Conserved gene clusters in the highly rearranged chloroplast genomes of *Chlamydomonas moewusii* and *Chlamydomonas reinhardtii*

Eric Boudreau, Christian Otis and Monique Turmel*
*Département de biochimie, Faculté des sciences et de génie, Université Laval, Québec (Québec) G1K 7P4,
Canada (* author for correspondence)*

Abstract

We have extended to about 75 the number of genes mapped on the *Chlamydomonas moewusii* and
Chlamydomonas reinhardtii chloroplast DNAs (cpDNAs) by partial sequencing of the very closely related
C. eugametos and *C. moewusii* cpDNAs and by hybridizations with *Chlamydomonas* chloroplast gene-
specific sequences. Only four of these genes (*tscA* and three reading frames) have not been identified in
any other algal cpDNAs and thus may be specific to *Chlamydomonas*. Although the *C. moewusii* and
C. reinhardtii cpDNAs differ by complex sequence rearrangements, 38 genes scattered throughout the
genome define 12 conserved clusters of closely linked loci. Aside from the rRNA operon, four of these
gene clusters share similarity to evolutionarily primitive operons found in other cpDNAs, representing
in fact remnants of these operons. Our results thus indicate that most of the ancestral bacterial operons
that characterize the chloroplast genome organization of land plants and early-diverging photosynthetic
eukaryotes have been disrupted before the emergence of the polyphyletic genus *Chlamydomonas*. All gene
rearrangements between the *C. moewusii* and *C. reinhardtii* cpDNAs, with the exception of those
accounting for the relocations of *atpA*, *psbI* and *rbcL*, occurred within corresponding regions of the
genome. One of these rearrangements seems to have led to disruption of the ancestral region contain-
ing *rpl23*, *rpl2*, *rps19*, *rpl16*, *rpl14*, *rpl5*, *rps8* and the *psaA* exon 1. This gene cluster, which bears striking
similarity to the *Escherichia coli* S10 and *spc* operons, spans a continuous DNA segment in *C. reinhardtii*,
while it maps to two separate fragments in *C. moewusii*.

Introduction

It is well documented that the chloroplasts of
green algae and land plants arose from a common
bacterial endosymbiont, most probably a cyano-
bacterium [19]. However, despite a common evo-
lutionary origin, land plant and green algal chlo-
roplast genomes appear to differ dramatically in
their evolutionary trends, with the latter genome
having evolved under more relaxed constraints.

Extensive studies of cpDNA from over 1000
photosynthetic land plant species [reviewed in
36], including the complete sequencing of the
cpDNAs from tobacco [45], rice [23] and the
liverwort *Marchantia polymorpha* [34], have
shown that the land plant chloroplast genome is

remarkably conserved in structure, size, gene content, primary sequence and overall gene arrangement. In most land plants, this genome consists of a circular DNA molecule of 120 to 160 kb, which is divided into small and large single-copy regions by a large inverted repeat. Another dominant feature of land plant cpDNA is the tight linkage of the 110–118 encoded genes, whose products are primarily involved in gene expression and photosynthesis. The majority of chloroplast genes are grouped into multicistronic operons, several of which share striking similarities with those found in cyanobacteria. A few chloroplast genes, including *rpl22*, *tufA* and three loci encoding proteins involved in chlorophyll synthesis (*chlB*, *chlL* and *chlN*), are absent from some land plant groups; for *rpl22* and *tufA*, such variable distribution has been attributed to gene transfer to the nucleus. All land plant species investigated so far, with few exceptions, show great similarities in their chloroplast gene order, with occasional variations being due to inversions of large sequences [36].

