

Research article

Open Access

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

Jean-Charles de Cambiaire, Christian Otis, Claude Lemieux and Monique Turmel*

Address: Département de biochimie et de microbiologie, Université Laval, Québec, Canada

Email: Jean-Charles de Cambiaire - Jean-Charles.de_Cambiaire@rsvs.ulaval.ca; Christian Otis - Christian.Otis@rsvs.ulaval.ca;

Claude Lemieux - Claude.Lemieux@rsvs.ulaval.ca; Monique Turmel* - Monique.Turmel@rsvs.ulaval.ca

* Corresponding author

Published: 25 April 2006

Received: 14 February 2006

BMC Evolutionary Biology 2006, 6:37 doi:10.1186/1471-2148-6-37

Accepted: 25 April 2006

This article is available from: <http://www.biomedcentral.com/1471-2148/6/37>

© 2006 de Cambiaire et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The phylum Chlorophyta contains the majority of the green algae and is divided into four classes. While the basal position of the Prasinophyceae is well established, the divergence order of the Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC) remains uncertain. The five complete chloroplast DNA (cpDNA) sequences currently available for representatives of these classes display considerable variability in overall structure, gene content, gene density, intron content and gene order. Among these genomes, that of the chlorophycean green alga *Chlamydomonas reinhardtii* has retained the least ancestral features. The two single-copy regions, which are separated from one another by the large inverted repeat (IR), have similar sizes, rather than unequal sizes, and differ radically in both gene contents and gene organizations relative to the single-copy regions of prasinophyte and ulvophyte cpDNAs. To gain insights into the various changes that underwent the chloroplast genome during the evolution of chlorophycean green algae, we have sequenced the cpDNA of *Scenedesmus obliquus*, a member of a distinct chlorophycean lineage.

Results: The 161,452 bp IR-containing genome of *Scenedesmus* features single-copy regions of similar sizes, encodes 96 genes, i.e. only two additional genes (*infA* and *rpl12*) relative to its *Chlamydomonas* homologue and contains seven group I and two group II introns. It is clearly more compact than the four UTC algal cpDNAs that have been examined so far, displays the lowest proportion of short repeats among these algae and shows a stronger bias in clustering of genes on the same DNA strand compared to *Chlamydomonas* cpDNA. Like the latter genome, *Scenedesmus* cpDNA displays only a few ancestral gene clusters. The two chlorophycean genomes share 11 gene clusters that are not found in previously sequenced trebouxiophyte and ulvophyte cpDNAs as well as a few genes that have an unusual structure; however, their single-copy regions differ considerably in gene content.

Conclusion: Our results underscore the remarkable plasticity of the chlorophycean chloroplast genome. Owing to this plasticity, only a sketchy portrait could be drawn for the chloroplast genome of the last common ancestor of *Scenedesmus* and *Chlamydomonas*.

Background

The complete chloroplast DNA (cpDNA) sequences currently available for green plants (green algae and land plants) point to radically divergent evolutionary trends of the chloroplast genome in the phyla Streptophyta and Chlorophyta. The Streptophyta [1] comprises all land plants and their closest green algal relatives, the members of the class Charophyceae sensu Mattox and Stewart [2]. In this phylum are currently available the chloroplast genome sequences of about 35 land plants and six or seven charophycean green algae (six algae if the controversial phylogenetic position of *Mesostigma viride* at the base of the Streptophyta and Chlorophyta [3-5] proves to be correct and seven algae if its association with the Streptophyta is confirmed [6-8]). The Chlorophyta [9] comprises the Prasinophyceae, Ulvophyceae, Trebouxiophyceae and Chlorophyceae. The Prasinophyceae represent the most basal divergence of the Chlorophyta [10,11] and, although the branching order of the Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC) remains uncertain [12], chloroplast and mitochondrial genome data suggest that the Trebouxiophyceae emerged before the Ulvophyceae and Chlorophyceae [13-15]. Complete chloroplast genome sequences have been reported for five chlorophytes: the prasinophyte *Nephroselmis olivacea* [16], the trebouxiophyte *Chlorella vulgaris* [17], the ulvophytes *Oltmannsiellopsis viridis* [13] and *Pseudendoclonium akinetum* [15] and the chlorophycean green alga *Chlamydomonas reinhardtii* [18].

In nearly all photosynthetic lineages investigated thus far in the Streptophyta, the chloroplast genome harbours the same quadripartite structure and the same gene partitioning pattern, genes are densely packed and most of the genes are organized into conserved clusters [19,20], the origin of which dates back to the common ancestor of all chloroplasts [16]. The typical 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 (LSC) region. The rRNA operon always resides in the IR and is transcribed toward the SSC region. Although the IR readily expands or contracts by gaining or losing genes from the neighbouring single-copy regions [21], each of the three genomic partitions (IR, SSC and LSC) shows a distinctive and highly conserved gene content. Including 106 to 137 genes, the gene repertoire appears to have progressively shrunk from charophycean green algae to land plants [20,22]. Slight changes in intron composition of the chloroplast genome also occurred during streptophyte evolution [19,20,22]. The vast majority of introns were likely acquired early during the evolution of charophycean green algae.

In the Chlorophyta, the chloroplast genome shows extraordinary variability at the levels of its quadripartite structure, global gene organization and intron composition. The cpDNA of the prasinophyte *Nephroselmis* features the largest gene repertoire (128 genes) and the most ancestral features, including the quadripartite structure and gene partitioning pattern observed in streptophytes [16]. In contrast, all four completely sequenced UTC algal cpDNAs encode fewer genes (94-112) and are substantially rearranged [13,15,17,18]. Moreover, genes in these cpDNAs are more loosely packed than in *Nephroselmis* and most streptophyte cpDNAs, intergenic spacers usually contain short dispersed repeats (SDRs) and the coding regions of some protein-coding genes are expanded [13,15,18]. Of the four UTC cpDNAs, that of the trebouxiophyte *Chlorella* has retained the highest degree of ancestral characters; it lacks an IR but has retained many ancestral gene clusters. Both ulvophyte cpDNAs feature an atypical quadripartite structure that deviates from the ancestral type displayed by *Nephroselmis* and streptophyte cpDNAs. In each genome, one of the single-copy regions features many genes characteristic of both the ancestral SSC and LSC regions, whereas the opposite single-copy region features only genes characteristic of the ancestral LSC region. Moreover, the rRNA genes in the IR are transcribed toward the latter single-copy region. From their observations, Pombert *et al.* [13] concluded that a dozen genes were transferred from the LSC to the SSC region before or soon after emergence of the Ulvophyceae and that the transcription direction of the rRNA genes changed. In the chlorophycean green alga *Chlamydomonas*, the single-copy regions are similar in size and both their gene contents and gene organizations display tremendous differences relative to the same cpDNA regions in ulvophytes, implying that numerous genes were exchanged between opposite single-copy regions during the evolutionary period separating the Ulvophyceae and the chlorophycean clade represented by *Chlamydomonas* (a clade known as the Chlamydomonadales or CW clade [11]). Gene reshuffling was so extensive that no reliable scenario of gene rearrangements can be predicted to explain the observed differences.

To gain insights into the various changes that underwent the chloroplast genome in the Chlorophyceae, we have undertaken the complete sequencing of the chloroplast genome from distinct lineages of this class. We report here the 161,452 bp chloroplast genome sequence of *Scenedesmus obliquus*, a member of the lineage that appears to share a sister relationship with the Chlamydomonadales (Sphaeropleales or DO clade) [23,24]. All swimming cells in this lineage are biflagellates with a directly opposed (DO) arrangement of basal bodies, instead of the clockwise (CW) arrangement seen in the Chlamydomonadales. *Scenedesmus* cpDNA was found to be a compact genome

that carries as many derived features as its *Chlamydomonas* homologue. It shares with *Chlamydomonas* cpDNA single-copy regions of similar sizes, an almost identical gene repertoire and several derived gene clusters; however, the sets of genes in the single-copy regions of these chlorophycean genomes are very different. These extensive differences in global gene arrangement underscore the remarkable plasticity of the chloroplast genome in the Chlorophyceae.

Results

General features

The *Scenedesmus* cpDNA sequence assembles as a circular molecule of 161,452 bp encoding a total of 96 genes (not counting intron ORFs, free-standing ORFs and duplicated genes) (Fig. 1 and Table 1). With an overall A+T content of 73.1%, this chloroplast genome is the most A+T rich among completely sequenced chlorophyte cpDNAs. Two identical copies of an IR sequence of 12,022 bp are separated from one another by single-copy regions differing by only 7.5 kbp in size (SC1 and SC2). For both isomeric forms of the genome, a remarkably strong bias is observed in the distribution of genes between the two DNA strands. In the isomer shown in Fig. 1, 82 genes occupy one strand whereas, only 20, including the six present in the IR, reside on the opposite strand. Furthermore, 64 consecutive genes, encompassing more than half of the genome, feature the same polarity. The genes in *Scenedesmus* cpDNA are more tightly packed than in the four other completely sequenced UTC algal cpDNAs, their density

(67.2%) being comparable to that found in *Nephroselmis* cpDNA (68.7%). Intergenic spacers have an average size of 465 bp and feature short dispersed repeats (SDRs). A total of nine introns, seven group I and two group II introns, were identified in *Scenedesmus* cpDNA; five of these introns display ORFs.

