Extensive Gene Rearrangements in the Chloroplast DNAs of Chlamydomonas Species Featuring Multiple Dispersed Repeats

Eric Boudreau and Monique Turmel

Program in Evolutionary Biology, Canadian Institute for Advanced Research, Département de biochimie, Faculté des sciences et de génie, Université Laval, Québec, Quebec, Canada. E-mail: mturmel@bcm.ulaval.ca.

We have constructed a physical and gene map for the chloroplast DNA (cpDNA) of the unicellular green alga Chlamydomonas gelatinosa, a close relative of Chlamydomonas reinhardtii. At 285 kb, the C. gelatinosa cpDNA is 89 kb larger than its C. reinhardtii counterpart. The alterations in the order of ORF(s) genes on the cpDNAs of these green algae are attributable to nine inversions and one event of expansion/contraction of the inverted repeat. These rearrangements are much more extensive than those previously reported between the cpDNAs of the closely related Chlamydomonas moewusii and Chlamydomonas pitschmannii. Because the divergence level of the C. gelatinosa and C. reinhardtii chloroplast-encoded large subunit rRNA gene sequences is equivalent to that of the corresponding C. moewusii and C. pitschmannii sequences, our results may suggest that, in the same period of time, there have been more numerous rearrangements in the lineage comprising C. gelatinosa and C. reinhardtii than in the lineage comprising C. moewusii and C. pitschmannii. Alternatively, given that substitution rates in chloroplast genes are not necessarily uniform across lineages, the extensive rearrangements between the C. gelatinosa and C. reinhardtii cpDNAs may reflect a longer divergence period for this pair of Chlamydomonas species compared to that for the C. moewusii/C. pitschmannii pair. We have also found that, like its C. reinhardtii homologue but unlike its C. moewusii and C. pitschmannii counterparts, the C. gelatinosa cpDNA features a large number of dispersed repeated sequences that are readily detectable by Southern blot hybridization with homologous fragment probes. Assuming that the two pairs of closely related Chlamydomonas species diverged at about the same time, these data suggest that the susceptibility of Chlamydomonas cpDNAs to rearrangements is correlated with the abundance of repeated sequences. Preliminary characterization of a 345-bp C. gelatinosa cpDNA region containing a repeated sequence by both DNA sequencing and Southern blot analysis has revealed no sequence homology between this region and the cpDNAs of C. reinhardtii and other Chlamydomonas species.

Introduction

Evolution of the chloroplast genome is very conservative in land plants (Palmer 1991; Downie and Palmer 1992), with alterations in both order and genome complexity being generally infrequent. Inversions of long sequences and/or events of contraction/expansion of the large inverted repeat account for most of the gene rearrangements observed: either one inversion or three inversions were identified relative to the ancestral form in the majority of rearranged cpDNAs from land plants. Complex cases of structural rearrangements have been described in only five groups of species: two lineages of legumes (Palmer et al. 1987; Palmer, Osorio, and Thompson 1988; Milligan, Hampton, and Palmer 1989), conifers (Strauss et al. 1988), geraniums (Palmer, Nugent, and Herbon 1987), Lobeliaceae (Knox, Downie, and Palmer 1993), and Anemone (Hoot and Palmer 1994). All of the gene rearrangements, except for two events in legumes, have not disrupted any of the groups of genes that were found to be cotranscribed as operons in angiosperms (Palmer, Osorio, and Thompson 1988; Milligan, Hampton, and Palmer 1989). Several of these operons are considered to be evolutionarily primitive as they are present in identical or similar form in bacteria, especially cyanobacteria (Palmer 1991). Although the mechanistic basis for cpDNA inversions remains unknown, it is thought that these events are caused by homologous or illegitimate recombination between repeated sequences (Palmer 1991). Such sequences (>50 bp) are particularly abundant in highly rearranged cpDNAs from land plants. While homologous recombination between inversely oriented repeat sequences is known to yield inversion of the intervening DNA segment, the exact relationships between inversion endpoints and repeated sequences in the chloroplast genome have been little investigated. Some inversion endpoints have also been found to be associated with short, repeated sequences near or within tRNA genes (Palmer 1991; Knox, Downie, and Palmer 1993; Hoot and Palmer 1994).

