The Chloroplast Genome of the Green Alga *Schizomeris leibleinii* (Chlorophyceae) Provides Evidence for Bidirectional DNA Replication from a Single Origin in the Chaetophorales

Jean-Simon Brouard, Christian Otis, Claude Lemieux, and Monique Turmel*

Département de biochimie, de microbiologie et de bio-informatique, Université Laval, Ville de Québec, Québec, Canada

The annotated sequences of the *Schizomeris leibleinii* chloroplast genome and of the *Uronema belkai* psbF-rrs spacer have been deposited in GenBank under the accession numbers HQ700713 and HQ700714, respectively.

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

Accepted: 12 April 2011

Abstract

In the Chlorophyceae, the chloroplast genome is extraordinarily fluid in architecture and displays unique features relative to other groups of green algae. For the Chaetophorales, 1 of the 5 major lineages of the Chlorophyceae, it has been shown that the distinctive architecture of the 223,902-bp genome of *Stigeoclonium helveticum* is consistent with bidirectional DNA replication from a single origin. Here, we report the 182,759-bp chloroplast genome sequence of *Schizomeris leibleinii*, a member of the earliest diverging lineage of the Chaetophorales. Like its *Stigeoclonium* homolog, the *Schizomeris* genome lacks a large inverted repeat encoding the rRNA operon and displays a striking bias in coding regions that is associated with a bias in base composition along each strand. Our results support the notion that these two chaetophoralean genomes replicate bidirectionally from a putative origin located in the vicinity of the small subunit ribosomal RNA gene. Their shared structural characteristics were most probably inherited from the common ancestor of all chaetophoralean algae. Short dispersed repeats account for most of the 41-kb size variation between the *Schizomeris* and *Stigeoclonium* genomes, and there is no indication that homologous recombination between these repeated elements led to the observed gene rearrangements. A comparison of the extent of variation sustained by the *Stigeoclonium* and *Schizomeris* chloroplast DNAs (cpDNAs) with that observed for the cpDNAs of the chlamydomonadalean *Chlamydomonas* and *Volvox* suggests that gene rearrangements as well as changes in the abundance of intergenic and intron sequences occurred at a slower pace in the Chaetophorales than in the Chlamydomonadales.

Key words: plastid genome evolution, chloroplast gene rearrangements, introns, replication origin, short dispersed repeats, Chlamydomonadales.

Introduction

The chloroplast genome is evolving in a dynamic fashion in the chlorophyte class Chlorophyceae (Chlorophyta). Comparative analyses of the complete chloroplast genome sequences from *Chlamydomonas reinhardii* (Chlamydomonadales), *Scenedesmus obliquus* (Sphaeropleales), *Stigeoclonium helveticum* (Chaetophorales), *Oedogonium cardiacum* (Oedogoniales), and *Floydiella terrestris* (Chaetopeltidales) revealed that each of the five orders recognized in this green algal class is characterized by a distinctive genome architecture (Maul et al. 2002; Bélanger et al. 2006; de Cambiaire et al. 2006; Brouard et al. 2008, 2010). In addition to uncovering the dynamic evolution of the chlorophycean chloroplast genome, these comparative studies resolved the branching order of the five chlorophycean lineages and allowed the reconstruction of a hypothetical scenario depicting the origins of some of the observed genomic changes (Brouard et al. 2010). Two major clades were identified, the CS clade (Chlamydomonadales and Sphaeropleales) and the OCC clade (Oedogoniales, Chaetopeltidales, and Chaetophorales), and the Oedogoniales was
identified as an early diverging branch of the OCC clade (Turmel et al. 2008; Brouard et al. 2010). These phylogenetic results received strong independent support from structural characters at the levels of gene content, intron content, gene order, and gene structure. More recently, the collection of chlorophycean chloroplast genomes was further enriched with the nearly complete chloroplast DNA (cpDNA) sequence of Volvox carteri (Chlamydomonadales) (Smith and Lee 2009) and the complete cpDNA sequence of Dunaliella salina (Chlamydomonadales) (Smith et al. 2010).