Gene content and gene structure

The gene repertoire of the *Scenedesmus* genome differs from that of its *Chlamydomonas* homologue only by the presence of two additional genes, *infA* and *rpl12* (Table 1). As in *Chlamydomonas* cpDNA, two identical copies of the *trnE(uuc)* gene are found on opposite strands outside the IR. Five ORFs with more than 65 codons were identified in intergenic regions (Fig. 1). The largest one, ORF932, resides in SC1 and is part of the long segment carrying genes with identical polarity. The protein encoded by this ORF shows limited sequence similarity with bacterial reverse transcriptases, the observed similarity being restricted to domain X. The four remaining ORFs display no homology with any known DNA sequences. All five ORFs differ from the conserved protein-coding genes at the levels of codon usage and nucleotide composition.

As is the case for *Chlamydomonas* cpDNA, the *rpoB* and *rps2* genes in *Scenedesmus* cpDNA each occur as two contiguous ORFs (Fig. 1) and two separate genes, *clpP* and *rps3*, have extensions in their coding sequences that are absent from other chlorophyte and streptophyte cpDNAs (Table

Table 1: General features of *Scenedesmus* and other UTC algal cpDNAs

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

^aBecause *Chlorella* cpDNA lacks an IR, only the total size of this genome is given.

^bIn this study, this region is designated SC1 rather than LSC because it displays major differences in gene content relative to the LSC region in *Mesostigma* and *Nephroselmis* cpDNAs [13].

^cThis region is designated SC2 rather than SSC because it displays major differences in gene content relative to the SSC region in *Mesostigma* and *Nephroselmis* cpDNAs [13].

^dConserved genes, unique ORFs and introns were considered as coding sequences.

^eGenes present in the IR were counted only once. Unique ORFs and intron ORFs were not taken into account.

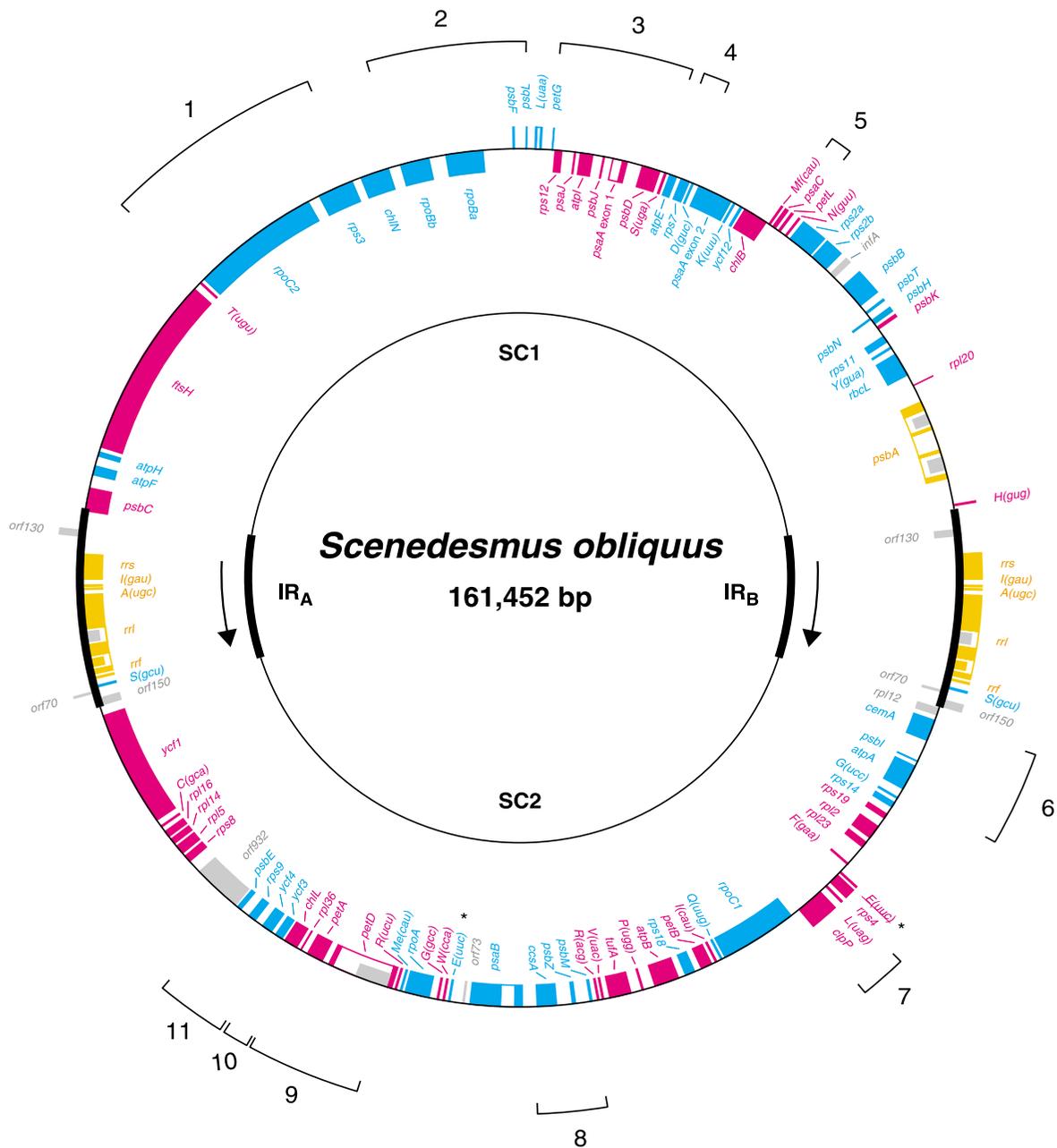


Figure 1
Gene map of *Scenedesmus* cpDNA and compared patterns of gene partitioning in *Chlamydomonas* and *Scenedesmus* cpDNAs. The two copies of the rRNA operon-containing IR (IR_A and IR_B) are represented by thick lines; the transcription direction of the rRNA genes is indicated by arrows. Genes (filled boxes) on the outside of the map are transcribed in a clockwise direction; those on the inside of the map are transcribed counterclockwise. The colour-code denotes the genomic regions containing the homologous genes in *Chlamydomonas* cpDNA: cyan, SC1; magenta, SC2; yellow, IR. Genes and ORFs absent from *Chlamydomonas* cpDNA are shown in grey. Labelled brackets denote the gene clusters shared specifically by *Scenedesmus* and *Chlamydomonas* cpDNAs (see Table 4 for the gene content of these clusters). tRNA genes are indicated by the one-letter amino acid code followed by the anticodon in parentheses (Me, elongator methionine; Mf, initiator methionine). Identical copies of the *trnE(uuc)* genes are denoted by asterisks. Introns are represented by open boxes and intron ORFs are denoted by narrow, filled boxes. The intron sequences bordering the *psaA* exons (*psaA* exon 1 and *psaA* exon 2) are spliced in *trans* at the RNA level. Note that only one of the two isomeric forms of the genome is shown here; these isomers differ with respect to the relative orientation of the single-copy regions.

Table 2: Expanded genes in *Scenedesmus* and other UTC algal cpDNAs

Gene	<i>Chlorella</i>		<i>Oltmannsiellopsis</i>		<i>Pseudendoclonium</i>		<i>Scenedesmus</i>		<i>Chlamydomonas</i>	
	Size (bp)	Factor ^a	Size (bp)	Factor ^a	Size (bp)	Factor ^a	Size (bp)	Factor ^a	Size (bp)	Factor ^a
<i>cemA</i> ^b	801	1.6	1059	2.2	909	1.8	1278	2.6	1503	3.1
<i>clpP</i>	606	0.9	588	0.9	597	0.9	1614	2.3	1575	2.3
<i>ftsH</i> ^b	5163	1.9	6879	2.6	7791	2.9	10998	4.1	8916	3.3
<i>infA</i> ^b	240	1.2	222	1.1	330	1.6	306	1.5	– ^c	– ^c
<i>rpoA</i>	837	0.9	1527	1.6	1734	1.8	1437	1.5	2213 ^d	2.3
<i>rpoB</i>	3906	1.2	4251	1.3	6537	2.0	4896 ^e	1.5	4967 ^e	1.5
<i>rpoC1</i>	2511	1.3	3066	1.5	4737	2.4	4590	2.3	5739 ^e	2.9
<i>rpoC2</i>	4689	1.3	5580	1.5	10389	2.8	7659	2.1	9363	2.5
<i>rps2</i>	804	1.2	717	1.0	777	1.1	2747 ^e	4.0	2731 ^e	4.0
<i>rps3</i>	696	1.1	708	1.1	690	1.1	2103	3.3	2139	3.3
<i>rps18</i> ^b	312	1.4	237	1.1	258	1.2	567	2.6	414	1.9
<i>ycf1</i> ^b	2460	2.0	2427	2.0	2505	2.0	7008	5.7	5988	4.9

^a Expansion factor relative to the corresponding *Mesostigma* gene. Each value was obtained by dividing the size of the UTC algal gene by the size of the corresponding *Mesostigma* gene.