Although the chloroplasts of green algae and land plants share a common origin, their genomes appear to have followed very different evolutionary pathways. The few green algal cpDNAs that have been studied...
(from the Chlorophyceae, Charophyceae, and Ulvophyceae) are structurally similar to their land plant counterparts; however, they are highly variable in size (89–234 kb compared to 120–160 kb for the majority of land plants), gene organization, and genome complexity, and none of them appears to show strong similarity with the consensus gene order found in vascular plant cpDNAs (reviewed by Palmer 1991). The majority of evolutionarily primitive operons present in land plant cpDNAs were lost from some lineages of green algae. For example, in the polyphyletic genus *Chlamydomonas* (Chlorophyceae), 40 of the approximately 76 genes that were mapped on the cpDNAs of five species (the interfertile species *C. eugametos* [292 kb]/ *C. moewusii* [243 kb] and *C. reinhardtii* [196 kb]/ *C. smithii* [195 kb] as well as *C. pitschmannii* [187 kb]) constitute 15 conserved gene clusters, only 5 of which share similarity to primitive operons found in land plant cpDNAs (Boudreau, Otis, and Turmel 1994; Boudreau and Turmel 1995).

It is of interest to study the mode and tempo of cpDNA evolution in the order Chlamydomonadales considering that the cpDNAs of *Chlamydomonas* display considerable structural evolution and that the relationships among representatives of this large genus and other genera from the Chlamydomonadales are well documented. Comparative analyses of chloroplast-encoded large-subunit rRNA sequences from 31 flagellate taxa of the Chlamydomonadales, including 23 *Chlamydomonas* species representing all of the 15 sporangial wall autolytic groups sensu Schlösser (1984) and 8 of the 9 morphological groups sensu Ettl (1976), provide strong support for two major lineages that include taxa from the biflagellate genus *Chlamydomonas* and a basal lineage that comprises taxa from the quadriflagellate genus *Carteria* (Turmel et al. 1993, 1995a; Buchheim et al. 1995). The *Chlamydomonas* species that have been previously examined for their cpDNA include representatives of each of the two major Chlamydomonad lineages: *C. eugametos*, *C. moewusii*, and *C. pitschmannii* fall within the same lineage, while *C. reinhardtii* and *C. smithii* belong to the other Chlamydomonad lineage. The cpDNAs of *C. eugametos*, *C. moewusii*, and *C. reinhardtii* have accumulated so many gene rearrangements during their evolution that the nature of the mutational events responsible for these rearrangements could not be deciphered (Boudreau, Otis, and Turmel 1994). In contrast, in each of the two distantly related pairs of interfertile species *C. eugametos*, *C. moewusii* and *C. reinhardtii*, *C. smithii*, the cpDNAs are colinear, except for a large number of deletions/additions exceeding 100 bp (Turmel, Bellemare, and Lemieux 1987; Boynton et al. 1992). Recently, our comparative analysis of cpDNAs in the closely related *C. moewusii* and *C. pitschmannii* has revealed that the same kinds of gene rearrangements occur in Chlamydomonad and land plant cpDNAs. These two *Chlamydomonas* cpDNAs are little rearranged, featuring only one or two inversions and possibly one or three event(s) of expansion/contraction of the inverted repeat (Boudreau and Turmel 1995). Considering that the *C. eugametos*, *C. moewusii* and *C. pitschmannii* cpDNAs contain no dispersed repeated sequences that are detectable by Southern blot hybridization with fragment probes (Lemieux et al. 1985, Turmel, Bellemare, and Lemieux 1987; Boudreau and Turmel 1995) and that such sequences are readily detected in the *C. reinhardtii* and *C. smithii* cpDNAs (Rochaix 1978; Boynton et al. 1992), the possibility is raised that chloroplast genes rearrange at different rates in the two major Chlamydomonad lineages. To gain insight into this question and learn more about the patterns of gene rearrangements in Chlamydomonad cpDNAs, we have investigated the chloroplast gene organization of *C. gelatinosa*, a close relative of *C. reinhardtii* which is separated from this green alga by an evolutionary distance equivalent to that between *C. moewusii* and *C. pitschmannii*, as estimated from the divergence of the chloroplast-encoded large subunit rRNA gene sequences.

