Mol Gen Genomics (2006) 276:464–477
DOI 10.1007/s00438-006-0156-2

ORIGINAL PAPER

Distinctive architecture of the chloroplast genome in the chlorophycean green alga *Stigeoclonium helveticum*

Anne-Sophie Bélanger · Jean-Simon Brouard · Patrick Charlebois · Christian Otis · Claude Lemieux · Monique Turmel

Received: 26 May 2006 / Accepted: 29 July 2006 / Published online: 31 August 2006

Abstract The chloroplast genome has experienced many architectural changes during the evolution of chlorophyte green algae, with the class Chlorophyceae displaying the lowest degree of ancestral traits. We have previously shown that the completely sequenced chloroplast DNAs (cpDNAs) of *Chamydomonas reinhardtii* (Chlamydomonadales) and *Scenedesmus obliquus* (Sphaeropleales) are highly scrambled in gene order relative to one another. Here, we report the complete cpDNA sequence of *Stigeoclonium helveticum* (Chaetophorales), a member of a third chlorophycean lineage. This genome, which encodes 97 genes and contains 21 introns (including four putatively *trans*-spliced group II introns inserted at novel sites), is remarkably rich in derived features and extremely rearranged relative to its chlorophycean counterparts. At 223,902 bp, *Stigeoclonium* cpDNA is the largest chloroplast genome sequenced thus far, and in contrast to those of *Chlamydomonas* and *Scenedesmus*, features no large inverted repeat. Interestingly, the pattern of gene distribution between the DNA strands and the bias in base composition along each strand suggest that the *Stigeoclonium* genome replicates bidirectionally from a single origin. Unlike most known *trans*-spliced group II introns, those of *Stigeoclonium* exhibit breaks in domains I and II. By placing our comparative genome analyses in a phylogenetic framework, we inferred an evolutionary scenario of the mutational events that led to changes in genome architecture in the Chlorophyceae.

Keywords Chlorophyta · Plastid genome evolution · Gene order · Origin of replication · Group II introns · Repeated sequences

Introduction

As revealed by the complete chloroplast DNA (cpDNA) sequences that have been reported so far for green plants, the chloroplast genome has evolved much less conservatively in the phylum Chlorophyta than in the Streptophyta. The Chlorophyta (Sluiman 1985) comprises the majority of extant green algae and is divided into four classes: the Prasinophyceae, Ulvophyceae, Trebouxiophyceae and Chlorophyceae. The Prasinophyceae represent the most basal divergence of the Chlorophyta (Friedl 1997; Lewis and McCourt 2004) and, although the branching order of the Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC) remains uncertain (Friedl and O'Kelly 2002), analyses of chloroplast genomic features and phylogenetic data derived from mitochondrial genome sequences suggest that the Trebouxiophyceae emerged before the Ulvophyceae and Chlorophyceae (Pombert et al. 2004, 2005, 2006). Complete chloroplast genome

Nucleotide sequence data reported are available in the GenBank database under the accession number DQ630521.

Communicated by R. Herrmann.

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00438-006-0156-2 and is accessible for authorized users.

A.-S. Bélanger · J.-S. Brouard · P. Charlebois · C. Otis · C. Lemieux · M. Turmel (✉)
Département de biochimie et de microbiologie,
Pavillon Charles-Eugène Marchand, Université Laval,
Quebec City, QC, Canada G1K 7P4
e-mail: Monique.Turmel@rsvs.ulaval.ca

sequences have been reported for only six chlorophytes: the prasinophyte *Nephroselmis olivacea* (Turmel et al. 1999b), the trebouxiophyte *Chlorella vulgaris* (Wakasugi et al. 1997), two green algae representing distinct basal lineages of the Ulvophyceae, *Oltmannsiellopsis viridis* (Pombert et al. 2006) and *Pseudendoclonium akinetum* (Pombert et al. 2005), and also representatives of two different lineages of the Chlorophyceae, *Chlamydomonas reinhardtii* (Maul et al. 2002) and *Scenedesmus obliquus* (de Cambiaire et al. 2006). The Streptophyta (Bremer 1985), on the other hand, unites all embryophytes (land plants) and their closest green algal relatives, the members of the class Charophyceae sensu Mattox and Stewart (1984). The currently available chloroplast genome sequences of about 35 photosynthetic land plants and seven charophycean green algae disclosed a high degree of conservation in overall structure and overall gene arrangement (Palmer 1991; Turmel et al. 2002, 2005, 2006). The vast majority of these genomes harbour the same quadripartite structure and gene partitioning pattern, their genes (106–137) are tightly packed, and most of them are grouped into multicistronic operons, several of which are evolutionarily related to those found in cyanobacteria, the progenitors of chloroplasts.

In the Chlorophyta, the chloroplast genome appears to have been progressively remodelled and to have gradually lost the many ancestral features observed in the Streptophyta, with the Prasinophyceae and Chlorophyceae exhibiting the highest and lowest levels, respectively. The gene-rich (128 genes) and compact cpDNA of the prasinophyte *Nephroselmis* displays the characteristic quadripartite structure and gene partitioning pattern found in streptophyte genomes as well as the great majority of their ancestral operons (Turmel et al. 1999b). This quadripartite structure is characterized by the presence of two copies of a large inverted repeat sequence (IR) separating a small single-copy (SSC) and a large single-copy region (LSC). The chloroplast genome of the trebouxiophyte *Chlorella*, which encodes 112 genes, has lost the IR, (Wakasugi et al. 1997) but the genes usually found in the IR and each of the single-copy regions have remained clustered together (Pombert et al. 2006). The chloroplast genomes of the two ulvophytes and of the two chlorophycean green algae feature an atypical quadripartite structure. In each ulvophyte genome, one of the single-copy regions features genes characteristic of both the ancestral SSC and LSC regions, whereas the opposite single-copy region contains exclusively genes that are characteristic of the ancestral LSC region (Pombert et al. 2005, 2006). Moreover, the rRNA genes in the IR are transcribed toward the latter region, instead of the SSC region as in the usual quadripartite architecture. From their observations, Pombert et al. (2006) concluded that a dozen genes were transferred from the LSC to the SSC region before or soon after the emergence of the Ulvophyceae and that the transcription direction of the rRNA genes changed. In the chloroplast genomes of the chlorophycean green algae *Scenedesmus* and *Chlamydomonas*, single-copy regions of similar sizes harbour sets of genes that are very different from those seen in other green algal genomes, indicating that genes were extensively shuffled between the two ancestral single-copy regions (Maul et al. 2002; de Cambiaire et al. 2006). Although the two chlorophycean genomes differ dramatically in their gene partitioning patterns, they share nearly identical gene repertoires and 11 derived gene clusters containing a total of 32 genes (de Cambiaire et al. 2006). Some of their genes, notably *rps3*, *clpP* and *rpoB*, display novelties (insertion sequences or discontinuities) in their structure. Unlike all other completely sequenced UTC algal cpDNAs that are characterized by the lower density of their genes relative to their *Nephroselmis* and streptophyte counterparts, the *Scenedesmus* genome is almost as compact as the *Nephroselmis* genome (de Cambiaire et al. 2006). Of all the UTC algal cpDNAs examined thus far, *Scenedesmus* cpDNA features the lowest proportion of short dispersed repeats in intergenic regions (only 8.7%); moreover, another singularity of this genome is the strong tendency of adjacent genes to occur on the same DNA strand (de Cambiaire et al. 2006). Given that *Scenedesmus* and *Chlamydomonas* have extremely rearranged genomes and do not represent basal lineages in the phylogeny of the Chlorophyceae (Buchheim et al. 2001; Shoup and Lewis 2003), the ancestral condition of the chloroplast genome could not be inferred for this class.

