequenced organelle genomes of two *Gloeotilopsis* species, our comparative analyses have provided a better understanding of the evolutionary histories of the mitochondrial and chloroplast genomes in the Ulvales–Ulotrichales, revealing novel genomic features in both genomes—LAGLIDADG ORF-containing group II introns, group II introns in non-coding regions, and pseudogenes resulting from duplications—as well as unanticipated phenomena that contributed to the expansion of the chloroplast genome in the *Gloeotilopsis* lineage. We showed that sequences accumulated in the chloroplast genome in this lineage through intracellular mitochondrion-to-chloroplast DNA transfers and intragenomic proliferation of group II introns through EBS sequence divergence and retrohoming.

Moreover, our results revealed that the chloroplast genome experienced a minimum of two independent events of IR loss in the Ulvales–Ulotrichales, one during the evolutionary period separating the *Pseudendoclonium* and *Gloeotilopsis* lineages and the other during the evolution of ulvalean green algae. The comparison of the *Pseudendoclonium* and *Gloeotilopsis* cpDNAs offered clues regarding the mechanism of IR loss in the Ulotrichales, suggesting that internal sequences from the rDNA operon were differentially lost from the two original IR copies during this process. Analysis of chloroplast genomes from additional representatives of the Ulotrichales will be necessary to test the hypothesis that breakage of the rDNA operon was intimately linked with IR loss in this lineage.

Supplementary Material

Supplementary table S1 and supplementary figures S1 and S2 are available at *Genome Biology* and *Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgment

This work was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada [2830-2007 to M.T. and C.L.].

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Andersen RA. 2005. Algal Culturing Techniques. Boston (MA): Elsevier/Academic Press.

Bao Z, Eddy SR. 2002. Automated *de novo* identification of repeat sequence families in sequenced genomes. Genome Res. 12:1269–1276.

Belfort M, Derbyshire V, Parker MM, Cousineau B, Lambowitz AM. 2002. Mobile introns: pathways and proteins. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington (DC): American Society of Microbiology Press.

Bennett MS, Triemer RE. 2015. Chloroplast genome evolution in the Euglenaceae. J Eukaryot Microbiol. 62:773–785.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27:573–580.

Blazier JC, et al. 2016. Variable presence of the inverted repeat and plastome stability in *Erodium.* Ann Bot 117:1209–1220.

Bock R. 2010. The give-and-take of DNA: horizontal gene transfer in plants. Trends Plant Sci. 15:11–22.

Boisvert S, Laviolette F, Corbeil J. 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol. 17:1519–1533.

Brouard JS, Otis C, Lemieux C, 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.

Cerutti H, Jagendorf A. 1995. Movement of DNA across the chloroplast envelope: implications for the transfer of promiscuous DNA. Photosynth Res. 46:329–337.

Cocquyt E, Verbruggen H, Leliaert F, De Clerck O. 2010. Evolution and cytological diversification of the green seaweeds (Ulvophyceae). Mol Biol Evol. 27:2052–2061.

de Cambiaire J-C, Otis C, Lemieux C, Turmel M. 2006. The complete chloroplast genome sequence of the chlorophycean green alga *Scenedesmus obliquus* reveals a compact gene organization and a biased distribution of genes on the two DNA strands. BMC Evol Biol. 6:37.

Dickson L, et al. 2001. Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites. Proc Natl Acad Sci U S A. 98:13207–13212.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Friedl T, Rybalka N. 2012. Systematics of the Green Algae: A Brief Introduction to the Current Status. In: Luttge U, Beyschlag W, Budel B, Francis D, editors. Progress in Botany, Vol. 73 Berlin, Germany: Springer-Verlag. p. 259–280.

Frith MC, Hamada M, Horton P. 2010. Parameters for accurate genome alignment. BMC Bioinformatics 11:1–14.

Fucikova K, et al. 2014. New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. Front Ecol Evol. 2:63.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8:195–202.