### Materials and Methods

**Isolation of cpDNA and Southern Blot Hybridizations**

cpDNA-enriched fractions from *C. gelatinosa* (SAG 69.72), *C. reinhardtii* (SAG 11-32b) and seven other *Chlamydomonas* species (*eugametos* [UTEX 9], *moewusii* [UTEX 97], *pitschmannii* [SAG 14.73], *humicola* [SAG 11-9], *peteri* [SAG 70.72], *iyengari* [SAG 25.72], *frankii* [SAG 19.72]) were prepared as described by Turmel et al. (1993). These DNAs were digested with various restriction endonucleases (Sambrook, Fritch, and Maniatis 1989) and transferred onto Hybond-N nylon membranes (Amersham, Arlington Heights, Ill.) (Sambrook, Fritch, and Maniatis 1989). The DNA blots were hybridized with cpDNA fragments and oligonucleotides as described previously (Boudreau and Turmel 1995). All of the gene probes except four were described previously (Boudreau, Otis, and Turmel 1994; Boudreau and Turmel 1995). These four probes include one protein-coding gene (*C. eugametos* ORFA [Boudreau, Otis, and Turmel 1994], equivalent to *C. reinhardtii* ORFS [EMBL/GenBank accession no. X76934]; this probe was produced by PCR) and three *C. reinhardtii* tRNA genes (trnH GUG, trnR UCU, and trnT UGU [X13879 and X75037]; these probes were gene-specific oligonu-
cleotides), and the details about their construction are available upon request from the authors.

Cloning

The C. gelatinosa cpDNA-enriched preparation was digested singly with the EcoRI and PstI restriction endonucleases and was ligated with the plasmid vector pBluescript KS+ (Stratagene, La Jolla, Calif.) which had been similarly cut. The EcoRI-digested plasmid was also dephosphorylated before ligation. Competent cells of Escherichia coli strain DH5α (Bethesda Research Laboratories, Gaithersburg, Md.) were transformed with the pBluescript plasmid containing the 2.4-kb PstI fragment was partly sequenced by the dideoxy chain termination method using the T7 sequencing kit of Pharmacia (Uppsala, Sweden) and a double-stranded DNA template. The sequencing reaction for one of the DNA strands was initiated with the T3 primer (Stratagene, La Jolla, Calif.), whereas that for the other strand was initiated with the synthetic oligonucleotide 5’-GOT-AATGTGTTACATTCTA-3’. Sequence analysis was carried out with the Genetics Computer Group (1991) software package. A database search for homology with the sequence determined was conducted at the National Center for Biotechnology Information using the BLAST network service.

DNA Sequencing and Analysis

The pBluescript plasmid containing the 2.4-kb PstI fragment was partly sequenced by the dideoxy chain termination method using the T7 sequencing kit of Pharmacia (Uppsala, Sweden) and a double-stranded DNA template. The sequencing reaction for one of the DNA strands was initiated with the T3 primer (Stratagene, La Jolla, Calif.), whereas that for the other strand was initiated with the synthetic oligonucleotide 5’-GOT-AATGTGTTACATTCTA-3’. Sequence analysis was carried out with the Genetics Computer Group (1991) software package. A database search for homology with the sequence determined was conducted at the National Center for Biotechnology Information using the BLAST network service.

Analysis of Gene Rearrangements

Specific inversions accounting for the gene rearrangements between the C. gelatinosa and C. reinhardtii cpDNAs were identified using the ReversalSort program (Bafna and Pevzner 1995).

Results

Gene and Physical Maps of the C. gelatinosa cpDNA

As shown in figure 1, the EcoRI, MluI, and PstI restriction fragments of the C. gelatinosa cpDNA were ordered together with several chloroplast genes onto a circular map of 285 kb. The physical and gene maps were generated concurrently by using the strategy we previously employed to construct the physical and gene maps of the C. pitschmannii cpDNA (Boudreau and Turmel 1995). This approach consists mainly in hybridizing a collection of chloroplast gene-specific probes to Southern blots containing single and double digests of the cpDNA being studied.

Hybridizations of 89 Chlamydomonas gene-specific probes to single (EcoRI, MluI, PstI, NciI, and PvuII) and double (EcoRI/MluI, EcoRI/PstI, and MluI/PstI) digests of the C. gelatinosa cpDNA allowed us to order most of the restriction fragments on a single segment covering up to 97% of the chloroplast genome. Filling of a gap situated between the ORFB and psbD loci was necessary to circularize the resulting physical map. The only restriction fragments greater than 2 kb that could not be mapped with the aid of gene probes were the 3.6-kb EcoRI fragment, the 9.1-kb MluI fragment, and the 7.0-kb PstI fragment; the remaining unmapped fragments consisted of EcoRI and MluI fragments smaller than 1.7 kb. Southern blot and colony hybridizations with the C. reinhardtii EcoRI fragment 26 (according to the nomenclature of Grant, Gillham, and Boynton 1980), which is situated between ORFB and psbD, enabled us to localize the C. gelatinosa 3.6-kb EcoRI fragment in this same region of the chloroplast genome (see fig. 1) and to recover a clone containing this fragment. Subsequent Southern blot hybridizations with the C. gelatinosa 3.6- and 9.1-kb EcoRI fragments permitted the mapping of the 7.0 kb PstI fragment (fig. 1) and the circularization of the C. gelatinosa cpDNA map. The 9.1-kb MluI fragment was located on the C. gelatinosa cpDNA by using as hybridization probes a cloned 7.4-kb HindIII fragment recovered by colony hybridization with the rpoC2b gene probe and a 1.8-kb MluI/HindIII subfragment of this fragment. The remaining C. gelatinosa EcoRI and MluI fragments that were not originally identified with gene probes were mapped by fine restriction analysis of the 3.6- and 3.8-kb PstI fragments and of the two 11.0-kb EcoRI fragments, respectively.