In contrast, the limited studies on green algal cpDNA have revealed substantial variations in size and gene order [reviewed in 36]. The low-resolution gene maps reported for this genome (89–400 kb) in a small number of green algae representing three (Charophyceae, Chlorophyceae and Ulvophyceae) of the five classes proposed by Mattox and Stewart [32] have indicated extensive sequence rearrangements not only between different classes and genera of green algae, but also within the highly diversified, polyphyletic genus *Chlamydomonas* (Chlorophyceae). Indeed heterologous hybridizations with large cpDNA restriction fragments have shown that the cpDNAs of the unicellular green flagellates *C. eugametos* and *C. reinhardtii*, two taxa differing at the morphological, physiological and reproductive levels [17, 43], display numerous rearrangements that cannot be simply explained by inversions [28]. Besides the rRNA operon, no conserved gene clusters were found among the twenty to sixty genes that were mapped on these cpDNAs [21, 51]. In view of recent rRNA sequence analyses indicating that *C. eugametos* and *C. reinhardtii*

represent the two major lineages in a group of green flagellates belonging to *Chlamydomonas* and closely related genera and that their level of rRNA sequence divergence is more than twice that found among all land plants [8, 54], this considerable variation in chloroplast genome organization may be not surprising. Consistent with this notion is the finding that the interfertile taxa *C. eugametos/C. moewusii* and *C. reinhardtii/ C. smithii* share essentially the same chloroplast gene order [6, 50].

To gain insight into the mode and tempo of *Chlamydomonas* cpDNA evolution, we have undertaken the construction of high-resolution gene maps for taxa representing the various lineages identified in this genus [54]. Comparison of such maps in closely related taxa should allow us to discern the nature of mutations accounting for changes in gene order, while their comparison in distantly related taxa should allow the identification of cpDNA regions with conserved gene arrangements. In the present study, we report the high-resolution gene maps of the *C. moewusii* and *C. reinhardtii* cpDNAs. Our comparative analysis of 75 gene loci has revealed that these highly rearranged cpDNAs share, besides the rRNA operon, 11 conserved clusters of closely linked genes, four of which represent segments of land plant chloroplast operons.

Materials and methods

Cloning of randomly fragmented cpDNA from C. eugametos

A cpDNA-enriched fraction from the wild-type *mt*$^+$ strain of *C. eugametos* (UTEX 9) was prepared as described by Turmel *et al.* [54]. A solution (500 μl) of this DNA at a concentration of 10 ng/μl in 1 mM Tris-HCl pH 8.0, 10 mM NaCl and 25% glycerol was passed through a nebulizer at 20 PSI for 2 min to generate fragments of 200 to 1500 bp (J.E. Surzycki, personal communication). After ethanol precipitation, 1 μg of DNA was incubated in the presence of 0.2 mM of each dNTP, 50 mM Tris-HCl pH 7.6, 10 mM MgCl$_2$, 3 units of Klenow fragment of *E. coli* DNA poly-

merase I and 3 units of T7 DNA polymerase in a total volume of 10 μl. The resulting DNA fragments with repaired termini were ligated to the plasmid pBluescript KS + (Stratagene, La Jolla, CA), which had been digested with Sma I and dephosphorylated. Competent cells of E. coli strain DH5αF'IQ (Bethesda Research Laboratories, Gaithersburg, MD) were directly transformed with the ligated DNA mixture. Plasmids from 500 randomly selected transformants were isolated using the alkaline extraction procedure of Birnboim and Doly [5] and screened for the presence of inserts homologous to the C. reinhardtii cpDNA as described below.

Subcloning and DNA sequencing

After purification, the C. moewusii chloroplast Eco RI fragments 3, 6', 10', 12, 15, 16, 18, 19 and 20 [49] as well as the C. eugametos chloroplast Eco RI fragments 2, 8 and 9' [29] were either grouped in various pools (at equimolar concentrations) for subcloning or were individually subcloned. Pooled or individual fragments were digested with Taq I or Sau 3AI, and the resulting subfragments were cloned into the dephosphorylated Acc I or Bam HI sites of the pBluescript

KS-plasmid vector (Stratagene, La Jolla, CA). The C. moewusii chloroplast Eco RI fragments 25 and 30 [50] were gel-purified and cloned into the Eco RI site of pBluescript KS -. E. coli strain DH5αF'IQ (Bethesda Research Laboratories, Gaithersburg, MD) was used as the host and recombinant clones were randomly selected. Plasmids carrying restriction or nebulizer-generated fragments were sequenced by the dideoxy chain termination method using the T7 sequencing kit of Pharmacia (Uppsala, Sweden). Sequencing reactions were initiated with the T7 or T3 primer (Stratagene, La Jolla, CA) or with synthetic oligonucleotides. Sequence analysis was performed using the University of Wisconsin GCG software package [11]. Genes were identified by database search at the National Center for Biotechnology Information using the BLAST network service.