Among all completely sequenced green algal genomes, those of chlorophycean green algae display the largest sizes (size ranges from 161 kb in Scenedesmus to 521 kb in Floydiella) and the lowest retention of ancestral genomic features (Maul et al. 2002; Bélanger et al. 2006; de Cambiaire et al. 2006; Brouard et al. 2008; Smith and Lee 2009; Brouard et al. 2010; Smith et al. 2010). The gene repertoire of chlorophycean genomes (94–100 genes) is consistently reduced relative to streptophyte (chlorophycean and land plant) and other chlorophycean cpDNAs, with all chlorophycean cpDNAs lacking eight conserved genes (accD, chlL, minD, psal, rpl19, ycf20, ycf62, and trnR(ccg)) relative to the two ulvophycean cpDNAs sequenced thus far (Pombert et al. 2005, 2006). Moreover, chlorophycean genomes have retained a very limited number of ancestral gene clusters from the bacterial progenitor of primary chloroplasts. Like most green plant cpDNAs, five of the completely sequenced chlorophycean genomes (those of the Chlamydomonadales, Sphaeropleales, and Oedogoniales) exhibit two copies of a large inverted repeat (IR) sequence separated by single-copy regions; however, gene contents of the single-copy regions differ remarkably between distinct chlorophycean orders and strongly deviate from the ancestral patterns observed for the prasinophyceans Nephrosemis and Pyramimonas, the trebouxiophyceans Pedinomonas, Parachlorella and Oocystis, and all streptophytes having an IR in their chloroplast genome (Turmel, Otis, and Lemieux 1999, 2009; de Cambiaire et al. 2006; Brouard et al. 2008; Turmel, Gagnon, et al. 2009). Several alterations at the level of gene structure, notably the fragmentation of the rpoB gene in two separate open reading frames (ORFs) and the expansion of the coding regions of clpP rps3, and rpoC1 (Bélanger et al. 2006; de Cambiaire et al. 2006; Brouard et al. 2008, 2010) distinguish all chlorophycean cpDNAs from their homologs in the Chlorophyta. In contrast, genomic changes such as the breakup of four genes (petD, psaA, psaC, rbcL) by putatively trans-spliced group II introns and the fragmentation of rpoC1 and rps2 arose following the emergence of specific lineages (Belanger et al. 2006; de Cambiaire et al. 2006; Brouard et al. 2008, 2010). Unlike most of their chlorophyte counterparts, several chlorophycean cpDNAs are intron rich, although the distribution and abundance of both group I and group II introns are highly variable. Similarly, the prevalence of short dispersed repeats in the chloroplast genomes of Chlamydomonas, Volvox, Stigeoclonium, and Floydiella contrasts sharply with the low frequency observed for such sequences in the Scenedesmus and Oedogonium chloroplasts.

The extremely high variability in chloroplast genome architecture observed among the chlorophyceans examined so far emphasizes the need for sampling additional taxa in order to document the extent of changes found within each order, validate the evolutionary scenario inferred by Brouard et al. (2010), and better understand the evolutionary forces shaping the genome. The focus of the present study is on the Chaetophorales and Chlamydomonadales. The chloroplast genome of Schizomeris leibleini, a representative of the earliest diverging lineage (Schizomeridaceae) of the Chaetophorales (Buchheim et al. 2001; Caisova et al. 2011) was fully sequenced and compared with that of its distant chaetophoralean relative S. helveticum (which belongs to a clade containing branched taxa of the Chaetophoraceae) as well as with other previously sequenced chloroplast genomes from the Chlorophyceae, including those of the chlamydomonadalean algae Chlamydomonas and Volvox.

The IR-lacking chloroplast genome of Stigeoclonium is rich in introns and dispersed short repeats and is presumed to share a common loss of the IR with the cpDNA of Floydiella (Chaetopeltidales) (Brouard et al. 2010). It displays a remarkable pattern of gene distribution in which genes on one half of the genome are encoded by the same strand and those on the other half are encoded by the alternative strand (Bélanger et al. 2006). As found in prokaryotic genomes that replicate bidirectionally from a single origin (Grigoriev 1998; Tillier and Collins 2000a, 2000b; Guy and Roten 2004), this strand bias in coding regions is closely associated with a bias in GC composition along each strand. Analysis of the cumulative GC skew has proven useful to identify the origin and terminus of replication in prokaryotic genomes (Grigoriev 1998; Guy and Roten 2004), and the application of this method to the Stigeoclonium chloroplast genome disclosed a putative replication origin in the trnS(gca)-rrs intergenic region and a putative terminus in the psbD-tufA intergenic region (Bélanger et al. 2006).