The C. gelatinosa cpDNA is divided into nearly equal single-copy regions by an inverted repeat sequence whose size is estimated to be at least 18 kb but less than 26 kb (fig. 1). This sequence lacks any PstI sites. Based on previous inferences of flip-flop isomerization in cpDNAs containing an inverted repeat (Palmer 1985), one can predict the existence of four large PstI fragments (fragments of 45 and 27 kb, and of 32 and 40 kb) containing this C. gelatinosa cpDNA region, two on each of the inversion isomers produced via intramolecular recombination between the two segments of the inverted repeat. Smearly PstI fragments with sizes consistent with but not proving the occurrence of flip-flop isomerization were detected during our Southern blot hybridizations with probes mapping within the inverted repeat.
Fig. 1.—Physical and gene maps of the C. gelatinosa cpDNA. The three circles from the inside to the outside represent the EcoRI, MluI, and PstI restriction maps, respectively. Partial PvuII and NciI restriction maps are also represented; juxtaposed to the innermost circle is the PvuII restriction map, which is itself juxtaposed to the NciI restriction map. These two maps have not been precisely aligned relative to each other and relative to the EcoRI, MluI, and PstI restriction maps. Fragments are identified by their size in kb and those that were cloned are denoted by the letter c. The sizes of all restriction fragments less than 24 kb are based on their electrophoretic mobility in agarose gels; the sizes of fragments exceeding 24 kb were estimated from those of subfragments produced with restriction endonucleases. This figure shows only one of the two possible isomers of the chloroplast genome; in the other conformation PstI fragments of 32 and 40 kb replace those of 27 and 45 kb. The minimal extent of the inverted repeat is denoted by the thick lines outside the circles, whereas the maximal extent is marked by thin extensions of these lines. Note that the minimal extent of the inverted repeat on the utpB side is based on the position of an oligonucleotide specific to this gene, which was determined by restriction analysis of a PCR-amplified fragment spanning the region between the ORFA and atpB (data not shown). The short bars outside the circles denote gene loci that were located by Southern blot hybridizations; the relative order of genes assigned to a common bar has not been determined. Mapping with additional restriction endonucleases allowed us to separate rp12 from the rp16-rps19 cluster, petG from psbL, and psbH from trnE(UUC) (data not shown). The asterisks denote repeated sequences that were identified by Southern blot hybridizations with the C. gelatinosa chloroplast PstI/MluI fragment of 345 bp located between atpA and trnG(UCC) (see fig. 4). The positions of genes and repeated sequences coincide with the middle of the shortest DNA segments that revealed hybridization signals. vcf genes are hypothetical chloroplast reading frames that were shown to be conserved between land plants, or between land plants, E. gracilis, or algae. Reading frames unique to Chlamydomonas chloroplasts are designated as ORFs (Boudreau, Otis, and Turmel 1994). Note that the psbI gene was previously designated as ycf8 (Monod et al. 1994).

A total of 77 genes were located on the C. gelatinosa cpDNA with the aid of our collection of Chlamydomonas chloroplast gene probes representing 78 genes. The relative order of most genes was determined from the hybridization patterns to EcoRI, MluI, and PstI single and double digests. The hybridization patterns to PvuII and NciI single digests were used to establish the relative order of genes that hybridized to common EcoRI, MluI, and PstI fragments (fig. 1). The only chloroplast gene that could not be mapped is ycf12. The
absence of this gene from the *C. gelatinosa* cpDNA would be surprising given that the probe used (a PCR-amplified fragment from the *C. moewusii* cpDNA) was previously found to recognize both the *C. reinhardtii* and *C. pitschmannii* cpDNAs (Boudreau, Otis, and Turmel 1994; Boudreau and Turmel 1995). The *C. moewusii* ycf12 probe hybridized strongly to the *C. pitschmannii* cpDNA under stringent conditions, but only weakly to the *C. reinhardtii* cpDNA under less stringent conditions (our unpublished results). It probably failed to recognize the *C. gelatinosa* cpDNA under the latter conditions because the ycf12 genes of these green algae are too divergent in sequence.