DNA amplifications

Internal regions of C. eugametos, C. moewusii or C. reinhardtii chloroplast genes were amplified by the polymerase chain reaction (PCR) from cpDNA-enriched preparations (100 ng) or from recombinant plasmids (50 ng), using the primers listed in Table 1. Thirty-five cycles of amplifica-

Table 1. Oligodeoxyribonucleotide primers used for amplifying coding regions of chloroplast genes.

Gene	5' Primer (5'→3')[b]	3' Primer (5'→3')[b]
ORFB	AGTATTAGCTAAACCTATG	TTCACTTTGACTAGGGAAG
psaJ	CGTTACTTTAAGAATTAAG	TCAGGTTATCCACAATTTT
psbH	ATGGCAACAGGAACTTCTA	AGCTAAAGTTTCCCAACTC
psbI	GGTAACAATCTCTTAGCTA	CCAAAGTACTTGTAATAGT
psbK	ATGACAACTTTAGCACT	CGGAAACTAACAGCTGC
rpl20	ATGACTCGTGTTAAACGTG	GCATAAATGCTTCTGGATC
rpl23	AAATATCCTGTAATTACAC	GTGAATAATTGATTGACTG
rpoCIa	CCTCATATTCAAAATTGC	GTTTTACTCGACAATTG
rps4	ATGTCACGTTATTTAGG	CGTGCTGCTGGAATTGT
rps11	TCATATTCAAGCAGGCC	GGACGACAACCATTGTG
rps19	ATGTCACGTTCTCTTAAAA	GCCACGATATGTACGTG
ycf4[a]	AAAATACGTACAAAACC	TTGTAATACGACTCACTATAG
ycf12	ATGAAGCTTTAGTATATACT	CTGGATCTTAAAATGTTAG

[a] The coding region of ycf4 was amplified from a recombinant plasmid, using a 5' primer specific to a coding sequence present in the insert and a 3' primer specific to the T7 promoter of the pBluescript vector.

[b] Sequence coordinates of the gene loci amplified with all pairs of primers are given in Table 2.

tion (1 min at 94 °C, 1 min at 50 °C, and 3 min at 72 °C) were carried out in a DNA thermal cycler (Perkin Elmer, Norwalk, CT). Reactions were run in the presence of 0.2 mM of each dNTP, 1.3 μM of each of the primers, 10 mM Tris-HCl pH 8.4, 50 mM KCl, 1.5 mM MgCl$_2$, 0.002% (w/v) gelatin, and 2.5 units of AmpliTaq DNA polymerase (Perkin Elmer, Norwalk, CT) in 100 μl total volume. PCR-amplified fragments were electrophoresed on 1% agarose gels and purified by eletroelution or with the QIAEX gel extraction kit (QIAGEN, Chatsworth, CA).