We report here that, although the Schizomeris genome is more compact and substantially rearranged relative to the Stigeoclonium cpDNA, it also lacks an IR and, as revealed by a cumulative GC skew analysis, displays a putative origin of replication in the same region. The two chaetophoralean genomes share very similar noncoding sequences in this region, supporting their putative role as an origin of replication and as anticipated, these sequences are also conserved at the corresponding cpDNA locus in the unbranched alga Uronema belkæ (Chaetophoraceae). Moreover, our comparative analyses of the two available pairs of closely related chlorophycean cpDNAs suggests that the chloroplast genome...
evolved more conservatively in the Chaetophorales than in the Chlamydomonadales.

Materials and Methods

Strains and Culture Conditions

*Schizomeris leblei*ni (UTEX LB 1228) and *U. belkae* (UTEX 1179) were obtained from the Culture Collection of Algae at the University of Texas at Austin and were grown in C medium (Andersen et al. 2005) and the modified *Volvox* medium (McCracken et al. 1980), respectively. Both cultures were subjected to alternating 12-h light–dark periods.

Cloning and Sequencing of the *Schizomeris* Chloroplast Genome

An A + T-rich organellar DNA fraction was obtained by CsCl-bisbenzimide isopycnic centrifugation as described earlier (Turmel et al. 1999). This DNA fraction was sheared by nebulization to produce 2,000- to 4,000-bp fragments that were subsequently cloned into the pSMART-HKCan plasmid (Lucigen Corporation). Positive clones were selected by hybridization of the plasmid library with the original DNA used for cloning. DNA templates were amplified using the TempliPhi Amplification Kit (GE Healthcare) and sequenced as described previously (Turmel et al. 2005). The DNA sequences were edited and assembled using SEQUENCER 4.7 (Gene Codes Corporation). Genomic regions absent in the clones analyzed were directly sequenced from polymerase chain reaction (PCR)-amplified fragments using internal primers. Alternatively, PCR-amplified fragments were subcloned using the TOPO TA cloning kit (Invitrogen) before sequencing.

Amplification of the Putative Replication Origin of *Uronema* cpDNA

We wished to determine whether the conserved non-coding sequences we identified in the cpDNA regions containing the putative origins of replication of *Schizomeris* and *Stigeoclonium* are also found at the corresponding loci in the chaetophoralean *Uronema*. We thus amplified by PCR the spacer between the psbF and *rrs* genes of *Uronema* using total cellular DNA and the following primers: 5'-'CGTTGAAATAATTGCATAGCAG-3' and 5'-'CAACTAGC-TAATCAGACCAGA-3'. The resulting fragment of 4,385 bp was sequenced using the latter primers as well as internal primers.

Sequence Analyses

Genes and ORFs were identified as described previously (Brouard et al. 2010). Boundaries of introns were located by modeling intron secondary structures (Michel et al. 1989; Michel and Westhof 1990) and by comparing the sequences of intron-containing genes with those of intronless homologs using diverse programs of the FASTA package (Pearson and Lipman 1988; Pearson et al. 1997). Regions of the genome sequence containing nonoverlapping repeated elements were mapped with REPEATMASKER (http://www.repeatmasker.org) running under the WU-BLAST 2.0 search engine (http://blast.wustl.edu), using the repeats ≥30 bp identified with REPFIND of the REPuter 2.74 program (Kurtz et al. 2001) as input sequences.

The sidedness index (Cs) was calculated as described by Cui et al. (2006) using the formula Cs = (n − nSB)/(n − 1), where n is the total number of genes in the genome and nSB is the number of sided blocks, that is, the number of blocks including adjacent genes on the same strand. The strand bias in base composition was calculated for the whole genome. For the entire genome sequence, the GC skew, that is, 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 150,000–155,000 and continuing by shifting 500 nucleotides downstream along the strand for each new window.

Homology searches in intergenic regions were performed with the BlastN tool of the NCBI web server against the nucleotide collection (nr/nt) using the Schizomeris intergenic sequences as queries and the default parameters.