^b Genes also expanded in streptophyte cpDNAs. Note that the *ftsH* gene is designated as *ycf2* in *Chlorella*, *Chlamydomonas* and land plants cpDNAs.

^c *infA* is missing in *Chlamydomonas* cpDNA

^d Size derived from our unpublished sequence data. A portion of the *rpoA* region is not annotated in GenBank:NC_005353 as a result of a sequencing error introducing a frameshift mutation.

^e Gene occurring as two independent ORFs. The indicated size includes the spacer separating the ORFs.

2). The extra coding sequences in each of these genes share the same insertion site in the two chlorophycean algae. Note that the intein gene previously identified within the *Chlamydomonas eugametos clpP* gene [25] lies at a position different from the insertion sequence reported here for *Scenedesmus clpP*. Six additional protein-coding genes in both *Scenedesmus* and *Chlamydomonas* cpDNAs resemble their *Chlorella* and/or ulvophyte homologues in exhibiting expanded coding regions (Table 2). In sharp contrast, the rRNA and tRNA genes of UTC algae show less than 1% deviation in size relative to their homologues in *Nephroselmis* and *Mesostigma*.

Gene partitioning and gene clustering

Our comparison of the gene complements found in the three genomic regions of *Scenedesmus* cpDNA with those observed in the *Chlamydomonas* genome reveals dramatic differences in the gene composition of the single-copy regions (Fig. 1). In each single-copy region of *Scenedesmus* cpDNA, we find numerous genes whose homologues map to the opposite single-copy region in the *Chlamydomonas* genome. Of the 43 genes displayed by *Scenedesmus* SC1 (largest single-copy region), 24 are located in the SC1 region (largest single-copy region) of *Chlamydomonas* (genes shown in cyan in Fig. 1), whereas all the others map to the alternate SC2 region (genes shown in magenta). Similarly, 19 of the 47 genes present in *Scenedesmus* SC2 reside in the SC1 region of *Chlamydomonas*, whereas all the others lie in the opposite SC2 region. Note here that the single-copy regions of both *Scenedesmus* and *Chlamydomonas* were arbitrarily disig-

nated (see Table 1) and that the genes shared by the SC1 or SC2 regions of these algae were not necessarily confined to the same single-copy region in the chloroplast genome of the last common ancestor of the two algae. Assuming that the SC1 or SC2 regions in *Scenedesmus* and *Chlamydomonas* cpDNAs are equivalent, it would be necessary to propose that the transcription direction of the rRNA operon was altered during the evolution of chlorophycean green algae concurrently with the extensive exchanges of genes that took place between the single-copy regions.

To identify the ancestral clusters carried by *Scenedesmus* cpDNA as well as the derived clusters that are shared with other UTC algal cpDNAs, we carried out a detailed comparative analysis of gene order. As a first step, we investigated the 24 gene clusters present in both *Mesostigma* and *Nephroselmis* cpDNAs and found that *Scenedesmus* cpDNA is identical to its *Chlamydomonas* counterpart in terms of composition of ancestral clusters (Fig. 2). Both chlorophycean green algal cpDNAs harbour a single, intact ancestral cluster (*psbB-T-/N-/H*, with the slash indicating a change in gene polarity) and the remains of four other ancestral clusters: altogether, these conserved clusters encode 20 genes. These observations confirm the notion that the lowest degree of ancestral clusters among UTC algal cpDNAs is found in the chlorophycean lineage [18,26]. As previously reported by Pombert *et al.* [13], the chloroplast genome of the trebouxiophyte *Chlorella* exhibits the highest conservation of ancestral clusters among completely sequenced UTC algal cpDNAs; it has retained

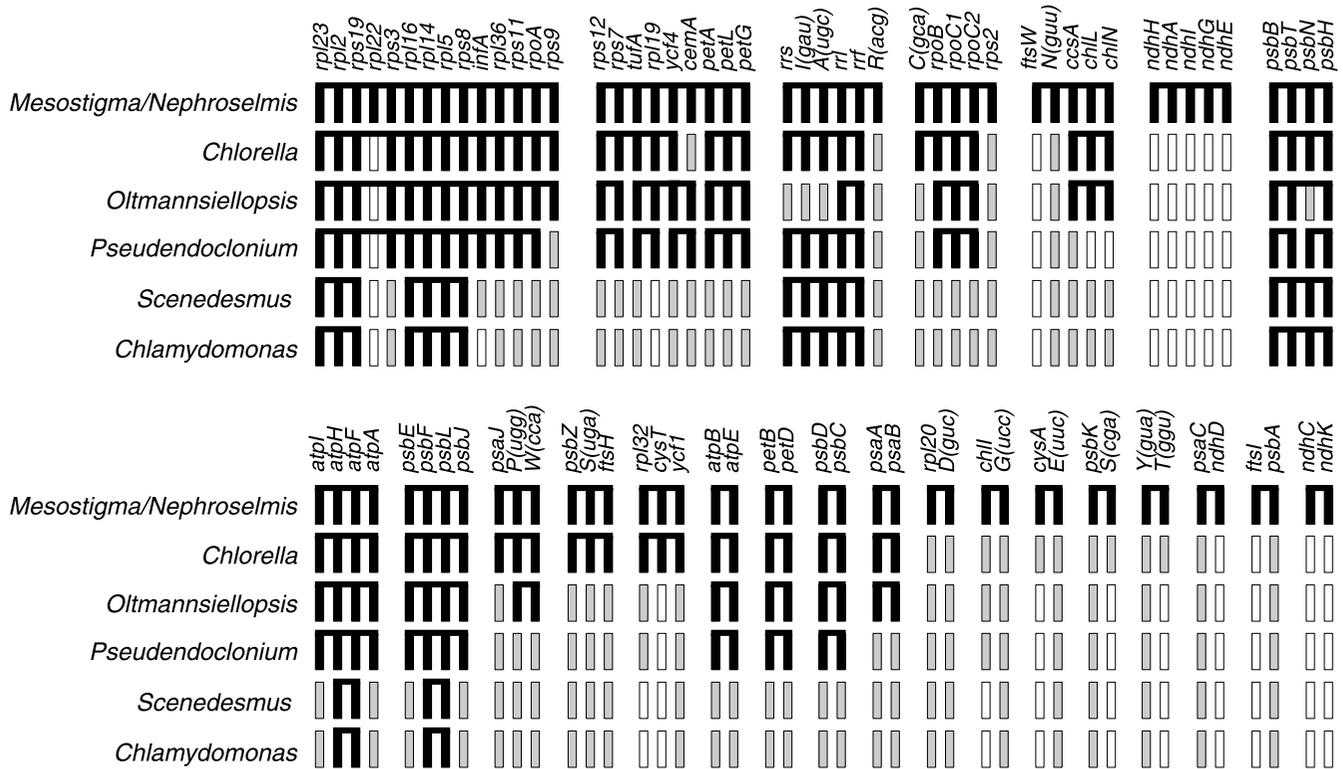


Figure 2
Conservation of ancestral gene clusters in *Scenedesmus* and other UTC algal cpDNAs. Black boxes represent the 89 genes found in the 24 clusters shared by *Mesostigma* and *Nephroselmis* cpDNAs as well as the genes in UTC algal cpDNAs that have retained the same order as those in these ancestral clusters. For each genome, the set of genes making up each of the identified clusters (either an intact or fragmented ancestral cluster) is shown as black boxes connected by a horizontal line. Black boxes that are contiguous but not linked together indicate that the corresponding genes are not adjacent on the genome. Gray boxes denote genes in UTC algal cpDNAs that have been relocated elsewhere on the chloroplast genome; open boxes denote genes that have disappeared from the chloroplast genome. Although the *rpl22* gene is missing from *Nephroselmis* cpDNA, it is shown as belonging to the large ribosomal protein cluster equivalent to the contiguous *S10*, *spc* and α operons of *Escherichia coli* because it is present in this cluster in the cpDNAs of *Mesostigma*, streptophytes and algae from other lineages. Note also that the *psbB* cluster of *Oltmannsiellopsis* and *Pseudendoconium* cpDNAs differs from the ancestral cluster found in other genomes by the presence of *psbN* on the alternate DNA strand.