**Differences in Gene Order Between the *C. gelatinosa* and *C. reinhardtii* cpDNAs**

The chloroplast gene organization of *C. gelatinosa* differs quite extensively from that of its close relative *C. reinhardtii*. As shown in figure 2, 18 blocks of conserved gene sequences are altered in order and/or direction in these two green algal cpDNAs. Figure 3 presents the most parsimonious scenario accounting for the differences in gene order between the *C. gelatinosa* and the *C. reinhardtii* cpDNAs. This scenario was determined with the ReversalSort program (Bafna and Pevzner 1995). A minimum of 10 mutational events, consisting of a single event of expansion/contraction of the inverted repeat and nine inversions, are required for interconverting the gene arrangements of these two *Chlamydomonas* cpDNAs. The expansion/contraction of the inverted repeat accounts for the gain/loss of one copy of block 7 which contains the *atpR* gene. Six inversions and three inversions are responsible for the gene rearrangements in the single-copy regions bordering *rrnS* and *psbA*, respectively. Ten of the 18 inversion endpoints are implicated in single mutational events, while four endpoints each are implicated in two separate inversions.

**Preliminary Characterization of a Family of Dispersed Repeated Sequences in the *C. gelatinosa* cpDNA**

Hybridizations of most of the cloned restriction fragments from the *C. gelatinosa* cpDNA to single and double digests of this green algal cpDNA revealed the presence of dispersed repeated sequences. Figure 4B shows the complex patterns we observed when the 345-bp *PstI/MluI* fragment was hybridized to double (*EcoRI/MluI, EcoRI/PstI*, and *MluI/PstI*) digests under stringent conditions. This subfragment recognizes all of the 22 *C. gelatinosa* chloroplast *PstI* restriction fragments, with the exception of the 1.0-, 1.2-, and 7.0-kb fragments (data not shown). From our hybridization data, we estimate that the repeated sequence found in the 345-bp *PstI/MluI* fragment resides in at least 35 other loci of the *C. gelatinosa* cpDNA (see fig. 1).

To determine the nature and boundaries of this repeated sequence, we sequenced the 345-bp *PstI/MluI* fragment and hybridized three oligonucleotides specific to distinct regions of the sequence to Southern blots containing single and double digests of *C. gelatinosa* cpDNA. As shown in figure 4A, the sequence of the *PstI/MluI* fragment features four short repeats of at least 14 bp in size, designated hereafter repeats A through D. Each of these four repeated sequences is present in two copies that are not immediately contiguous to each other. With the exception of the repeat B, the copies of each repeated sequence are separated by a stretch of sequence containing a different repeat motif. The repeated sequences A (18 bp), C (15 bp), and D (18 bp) are oriented in the same direction, whereas the repeated sequences B (14 bp) are inverted relative to one another. A database search for homology with the complete sequence of the 345-bp *PstI/MluI* fragment using BLAST revealed that a 30-bp segment featuring the repeat C (between positions 178 and 207, see fig. 4A) is identical to a region of the *C. gelatinosa* chloroplast *tscA* gene which corresponds to the domain III of the *tscA* RNA (Turmel et al. 1995b). No significant similarity to any other known DNA sequences was identified. Southern blot hybridizations of single and double digests of the *C. gelatinosa* cpDNA with 18-mer oligonucleotide probes specific to the repeats A and D revealed complex hybridization patterns that differed from one another and from the pattern obtained with the 345-bp *PstI/MluI* fragment probe by a number of hybridization fragments (fig. 4B). Hybridization of an 18-mer oligonucleotide (5'-CCCTTCCTCCTTTTGCCC-3') specific to the region of the *C. gelatinosa* 345-bp *PstI/MluI* fragment immediately precedes the first copy of the repeat element A (region comprised between positions 72 and 89) also produced a pattern that resembled those obtained with the oligonucleotides specific to repeats A and D; however, a slightly reduced number of hybridizing fragments were observed (fig. 4B). The observation that the repeat-A and-D probes hybridized to a number of common restriction fragments suggests that larger repeated sequences containing these two small repeats are dispersed throughout the *C. gelatinosa* cpDNA. Variations in size and/or sequence of these large repeated sequences can account for the observed differences in complexity of the hybridization patterns.