Phylogenetic analyses of the nuclear-encoded small subunit and large subunit rRNA genes indicate that the Chlorophyceae comprise at least five major groups that generally correspond to currently recognized orders of families (Buchheim et al. 2001; Shoup and Lewis 2003). The Chlamydomonadales and Sphaeropleales [also designated as the clockwise (CW) and directly opposed (DO) flagellar apparatus clades], which are represented by *Chlamydomonas* and *Scenedesmus* respectively, apparently share a sister-relationship. The Chaetophorales, Oedogoniales and Chaetopeltidales are basal relative to the Chlamydomonadales and Sphaeropleales; however, the precise divergence order of these three monophyletic groups remains unknown (Buchheim et al. 2001; Shoup and Lewis 2003). To identify some of the forces and major events that

shaped the chloroplast genome during the evolution of chlorophyceans, we have determined the complete cpDNA sequence of *Stigeoclonium helveticum*, a member of the Chaetophorales. Motile cells in this group are quadriflagellated and polymorphic for flagellar orientation (DO + CW) (Watanabe and Floyd 1989). We found that the *Stigeoclonium* genome is extremely rearranged relative to its *Scenedesmus* and *Chlamydomonas* homologues and harbours the fewest ancestral features among all completely sequenced cpDNAs. This IR-lacking genome, which represents the largest chloroplast genome ever sequenced, displays a number of distinctive traits, including a strong bias in gene content and base composition of the DNA strands that is consistent with bidirectional replication from a single origin.

Materials and methods

Strain and culture conditions

Stigeoclonium helveticum was obtained from the Culture Collection of Algae at the University of Texas at Austin (UTEX 441) and grown in modified Volvox medium (McCracken et al. 1980) under 12 h light/dark cycles.

Isolation and sequencing of cpDNA

A + T-rich organelle DNA was separated from nuclear DNA by CsCl-bisbenzimide isopycnic centrifugation (Turmel et al. 1999a). Both the chloroplast and mitochondrial genomes were completely sequenced as described previously (Pombert et al. 2004), using as templates plasmid clones originating from the organelle DNA fraction as well as PCR fragments spanning uncloned regions. Sequences were edited and assembled with SEQUENCER 4.2.1 (GeneCodes, Ann Arbor, MI, USA). To ensure that the sequence assembly of each genome is correct, we ascertained that the sizes of overlapping regions encompassing the whole genome sequence matched perfectly those of the corresponding regions amplified by PCR.

Analyses of genome sequence

Gene content was determined by BLAST homology searches (Altschul et al. 1990) against the nonredundant database of the National Center for Biotechnology and Information (NCBI) server. Protein-coding genes and open reading frames (ORFs) were localized precisely using ORFFINDER at NCBI, various programs of the Wisconsin package version 10.3 (Accelrys, San Diego, CA, USA) and other applications from the EMBOSS version 2.9.0 package (Rice et al. 2000). Genes coding for tRNAs were localized using tRNAscan-SE 1.23 (Lowe and Eddy 1997). Intron boundaries were determined by modelling intron secondary structures (Michel et al. 1989; Michel and Westhof 1990) and by comparing intron-containing genes with intronless homologues using FRAMEALIGN of the Wisconsin package. Homologous introns were detected by BLASTN searches (Altschul et al. 1990) against the non-redundant database of NCBI.

Repeated sequences were mapped with PipMaker (Schwartz et al. 2000). Repeats were identified with REPuter 2.74 (Kurtz et al. 2001) using the *–f* (forward), *–p* (palindromic) and *–allmax* options at minimum lengths (*–l*) of 30 and 45 bp and were classified with REPEATFINDER (Volfovsky et al. 2001). Number of copies of each repeat unit was determined with FINDPATTERNS of the Wisconsin package. Stem-loop structures and direct repeats were identified using PALINDROME and ETANDEM in EMBOSS 2.9.0 (Rice et al. 2000), respectively. Genomic regions containing non-overlapping repeated elements were identified with RepeatMasker (http://www.repeatmasker.org) running under the WU-BLAST 2.0 (http://www.blast.wustl.edu) search engine.

The sidedness index (C_s) was determined as described by Cui et al. (2006) using the formula $C_s = (n - n_{SB})/(n - 1)$, where n is the total number of genes in the genome and n_{SB} is the number of sided blocks, i.e. the number of blocks including adjacent genes on the same strand. The strand bias in base composition was calculated for the whole genome and for intergenic regions. For the entire genome sequence (GenBank accession number DQ630521), the sum of values (G − C)/(G + C), where C and G represent the number of occurrences of these two nucleotides, was calculated for windows of length 5,000, starting with nucleotides 50,000 to 55,000 and continuing by shifting 500 nucleotides downstream along the strand for each new window. For intergenic regions, the value (G − C)/(G + C) was calculated separately for each region.

All conserved gene pairs exhibiting identical gene polarities in green algal cpDNAs were identified using a custom-built program. The GRIMM web server (Tesler 2002) was used to infer the minimal number of gene permutations by inversions in pairwise comparisons of chloroplast genomes. Because GRIMM cannot deal with duplicated genes and requires that the compared genomes have the same gene content, genes within one of the two copies of the IR were excluded

and only the genes common to all the compared genomes were analysed. The data set used in the comparative analyses reported in Supplementary Table S3 contained 89 genes; pieces of *rpoB* and all exons of the genes containing *trans*-spliced introns were coded as distinct fragments (for a total of 96 gene loci).

Results

General features

The *Stigeoclonium* chloroplast genome sequence maps as a circular molecule of 223,902 bp containing a total of 97 genes, each present in single copy (Fig. 1). No remnant of an IR sequence was identified in *Stigeoclonium* cpDNA. Table 1 compares the general features of *Stigeoclonium* cpDNA with other completely sequenced chlorophyte cpDNAs. With an A + T content of 71.1%, *Stigeoclonium* cpDNA ranks at the second position, after its *Scenesdesmus* homologue, with respect to the abundance of these bases. The 97 conserved genes, 21 introns, and the two free standing ORFs of more than 100 codons (*orf101* and *orf107*) account for 55.8% of the total genome sequence of *Stigeoclonium*, with the introns representing 11% of the sequence. Sixteen group I introns and five group II introns, four of which are likely *trans*-spliced at the RNA level, are present in the *Stigeoclonium* genome. Intergenic spacers vary from 46 to 3,612 bp for an average size of 950 bp, a value that is comparable to that observed for *Chlamydomonas* cpDNA (average size of 941 bp). The *Stigeoclonium* genome is rich in dispersed repeated sequences, these elements accounting for 40.3% of the intergenic regions.

Gene content and gene structure

Relative to *Scenedesmus* and *Chlamydomonas* cpDNAs, *Stigeoclonium* cpDNA encodes four additional genes [*rpl32*, *psaM*, *trnL*(caa) and *trnS*(gga)] but lacks *petA*, a gene present in all previously sequenced chlorophyte cpDNAs (Supplementary Table S1). Like *Chlamydomonas* cpDNA, it is missing the *infA* and *rpl12* genes that are present in *Scenedesmus* and other chlorophyte cpDNAs. All three chlorophycean cpDNAs lack six genes (*accD, chlI, minD, psaI, rpl19* and *ycf20*) that have been retained in the genomes of the three other UTC algae examined thus far. Moreover, like their two ulvophyte homologues, they are missing four genes