Gene Order Analyses

Gene pairs exhibiting identical gene polarities in *Schizomeris* and other chlorophycean cpDNAs were identified using a custom-built program. The GRIMM web server (Tesler 2002) was used as described previously (Bélangier et al. 2006) to infer the minimal number of gene permutations by inversions in all possible pairwise comparisons of chlorophycean cpDNAs. The data set used in the latter analyses consisted of 93 genes; note that pieces of the *rpoB* gene and all exons of the genes containing trans-spliced introns were coded as distinct fragments (for a total of 98 gene loci). Because GRIMM requires that the compared genomes have exactly the same gene content, duplicated genes, including those within 1 of the 2 copies of the IR, as well as genes lacking in some of the genomes analyzed were excluded from the data set and the genomes were considered as linear molecules. In the case of the duplicated *trnS*(gcu) gene in the *Schizomeris* genome, the copy being part of the gene pair shared with the *Stigeoclonium* genome (*trnS*(gcu)-ycf1) was retained for analysis. The data set used for inferring the scenario of gene rearrangements between the IR-lacking *Schizomeris* and *Stigeoclonium* cpDNAs was constructed in the same manner and comprised 99 genes, for a total of 104 gene loci. In this analysis, the two compared genomes were considered as circular molecules.
Phylogenetic Analyses of Sequence Data

The deduced amino acid sequences of individual chloroplast protein-coding genes shared by eight chlorophyceans (for names of taxa and accession numbers of chloroplast genome sequences, see footnote of table 1) and the ulvophycean *Phycomonas* (for names of taxa and accession numbers of chloroplast protein-coding genes shared by eight chlorophyceans) were obtained using the amino acid alignment tool Muscle 3.7 (Edgar 2004). A total of 61 genes met this criterion: atpA, B, E, F, H, I, ccsA, cemA, cpIP, ftsH, petB, D, G, L, psaA, B, C, J, psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z, rbcL, rpl2, 5, 14, 16, 20, 23, 36, rpoA, B, C1, C2, rps2, 3, 4, 7, 8, 9, 11, 12, 14, 18, 19, tufA, ycf1, 3, 4, 12. The amino acid alignments were then converted into alignments of codons. The poorly aligned and divergent regions in each codon alignment were removed using GBLOCKS 0.91b (Castresana 2000) and the option --f = c for exclusion of complete codons. After concatenation of the codon alignments, a maximum likelihood tree was inferred from the resulting data set using Treefinder (version of October 2008) and the GTR + I4 model of nucleotide substitutions. Evolutionary distances were estimated with CODEML in PAML (Yang 1997, 2007) using the same data set and the tree generated by Treefinder as a constraint. The analysis was run using all default parameters, except that the free-ratio model was selected to allow the ratio of synonymous and nonsynonymous substitution rates to vary along each branch in the tree.

Data Deposition

The annotated sequences of the *S. leibleinii* chloroplast genome and of the *U. belkae* psbF-rrs spacer have been deposited in GenBank under the accession numbers HQ700713 and HQ700714, respectively.

Results and Discussion

The Schizomeris Chloroplast Genome Shares a Distinctive Architecture with the Stigeoclonium Genome

At 182,759 bp, the circular-mapping chloroplast genome of *Schizomeris* is 41-kb smaller than the *Stigeoclonium* genome (fig. 1). Table 1 summarizes the general features of the *Schizomeris* genome and compares them with those of other chlorophycean cpDNAs. Like its homologs in *Stigeoclonium* and the more distantly related *Floydiella*, the *Schizomeris* chloroplast genome lacks an IR, thus strengthening the hypothesis that the IR was lost before the divergence of the Chaetophorales and the Chaetopeltidales (Brouard et al. 2010). The repertoire of conserved
chloroplast genes in Schizomeris only differs from that of Stigeoclonium by the presence of trnT(ggu) (supplementary table S1, Supplementary Material online); other distinguishing features of the Schizomeris genome include the complete duplication of the trnS(gcu) and partial duplication of psaA. In the supplementary section S1 of the Supplementary Material online, we discuss about the possible origin of trnT(ggu) and report other cases of gene duplications in green plant chloroplast genomes. The Schizomeris genome contains 12 extra introns compared with its Stigeoclonium counterpart (table 1). Nine of the 33 introns it displays (for the characteristics of all introns, see supplementary table S2 and section S2, Supplementary Material online) occur at positions that have not been reported so far in green plants. An updated compilation of known insertion sites for group I introns in completely sequenced chlorophyte genomes is presented in supplementary figure S1 (Supplementary Material online). In contrast to the Stigeoclonium and Floydiella genomes (table 1), the intergenic regions of the Schizomeris cpDNA display a small proportion of short dispersed repeats (see supplementary section S3 and fig. S2, Supplementary Material online).
The Schizomeris genome resembles its Stigeoclonium counterpart in displaying a striking bias in the distribution of genes between the two DNA strands. In the Schizomeris cpDNA, all the conserved genes in the 120-kb segment extending from trnS(gga) to trnG(gcc) are located on the same strand, whereas most of the remaining genes are encoded on the opposite strand (fig. 1). Two large blocks of genes encoded by distinct strands are also observed in the Stigeoclonium cpDNA; however, each block occupies approximately one half of the genome (fig. 1). The sidedness index (Cs) is a helpful parameter to estimate the propensity of adjacent genes to be located on the same DNA strand in a given genome (Cui et al. 2006). The calculated values for the Schizomeris and Stigeoclonium cpDNAs (Cs = 0.97 and Cs = 0.95, respectively) are substantially higher than those reported for other chlorophycean cpDNAs (table 1).