Goremykin VV, Salamini F, Velasco R, Viola R. 2009. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol Biol Evol. 26:99–110.

Goulding SE, Olmstead RG, Morden CW, Wolfe KH. 1996. Ebb and flow of the chloroplast inverted repeat. Mol Gen Genet. 252:195–206.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol. 28:583–600.

Hallick RB, et al. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res. 21:3537–3544.

Ichiyanagi K, et al. 2002. Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. Mol Microbiol. 46:1259–1272.

Iorizzo M, et al. 2012a. Against the traffic: the first evidence for mitochondrial DNA transfer into the plastid genome. Mob Genet Elements 2:261–266.

Iorizzo M, et al. 2012b. *De novo* assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol. 12:61.

Kurtz S, et al. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29:4633–4642.

Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35:3100–3108.

Lambowitz AM, Belfort M. 2015. Mobile bacterial group II introns at the crux of eukaryotic evolution. Microbiol Spectr 3:MDNA3-0050-2014.

Leclercq S, Giraud I, Cordaux R. 2011. Remarkable abundance and evolution of mobile group II introns in *Wolbachia* bacterial endosymbionts. Mol Biol Evol. 28:685–697.

Leliaert F, Lopez-Bautista JM. 2015. The chloroplast genomes of *Bryopsis plumosa* and *Tydemania expeditiones* (Bryopsidales, Chlorophyta): compact genomes and genes of bacterial origin. BMC Genomics 16:204.

Leliaert F, et al. 2012. Phylogeny and molecular evolution of the green algae. CRC Crit Rev Plant Sci. 31:1–46.

Lemieux C, Otis C, Turmel M. 2000. Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. Nature 403:649–652.

Lemieux C, Otis C, 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.

Lemieux C, Otis C, 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.

Lohse M, Drechsel O, 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.

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.

Lu F, et al. 2011. The *Bryopsis hypnoides* plastid genome: multimeric forms and complete nucleotide sequence. PLoS One 6:e14663.

Magoc T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963.

Margulies M, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17:10–12.

Melton JT, 3rd, Leliaert F, Tronholm A, Lopez-Bautista JM. 2015. The complete chloroplast and mitochondrial genomes of the green macroalga *Ulva* sp. UNA00071828 (Ulvophyceae, Chlorophyta). PLoS One 10:e0121020.

Melton JT, 3rd, Lopez-Bautista JM. 2016. *De novo* assembly of the mitochondrial genome of *Ulva fasciata* Delile (Ulvophyceae, Chlorophyta), a distromatic blade-forming green macroalga. Mitochondrial DNA 27:3817–3819.

Melton JT, Lopez-Bautista JM. 2015. The chloroplast genome of the marine green macroalga *Ulva fasciata* Delile (Ulvophyceae, Chlorophyta). Mitochondrial DNA 1–3.

Michel F, Umesono 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.

Mohr G, Ghanem E, Lambowitz AM. 2010. Mechanisms used for genomic proliferation by thermophilic group II introns. PLoS Biol. 8:e1000391.

Monteiro-Vitorello CB, et al. 2009. The *Cryphonectria parasitica* mitochondrial *rns* gene: plasmid-like elements, introns and homing endonucleases. Fungal Genet Biol. 46:837–848.

Mullineux ST, Costa M, Bassi GS, Michel F, Hausner G. 2010. A group II intron encodes a functional LAGLIDADG homing endonuclease and self-splices under moderate temperature and ionic conditions. RNA 16:1818–1831.

Paquin B, Laforest MJ, Lang BF. 1994. Interspecific transfer of mitochondrial genes in fungi and creation of a homologous hybrid gene. Proc Natl Acad Sci U S A. 91:11807–11810.

Perrineau MM, Price DC, Mohr G, Bhattacharya D. 2015. Recent mobility of plastid encoded group II introns and twintrons in five strains of the unicellular red alga *Porphyridium*. Peer J. 3:e1017.