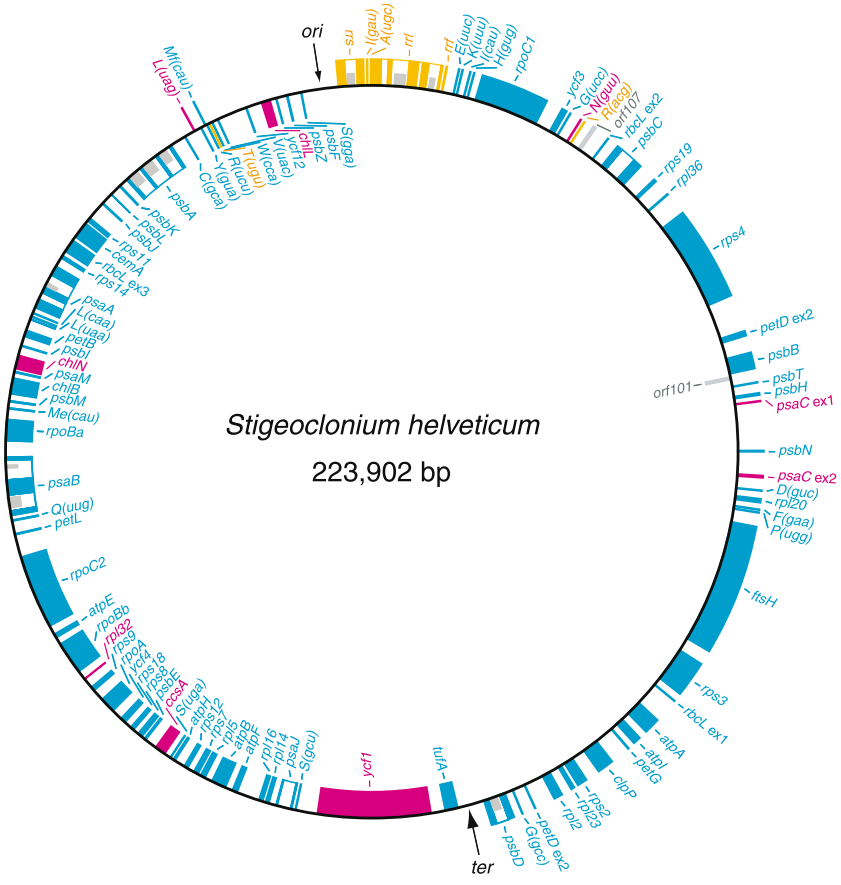

Fig. 1 Gene map of *Stigeoclonium* cpDNA. Genes (*filled boxes*) on the outside of the map are transcribed in a clockwise direction. Genes shown in *yellow*, *cyan* and *magenta* map to the IR, LSC and SSC regions of *Mesostigma* cpDNA. Genes and ORFs absent from *Mesostigma* cpDNA are shown in *grey*. Introns are represented by *open boxes* and intron ORFs are denoted by *narrow*, *filled boxes*. *Arrows* mark the putative origin (*ori*) and terminus (*ter*) of replication. tRNA genes are indicated by the one-letter amino acid code followed by the anticodon in parentheses (*Me*, elongator methionine; *Mf*, initiator methionine)

Table 1 General features of *Stigeoclonium* and other UTC algal cpDNAs

Feature	*Chlorella*	*Oltmannsiellopsis*	*Pseudendoclonium*	*Stigeoclonium*	*Scenedesmus*	*Chlamydomonas*
Size (bp)						
Total	150,613	151,933	195,867	223,902	161,452	203,827
IR	–[a]	18,510	6,039	–[a]	12,022	22,211
LSC	–[a]	33,610	140,914	–[a]	72,440[b]	81,307[b]
SSC	–[a]	81,303	42,875	–[a]	64,968[c]	78,088[c]
A + T (%)	68.4	59.5	68.5	71.1	73.1	65.5
Coding sequences (%)[d]	60.9	59.2	62.3	55.8	67.2	50.1
Genes (no.)[e]	112	105	105	97	96	94
Introns (no.)						
Group I	3	5	27	16	7	5
Group II	0	0	0	5	2	2

[a] Because *Chlorella* and *Stigeoclonium* cpDNAs lack an IR, only the total sizes of theses genomes are given

[b] This region was designated as SSC1 in the study of de Cambiaire et al. (2006)

[c] This region was designated as SSC2 in the study of de Cambiaire et al. (2006)

[d] Conserved genes, unique ORFs and introns were considered as coding sequences

[e] Genes present in the IR were counted only once. Unique ORFs and intron ORFs were not taken into account

[*cysA*, *cysT*, *trnL*(gag) and *trnT*(ggu)] relative to the chloroplast genome of the trebouxiophyte *Chlorella*.

Numerous genes in the *Stigeoclonium* genome (*cemA*, *clpP*, *ftsH*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps18*, *rps3*, *rps4* and *ycf1*) have expanded coding regions relative to their *Mesostigma* and *Nephroselmis* homologues. Most of these genes have been previously identified in other UTC algae (Pombert et al. 2005, 2006; de Cambiaire et al. 2006). Three genes (*clpP*, *rps3* and *rps4*) display enlarged coding regions only in members of the Chlorophyceae (Supplementary Table S2). The *Stigeoclonium rps4* gene is unusual in carrying an insertion sequence that is about 12-fold larger than those present in *Scenedesmus* and *Chlamydomonas* cpDNAs. Owing to its considerable size (340 kDa), the full-length protein sequence predicted from *Stigeoclonium rps4* is not likely to represent a functional ribosomal protein. On the other hand, our findings that the 5′ and 3′ termini of this gene share sequence homology with virtually the entire *Escherichia coli rpsD* gene and that its reading frame is maintained over more than 8 kb argue against the idea that *Stigeoclonium rps4* is a pseudogene. If this green algal gene is functional, then the sequence of its large expansion element would be expected to be excised at the RNA or protein level. Obviously, in the absence of evidence for a putative intron or intein element in *Stigeoclonium rps4*, no firm conclusion can be drawn regarding the functional status of this gene.

Like its *Scenedesmus* and *Chlamydomonas* counterparts, the *rpoB* gene in *Stigeoclonium* cpDNA consists of two separate ORFs that are not associated with sequences typical of group I or group II introns; however, instead of being contiguous, these ORFs are distant from one another in the *Stigeoclonium* genome (Fig. 1). In contrast to the *Scenedesmus* and *Chlamydomonas rps2* genes and the *Chlamydomonas rpoC1*, which also consist of distinct ORFs bordered by sequences unrelated to conventional introns, the corresponding genes in *Stigeoclonium* display a continuous structure. In addition to *rpoB*, the *petD*, *psaC* and *rbcL* genes occur as dispersed pieces in *Stigeoclonium* cpDNA (Fig. 1); in all three cases, each gene piece consists of an exon bordered by the 5′ or 3′ portion of a putatively *trans*-spliced group II intron.

Bias in gene coding regions and base composition of the two DNA strands

Like their *Scenedesmus* homologues, genes in *Stigeoclonium* cpDNA show a remarkably strong bias in their distribution between the two DNA strands (Fig. 1). The 59 consecutive genes in the 113.6 kb segment extending from *tufA* to *trnS*(gga), with the exception of *trnL*(uag) and *trnMf*(cau), are located on one strand, whereas all the other genes reside on the other strand. The sidedness index (C_s), i.e. the propensity of adjacent genes to be located on the same strand (Cui et al. 2006), is significantly higher in *Stigeoclonium* cpDNA ($C_s = 0.9479$) than that reported for *Scenedesmus* cpDNA ($C_s = 0.8842$).

The coding strand bias in the *Stigeoclonium* genome is closely associated with a strand bias in base composition. The cumulative GC skew diagram shown in Fig. 2a has a V-shape, with the minimum and maximum separated by half of the genome length and coinciding with the loci displaying a switch in coding strand. Starting from the minimum, i.e. a point in the region separating *trnS*(gga) and *rrs*, genes on each half of the genome are encoded on the strand displaying more G

than C residues. The GC skew is readily detectable in intergenic regions (Fig. 2b); as observed for the overall genome, the skew switches polarity in the vicinity of the two sites showing a switch in coding strand, with the coding strand manifesting a positive skew.

The cumulative GC skew analyses of prokaryotic genomes display the same profile as that reported here for the *Stigeoclonium* chloroplast genome (Grigoriev 1998). For prokaryotic genomes, it has been shown that the minimum and maximum coincides with the origin and terminus of replication (Grigoriev 1998) and that a majority of genes are encoded on the leading strand and are therefore transcribed in the same direction as the genome replication, a property termed the coorientation rule. The leading strand is richer in G than in C relative to the opposite strand most probably because it is subject to more frequent C deaminations during the time it remains temporarily single-stranded during gene transcription and chromosome replication (Guy and Roten 2004). Given the striking similarity between the plots of cumulative GC skew obtained for the *Stigeoclonium* and prokaryotic genome sequences, it is likely that the *Stigeoclonium* genome replicates bidirectionally from a single origin situated in the *trnS*(gga)-*rrs* spacer. It should be noted that our analysis of the cumulative GC skew for the IR-containing cpDNAs of *Scenedesmus* and *Chlamydomonas* did not disclose any putative origin and terminus of replication that are consistent with a bidirectional mode of replication, although adjacent genes tend to be encoded on the same DNA strand (Cui et al. 2006; de Cambiaire et al. 2006). The high level of strandedness in the latter chlorophycean chloroplast genomes has probably been generated by selection to regulate gene expression by favouring the formation of long, multicistronic transcripts.