11 intact clusters and four partially conserved ones that include 62 genes. The chloroplast genomes of the ulvophytes *Oltmannsiellopsis* and *Pseudendoconium* have retained only five of the 24 ancestral clusters in an intact state as well as 13 and 10 partially conserved clusters, respectively.

The overall gene order in *Scenedesmus* cpDNA was also compared to the global gene arrangements of other completely sequenced UTC algal cpDNAs. As expected, we found that *Scenedesmus* cpDNA most closely resembles its *Chlamydomonas* counterpart at the level of derived gene clusters. The two chlorophycean genomes share 11 clusters that are not conserved in *Chlorella*, *Oltmannsiellopsis* and *Pseudendoconium* cpDNAs (Fig. 1 and Table 4). Three of these clusters (clusters 9, 10 and 11) are adjacent on the

Scenedesmus genome, whereas no shared derived clusters show contiguity on *Chlamydomonas* cpDNA. Only the *rpoB*-/*psbF*-*psbL* cluster originated by fusion of a fragment of ancestral cluster (*psbF*-*psbL*) with other genes. Two of the 11 derived clusters, *petA*-*petD* and *trnL(uag)*-*clpP* clusters, as well segments of the *atpA*-*psbI*-*cemA* and *psbE*-*rps9*-*ycf4*-*ycf3* clusters (relevant segments are underlined) have been previously identified not only in *C. reinhardtii* and its close relative *C. gelatinosa* [27] but also in the distantly pair of interfertile algae *C. eugametos* and *C. moewusii* [26] and their close relative *C. pitschmannii* [28], indicating their conservation in the Chlamydomonadales. The *Scenedesmus* chloroplast genome shares no specific gene clusters with *Oltmannsiellopsis* and *Pseudendoconium* cpDNA and only one pair of genes (*psaJ*-*rps12*, a subset of cluster 3 in Table 4) with *Chlorella* cpDNA.

Table 3: Introns in *Scenedesmus* cpDNA and homologous introns at identical gene locations in other green algal cpDNAs

Designation	Scenedesmus introns			Homologous introns	
	Subgroup ^a	ORF location ^b	ORF type ^c	Green alga ^d /Intron number ^e	Accession no.
So.L(uaa).1	IC3	–	–	<i>Bryopsis plumosa</i> (U) <i>Chlorella vulgaris</i> (T)	[GenBank: M61159] [GenBank: NC_001865]
So.psaB.1	IA1	–	–	<i>Chlamydomonas moewusii</i> (C)	[GenBank: M90641]
So.psbA.1	IA1	L5	H-N-H	–	–
So.psbA.2	IA1	–	–	–	–
So.psbA.3	IA1	L5	H-N-H	<i>Pseudoclonium akinetum</i> i6 (U)	[GenBank: AY835431]
So.rrl.1	IB4	L6	LAGLIDADG	<i>Chlamydomonas eugametos</i> i5 (C)	[GenBank: Z17234]
So.rrl.2	IA3	L6	LAGLIDADG	<i>Chlamydomonas reinhardtii</i> (C) <i>Chlorella vulgaris</i> (T) <i>Oltmannsiellopsis viridis</i> i3 (U) <i>Pseudoclonium akinetum</i> (U)	[GenBank: NC_005353] [GenBank: NC_001865] [GenBank: DQ291132] [GenBank: AY835431]
So.petD.1	IIA	DIV	RT-X-Zn	–	–
So.psaA.1	IIB	–	–	<i>Chlamydomonas reinhardtii</i> i2 (C)	[GenBank: NC_005353]

^a Group I introns were classified according to Michel and Westhof [66], whereas classification of group II introns was according to Michel *et al.* [65].

^b L followed by a number refers to the loop extending the base-paired region identified by the number; D refers to a domain of group II intron secondary structure.

^c For group I intron ORFs, the conserved motif in the predicted homing endonuclease is given; for the *petD* intron ORF, RT, X and Zn refer to the reverse transcriptase, maturase and nuclease domains of reverse transcriptases, respectively.

^d For the two *rrl* introns, only the homologues identified in completely sequenced genomes are given. The complete lists of introns homologous to *So. rrl.1* and *So.rrl.2* can be obtained in Turmel *et al.* [35] and Pombert *et al.* [13], respectively. The letter in parentheses indicates the chlorophyte lineage comprising the green alga indicated. U, Ulvophyceae; T, Trebouxiophyceae; C, Chlorophyceae.

^e The intron number is given when more than one intron is present.

Introns

The seven group I introns of *Scenedesmus* interrupt four genes: three introns occur in *psbA*, two in *rrl* and the two others in *psaB* and *trnL(uaa)*. These introns fall within

Table 4: Derived gene clusters shared by *Scenedesmus* and *Chlamydomonas* cpDNAs

Cluster no. ^a	Gene composition
1	<i>rps3-rpoC2</i>
2	<i>rpoBb-rpoBa-lpsbF-psbL^b</i>
3	<i>psbD-psaAa-psbJ-atpI-psaJ-rps12</i>
4	<i>rps7-atpE</i>
5	<i>psaC-petL-trnN(guu)</i>
6	<i>atpA-psbI-cemA</i>
7	<i>trnL(uag)-clpP</i>
8	<i>ccsA-psbZ-psbM</i>
9	<i>petA-petD</i>
10	<i>chlL-rpl36</i>
11	<i>psbE-rps9-ycf4-ycf3</i>

^a Clusters are labelled as in Fig. 1.

^b The slash indicates a change in gene polarity.

four different subgroups (IA1, IA3, IB4 and IC3), with the IA1 subgroup including the four introns present in *psaB* and *psbA* (Table 3). At 255 bp, the IC3 intron in *trnL(uaa)* is the smallest of the *Scenedesmus* introns. As homologous introns are inserted at the same position not only in the chloroplast *trnL(uaa)* genes of the chlorophytes *Bryopsis* and *Chlorella* (Table 3) but also in the *trnL(uaa)* genes of streptophytes and algae from other lineages, this intron is thought to have been inherited by vertical inheritance from the common ancestor of all chloroplasts [29]. The IA3 and IB4 introns in *Scenedesmus rrl* are also positionally and structurally homologous to previously reported introns in green plant cpDNAs (Table 3). Although the four IA1 introns revealed relatively poor sequence similarity with one another, two of these introns, *So.psaB.1* and *So.psbA.3*, were found to be clearly homologous to introns inserted at identical gene locations in *Chlamydomonas moewusii* and *Pseudoclonium* cpDNAs, respectively (Table 3). *So. psbA.3* and its *Pseudoclonium* homologue display not only similar primary sequences and secondary structures, but also similar ORFs encoding potential hom-

ing endonucleases carrying the H-N-H motif (44% identity at the protein sequence level). The two other IA1 introns of *Scenedesmus*, *So.psbA.1* and *So.psbA.2*, represent unique insertion positions in the *psbA* gene; they are located only 5 bp and 6 bp away from the second and fourth introns in *Pseudendoclonium psbA*, respectively. For these two pairs of closely linked introns, similarity was found to be limited to the *So.psbA.1* intron-encoded H-N-H homing endonuclease, which shares 33% sequence identity with the protein encoded by second intron in *Pseudendoclonium psbA*.

One of the two group II introns of *Scenedesmus* is spliced in *trans* at the RNA level. This intron occurs in *psaA* and is inserted at exactly the same site as the second of the two *trans*-spliced introns in *Chlamydomonas psaA* [30] (Table 3). Like its *Chlamydomonas* homologue, it has no ORF. These positionally identical *Scenedesmus* and *Chlamydomonas psaA* introns both belong to the IIB subgroup, share poor sequence similarity and are both fragmented within domain IV (Fig. 3). The second group II intron in *Scenedesmus* cpDNA lies within *petD* and has no known homologue (Table 3). This intron harbours in its domain IV an ORF encoding a reverse transcriptase [31] with the typical reverse transcriptase, maturase and nuclease domains [32].

Short dispersed repeats

SDR elements were identified in many intergenic spacers and some coding regions of *Scenedesmus* cpDNA; however, they are less numerous than those found in other UTC algal cpDNAs (Table 5). The repeats ≥ 30 bp in the *Scenedesmus* genome represent 8.7% of the total size of the intergenic spacers, whereas the fraction of the intergenic regions represented by such repeats in *Chlamydomonas* cpDNA reaches 31.9%. Analysis of the most abundant repeat elements in the *Scenedesmus* genome using Repeat-Finder revealed three distinct groups of repeat units featuring sequences of 15 or 16 bp (Table 6). With a total of 41 copies, repeat unit B represents the most abundant group of repeats. Repeat units A and C are restricted to intergenic regions, whereas some copies of unit B are also found within the coding regions of five genes that are expanded relative to their *Mesostigma* counterparts, *i.e.* *cemA*, *ftsH*, *infA*, *rpoBa* and *rpoC2* (Fig. 4 and see Table 2).