Southern blot hybridizations

C. moewusii (UTEX 97) and *C. reinhardtii* (137c) cpDNA-enriched preparations [51] as well as recombinant plasmids carrying randomly fragmented cpDNA from *C. eugametos* were digested with appropriate restriction endonucleases. The resulting fragments were electrophoresed on 0.8% agarose gels and transferred onto Hybond-N nylon membranes (Amersham, Arlington Heights, IL) as recommended by the manufacturer. Hybridizations of *Chlamydomonas* cpDNA blots with ^{32}P-labelled gene-specific fragments (see Table 2 for their descriptions) as well as hybridizations of plasmid DNA blots with ^{32}P-labelled *C. reinhardtii* cpDNA were performed according to Turmel *et al.* [51]. These DNA probes were labelled with [α-^{32}P]dCTP or [α-^{32}P]dATP (3000 Ci/mmol) using the Multiprime DNA labelling system (Amersham, Arlington Heights, IL). Pre-hybridizations and hybridizations of *Chlamydomonas* cpDNA blots with ^{32}P-end-labelled oligodeoxyribonucleotides (see Table 2) were carried out at 37 °C in 5 × SSC, 0.1% SDS, 0.02 M sodium phosphate pH 7.0, 0.2% Ficoll, 0.2% bovine serum albumin (BSA), 0.2% polyvinylpyrrolidone (PVP) and 100 μg/ml denatured salmon sperm DNA; post-hybridization washes were done at 37 °C for 30 min in 3 × SSC, 0.1% SDS, 0.01 M sodium phosphate pH 7.0, 0.2% Ficoll, 0.2% BSA and 0.2% PVP, and then at 37 °C for 20 min in 1 × SSC and 0.1% SDS.

Results

Identification and mapping of genes on the C. eugametos *and* C. moewusii *cpDNAs by partial DNA sequencing*

To study the gene organization of *Chlamydomonas* cpDNAs, we first attempted, with little success, to map several genes by carrying out Southern blot hybridizations under low-stringency conditions using gene-specific fragments from the tobacco cpDNA as probes. Because most land plant chloroplast gene sequences employed proved to be ineffective probes in hybridizations with the *C. eugametos* (243 kb) and *C. moewusii* (292 kb) cpDNAs, we undertook the partial sequencing of these colinear cpDNAs. This strategy allowed us to investigate the gene content and organization of these two green algal cpDNAs and also to generate gene-specific probes that proved useful for the precise mapping of some loci and for the analysis of the divergent *C. reinhardtii* cpDNA (196 kb). Using both random and directed sequencing approaches, we have sequenced so far about 140 kb of the *C. eugametos* or *C. moewusii* cpDNA, and identified 32 additional genes on these DNAs (see Fig. 1). All of these genes, with the exception of the ORF715, revealed similarity with known DNA sequences. Continuous stretches of sequences encompassing more than one gene have been assembled for five separate cpDNA regions encoding the following genes: (1) *rrnS* [12], *trnI, trnA, rrnL* and 5S rDNA [53]; (2) *chlN, psbF, psbL, petG, rps3, rps7, psaA* exon 3 and *rps18*; (3) *psbC* [55], *rps12, psaA* exon 1, *rps8, rpl5, rpl14, rpl16* and an ORF (ORFA) homologous to the *C. reinhardtii* ORF1 [58]; (4) *clpP* and *psaA* exon 2; (5) *rpoC1a, rpoC1b, petA, petD* [2, 3] and *trnR*-UCU [2]. In addition to these gene loci, several individual genes (most of which have been partially sequenced) were identified in the course of sequencing both extremities of random *Taq* I or *Sau* 3AI subfragments derived from selected *Eco* RI clones as well as by sequencing the ends of randomly generated *C. eugametos* cpDNA fragments (200–1500 bp) found to cross-hybridize with radiola-

Table 2. Description of chloroplast gene probes.