We examined the cpDNAs of eight other *Chlamydomonas* species (*C. reinhardtii, eugametos, moewusii, pitschmannii, humicola, peterfii, iyengarii, and frankii*) for the presence of repeat elements that are similar to those identified here in *C. gelatinosa*. Neither the 345-
Fig. 1. Comparative gene organization of the C. reinhardtii and C. gelatinosa cpDNAs. Genes are numbered, and gene sequences and their counterparts in C. reinhardtii are shown by thick lines. The gene sequences of C. gelatinosa are shown by thin lines. Note the inversion in the lower-right corner of the figure. The blocks of sequence whose gene arrangements are conserved between the C. reinhardtii and C. gelatinosa cpDNAs are shown by thick lines, and the single-copy regions between the two genomes are shown by thin lines.
Chloroplast Genome Evolution in Chlamydomonas 239

CpDNA are confined to species interfertile with this green alga. Similar hybridizations by Boynton et al. (1997) had previously revealed that the AatII repeat is absent from *C. eugametos*.

**Discussion**

The results of our study indicate that the gene organization of the *C. gelatinosa* cpDNA is highly divergent from that of the closely related species *C. reinhardtii*, with the differences between these cpDNAs being attributable to nine inversions and one event of expansion/contraction of the inverted repeat. The gene rearrangements between the cpDNAs of the closely related *C. moewusii* and *C. pitschmannii* are much less extensive, being accounted for by one or two inversions and one or three events of expansion/contraction of the inverted repeat (Boudreau and Turmel 1995). The same number of chloroplast genes (77) were mapped in the two pairs of closely related green algae, which precludes that the variance in the extent of chloroplast gene rearrangements is due to different degrees of resolution of the gene maps constructed. Because the divergence level of the *C. gelatinosa* and *C. reinhardtii* chloroplast-encoded large subunit rRNA gene sequences (1.36%, excluding variable regions) is equivalent to that (1.96%) of the corresponding *C. moewusii* and *C. pitschmannii* sequences (Turmel et al. 1993), our results may suggest that, in the same period of time, there have been more numerous rearrangements in the Chlamydomonas lineage comprising *C. gelatinosa* and *C. reinhardtii* than in the other comprising *C. moewusii* and *C. pitschmannii*. Alternatively, given that substitution rates in chloroplast genes are not necessarily uniform across lineages (Clegg, Larn, and Golembert 1991; Bousquet et al. 1992; Gaut et al. 1992), the extensive rearrangements between the *C. gelatinosa* and *C. reinhardtii* cpDNAs may reflect a longer divergence period for this pair of *Chlamydomonas* species compared to that for the *C. moewusii/C. pitschmannii* pair. In the two pairs of closely related *Chlamydomonas* species, the cpDNA region that underwent the most changes in gene order is the single-copy region bordering the *rrnS* gene. The one or two inversions in the *C. moewusii/C. pitschmannii* pair and six of the nine inversions in the *C. gelatinosa/C. reinhardtii* pair have occurred in this single-copy region.

The *C. gelatinosa* and *C. reinhardtii* cpDNAs feature a large number of repeated sequences that are readily detectable by Southern blot hybridization with fragment probes (Boynton et al. 1992; Gaut et al. 1992). Although such elements have not been identified by hybridization in the less scrambled *C. moewusii* and *C. pitschmannii* cpDNAs (Turmel, Bellemare, and Lemieux 1987; Boudreau and Turmel 1995), very short repeated