Although repeat units A, B and C occur on both strands of *Scenedesmus* cpDNA, they are not evenly distributed throughout the genome (Fig. 4). Many intergenic spacers entirely lack copies of these repeat units and tend to be clustered in distinct cpDNA regions (e.g., the regions in the vicinity of *petD*, *psbA* and *rpoC1*). On the other hand, numerous intergenic spacers are populated by two or more copies of the same repeat unit and/or by several copies representing different units (Fig. 4). The repeats in the

latter spacers often form longer repeated sequences that are not randomly distributed on the *Scenedesmus* genome. In Fig. 4, it can be seen that the great majority of the repeats exceeding 30 bp in size are confined to one half of the genome. While most of the intergenic spacers harbouring SDRs reside in regions that differ in gene order relative the *Chlamydomonas* genome, some occur in shared, derived gene clusters (clusters 1, 2, 3, 5, 6 and 7).

In the intergenic spacers displaying copies of the same repeat unit, these copies are often arranged in direct orientation (Fig. 4) and separated by 23–25 bp; identical repeats also occur on different strands but, in this configuration, their distances are highly variable (8–116 bp). In intergenic regions of *Chlamydomonas* cpDNA, repeated elements appear to show arrangements similar to those reported here for *Scenedesmus* cpDNA. In contrast, in *Oltmannsiellopsis* and *Pseudendoclonium* cpDNAs, SDRs occur predominantly as stem-loop structures [13,15]. None of the repeat units of *Scenedesmus* cpDNA was identified as being part of SDRs in other UTC algal cpDNAs.

Discussion

Our comparative analyses of the *Scenedesmus* chloroplast genome with previously sequenced chlorophyte cpDNAs highlight the remarkable plasticity displayed by the chloroplast genome in the Chlorophyceae. As expected, we found that *Scenedesmus* cpDNA shows the most similarities with *Chlamydomonas* cpDNA. The almost identical gene repertoires displayed by these chlorophycean green algal cpDNAs contrasts with the tremendous differences they exhibit at the level of gene order and pattern of gene partitioning between the single-copy regions. This highly variable gene organization is not the only surprising result that emerged from our study. Three other features of the *Scenedesmus* genome were found to be peculiar: (1) its high gene density, which mirrors that found for *Nephroselmis* cpDNA and diverges from the tendency of previously studied UTC algal cpDNAs to grow in size by gaining sequences in intergenic regions and selected gene coding regions [13,15,18], (2) the low abundance of its repeated sequences, which represents the lowest level identified thus far in a UTC algal cpDNA and (3) the strongly biased distribution of its genes between the two DNA strands.

The features shared by *Scenedesmus* and *Chlamydomonas* cpDNAs provide information on the cpDNA of the last common ancestor of DO and CW green algae; however, the portrait that could be drawn for this ancestral genome is rather sketchy owing to the major differences observed at the levels of gene order and intron content. We infer that the chloroplast genome of the last common ancestor of DO and CW green algae harboured a total of 96 genes, including a duplicated *trnE(uuc)* gene, that one third of

Table 5: Abundance of SDRs in *Scenedesmus* and other UTC algal cpDNAs

cpDNA	Maximal size of repeats (bp)	Number of repeats ^a		Total size (bp)	Non-overlapping repeats ≥ 30 bp ^b	
		≥ 30 bp	≥ 45 bp		Fraction of genome (%)	Fraction of intergenic regions (%)
<i>Chlorella</i>	84	269	44	11,743	7.8	20.8
<i>Oltmannsiellopsis</i>	172	1,205	161	18,033	11.9	30.1
<i>Pseudendozonium</i>	171	1,047	203	10,073	5.1	13.6
<i>Scenedesmus</i>	112	86	21	4,817	3.0	8.7
<i>Chlamydomonas</i>	221	3,247	551	32,244	15.8	31.9

^aRepeats with identical sequences on the same DNA strand or different DNA strands were identified using REPuter.

^bThe repeats identified with REPuter were annotated and masked on the genome sequence using RepeatMasker.

these genes were organized in the same order as those found in the 11 gene clusters specifically shared by *Scenedesmus* and *Chlamydomonas* cpDNAs, that both *rps2* and *rpoB* were fragmented in two pieces and that *clpP* and *rps3* each displayed an insertion sequence. During the evolution of chlamydomonads, *infA* and *rpl12* genes disappeared from the chloroplast genome and *rpoC1* was broken into two separate reading frames. We also predict with confidence that the ancestral genome contained introns in *rrl* and *psaA* at the same positions as those shared by *Scenedesmus* and *Chlamydomonas* cpDNAs as well as introns in *trnL(uaa)*, *psaB*, and *rrl* at the same positions as those shared by *Scenedesmus* and other chlorophyte cpDNAs. Homologues of all these introns, with the exception of the *trnL(uaa)* intron, have been identified in chlamydomonads distantly related to *C. reinhardtii* (*rrl*, [33-35]; *psaA*, [26,36]; *psaB*, [37]). In contrast, the *trnL(uaa)* intron shows a broader distribution among green algae and is thought to have been inherited from the common ancestor of all chloroplasts [29]. Undoubtedly, as reported for *C. reinhardtii* cpDNA [38], the *psaA* intron of the last common ancestor of DO and CW green algae featured a break in domain IV, the two *psaA* exons were unlinked and transcribed independently along with an intron fragment, and the intron was spliced *in trans*. In the CW lineage, a second *trans*-spliced group II intron (a tripartite intron comprising the RNA species encoded by the chloroplast *tscA* gene) took residence within *psaA* [38], group I introns inserted at several new sites within *rrl*

[33,34,39] and members of the group I family also invaded multiple sites of the *rrs* [40,41], *psbA* [42] and *psbC* [37] genes.

The unusual structures displayed by the expanded *clpP* and *rps3* genes and the fragmented *rps2* and *rpoB* genes are inventions that arose in the Chlorophyceae. For both *clpP* and *rps3*, it has been shown that the insertion sequence is not removed at the RNA level [43,44]. Characterization of the chloroplast ClpP/R protease complex of *C. reinhardtii* revealed that the approximately 30 kDa insertion sequence in *clpP*, designated as IS1, could be a new type of intein [45]. Two distinct proteins derived from the chloroplast *clpP* gene, a long version containing IS1 and a shorter version lacking this sequence element, were found to be stable components of this complex. IS1 has been hypothesized to prevent interaction with the HSP100 chaperone and to be localized in only one of the two heptamers forming the complex, thus prohibiting access of protein substrates to the proteolytic chamber of the ClpP/R complex via one of its axial pores. In contrast, a proteomic analysis of ribosomal proteins in the small subunit of the chloroplast ribosome from *C. reinhardtii* revealed that the insertion sequence in *rps3* is an integral part of the mature product of this gene [46]. In this same analysis, Rps2 was also found to be an unusually large ribosomal protein; this protein of 570 amino acid residues encoded by *rps2b* (the largest of the two ORFs showing sequence similarity to the bacterial *rps2* genes), contains an N-terminal exten-

Table 6: SDR repeat units in *Scenedesmus* cpDNA

Designation	Size (bp)	Sequence	Copy number ^a	
			Forward strand	Reverse strand
A	15	TTTACGCTTTTTTTC	17	10
B	15	TTCTTCTTCATTTTT	22	19
C	16	TGCTTTGCTGCTTTTT	16	22

^aCopies of each SDR unit were identified using FINDPATTERNS (one mismatch was allowed).