Pfeifer A, Martin B, Kamper J, Basse CW. 2012. The mitochondrial LSU rRNA group II intron of *Ustilago maydis* encodes an active homing endonuclease likely involved in intron mobility. PLoS One 7:e49551.

Pombert JF, James ER, Janouskovec J, Keeling PJ. 2012. Evidence for transitional stages in the evolution of euglenid group II introns and twintrons in the *Monomorphina aenigmatica* plastid genome. PLoS One 7:e53433.

Pombert JF, Lemieux C, Turmel M. 2006. The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. BMC Biol. 4:3.

Pombert JF, Otis C, Lemieux C, Turmel M. 2005. The 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.

Pombert JF, Otis C, Lemieux C, Turmel M. 2004. The complete mitochondrial DNA sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) highlights distinctive evolutionary trends in the chlorophyta and suggests a sister-group relationship between the Ulvophyceae and Chlorophyceae. Mol Biol Evol. 21:922–935.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European molecular biology open software suite. Trends Genet. 16:276–277.

Ruhlman TA, Jansen RK. 2014. The plastid genomes of flowering plants. In: Maliga P, editor. Chloroplast Biotechnology. Totowa (NJ): Humana Press. p. 3–38.

Salman V, Amann R, Shub DA, Schulz-Vogt HN. 2012. Multiple self-splicing introns in the 16S rRNA genes of giant sulfur bacteria. Proc Natl Acad Sci U S A. 109:4203–4208.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864.

Škaloud P, Nedbalová L, Elster J, Komárek J. 2013. A curious occurrence of *Hazenia broadyi* spec. nova in Antarctica and the review of the genus *Hazenia* (Ulotrichales, Chlorophyceae). Polar Biol. 36:1281–1291.

Smith DR. 2011. Extending the limited transfer window hypothesis to inter-organelle DNA migration. Genome Biol Evol. 3:743–748.

Smith DR. 2014. Mitochondrion-to-plastid DNA transfer: it happens. New Phytol 202:736–738.

Smith DR, et al. 2010. The *Dunaliella salina* organelle genomes: large sequences, inflated with intronic and intergenic DNA. BMC Plant Biol. 10:83.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Straub SC, Cronn RC, Edwards C, Fishbein M, Liston A. 2013. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (apocynaceae). Genome Biol Evol. 5:1872–1885.

Sun L, et al. 2016. Chloroplast phylogenomic inference of green algae relationships. Sci Rep 6:20528.

Tesler G. 2002. GRIMM: genome rearrangements web server. Bioinformatics 18:492–493.

Toor N, Hausner G, Zimmerly S. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. RNA 7:1142–1152.

Toor N, Zimmerly S. 2002. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. RNA 8:1373–1377.

Turmel M, de Cambiaire JC, Otis C, 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.

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.

Turmel M, Otis C, Lemieux C. 2002. 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. 2015. Dynamic evolution of the chloroplast genome in the green algal classes Pedinophyceae and Trebouxiophyceae. Genome Biol Evol. 7:2062–2082.

Valach M, Burger G, Gray MW, Lang BF. 2014. Widespread occurrence of organelle genome-encoded 5S rRNAs including permuted molecules. Nucleic Acids Res. 42:13764–13777.

Wang RJ, et al. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol Biol. 8:36.

Zhou L, Wang L, Zhang J, Cai C, He P. 2016a. Complete mitochondrial genome of *Ulva linza*, one of the causal species of green macroalgal blooms in Yellow Sea, China. Mitochondrial DNA Part B. 1:31–33.

Zhou L, Wang L, Zhang J, Cai C, He P. 2016b. Complete mitochondrial genome of *Ulva prolifera*, the dominant species of green macroalgal blooms in Yellow Sea, China. Mitochondrial DNA Part B. 1:76–78.

Zimmerly S, Semper C. 2015. Evolution of group II introns. Mob DNA 6:7.

Associate editor: John Archibald