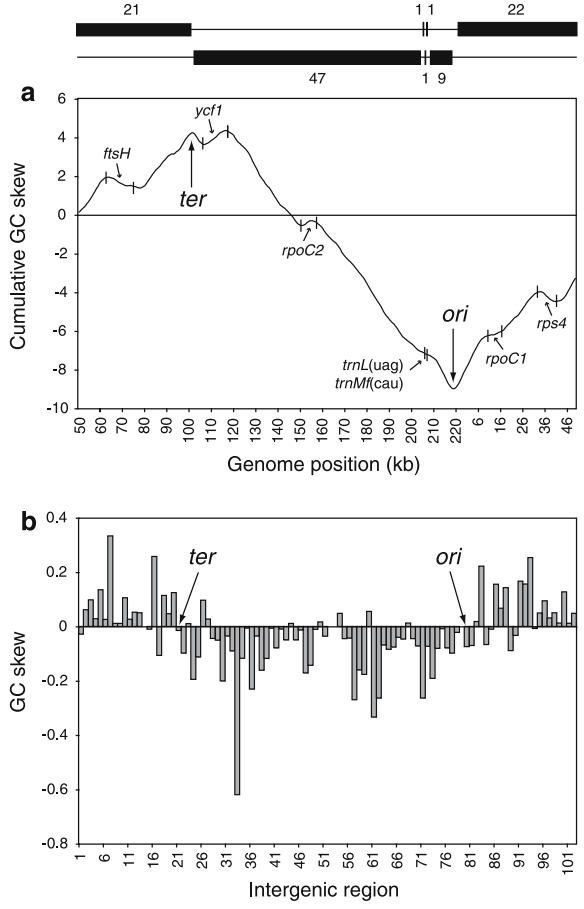

Fig. 2 Analyses of GC skew in *Stigeoclonium* cpDNA. **a** Plot of cumulative GC skew for the whole genome sequence (GenBank accession number DQ630521). The cumulative GC skew was calculated as indicated in the Materials and methods; base at position 50,000 was arbitrarily selected to represent the starting point. The putative origin and terminus of replication were located in the *trnS*(gga)-*rrs* and *psbD-tufA* intergenic regions, respectively. Boundaries of genes associated with local distortions are denoted by *vertical lines*. Above the plot is a representation of the coding strand, either the strand whose sequence is reported in GenBank (*the top strand*) or the alternate strand (*bottom strand*). The numbers of known genes and gene pieces encoded in the *filled boxes* are indicated. **b** GC skew diagram for the intergenic regions in the genome sequence. The GC skew was calculated as indicated in the Materials and methods, starting with the *psbH-psaC* intergenic region located between positions 50,839 and 51,388. The intergenic regions are numbered from 1 to 102

Disruptions of linearity, detected as local minima and maxima, are visible in the plot of cumulative GC skew of the *Stigeoclonium* genome (Fig. 2a). Interestingly, these distortions correspond to expanded regions in the *ftsH*, *rpoC1*, *rpoC2*, *rps4* and *ycf1* genes. As demonstrated for two *E. coli* strains (Grigoriev 1998), they possibly represent recent genome rearrangements such as inversions or horizontally acquired sequences.

Gene order

As observed previously for *Scenedesmus* and *Chlamydomonas* cpDNAs (Maul et al. 2002; de Cambiaire et al. 2006), the chloroplast genome of *Stigeoclonium* does not reveal any remnant of the ancestral gene partitioning pattern displayed by *Mesostigma, Nephroselmis* and streptophyte cpDNAs. In Fig. 1, it can be seen that homologues of the genes residing in the SSC and LSC regions of the *Mesostigma* genome are widely dispersed throughout the *Stigeoclonium* genome. In contrast, most of these genes in *Chlorella* and *Pseudendoclonium* cpDNAs have remained clustered together despite significant changes in genome architecture (Pombert et al. 2006).

The *Stigeoclonium* chloroplast genome is poor in ancestral gene clusters and its gene organization differs

remarkably from those of its chlorophyte counterparts. Figure 3 compares all gene pairs present in UTC algal cpDNAs with those present in *Mesostigma* and *Nephroselmis* cpDNAs and clearly illustrates the erosion of ancestral clusters that took place during the evolution of chlorophytes. It should be noted that, in this analysis, the unlinked exons of genes displaying putatively *trans*-spliced group II introns as well as the two *rpoB* gene pieces (*rpoBa* and *rpoBb*) were considered as separate gene loci. The trebouxiophyte *Chlorella* has retained almost all ancestral gene pairs shared by *Mesostigma* and *Neproselmis*, the ulvophytes *Oltmannsiellopsis* and *Pseudendoclonium* have lost a number of ancestral clusters present in *Chlorella*, and the chlorophycean green algae have retained only a few ancestral clusters. Apart from the rRNA operon, the three chlorophycean genomes share only three gene pairs that represent remnants of distinct ancestral operons (*psbB-psbT*, *rpl16-rpl14* and *rpl23-rpl2*). Both *Scenedesmus* and *Chlamydomonas* cpDNAs have retained longer versions of the latter protein operons (*psbB-psbT-psbN-psbH*, *rpl14-rpl16-rpl5-rps8* and *rpl23-rpl2-rps19*). In addition, these two algal cpDNAs display two ancestrally inherited gene pairs (*atpF-atpH* and *psbF-psbL*), whereas the *Stigeoclonium* genome has retained two ancestral gene pairs that represent fragments of separate operons (*rpl12-rsp7* and *psbL-psbJ*). When the derived gene pairs, i.e. gene pairs that are shared specifically by UTC algal cpDNAs, are taken into account, we find that ulvophyte cpDNAs share more derived traits with *Chlorella* cpDNA than with chlorophycean green algal cpDNAs and that the *Stigeoclonium* genome is highly rearranged relative to its *Scenedesmus* and *Chlamydomonas* homologues, which share 17 derived gene pairs accounting for 11 clusters (de Cambiaire et al. 2006). None of these derived gene pairs is present in *Stigeoclonium* cpDNA (Fig. 3). This genome shares only two derived gene pairs [*rps8-psbE* and *trnS*(gcu)-*ycf1*] with its *Scenedesmus* homologue, one [*psaA*ex3-*trnL*(caa)] with *Pseudendoclonium* cpDNA and one (*rbcL*ex3-*rps14*) with *Chlorella* cpDNA.

An alternative approach for comparing the degrees of similarity displayed by different genomes with respect to their gene order is to estimate the number of gene permutations that would be required to convert the gene order of a given genome to that of another genome. The data obtained with this approach corroborate the notion that the gene organization of *Stigeoclonium* cpDNA diverges radically from those of