bp *PstI*/*MluI* fragment of this green alga nor the oligonucleotides specific to the repeats A and D hybridized to the *EcoRI* digests of these *Chlamydomonas* cpDNAs under low- to moderate-stringency conditions (data not shown), indicating that the repeated sequences A and D are unique to the *C. gelatinosa* cpDNA. In addition, we tested whether the short dispersed repeated sequences that have been identified in the cpDNAs of *C. reinhardtii* and species interfertile with this green alga (Boynton et al. 1992) are present in the *C. gelatinosa* cpDNA. No positive hybridization signals were observed when *C. gelatinosa* cpDNA blots were hybridized under low- to moderate-stringency conditions with two probes that are complementary to *C. reinhardtii* repeated sequences: a 281-bp AatII fragment from the spacer between the chloroplast small subunit and large subunit rRNA genes of *C. reinhardtii*, and the oligonucleotide 5'-GAGCTCCCTTCGGGCAATAAA-3' that is complementary to one of the two shorter repeat elements (AatII repeat) found in this fragment (see Boynton et al. 1992) (data not shown). The two probes also failed to hybridize to cpDNA blots from more distantly related *Chlamydomonas* species (*pitschmannii, humicola, petersfi, iyengarii, and frankii*), indicating that the repeated sequences reported in the *C. reinhardtii*
sequences that can form stem-loop structures have been found at various loci of these green algal cpDNAs (Durocher et al. 1989; Turmel, Boulanger, and Bergeron 1989; Turmel, Boulanger, and Lemieux 1989; Turmel et al. 1991; Turmel and Otis 1994). Assuming that the two pairs of closely related Chlamydomonas species diverged at about the same time, these observations suggest that the susceptibility of Chlamydomonas cpDNAs to rearrangements is correlated with the abundance of repeated sequences. Such a correlation has been reported for land plant cpDNAs (Palmer 1991). It will be of interest to examine additional green algae representing the two major Chlamydomonad lineages to see if these differences in the extent of gene rearrangements and abundance of dispersed repeated sequences are lineage-specific. Whether most or all of the endpoints of rearranged segments in the C. gelatinosa and C. reinhardtii cpDNAs coincide with dispersed repeated sequences is unknown. No repeated sequences have been mapped with precision in the C. gelatinosa cpDNA; only the approximate positions of one family of repeated sequences have been determined in the course of the present study (see fig. 1). In the C. reinhardtii cpDNA, sequence analyses of various parts of the genome have suggested that repeated sequences are ubiquitous in the intergenic regions (Boynton et al. 1992; our analysis of more recently reported sequences in databases also supports this idea).

Some of the endpoints of rearranged sequences in the C. gelatinosa and C. reinhardtii cpDNAs also appear to lie in the proximity of tRNA genes. Similar observations in land plant cpDNAs have led to the hypothesis that tRNA or tRNA-like gene sequences are implicated in gene shuffling (Palmer 1991; Knox, Downie, and Palmer 1993; Hoot and Palmer 1994). Because this role is supported by only circumstantial evidence, the finding of tRNA gene sequences near inversion endpoints might just be fortuitous. In contrast to the situation with the P. pischmannii cpDNA where a trnG(UCC) gene sequence has been located near each of the endpoints of a rearranged segment (Boudreau and Turmel 1995), we have detected no duplications of tRNA gene sequences at the borders of shuffled sequences in the C. gelatinosa and C. reinhardtii cpDNAs.

Little is known about the origin and evolution of dispersed repeated sequences in cpDNAs (Palmer 1991). If we assume that the dispersed repeats in the C. gelatinosa and C. reinhardtii cpDNAs have a common origin, then these sequences must have diverged very rapidly and their similarity in each green algal cpDNA must have been maintained by a yet unknown mechanism. Alternatively, the C. gelatinosa and C. reinhardtii repeats could be of independent recent origin and not be evolving in concert. We have observed no positive signals following Southern blot hybridizations of C. gelatinosa and C. reinhardtii cpDNA digests with chloroplast probes specific to the C. reinhardtii and C. gelatinosa dispersed repeated sequences, respectively. Moreover, we have detected no significant homology between the C. reinhardtii repeated sequences that have been
previously reported and the C. gelatinosa repeated sequence found in the 345-bp PstI/MluI fragment mapping near the endpoint of an inversion in the region between atpA and rrnS(UCU). A 9-bp stretch is shared between the motifs of the C. gelatinosa repeat A (18 bp) and the C. reinhardtii AatII repeat (minimal size of 30 bp); however, it was considered to be too short to reflect unambiguous homology. Like their homologues in C. reinhardtii (Boynton et al. 1992), the dispersed repeats in the C. gelatinosa cpDNA appear to be composed of shorter repeat elements that occur in variable combinations in different loci. Three repeats of at least 14 bp reside in the 345-bp PstI/MluI fragment of C. gelatinosa (see fig. 4). Interestingly, a 30-bp sequence in the chloroplast tsaA gene of this green alga contains repeat C (Turmel et al. 1995b) and is identical with the 345-bp PstI/MluI fragment, indicating that isolated repeat motifs can also be found in some loci of the C. gelatinosa cpDNA. The two widespread motifs marked by AatII and KpnI sites in the short dispersed repeated sequences of C. reinhardtii (Boynton et al. 1992) are substantially longer (at least 30-bp long) than those characterized here in the C. gelatinosa cpDNA. Several observations are consistent with the idea that the short repeated sequences in the C. reinhardtii cpDNA participate in intra-and intermolecular recombination events to yield dimer and multimer cpDNA molecules, populations of isomeric molecules, and deletion and rearrangement mutants of cpDNA (Palmer et al. 1985; Boynton et al. 1992). Recombination events between short dispersed repeats have also been proposed to account for many of the insertions/deletions between the C. reinhardtii and C. smithii cpDNAs, and for the restoration of cpDNA sequences in transformation experiments using C. reinhardtii deletion mutants as recipients (Palmer et al. 1985; Boynton et al. 1992).