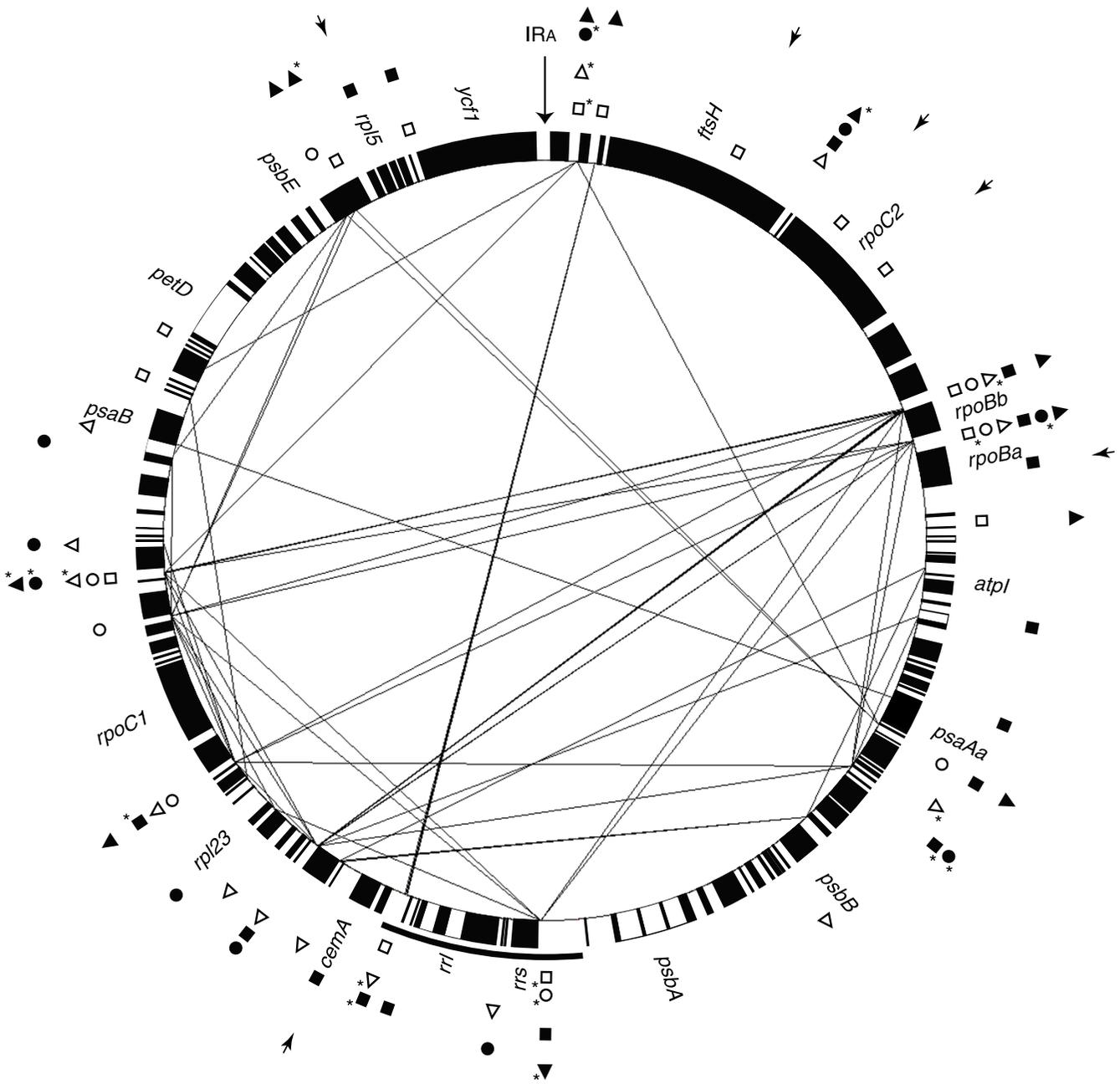


Figure 4
Positions of SDR elements in *Scenedesmus* cpDNA. Lines connect cpDNA loci displaying repeats ≥ 30 bp with identical sequences either on the same strand or different strands. For this analysis carried out with REPuter, one copy of the IR sequence (IR_A) was deleted; the location of this deleted copy is indicated by the long, vertical arrow. The loci containing repeat units A, B and C are represented by symbols of different shapes outside the gene map: triangles, repeat unit A; squares, repeat unit B; circles, repeat unit C. Filled symbols denote the repeats occupying the + strand; open symbols denote the repeats found on the alternate strand. A symbol accompanied by an asterisk indicates the presence of two or more copies. Small arrows point to gene coding regions containing copies of repeat unit B.

sion and a C-terminal half with homology to characterized Rps2 proteins from other organisms. No peptides were found to be derived from *rps2a*, indicating that the

latter sequence may be a pseudogene. The biological significance of the additional domains found in Rps3 and Rps2 remains uncertain. These domains, which are

Table 7: Minimal numbers of inversions accounting for gene rearrangements between green algal cpDNAs

Compared cpDNA	Number of inversions ^a					
	<i>Nephroselmis</i>	<i>Chlorella</i>	<i>Oltmannsiellopsis</i>	<i>Pseudendoclonium</i>	<i>Scenedesmus</i>	<i>Chlamydomonas</i>
<i>Mesostigma</i>	43	46	54	54	75	75
<i>Nephroselmis</i>		47	55	55	73	74
<i>Chlorella</i>			50	52	74	75
<i>Oltmannsiellopsis</i>				55	78	75
<i>Pseudendoclonium</i>					76	77
<i>Scenedesmus</i>						58

^a Numbers of gene permutations by inversions were computed using GRIMM.

exposed to the solvent side and are located near each other and around the neck of the 30S subunit, may be related to unique features of translational regulation, or they may be orthologues of nonribosomal proteins [46]. Finally, it is not yet clear how the fragmented *rpoB* gene is expressed at the protein level. This gene is undoubtedly essential for cell survival in view of the fact that, unlike their homologues in land plants, the *C. reinhardtii* nuclear genome does not appear to encode a chloroplast-targeted RNA polymerase [47,48].

The considerable differences in gene density and abundance of SDRs observed in *Scenedesmus* and *Chlamydomonas* cpDNAs raise questions about the status of the chloroplast genome of the common ancestor of DO and CW green algae with regards to these features. From the data derived from previously sequenced chlorophyte cpDNAs, Pombert *et al.* [13,15] proposed that proliferation of repeated sequences in intergenic regions and selected genes occurred progressively during the evolution of UTC algae, thereby accounting for the observation that the *Chlamydomonas* genome is the most rich in SDR elements and the least tightly packed with genes. The results reported here are compatible with this idea and support the presence of SDRs in the common ancestor of *Scenedesmus* and *Chlamydomonas* cpDNAs provided that specific loss of numerous SDRs occurred concurrently with streamlining of the genome in the *Scenedesmus* lineage. On the other hand, considering that no common SDRs have been identified in different UTC algal cpDNAs, the idea that these genetic elements were independently acquired in UTC lineages cannot be ruled out.

The single-copy regions of *Scenedesmus* and *Chlamydomonas* cpDNAs are almost equal in size but differ radically in gene content, indicating that many genes were exchanged between opposite single-copy regions during the evolution of the DO and CW algae. This observation contrasts with the situation reported for the cpDNAs of *C. reinhardtii* and *C. moewusii* [26]. These representatives of deeply branched chlamydomonad lineages also display

extensive gene rearrangements in their cpDNAs; however, these rearrangements are mainly confined to individual single-copy regions. Only two (*atpA* and *psbI*) of the 77 genes mapped on the *C. reinhardtii* and *C. moewusii* genomes [26-28] moved from one single-copy region to the other.

To compare the level of gene rearrangements displayed by *Scenedesmus* and *Chlamydomonas* cpDNAs with those exhibited by other pairs of chlorophyte cpDNAs, we examined the orders of the 90 genes common to seven green algal cpDNAs (Table 7). In these analyses, changes in gene order were assumed to occur only by inversions, an hypothesis that is strongly supported by previous mapping analyses of chloroplast genes from closely related chlamydomonads [27,28]. We estimated that a minimum of 58 inversions would be required to convert the gene order of *Scenedesmus* cpDNA to that of *Chlamydomonas* cpDNA. Although extensive, this level of gene rearrangements is less important than those observed in the comparisons involving each chlorophycean genome and a genome from another green algal class (Table 7). The similar levels of gene rearrangements observed in these inter-class comparisons suggest that the gene organizations of both *Chlamydomonas* and *Scenedesmus* diverged considerably from that of their last common ancestor.

The high gene density and strongly biased distribution of genes between the two DNA strands in the *Scenedesmus* genome most probably reflect the influence of natural selection on genome organization. A bias in clustering of adjacent genes on the same DNA strand has also been reported for the *Chlamydomonas* chloroplast genome [49]; however, this bias is less conspicuous than that observed for *Scenedesmus* cpDNA. For *Chlamydomonas* cpDNA, a parametric bootstrap approach was used to test if gene order evolves under selection [49]. In this analysis, the putative gene order in the common ancestor of *Chlamydomonas* and *Chlorella* was inferred and subjected to random rearrangements. It was found that the multiple gene rearrangements in the *Chlamydomonas* lineage

resulted in an increased sidedness, *i.e.* an increased propensity of adjacent genes to be located on the same strand. Sidedness indices of 0.6966 and 0.8710 were scored for the common ancestor and *Chlamydomonas*, respectively, and simulated genomes showed a significant decrease in sidedness relative to the ancestral genome. At 0.8842, the sidedness index we calculated for *Scenedesmus* cpDNA is slightly higher than that reported for its *Chlamydomonas* counterpart.