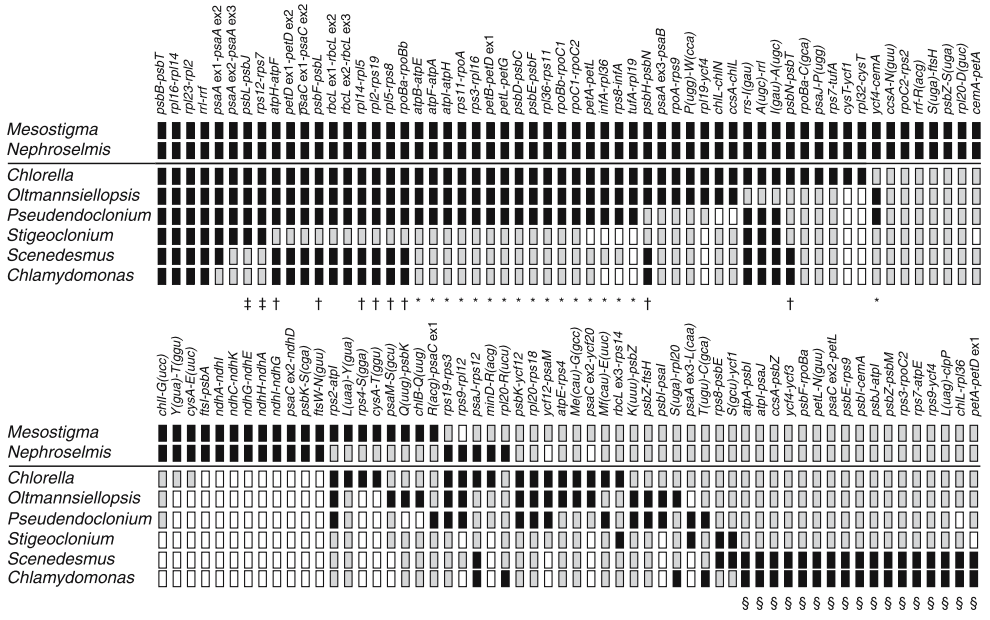

Fig. 3 Conservation of ancestral and derived gene pairs in *Stigeoclonium* and other UTC algal cpDNAs. *Filled boxes* indicate the presence of gene pairs with the same relative polarities in two or more genomes. *Grey or open boxes* indicate the absence of gene pairs. A *grey box* indicates that the two genes associated with a gene pair are found in the genome but are unlinked. An *open box* indicates that one or both genes associated with a gene pair are absent from the genome. For each gene pair, adjoining termini of the genes are indicated. *, ancestral gene pairs lost in the lineages leading to the three chlorophycean taxa; †, ancestral gene pairs lost in the lineage leading to *Stigeoclonium*; ‡, ancestral gene pairs lost in the lineages leading to *Chlamydomonas* and *Scenedesmus*; §, derived gene pairs gained in the *Chlamydomonas* and *Scenedesmus* lineages. Note that the pairs of coding regions that were lost following the acquisition of *trans*-spliced group II introns (*petD*ex1-*petD*ex2, *psaA*ex1-*psaA*ex2, *psaA*ex2-*psaA*ex3, *psaC*ex1-*psaC*ex2, *rbcL*ex1-*rbcL*ex2 and *rbcL*ex2-*rbcL*ex3) were not scored as losses of ancestral gene pairs

previously sequenced chlorophyte genomes (Supplementary Table S3). We estimated that more than 80 inversions would be required to convert the gene order of *Stigeoclonium* cpDNA into that of any other chlorophyte cpDNA. All the additional pairwise comparisons we carried out yielded reduced numbers of inversions, with the fewest (43 inversions) being obtained in the comparison of the *Mesostigma* and *Nephroselmis* genomes. With 58 inversions distinguishing the *Scenedesmus* and *Chlamydomonas* cpDNAs, these chlorophycean genomes are clearly more similar to one another than each of these genomes is to its *Stigeoclonium* homologue.

Group I introns

The 16 group I introns in *Stigeoclonium* cpDNA interrupt eight genes, range from 243 to 1,946 bp in size, and fall within subgroups IA1, IA2, IA3, IB and IC3 according to the classification system proposed by Michel and Westhof (1990) (Supplementary Table S4). The *psbC*, *psbD*, *rrs*, and *trnL*(uaa) genes each exhibit one group I intron, whereas the remaining four genes contain two (*psaB*), three (*psaA* and *psbA*) or four introns (*rrl*). Ten of these introns carry internal ORFs, eight of which code for putative homing endonucleases of the HNH, GIY-YIG and LAGLIDADG families (Stoddard 2005) (Supplementary Table S4). Eleven introns are positionally and structurally homologous to introns in other UTC algal chloroplast genomes (Fig. 4). Among these introns, the *rrl* intron inserted at site 2,593 exhibits the broadest distribution among UTC green algae, being present in all completely sequenced chloroplast genomes of these algae, except in *Stigeoclonium* cpDNA. The remaining ten introns have homologues in only one or two UTC algae. The insertion sites of the *psbD* intron, of two introns in *psaA* and of two others in *psbA* have not been previously documented and none of these introns shows high structural similarity with an intron inserted at a distinct site in the *Stigeoclonium* genome.

Group II introns

The five group II introns of *Stigeoclonium* vary from 654 to 1,918 bp in size and reside within *psaC*, *psaJ*, *petD* and *rbcL*. Each of these genes is interrupted by one intron, with the exception of *rbcL*. Positionally homologous introns have not been identified in other chloroplast genomes (Fig. 4); this is the first report indicating the presence of group II introns in *psaC*, *psaJ* and *rbcL*. All five *Stigeoclonium* group II introns lack an ORF ≥ 100 codons and all, except the *psaJ*

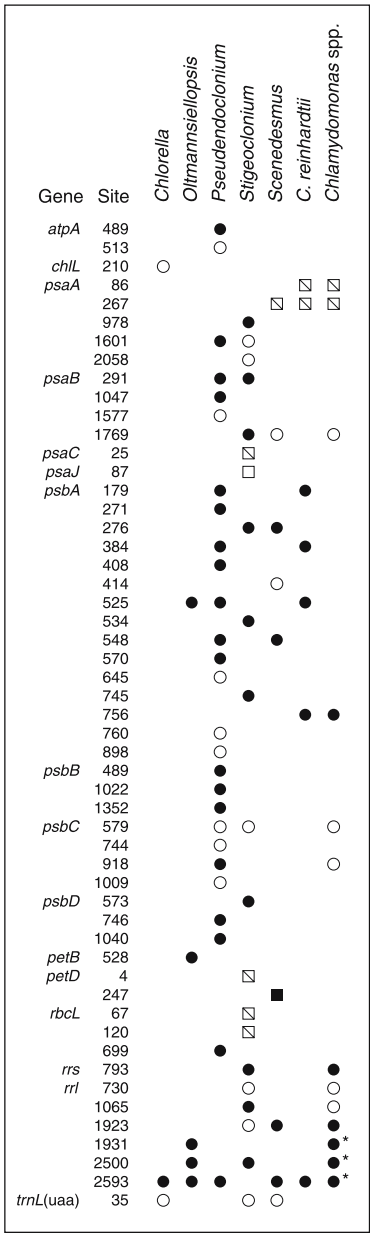

Fig. 4 Distribution of introns in *Stigeoclonium* and other UTC algal cpDNAs. *Circles* denote the presence of group I introns and *squares* denote the presence of group II introns. *Divided squares* represent *trans*-spliced group II introns. *Open symbols* denote the absence of intron ORFs, whereas *filled symbols* denote their presence. Intron insertion sites in genes coding for tRNAs and proteins are given relative to the corresponding genes in *Mesostigma* cpDNA; insertions 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. The column at the extreme right indicates the introns of *Chlamydomonas* species other than *C. reinhardtii* that are known to have homologues in completely sequenced UTC algal genomes. References for the latter introns are as follows: *psaB* (Turmel et al. 1993b); *psbA* (Turmel et al. 1989); *psbC* (Turmel et al. 1993b); *rrs* (Durocher et al. 1989); and *rrl* (Turmel et al. 1991; Côté et al. 1993; Turmel et al. 1993a, 1995b). An *asterisk* denotes the absence of the ORF in some *Chlamydomonas* species

intron, are discontinuous. The second intron in *rbcL* is split in domain II, whereas the sites of discontinuity of the other introns map to various locations within domain I (Supplementary Fig. S1). The second intron in *rbcL* and the *cis*-spliced *psaJ* intron were classified into the subgroup IIA according to the nomenclature proposed by Michel et al. (1989), whereas the *petD* intron was classified into the subgroup IIB. The two remaining introns could not be categorized into any of these subgroups because they exhibit characteristics of both subgroups. No close structural relationship was identified among the five group II introns.

Repeated sequences

Comparison of the *Stigeoclonium* cpDNA sequence against itself using PipMaker (Schwartz et al. 2000) disclosed the presence of repeats in many intergenic regions, some expanded genes (*cemA, ftsH*, *rpoC1, rpoC2, rps2* and *ycf1)*, and four introns (Sh.*psaA*.2, Sh.*psaB*.1, Sh.*psbA*.1 and Sh.*psbC*.1) (Supplementary Fig. S2). The intergenic regions of this genome display a higher proportion of repeats compared to those in the *Chlamydomonas* genome (Table 2).

The most abundant repeated sequences in the *Stigeoclonium* genome consist of dispersed repeats and can be classified into five groups of non-overlapping repeat units (A through E) on the basis of their primary sequences (Table 3). Each group features variants that differ slightly in primary sequence; for groups A, B, C, we identified some of these variants (e.g. A1, B1 and B2). The sequences of all identified repeat units form perfect palindromes or putative stem-loop structures with a loop of 2–8 bases. Their total sizes vary from 29 to 52 bp. Repeat unit C features exclusively A and T bases. Repeat units A and C represent the most important groups in term of copy number, and members of these groups are scattered all over the genome (Supplementary Fig. S2). Although the repeat units belonging to groups A, B and C occur mainly as palindromes or stem-loop structures, copies of these repeat units are found as reduced versions consisting of half-stems (i.e. sequences lacking a twofold axis of symmetry). Many intergenic regions feature larger repeats that are composed of two or more copies of the same repeat unit and/or of repeats representing different units (Supplementary Fig. S2). Segments of identical sequences containing such composite repeats are located in distinct loci of the *Stigeoclonium* genome. The largest repeat of this type is 625 bp long (Table 2). No repeats identical to those reported in Table 3 were detected in any other completely sequenced UTC algal cpDNA.