A Bidirectional Mode of cpDNA Replication from a Single Origin Appears to Have Been Maintained in the Chaetophorales

We investigated whether the Schizomeris genome resembles its Stigeoclonium counterpart in exhibiting a bias in GC composition along each strand. As shown in figure 2A, the plot of the cumulative GC skew in the Schizomeris genome has a V-shape, with the minimum and maximum lying on opposite ends of the genome. This profile is similar to those reported for the Stigeoclonium cpDNA and prokaryotic genomes that replicate bidirectionally from a single origin. The putative origin of replication in the Schizomeris cpDNA (corresponding to the minimum) was localized in the psbF-trnS(gga) spacer, a site that coincides with a switch in coding strand and actually marks one of the boundaries of the block containing the 62 consecutive genes on the same strand (figs. 1 and 2A). According to our cumulative GC skew analysis, the putative terminus of replication (corresponding to the maximum) resides in the psbD-trnS(gcu) spacer located in the 62-gene block (fig. 2A). This site is 12.5 kb away from the switch in coding strand observed between trnG(gcc) and the first exon of petD (fig. 1).

The putative origin and terminus of replication in the Schizomeris genome map at almost the same positions as those identified in the Stigeoclonium genome (Bélanger et al. 2006). The putative origin in the Stigeoclonium cpDNA was located in the trnS(gga)-rrs spacer, that is, in the spacer immediately adjacent to that carrying the putative origin in the Schizomeris cpDNA. In this context, it should be pointed out that the putative origins in the Schizomeris and Stigeoclonium genomes lie in a region of gene synteny, with the exception that the trnS(gga) gene differs in polarity (fig. 1). Given the relatively high degree of synteny reported here for the two chaetophoralen genomes (see below), the inversion responsible for the change in polarity of trnS(gga) might account for our observation that the Schizomeris and
Stigeoclonium origins map to distinct spacers. Like its Schizomeris counterpart, the terminus of replication in the Stigeoclonium cpDNA was mapped to the spacer downstream of psbD. Again here, it is obvious that this region was implicated in genomic rearrangements: in Schizomeris, trnG(gcc), and psbD lie on opposite borders of the syntenic block encoding trnS(gcau), ycf1, and tufA (block 6 in fig. 1), whereas in Stigeoclonium, these two genes are next to each other just beside the latter syntenic block. These rearrangements, which apparently included the inversion of the region containing the block 6 and trnG(gcau), might explain why the switch in coding strand associated with the replication terminus corresponds to the second highest peak in the cumulative GC skew diagram of the Schizomeris genome (fig. 2A). In support of this hypothesis, disruptions of linearity appearing at local minima and maxima in GC skew analysis of prokaryotic genomes have been proposed to represent recent genome rearrangements (Grigoriev 1998).