Coding strand biases have also been reported for the plastid genomes of the parasitic green alga *Helicosporidium* sp. [50], the euglenozoan alga *Euglena gracilis* [51], and apicomplexan parasites [52-55]. This feature is prominent in the highly reduced *Helicosporidium* genome where a symmetry in strand bias of coding regions has been observed, with nearly all genes on each half of the genome being encoded on one strand. The two strands of the *Helicosporidium* and *Euglena* genomes are also biased with regards to nucleotide composition and this compositional bias switches at the putative origin of DNA replication [50,51]. It has been proposed that the coding strand bias observed in these genomes is generated by selection to code highly expressed genes on the leading strand to limit collisions between RNA and DNA polymerases, thereby increasing the rates of both replication and transcription. Unlike their *Helicosporidium* and *Euglena* homologues, *Scenedesmus* and *Chlamydomonas* cpDNAs show no strand bias in nucleotide composition (our unpublished results and [49]), thus providing no support for the notion that gene orders in chlorophycean genomes are selected to maximize the rate of replication. The high degree of sidedness observed for *Scenedesmus* and *Chlamydomonas* cpDNAs could result mainly from selection of polycistronic transcription to coordinate gene expression [49].

Conclusion

Our study revealed that, although *Scenedesmus* and *Chlamydomonas* cpDNAs display nearly identical gene repertoires and a high level of sidedness in the distribution of their genes on the two DNA strands, their gene orders are highly scrambled. In future studies, it will be interesting to investigate whether remodelling of the chloroplast genome is subjected to different constraints in the DO and CW lineages and whether the derived state of *Scenedesmus* and *Chlamydomonas* cpDNAs arose early during the evolution of chlorophycean green algae. To test this hypothesis, it will be necessary to examine other representatives of the DO and CW clades as well as members of more basal lineages of the Chlorophyceae.

Methods

Strain and culture conditions

Scenedesmus obliquus (Turp.) Kütz was obtained from the Culture Collection of Algae at the University of Texas at

Austin (UTEX 393) and grown in modified Volvox medium [56] under 12 h light/dark cycles.

Isolation and sequencing of cpDNA

An A+T rich fraction containing cpDNA was isolated and sequenced as described in Pombert *et al.* [14]. Sequences were edited and assembled with AUTOASSEMBLER 2.1.1 (Applied Biosystems). The fully annotated chloroplast genome sequence has been deposited in [GenBank: [DQ396875](http://www.ncbi.nlm.nih.gov/nuccore/DQ396875)].

Analyses of genome sequence

Gene content was determined by Blast homology searches [57] 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 GCG version 10.2 package (Accelrys, Burlington, Mass.) and other applications from the EMBOSS version 2.6.0 package [58]. Genes coding for tRNAs were localized using tRNAscan-SE 1.23 [59]. Repeated sequences were identified using REPuter 2.74 [60] and classified using REPEATFINDER [61]. Numbers of SDR units were determined with FINDPATTERNS of the GCG Wisconsin Package version 10.2. The total length of genome sequences containing repeated elements was estimated with RepeatMasker [62] running under the WU-BLAST 2.0 search engine [63].

The GRIMM web server [64] 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 Table 7 contained 90 genes; the three exons of the *trans*-spliced *psaA* gene were coded as distinct fragments (for a total of 92 gene loci).

Abbreviations

cpDNA, chloroplast DNA; CW, clockwise; DO, directly opposed; IR, inverted repeat; LSC, large single copy; SDR, short dispersed repeat; SSC, small single copy; UTC, Ulvo-phyceae/Trebouxiophyceae/Chlorophyceae.

Authors' contributions

JCC participated in the conception of this study, carried out part of the genome sequencing, performed all sequence analyses, annotated the genome, generated the tables and figures, and drafted the manuscript. CO participated in the sequencing and contributed to the assembly of the genome sequence. CL and MT conceived and supervised the study, contributed to the interpretation of the

data and prepared the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Patrick Charlebois for his help with the analysis of conserved gene clusters and Jean-François Pombert for critical reading of the manuscript. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (to MT and CL).

References

- Bremer K: **Summary of green plant phylogeny and classification.** *Cladistics* 1985, **1**:369-385.
- Mattox KR, Stewart KD: **Classification of the green algae: a concept based on comparative cytology.** In *The Systematics of the Green Algae* Edited by: Irvine DEG, John DM. London: Academic Press; 1984:29-72.
- Lemieux C, Otis C, Turmel M: **Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution.** *Nature* 2000, **403(6770)**:649-652.
- Turmel M, Ehara M, Otis C, Lemieux C: **Phylogenetic relationships among Streptophytes as inferred from chloroplast small and large subunit rRNA gene sequences.** *J Phycol* 2002, **38**:364-375.
- Turmel M, Otis C, Lemieux C: **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* 2002, **19(1)**:24-38.
- Bhattacharya D, Weber K, An SS, Berning-Koch W: **Actin phylogeny identifies *Mesostigma viride* as a flagellate ancestor of the land plants.** *J Mol Evol* 1998, **47(5)**:544-550.
- Karol KG, McCourt RM, Cimino MT, Delwiche CF: **The closest living relatives of land plants.** *Science* 2001, **294**:2351-2353.
- Marin B, Melkonian M: **Mesostigmatophyceae, a new class of streptophyte green algae revealed by SSU rRNA sequence comparisons.** *Protist* 1999, **150(4)**:399-417.
- Sluiman HJ: **A cladistic evaluation of the lower and higher green plants (*Viridiplantae*).** *Plant Syst Evol* 1985, **149**:217-232.
- Friedl T: **The evolution of the green algae.** *Plant Syst Evol* 1997, **11(Suppl)**:87-101.
- Lewis LA, McCourt RM: **Green algae and the origin of land plants.** *Am J Bot* 2004, **91(10)**:1535-1556.
- Friedl T, O'Kelly CJ: **Phylogenetic relationships of green algae assigned to the genus *Planophila* (Chlorophyta): evidence from 18S rDNA sequence data and ultrastructure.** *Eur J Phycol* 2002, **37**:373-384.
- Pombert JF, Lemieux C, Turmel M: **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 Biology* 2006, **4**:3.
- Pombert JF, Otis C, Lemieux C, Turmel M: **The complete mitochondrial DNA sequence of the green alga *Pseudoclonium akinetum* (Ulvophyceae) highlights distinctive evolutionary trends in the Chlorophyta and suggests a sister-group relationship between the Ulvophyceae and Chlorophyceae.** *Mol Biol Evol* 2004, **21(5)**:922-935.
- Pombert JF, Otis C, Lemieux C, Turmel M: **The Chloroplast Genome Sequence of the Green Alga *Pseudoclonium akinetum* (Ulvophyceae) Reveals Unusual Structural Features and New Insights into the Branching Order of Chlorophyte Lineages.** *Mol Biol Evol* 2005, **22(9)**:1903-1918.
- Turmel M, Otis C, Lemieux C: **The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea* insights into the architecture of ancestral chloroplast genomes.** *Proc Natl Acad Sci USA* 1999, **96(18)**:10248-10253.
- 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: **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* 1997, **94(11)**:5967-5972.
- Maul JE, Lilly JW, Cui L, dePamphilis CW, Miller W, Harris EH, Stern DB: **The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats.** *Plant Cell* 2002, **14(11)**:2659-2679.
- Palmer JD: **Plastid chromosomes: structure and evolution.** In *The Molecular Biology of Plastids Cell Culture and Somatic Cell Genetics of Plants Volume 7A*. Edited by: Bogorad L, Vasil I. San Diego: Academic Press; 1991:5-53.
- Turmel M, Otis C, Lemieux C: **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* 2002, **99(17)**:11275-11280.
- Goulding SE, Olmstead RG, Morden CW, Wolfe KH: **Ebb and flow of the chloroplast inverted repeat.** *Mol Gen Genet* 1996, **252(1-2)**:195-206.
- Turmel M, Otis C, Lemieux C: **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 Biology* 2005, **3**:22.
- Buchheim MA, Michalopoulos EA, Buchheim JA: **Phylogeny of the Chlorophyceae with special reference to the Sphaeropleales: a study of 18S and 26S rDNA data.** *J Phycol* 2001, **37**:819-835.
- Shoup S, Lewis LA: **Polyphyletic origin of parallel basal bodies in swimming cells of chlorophycean green algae (Chlorophyta).** *J Phycol* 2003, **39**:789-796.
- Wang S, Liu XQ: **Identification of an unusual intein in chloroplast ClpP protease of *Chlamydomonas eugametos*.** *J Biol Chem* 1997, **272(18)**:11869-11873.
- Boudreau E, Otis C, Turmel M: **Conserved gene clusters in the highly rearranged chloroplast genomes of *Chlamydomonas moewusii* and *Chlamydomonas reinhardtii*.** *Plant Mol Biol* 1994, **24(4)**:585-602.
- Boudreau E, Turmel M: **Extensive gene rearrangements in the chloroplast DNAs of *Chlamydomonas* species featuring multiple dispersed repeats.** *Mol Biol Evol* 1996, **13(1)**:233-243.
- Boudreau E, Turmel M: **Gene rearrangements in *Chlamydomonas* chloroplast DNAs are accounted for by inversions and by the expansion/contraction of the inverted repeat.** *Plant Mol Biol* 1995, **27(2)**:351-364.
- Besendahl A, Qiu YL, Lee J, Palmer JD, Bhattacharya D: **The cyanobacterial origin and vertical transmission of the plastid tRNA(Leu) group-I intron.** *Curr Genet* 2000, **37(1)**:12-23.
- Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD: **A small chloroplast RNA may be required for trans-splicing in *Chlamydomonas reinhardtii*.** *Cell* 1991, **65(1)**:135-143.
- Fassbender S, Bruhl KH, Ciriacy M, Kuck U: **Reverse transcriptase activity of an intron encoded polypeptide.** *EMBO J* 1994, **13(9)**:2075-2083.
- Zimmerly S, Hausner G, Wu X: **Phylogenetic relationships among group II intron ORFs.** *Nucleic Acids Res* 2001, **29(5)**:1238-1250.
- Turmel M, Côté V, Otis C, Mercier J-P, Gray MW, Lonergan KM, Lemieux C: **Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion).** *Mol Biol Evol* 1995, **12(4)**:533-545.
- Turmel M, Gutell RR, Mercier J-P, Otis C, Lemieux C: **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* 1993, **232(2)**:446-467.
- Turmel M, Otis C, Cote V, Lemieux C: **Evolutionarily conserved and functionally important residues in the I-CeuI homing endonuclease.** *Nucleic Acids Res* 1997, **25(13)**:2610-2619.
- Turmel M, Choquet Y, Goldschmidt-Clermont M, Rochaix JD, Otis C, Lemieux C: **The trans-spliced intron I in the *psaA* gene of the *Chlamydomonas* chloroplast: a comparative analysis.** *Curr Genet* 1995, **27(3)**:270-279.
- Turmel M, Mercier JP, Cote MJ: **Group I introns interrupt the chloroplast *psaB* and *psbC* and the mitochondrial *rnl* gene in *Chlamydomonas*.** *Nucleic Acids Res* 1993, **21(22)**:5242-5250.
- Kuck U, Choquet Y, Schneider M, Dron M, Bennoun P: **Structural and transcription analysis of two homologous genes for the**