Discussion

Distinctive features of the *Stigeoclonium* chloroplast genome

Although the *Stigeoclonium* chloroplast genome shares several derived features with *Chlamydomonas* and *Scenedesmus* cpDNAs, it displays a number of distinctive traits. *Stigeoclonium* cpDNA is the largest chloroplast genome yet sequenced and in contrast to its two chlorophycean counterparts, features no IR. Genes that are usually part of ancestral clusters in green algal cpDNAs have been reshuffled to a significantly greater extent in the *Stigeoclonium* genome than in *Scenedesmus* and *Chlamydomonas* cpDNA and virtually all of the derived clusters identified in the latter algae are absent from the *Stigeoclonium* genome (Fig. 3, Supplementary Table S3). The distribution of the *Stigeoclonium* genes between the two DNA strands shows an

Table 2 Abundance of repeats in *Stigeoclonium* and other UTC algal cpDNAs

cpDNA	Maximal size of repeats (bp)	Number of repeats[a]		Non-overlapping repeats ≥ 30 bp[b]		
		≥30 bp	≥45 bp	Total size (bp)	Fraction of genome (%)	Fraction of intergenic regions (%)
Chlorella	84	269	44	11,743	7.8	20.8
Oltmannsiellopsis	172	1,205	161	18,033	11.9	30.1
Pseudendoclonium	171	1,047	203	10,073	5.1	13.6
Stigeoclonium	625	2,856	640	39,941	17.8	40.3
Scenedesmus	112	86	21	4,817	3.0	8.7
Chlamydomonas	221	3,247	551	32,244	15.8	31.9

[a] Number of repeats with identical sequences were estimated using REPuter (Kurtz et al. 2001). Note that these estimates include overlapping repeats

[b] Non-overlapping repeat elements were mapped on the genome with RepeatMasker using the repeats identified with REPuter as input sequences

Table 3 Repeat units in *Stigeoclonium* cpDNA

Designation[a]	Size (bp)	Sequence[b]	Copy number[c]	
			100%	90%[d]
A1	29	TTCCCCCGGAAGC-$(N)_3$-GCTTCCGGGGGAA	35 (105)	79 (195)
A2	29	TTCCCTCGGAATC-$(N)_3$-GATTCCGAGGGAA	18 (42)	21 (66)
B1	30	TCTCTTTTTGGGCTTAAGCCCAAAAAGAGA	5 (22)	19 (35)
B2	32	TCTTTTTTTTGGGCTTAAGCCCAAAAAAAAGA	2 (7)	6 (8)
B3	32	TTTCTCTTCTGGGCTTAAGCCCAGAAGAGAAA	1 (5)	4 (8)
C1	34	AAATTAAAAAATAAAT-$(N)_2$-ATTTATTTTTTAATTT	29 (68)	46 (102)
C2	36	AAATTAAAAAAATAAAT-$(N)_2$-ATTTATTTTTTTAATTT	12 (26)	28 (60)
D	46–48	TCTCTGTCGTTTAAATTGTTC-$(N)_{4-6}$-GAACAATTTAAACGACAGAGA	15 (30)	28 (58)
E	50–52	TTTTTCTTTATGCTAGCAAAAA-$(N)_{6-8}$-TTTTTGCTAGCATAAAGAAAAA	7 (14)	27 (54)

[a] Repeats sharing >75% sequence identity have been assigned to the same class of repeat units (A, B, C, D or E)

[b] The underlined sequences often occur as solitary sequences (half-stems) in the genome

[c] Copies of each repeat unit were identified in FINDPATTERNS searches using 100 or 90% sequence identity. The values in parentheses indicate the number of occurrences of the sequences corresponding to the half-stems

[d] Sequence redundancy can occur between the subclasses of a given class of repeats units

almost perfect symmetry (Fig. 1) and most remarkably, the gene-encoding strand on each half of the genome is richer in G than in C compared to the alternate strand (Fig. 2). Another distinctive feature of the *Stigeoclonium* chloroplast genome is its large set of introns (21 introns vs. 9 in *Scenedesmus* and 7 in *Chlamydomonas*), which includes four putatively *trans*-spliced group II introns that have no homologues in other green algal cpDNAs (Fig. 4). As each of these group II introns consists of two pieces that are far apart on the genome, two distinct precursor transcripts, each containing an intron piece, presumably assemble at the site of discontinuity of the intron via base-pairings and tertiary interactions to reconstitute the intron structure required for splicing.

Considering that the presence of an rDNA-encoding IR is a prominent feature of the chloroplast genome in diverse green algal and plant lineages and that its absence from some lineages has been attributed to independent losses (Palmer and Thompson 1981; Palmer et al. 1987; Lidholm et al. 1988; Strauss et al. 1988; Turmel et al. 2005), we infer that an IR was present in the chloroplast genome of the common ancestor of the green algae belonging to the Chlamydomonadales, Sphaeropleales, and Chaetophorales but was lost in the lineage leading to *Stigeoclonium* (Chaetophorales). As the IR is thought to play a major role in stabilizing gene order (Palmer and Thompson 1982; Strauss et al. 1988; Palmer 1991), it is perhaps not surprising that the *Stigeoclonium* chloroplast genome is extremely rearranged relative to *Scenedesmus* and *Chlamydomonas* cpDNAs. To account for the highly scrambled gene order observed in the great majority of previously documented green plant cpDNAs lacking an IR (Palmer and Thompson 1982; Strauss et al. 1988; Wakasugi et al. 1994; Turmel et al. 2005), it has been hypothesized that the loss of the IR enhances opportunities for intramolecular recombination between homologous sequence elements such as short dispersed repeats (Palmer 1991). Therefore, according to this hypothesis, both the absence of the IR and the great abundance of short dispersed repeats in the *Stigeoclonium* genome are important factors that influenced the order of genes and gene pieces.

The mode of DNA replication appears to be an additional factor that contributed to the unusual arrangement of genes in the *Stigeoclonium* genome, in particular to the strand bias in coding regions. Both the strand biases in coding regions and in GC composition displayed by this algal genome are typical of those observed in prokaryotic genomes that replicate bidirectionally from a single origin (Grigoriev 1998; Tillier and Collins 2000a, b; Guy and Roten 2004). Analysis of the cumulative GC skew has allowed us to map a putative replication origin in the *trnS*(gga)-*rrs* intergenic region and a putative terminus in the *psbD-tufA* intergenic region (Figs. 1, 2). Further work will be needed to determine whether the intergenic spacer upstream of the small subunit rRNA gene (*rrs*) functions as an origin and whether the unique direct repeats and potential stem-loop structure found at this locus are essential for replication. Evidence for bidirectional replication from a single origin based on GC skew analysis has been reported for only two other IR-lacking chloroplast genomes showing a coding strand bias, the genome of the euglenoid *Euglena gracilis* whose plastids were acquired by secondary endosymbiosis from a green alga (Morton 1999) and the genome of the parasitic green alga *Helicosporidium* sp. (Trebouxiophyceae) (de Koning and Keeling 2006). Consistent with the GC skew

analysis of *Euglena* cpDNA, previous electron microscopic analysis of replication intermediates had suggested that this genome is replicated bidirectionally from a single origin (near the repeated rRNA genes) to a terminus on the opposite side of the circular genome (Koller and Delius 1982; Ravel-Chapuis et al. 1982). As in *Stigeoclonium* cpDNA, the putative origin of bidirectional replication in the reduced genome of *Helicosporidium* has been located just upstream of the *rrs* gene. In contrast, studies of cpDNA replication in *Chlamydomonas* and various land plants indicate that these genomes replicate by a mechanism different than that used by prokaryotic genomes (Heinhorst and Cannon 1993; Kunnimalaiyaan and Nielsen 1997). Except for *Euglena* cpDNA, all chloroplast genomes that were examined have been found to contain multiple origins whose number and locations may vary in different organisms.