In an attempt to identify conserved motifs that may be involved in the replication of the Schizomeris and the Stigeoclonium cpDNAs, we performed BlastN searches against the NCBI database using Schizomeris intergenic sequences as queries. Only 6 of the 105 Schizomeris intergenic spacers (psbB-psbT, rps8-psbE, rpoBex2-rpl32, rps2-clpP, psbF-trnS(gcau), and trnS(gcau)-trs) displayed regions of significant homology (E values ranging from 7 × 10^{-5} to 8 × 10^{-17}) with known sequences. As expected, all significant hits matched with the Stigeoclonium genome. More importantly, our searches revealed that three short Schizomeris noncoding sequences (S1, S2, and S3) in the region extending from psbF to trs have counterparts in the Stigeoclonium cpDNA (fig. 2B). The 71-bp S1 is situated 247-bp upstream from psbF, whereas the 138-bp S2, also found within the psbF-trnS(gcau) spacer, is 1,321-bp away from S1 and 327-bp upstream from trnS(gcau). These two sequence elements represent the best candidates for a replication origin. Considering that the 187-bp S3 is located 40-bp upstream from trs in both genomes, it would seem more compatible with a role in the transcriptional regulation of the ribosomal RNA operon. In contrast, no putative termination site could be detected in the intergenic spacers corresponding to the maximum or second highest peaks in our cumulative GC skew analysis. All other intergenic sequences that were found to be conserved in our BlastN analyses are possibly involved in transcriptional or translational regulation.

To confirm that the above-mentioned noncoding sequence elements shared by Schizomeris and Stigeoclonium are also conserved in other chaetophoraleans, we sequenced the region spanning psbF and trs in U. belkae, a member of the clade comprising the unbranched taxa of the Chaetophoraceae (Caisova et al. 2011). As anticipated, all three noncoding elements were retrieved at the same relative positions in Uronema; however, the trnS(gcau) gene was found to be missing from the region analyzed. The conservation of S1, S2, and S3 in the Uronema psbF-trs spacer further supports the functional importance of these sequences and strengthens the above conclusions regarding the potential role of S1 and S2 in the initiation of DNA replication.

Our finding that the Stigeoclonium and Schizomeris cpDNAs share a putative replication origin in the vicinity of the ribosomal RNA operon mirrors the situation observed for the two other IR-lacking chloroplast genomes that were proposed to replicate bidirectionally from a single origin: the genome of the euglenoid Euglena gracilis (Morton 1999) and the genome of the parasitic green alga Helicosporidium sp. (Trebouxixophyceae) (de Koning and Keeling 2006). In the case of Euglena, it was shown that the replication origin inferred by the GC skew analysis (Morton 1999) matched the single locus previously identified by electron microscopic analysis of replication intermediates (Koller and Delius 1982; Ravel-Chapuis et al. 1982). In contrast, all other chloroplast genomes whose replication origin was studied by diverse biochemical methods were found to contain multiple origins either in the large IR, close to the ribosomal RNA operon, as in Pismum sativum (Meeker et al. 1988), Nicotiana tabacum (Kunnimalaiyan and Nielsen 1997a, 1997b), and Oenethera hookeri (Chiu and Sears 1992; Sears et al. 1996) or in the single-copy regions as in Chlamydomonas where oriA and oriB lie in the vicinity of rpl16 and chlL, respectively (Waddell et al. 1984; Chang and Wu 2000).

According to the standard model of cpDNA replication, which is based on electron microscopic studies of pea and maize cpDNA molecules, replication starts at two displacement loop initiation sites located about 7-kb apart on opposite strands (Kolodner and Tewari 1975). The displacement loops expand unidirectionally toward each other to form a Cairns (theta)-type structure, which expands bidirectionally until completion of two daughter molecules. Replication may then continue by a rolling circle mechanism (Heinhorst and Cannon 1993; Day and Madesis 2007). Krishnan and Rao (2009) recently tested the standard model of replication by investigating the accumulation of the adenine to guanine deaminations on the displaced single strands. In accordance with this model, local deamination gradients in virtually all the analyzed land plant cpDNAs were shown to increase bidirectionally from the center of each single-copy region toward the pairs of replication origins localized within the IR. In this study, the authors also reported on the basis of Blast searches that most completely sequenced chloroplast genomes bear at least one or more homologs of the replication origins found in the tobacco chloroplast genome. In our opinion, the Blast signal detected in chlorophyte genomes could simply reflect the conservation of the trnI(gcau) and orf35(yfc1) sequences that overlap the replication origins in the N. tabacum cpDNA.
Chloroplast genomes were treated as circular molecules using a data set of 104 gene loci. Here, the two IR-lacking yielded 20 and 24 inversions, respectively (table 2). Chlamydomonas belong to the chlamydomonadaleans, Stigeoclonium belongs to the chaetophoraleans and longing to different orders are compared, analyses involving other genomes. Although more than 80 inversions are required to convert the gene order of a given genome to those that is, the minimal number of inversions that would be re-...
genomes promote recombination and that we have failed to recognize this role because the *Schizomeris* and *Stigeoclonium* cpDNAs show extraordinary differences in their repeated sequences. Examination of chloroplast genomes from very closely related chlorophyceans will be needed to better understand how they evolve and whether they play a role in gene rearrangements.