- P700 chlorophyll a-apoproteins in *Chlamydomonas reinhardtii* evidence for in vivo trans-splicing.** *Embo J* 1987, **6(8)**:2185-2195.
39. Turmel M, Boulanger J, Schnare MN, Gray MW, Lemieux C: **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* 1991, **218(2)**:293-311.
 40. Durocher V, Gauthier A, Bellemare G, Lemieux C: **An optional group I intron between the chloroplast small subunit rRNA genes of *Chlamydomonas moewusii* and *C. eugametos*.** *Curr Genet* 1989, **15(4)**:277-282.
 41. Turmel M, Mercier JP, Cote V, Otis C, Lemieux C: **The site-specific DNA endonuclease encoded by a group I intron in the *Chlamydomonas pallidostigmatica* chloroplast small subunit rRNA gene introduces a single-strand break at low concentrations of Mg²⁺.** *Nucleic Acids Res* 1995, **23(13)**:2519-2525.
 42. Turmel M, Boulanger J, Lemieux C: **Two group I introns with long internal open reading frames in the chloroplast *psbA* gene of *Chlamydomonas moewusii*.** *Nucleic Acids Res* 1989, **17(10)**:3875-3887.
 43. Huang C, Wang S, Chen L, Lemieux C, Otis C, Turmel M, Liu XQ: **The *Chlamydomonas* chloroplast *clpP* gene contains translated large insertion sequences and is essential for cell growth.** *Mol Gen Genet* 1994, **244(2)**:151-159.
 44. Turmel M, Otis C: **The chloroplast gene cluster containing *psbF*, *psbL*, *petG* and *rps3* is conserved in *Chlamydomonas*.** *Curr Genet* 1994, **27(1)**:54-61.
 45. Majeran W, Friso G, van Wijk KJ, Vallon O: **The chloroplast ClpP complex in *Chlamydomonas reinhardtii* contains an unusual high molecular mass subunit with a large apical domain.** *Febs J* 2005, **272(21)**:5558-5571.
 46. Yamaguchi K, Prieto S, Beligni MV, Haynes PA, McDonald WH, Yates JR 3rd, Mayfield SP: **Proteomic characterization of the small subunit of *Chlamydomonas reinhardtii* chloroplast ribosome: identification of a novel S1 domain-containing protein and unusually large orthologs of bacterial S2, S3, and S5.** *Plant Cell* 2002, **14(11)**:2957-2974.
 47. Eberhard S, Drapier D, Wollman FA: **Searching limiting steps in the expression of chloroplast-encoded proteins: relations between gene copy number, transcription, transcript abundance and translation rate in the chloroplast of *Chlamydomonas reinhardtii*.** *Plant J* 2002, **31(2)**:149-160.
 48. Lilly JW, Maul JE, Stern DB: **The *Chlamydomonas reinhardtii* organellar genomes respond transcriptionally and post-transcriptionally to abiotic stimuli.** *Plant Cell* 2002, **14(11)**:2681-2706.
 49. Cui L, Leebens-Mack J, Wang L-S, Tang J, Rymarquis L, Stern DB, dePamphilis CW: **Adaptive evolution of chloroplast genome structure inferred using a parametric bootstrap approach.** *BMC Evol Biol* 2006, **6**:13.
 50. de Koning AP, Keeling PJ: **The complete plastid genome sequence of the parasitic green alga *Helicosporidium* sp. is highly reduced and structured.** *BMC Evol Biol* 2006 in press.
 51. Morton BR: **Strand asymmetry and codon usage bias in the chloroplast genome of *Euglena gracilis*.** *Proc Natl Acad Sci U S A* 1999, **96(9)**:5123-5128.
 52. Cai X, Fuller AL, McDougald LR, Zhu G: **Apicoplast genome of the coccidian *Eimeria tenella*.** *Gene* 2003, **321**:39-46.
 53. Denny P, Preiser P, Williamson D, Wilson I: **Evidence for a single origin of the 35 kb plastid DNA in Apicomplexans.** *Protist* 1998, **149**:51-59.
 54. Gardner MJ, Bishop R, Shah T, de Villiers EP, Carlton JM, Hall N, Ren Q, Paulsen IT, Pain A, Berriman M, Wilson RJ, Sato S, Ralph SA, Mann DJ, Xiong Z, Shallom SJ, Weidman J, Jiang L, Lynn J, Weaver B, Shoaibi A, Domingo AR, Wasawo D, Crabtree J, Wortman JR, Haas B, Angiuoli SV, Creasy TH, Lu C, Suh B, Silva JC, Utterback TR, Feldblyum TV, Pertea M, Allen J, Nierman WC, Taracha EL, Salzberg SL, White OR, Fitzhugh HA, Morzaria S, Venter JC, Fraser CM, Nene V: **Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes.** *Science* 2005, **309(5731)**:134-137.
 55. Wilson RJ, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW, Williamson DH: **Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*.** *J Mol Biol* 1996, **261(2)**:155-172.
 56. McCracken DA, Nadakavukaren MJ, Cain JR: **A biochemical and ultrastructural evaluation of the taxonomic position of *Glaucosphaera vacuolata* Korsch.** *New Phytol* 1980, **86**:39-44.
 57. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3)**:403-410.
 58. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16(6)**:276-277.
 59. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 1997, **25(5)**:955-964.
 60. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R: **REPuter: the manifold applications of repeat analysis on a genomic scale.** *Nucleic Acids Res* 2001, **29(22)**:4633-4642.
 61. Volfovsky N, Haas BJ, Salzberg SL: **A clustering method for repeat analysis in DNA sequences.** *Genome Biol* 2001, **2(8)**:Research0027.
 62. RepeatMasker 2001 [<http://www.repeatmasker.org>].
 63. WU-BLAST 2.0 [<http://blast.wustl.edu>]
 64. Tesler G: **GRIMM: genome rearrangements web server.** *Bioinformatics* 2002, **18(3)**:492-493.
 65. Michel F, Umesono K, Ozeki H: **Comparative and functional anatomy of group II catalytic introns—a review.** *Gene* 1989, **82(1)**:5-30.
 66. Michel F, Westhof E: **Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis.** *J Mol Biol* 1990, **216(3)**:585-610.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