Prior to our study, the only known *trans*-spliced group II introns in chlorophyte cpDNAs were the bipartite introns occupying the same site in the *Scenedesmus* and *Chlamydomonas psaA* genes (Kück et al. 1987; de Cambiaire et al. 2006) and the tripartite intron inserted at a distinct site in the *Chlamydomonas psaA* (Kück et al. 1987; Goldschmidt-Clermont et al. 1991; Turmel et al. 1995a). Most other *trans*-spliced group II introns are bipartite and have been documented mainly in land plant mitochondrial genomes. Interestingly, *cis*-spliced versions of these mitochondrial introns have been found in some land plant taxa, supporting the notion that disruption of ancestral *cis*-spliced introns gave rise to *trans*-spliced introns (Malek et al. 1997; Malek and Knoop 1998). Not only was the finding of four bipartite group II introns in *Stigeoclonium* cpDNA unexpected; it was also surprising that the sites of discontinuities of these introns lie within domain I or II, because the majority of reported *trans*-spliced group II introns are fragmented within domain III (Michel et al. 1989) or IV (Michel and Ferat 1995). Only the tripartite introns in *Chlamydomonas* chloroplast *psaA* (Goldschmidt-Clermont et al. 1991; Turmel et al. 1995a) and in *Oenothera* mitochondrial *nad5* (Knoop et al. 1997) are known to have a break within domain I; the central fragments of these introns encompass part of domain I, the entire domain II and III, and part of domain IV. To our knowledge, no discontinuity within domain II of group II introns has been documented thus far.

Evolution of the chlorophycean chloroplast genome

The addition of the *Stigeoclonium* chloroplast genome sequence to the collection of completely sequenced green algal cpDNAs sheds light into the architecture of the chloroplast genome from the last common ancestor of the green algae belonging to the Chaetophorales, Sphaeropleales and Chlamydomonadales; however, the portrait that can be drawn for this ancestral genome is rather sketchy (Fig. 5). This genome almost certainly featured an IR and contained a minimum of 100 genes, a few of which were probably organized as ancestral gene clusters. The coding regions of at least three genes (*clpP, rps3* and *rps4*) were already expanded in size and *rpoB* was split into two separate ORFs. The intron content cannot be predicted as the patchy distribution observed for these elements among UTC lineages (Fig. 4) may result from both horizontal transfers and losses of introns. Short dispersed repeats were also likely present because such sequences are found in the trebouxiophyte, ulvophyte and chlorophycean cpDNAs studied thus far.

When our comparative analysis of the *Stigeoclonium*, *Scenedesmus* and *Chlamydomonas* chloroplast genomes is placed in a phylogenetic framework, we find that a number of mutational events can be inferred during the evolution of chlorophycean green algae. Our recent phylogenetic analyses of genes and proteins derived from chloroplast genome sequences of green algae representing the four chlorophyte classes revealed that *Stigeoclonium* occupies a basal position

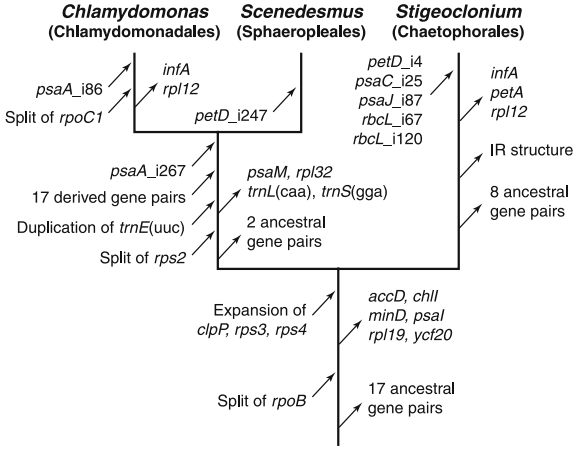

Fig. 5 Gains/losses of structural cpDNA features inferred from the phylogenetic tree placing *Stigeoclonium* at a basal position relative to *Chlamydomonas* and *Scenedesmus*. Gains of group II introns, derived gene pairs, expanded gene sequences, duplicated genes, and genes split into two distinct ORFs are indicated by *arrows pointing toward the tree*, whereas losses of genes, ancestral gene pairs and IR structure are indicated by *arrows pointing away from the tree*. Each group II intron is denoted by the name of the gene in which it resides followed by 'i' and its insertion position relative to the corresponding *Mesostigma* gene. The ancestral and derived gene pairs that were lost or gained on different branches of the tree are indicated in Fig. 3

relative to a clade uniting *Scenedesmus* and *Chlamydomonas* (our unpublished results). This topology, which was found to be very robust regardless of the methods of analysis used, is supported by several cpDNA features (Fig. 5). For example, the affiliation of *Chlamydomonas* and *Scenedesmus* to the same clade is supported by the five sets of traits that these algal cpDNAs have in common but that are lacking from *Stigeoclonium* cpDNA and other chlorophyte cpDNAs: (1) the absence of four genes, (2) the presence of a duplicated *trnE*(uuc) gene, (3) the presence of a *trans*-spliced group II intron at site 267 in *psaA*, (4) the absence of two ancestral gene pairs and the presence of 17 derived gene pairs (see Fig. 3) and (5) the split of *rps2* into two separate ORFs. Following the split of the Chlamydomonadales and Sphaeropleales, the chloroplast genome sustained no further changes in the *Scenedesmus* lineage, except the acquisition of a *cis*-spliced group II intron in *petD* (Kück 1989). In the *Chlamydomonas* lineage, a second *trans*-spliced group II intron was gained by *psaA* (Kück et al. 1987), two genes were lost and *rpoC1* was split into two separate ORFs. The distinctive traits displayed by the *Stigeoclonium* cpDNA probably reflect events that occurred specifically during the evolution of the Chaetophorales. These events include the insertion of five group II introns, the fragmentation of four of these introns, the loss of three genes, the loss of the IR as well as the losses of eight ancestral gene pairs (Fig. 5).

The branching order reported here for the Chaetophorales, Sphaeropleales and Chlamydomonadales is congruent with the current hypothesis for the divergence order of chlorophycean lineages as inferred from the nuclear-encoded small subunit and large subunit rRNA gene sequences (Buchheim et al. 2001; Shoup and Lewis 2003). According to this hypothesis, the evolution of a polymorphic DO + CW condition for the flagellar apparatus in the basal lineage represented by *Stigeoclonium* (Chaetophorales) became fixed for the CW condition in the Chlamydomonadales and for the DO condition in the Sphaeropleales. Of course, to better understand how the CW and DO organizations of basal bodies found in these chlorophycean lineages originated from the counterclockwise organization observed in trebouxiophytes and ulvophytes, a robust phylogeny encompassing all identified chlorophycean lineages will be required. Sequencing of the chloroplast genome from additional chlorophycean taxa would not only be useful to unravel the branching order of the major chlorophycean lineages but would also throw light into the most ancestral condition of this organelle genome in the Chlorophyceae.

Acknowledgments We thank Jean-François Pombert and Jean-Charles de Cambiaire for their help with the sequence analyses and for critical reading of the manuscript. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (to C.L. and M.T.).