The results of our analysis of the C. gelatinosa cpDNA are consistent with previous reports indicating that the cpDNAs of green algae belonging to the Chlamydomonadales are less constrained than their land plant counterparts to maintain a compact gene organization (Boudreau, Otis, and Turmel 1994; Boudreau and Turmel 1995). At 285 kb, the C. gelatinosa cpDNA is 89 kb larger than its C. reinhardtii counterpart (196 kb; Rochaix 1978) and only 7 kb smaller than the largest Chlamydomonad cpDNA (292-kb C. moewuissii cpDNA) characterized thus far (Turmel, Bellemare, and Lemieux 1987). As is the case for the C. moewuissii and C. pitschmannii cpDNAs (Boudreau and Turmel 1995), the substantial difference in size between the C. gelatinosa and C. reinhardtii cpDNAs is mainly attributable to multiple deletions/additions that have occurred in intergenic regions. The genes located in the single-copy regions are much more loosely packed in the C. gelatinosa cpDNA than in the C. reinhardtii cpDNA. The single-copy regions bordering the rrnS gene differ by about 38 kb, the opposite single-copy regions vary by approximately 51 kb, while the inverted repeats are about of the same size. In land plant cpDNAs, the expansion/contraction of the inverted repeat is the major cause of size variation (Palmer 1991; Downie and Palmer 1992). As proposed for the deletions/additions between the C. reinhardtii and C. smithii cpDNAs (Palmer et al. 1985), those between the C. reinhardtii and C. gelatinosa cpDNAs might have been caused by the proliferation and/or removal of existing sequences through recombination between short dispersed repeated sequences. The DNA segments exhibiting similarity in gene order in the latter cpDNAs could not be aligned on the basis of the EcoRI and PstI sites mapped, indicating substantial sequence divergence. It is thus likely that molecular characterization of the endpoints of the rearranged segments would fail to reveal the sequence elements that have participated in the gene rearrangements.

In the C. gelatinosa cpDNA, two of the chloroplast gene clusters that were most probably present in the common ancestor of green algae falling within the major Chlamydomonad lineages represented by C. moewuissii and C. reinhardtii have not remained intact. They consist of trnR(UCU)-trnT(UGU) and of rps18-rps2-trnD-psbB-psbT-psbN-psbH-trnE1 (rps18 cluster), with the latter sharing similarity to the chloroplast psbB operon (Boudreau, Otis, and Turmel 1994). An inversion in the C. gelatinosa cpDNA single-copy region bordering the psbA gene resulted in splitting of the rps18 gene cluster into two fragments (rps18-rps2 and trnD-psbB-psbT-psbN-psbH-trnE1), whereas a separate inversion in the other single-copy region disrupted the trnR(UCU)-trnT(UGU) cluster which is known to contain genes with the same polarity in the C. reinhardtii and C. eugametos cpDNAs (Khrebtukova and Spritzter 1994; see Boudreau, Otis, and Turmel 1994). These observations suggest that the ancestral clusters containing these genes have been reorganized after the divergence of the two major Chlamydomonad lineages, more specifically during the evolution of green algae belonging to the lineage represented by C. reinhardtii.

Acknowledgments

We thank C. Lemieux and two anonymous referees for valuable comments, J.-P. Mercier for providing the Chlamydomonas cpDNA-enriched preparations and C. Otis for preparing an HindIII clone bank of this DNA. We are grateful to J. E. Boynton, N. W. Gilham, and E. H. Harris for providing the C. reinhardtii cpDNA clones P-74A and P-230 containing the EcoRI 26 and 281-bp AatII fragments, respectively, and for sharing the
sequence of the latter fragment. We also acknowledge P. Pevzner for analyzing the gene rearrangements between the C. gelatinosa and C. reinhardtii cpDNAs with the ReversalSort program. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (GP0003293) and "Le Fonds pour la Formation de Chercheurs et l’Aide à la Recherche" (93-E-K-035U). M.I. is a Scholar in the Evolutionary Biology Program of the Canadian Institute for Advanced Research.

LITERATURE CITED


BARBARA A. SCHAAL, reviewing editor

Accepted September 7, 1995