**Mode and Tempo of Chloroplast Genome Evolution in the Chaetophorales and Chlamydomonadales**

The availability of cpDNA sequences from two pairs of closely related chlorophycean algae (*Chlamydomonas* and *Volvox* representing the Chlamydomonadales and *Schizomeris* and *Stigeoclonium* representing the Chaetophorales) offers the opportunity to test whether the chloroplast genome evolves under similar constraints in the OCC and CS lineages. The histograms shown in figure 3 compare the lengths of coding and noncoding (introns, repeats, and intergenic spacers) sequences in the four chlorophycean genomes. Clearly, the major change in size that the chloroplast genome experienced in the Chaetophorales is accounted for by dispersed repeats. In contrast, the chlamydomonadalean genomes sustained changes not only in the abundance of repeats but also in the abundance of intron and intergenic sequences. The higher variability observed for the Chlamydomonadales at the level of noncoding sequences is somewhat intriguing considering that the evolutionary distance between *Chlamydomonas* and *Volvox* (12,701 substitutions), as estimated by the numbers of accumulated synonymous and nonsynonymous substitutions in 61 protein-coding genes, is 1.6-fold smaller than that separating *Schizomeris* and *Stigeoclonium* (20,211 substitutions) (fig. 3). Interestingly, the higher variability in the abundance of noncoding sequences is also accompanied by a higher rate of genome rearrangements in the Chlamydomonadales. Using evolutionary distances as a means to normalize the number of genome reversals in the two pairs of related algae, we estimated that the *Chlamydomonas/Volvox* pair experienced almost twice as many inversions per distance unit (1.9 inversions/1,000 nucleotide substitutions) compared with the *Schizomeris/Stigeoclonium* pair (1.0 inversion/1,000 nucleotide substitutions) (fig. 3).

At this point, we can offer no compelling explanation as to why the chloroplast genome evolves more conservatively at the levels of gene order and genome size in the Chaetophorales than in the Chlamydomonadales. The presence/absence of a stabilizing IR structure cannot be invoked because the more highly conserved genomes found in the chaetophoralean lineage both lack an IR. Considering that short dispersed repeats can be gained or lost rapidly over evolutionary time, perhaps such repeats were more abundant in ancestral genomes of the Chlamydomonadales compared with those of the Chaetophorales, thus favoring a higher level of recombination in the former lineage.

**Conclusions**

By analyzing the chloroplast genome from a representative of the earliest diverging lineage of the Chaetophorales, we have gained insights into the ancestral structure and evolutionary dynamics of cpDNA in this major group of the Chlorophyceae. The *Schizomeris* genome was found to resemble its *Stigeoclonium* homolog with respect to the absence of
the IR and the presence of a strong bias in coding strand and nucleotide composition along each DNA strand. This distinctive architecture probably represents the ancestral condition of the chaetophoralean chloroplast genome and is most likely the result of evolutionary pressures to retain a bidirectional mode of DNA replication from a single origin. The conserved motifs we uncovered in the vicinity of the ribosomal RNA operon may play a role in the initiation of DNA replication. Our study also revealed that short dispersed repeats evolve at a rapid pace in the genomic landscape of the Chlorophyceae; however, the chaetophoralean genomes appear to be more conserved in gene order and display less variation in the abundance of intergenic and intron sequences than their chlamydomonadalean homologs. Clearly, chloroplast genome sequences from additional representatives of the OCC and CS lineages are needed to identify the factors underlying the tremendous plasticity of the chloroplast genome in the Chlorophyceae.

Supplementary Material

Supplementary figures S1–S3, tables S1 and S2, and sections S1–S3 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This study was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (to M.T. and C.L.).

Literature Cited


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.


**Associate editor:** Geoff McFadden