References

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

Bremer K (1985) Summary of green plant phylogeny and classification. Cladistics 1:369–385

Buchheim MA, Michalopulos EA, Buchheim JA (2001) Phylogeny of the Chlorophyceae with special reference to the Sphaeropleales: a study of 18S and 26S rDNA data. J Phycol 37:819–835

de Cambiaire JC, 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

Côté V, Mercier J-P, Lemieux C, Turmel M (1993) The single group-I intron in the chloroplast *rrnL* gene of *Chlamydomonas humicola* encodes a site-specific DNA endonuclease (I-*Chu*I). Gene 129:69–76

Cui L, Leebens-Mack J, Wang L-S, Tang J, Rymarquis L, Stern DB, dePamphilis CW (2006) Adaptive evolution of chloroplast genome structure inferred using a parametric bootstrap approach. BMC Evol Biol 6:13

Durocher V, Gauthier A, Bellemare G, Lemieux C (1989) An optional group I intron between the chloroplast small subunit rRNA genes of *Chlamydomonas moewusii* and *C. eugametos*. Curr Genet 15:277–282

Friedl T (1997) The evolution of the green algae. Plant Syst Evol 11(Suppl):87–101

Friedl T, O'Kelly CJ (2002) Phylogenetic relationships of green algae assigned to the genus *Planophila* (Chlorophyta): evidence from 18S rDNA sequence data and ultrastructure. Eur J Phycol 37:373–384

Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD (1991) A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. Cell 65:135–143

Grigoriev A (1998) Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res 26:2286–2290

Guy L, Roten CA (2004) Genometric analyses of the organization of circular chromosomes: a universal pressure determines the direction of ribosomal RNA genes transcription relative to chromosome replication. Gene 340:45–52

Heinhorst S, Cannon GC (1993) DNA replication in chloroplasts. J Cell Sci 104:1–9

Knoop V, Altwasser M, Brennicke A (1997) A tripartite group II intron in mitochondria of an angiosperm plant. Mol Gen Genet 255:269–276

Koller B, Delius H (1982) Origin of replication in chloroplast DNA of *Euglena gracilis* located close to the region of variable size. EMBO J 1:995–998

de Koning AP, Keeling PJ (2006) The complete plastid genome sequence of the parasitic green alga *Helicosporidium* sp. is highly reduced and structured. BMC Biol 4:12

Kück U (1989) The intron of a plastid gene from a green alga contains an open reading frame for a reverse transcriptase-like enzyme. Mol Gen Genet 218:257–265

Kück U, Choquet Y, Schneider M, Dron M, Bennoun P (1987) Structural and transcription analysis of two homologous

genes for the P700 chlorophyll a-apoproteins in *Chlamydomonas reinhardii*: evidence for in vivo trans-splicing. EMBO J 6:2185–2195

Kunnimalaiyaan M, Nielsen BL (1997) Chloroplast DNA replication: mechanism, enzymes and replication origins. J Plant Biochem Biotechnol 6:1–7

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29:4633–4642

Lewis LA, McCourt RM (2004) Green algae and the origin of land plants. Am J Bot 91:1535–1556

Lidholm J, Szmidt AE, Hällgren J-E, Gustafsson P (1988) The chloroplast genomes of conifers lack one of the rRNA-encoding inverted repeats. Mol Gen Genet 212:6–10

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

Malek O, Knoop V (1998) *Trans*-splicing group II introns in plant mitochondria: the complete set of *cis*-arranged homologs in ferns, fern allies, and a hornwort. RNA 4:1599–1609

Malek O, Brennicke A, Knoop V (1997) Evolution of *trans*-splicing plant mitochondrial introns in pre-Permian times. Proc Natl Acad Sci USA 94:553–558

Mattox KR, Stewart KD (1984) Classification of the green algae: a concept based on comparative cytology. In: Irvine DEG, John DM (eds) The systematics of the green algae. Academic, London, pp 29–72

Maul JE, Lilly JW, Cui L, dePamphilis CW, Miller W, Harris EH, Stern DB (2002) The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. Plant Cell 14:2659–2679

McCracken DA, Nadakavukaren MJ, Cain JR (1980) A biochemical and ultrastructural evaluation of the taxonomic position of *Glaucosphaera vacuolata* Korsch. New Phytol 86:39–44

Michel F, Ferat J-L (1995) Structure and activities of group II introns. Annu Rev Biochem 64:435–461

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

Michel F, Umesono K, Ozeki H (1989) Comparative and functional anatomy of group II catalytic introns—a review. Gene 82:5–30

Morton BR (1999) Strand asymmetry and codon usage bias in the chloroplast genome of *Euglena gracilis*. Proc Natl Acad Sci USA 96:5123–5128

Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil I (eds) The molecular biology of plastids. Cell culture and somatic cell genetics of plants. Academic, San Diego, pp 5–53

Palmer JD, Thompson WF (1981) Rearrangements in the chloroplast genomes of mung bean and pea. Proc Natl Acad Sci USA 78:5533–5537

Palmer JD, Thompson WF (1982) Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell 29:537–550

Palmer JD, Osorio B, Aldrich J, Thompson WF (1987) Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. Curr Genet 11:275–286

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

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, 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

Ravel-Chapuis P, Heizmann P, Nigon V (1982) Electron microscopic localization of the replication origin of *Euglena gracilis* chloroplast DNA. Nature 300:78–81

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

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W (2000) PipMaker—a web server for aligning two genomic DNA sequences. Genome Res 10:577–586

Shoup S, Lewis LA (2003) Polyphyletic origin of parallel basal bodies in swimming cells of chlorophycean green algae (Chlorophyta). J Phycol 39:789–796

Sluiman HJ (1985) A cladistic evaluation of the lower and higher green plants (*Viridiplantae*). Plant Syst Evol 149:217–232

Stoddard BL (2005) Homing endonuclease structure and function. Q Rev Biophys 38:49–95

Strauss SH, Palmer JD, Howe GT, Doerksen AH (1988) Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. Proc Natl Acad Sci USA 85:3898–3902

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

Tillier ER, Collins RA (2000a) Genome rearrangement by replication-directed translocation. Nat Genet 26:195–197

Tillier ER, Collins RA (2000b) The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. J Mol Evol 50:249–257

Turmel M, Boulanger J, Lemieux C (1989) Two group I introns with long internal open reading frames in the chloroplast *psbA* gene of *Chlamydomonas moewusii*. Nucleic Acids Res 17:3875–3887

Turmel M, Boulanger J, Schnare MN, Gray MW, Lemieux C (1991) Six group I introns and three internal transcribed spacers in the chloroplast large subunit ribosomal RNA gene of the green alga *Chlamydomonas eugametos*. J Mol Biol 218:293–311

Turmel M, Gutell RR, Mercier J-P, Otis C, Lemieux C (1993a) Analysis of the chloroplast large subunit ribosomal RNA gene from 17 *Chlamydomonas* taxa. Three internal transcribed spacers and 12 group I intron insertion sites. J Mol Biol 232:446–467

Turmel M, Mercier JP, Côté MJ (1993b) Group I introns interrupt the chloroplast *psaB* and *psbC* and the mitochondrial *rrnL* gene in *Chlamydomonas*. Nucleic Acids Res 21:5242–5250

Turmel M, Choquet Y, Goldschmidt-Clermont M, Rochaix JD, Otis C, Lemieux C (1995a) The *trans*-spliced intron 1 in the *psaA* gene of the *Chlamydomonas* chloroplast: a comparative analysis. Curr Genet 27:270–279

Turmel M, Côté V, Otis C, Mercier J-P, Gray MW, Lonergan KM, Lemieux C (1995b) Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion). Mol Biol Evol 12:533–545

Turmel M, Lemieux C, Burger G, Lang BF, Otis C, Plante I, Gray MW (1999a) The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*. Two radically different evolutionary patterns within green algae. Plant Cell 11:1717–1729

Turmel M, Otis C, 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 USA 96:10248–10253

Turmel M, Otis C, 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 USA 99:11275–11280

Turmel M, Otis C, Lemieux C (2005) The complete chloroplast DNA sequences of the charophycean green algae *Staurastrum* and *Zygnema* reveal that the chloroplast genome underwent extensive changes during the evolution of the Zygnematales. BMC Biol 3:22

Turmel M, Otis C, Lemieux C (2006) The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. Mol Biol Evol 23:1324–1338

Volfovsky N, Haas BJ, Salzberg SL (2001) A clustering method for repeat analysis in DNA sequences. Genome Biol 2:Research0027

Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc Natl Acad Sci USA 91:9794–9798

Wakasugi T, Nagai T, Kapoor M, Sugita M, Ito M, Ito S, Tsudzuki J, Nakashima K, Tsudzuki T, Suzuki Y, Hamada A, Ohta T, Inamura A, Yoshinaga K, Sugiura M (1997) Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division. Proc Natl Acad Sci USA 94:5967–5972

Watanabe S, Floyd GL (1989) Ultrastructure of the quadriflagellate zoospores of the filamentous green algae *Chaetophora incrassata* and *Pseudoschizomeris caudata* (Chaetophorales, Chlorophyceae) with emphasis on the flagellar apparatus. Bot Mag Tokyo 102:533–546