Gene	Source	Nature and size of probe[c]	Gene location[d]	Reference[g]
chlB	*C. moewusii*	*Xba* I–*Sal* I, 2314 bp	− 479, [1689], + + 146	[38]
chlL	*C. eugametos*	random, 417 bp	+ 635, + + 335[e]	–
ORF715[a]	*C. eugametos*	random, 593 bp	− 84, + 509	–
ORFA[a]	*C. eugametos*	*Sca* I, 657 bp	nd, + + 70	–
ORFB[a]	*C. reinhardtii*	PCR, 955 bp	+ 11, + 965	[41]
psaC	*C. reinhardtii*	*Sau* 3A1, 1.05 kb	nd, [246], nd	[49]
psaJ	*C. reinhardtii*	PCR, 175 bp	− 28, [126], + + 21	[31]
psbH	*C. reinhardtii*	PCR, 255 bp	+ 1, + 255	[25]
psbI	*C. reinhardtii*	PCR, 212 bp	− 72, [114], + + 26	[30]
psbK	*C. reinhardtii*	PCR, 137 bp	+ 1, + 137	[46]
rpl20	*C. reinhardtii*	PCR, 322 bp	+ 1, + 322	[61]
rpl23	*C. reinhardtii*	PCR, 256 bp	+ 13, + 268	X.-Q. Liu (pers. comm.)
rpoC1a	*C. moewusii*	PCR, 912 bp	+ 36, + 947[f]	–
rpoC1b	*C. moewusii*	*Bam* HI-*Sau* 3A1, 644 bp	+ 1574, + 2175[f]	–
rpoC2a	*C. eugametos*	random, 285 bp	+ 537, + 821[e]	–
rpoC2b	*C. eugametos*	random, 573 bp	+ 3459, nd[f]	–
rps2	*C. eugametos*	random, 190 bp	+ 482, + 657[f]	–
rps4	*C. reinhardtii*	PCR, 386 bp	+ 1, + 386	X.-Q. Liu (pers. comm.)
rps7 5′	*C. moewusii*	*Taq* I, 291 bp	− 57, + 234	–
rps7 3′	*C. moewusii*	*Sau* 3A1, 806 bp	+ 43, + + 374	–
rps11	*C. reinhardtii*	PCR, 306 bp	+ 66, + 371	S. J. Surzycki (pers. comm.)
rps19	*C. reinhardtii*	PCR, 246 bp	+ 1, + 246	X.-Q. Liu (pers. comm.)
trnC-GCA	*C. reinhardtii*	GAATAGAGGATTTGCAATC	+ 48, + 30	[35]
trnD-GUC	*C. eugametos*	CCGTGACAGGGCAGTGCTC	+ 40, + 22	–
trnE(1,2)-UUC	*C. reinhardtii*	CCGTGAAAGGGAGGTGTCC	+ 40, + 22	[35]
trnG-UCC	*C. reinhardtii*	ACTTGGAAGGATCGCACTC	+ 40, + 22	[62]
trnH-GUG	*C. eugametos*	GCATAATGGAGTCACAGTC	+ 48, + 30	–
trnI-CAU	*C. reinhardtii*	GATTATGAGTCGTTTGCCT	+ 40, + 22	–
trnL-UAG	*C. reinhardtii*	TTCCTAAAACCAGGATGTC	+ 39, + 21	[24]
trnfM-CAU	*C. eugametos*	GATTATGAGCCCCACGAGC	+ 40, + 22	–
trnM-CAU	*C. moewusii*	GCTTATGAAACGGACGCTC	+ 40, + 22	–
trnR-ACG	*C. reinhardtii*	GGTTCGTAGCCATGTGCTC	+ 40, + 22	[60]
trnR-UCU	*C. moewusii*	ATTTAGAAGATCGATGTCC	+ 40, + 22	–
trnS-GCU	*C. eugametos*	GATTAGCAATCAGCCGCTT	+ 40, + 22	–
trnS-UGA	*C. reinhardtii*	TTTCAAAATTGGTGCAATT	+ 38, + 20	[61]
trnT-UGU	*C. moewusii*	GTTTACAAAACCAAAGCTC	+ 40, + 22	[38]
trnW-CCA	*C. reinhardtii*	TTTTGGAAACCTGCGTTCT	+ 39, + 21	[62]
trnY-GUA	*C. eugametos*	GATTTACAATCCACCCCCA	+ 40, + 22	–
tscA	*C. reinhardtii*	PCR, 720 bp	− 154, [430 ± 20], + + 116	[16]
ycf3[b]	*C. moewusii*	*Taq* I, 1.0 kb	nd, + 1070[f]	–
ycf4[b]	*C. reinhardtii*	PCR, 223 bp	nd, + 395[f]	–
ycf5[b]	*C. eugametos*	*Taq* I, 0.8 kb	nd, + 659[f]	–
ycf8[b]	*C. reinhardtii*	ATGGAAGCTTTAGTATATACT	+ 1, + 21	[33]
ycf12[b]	*C. moewusii*	PCR, 131 bp	+ 1, + + 29	–

[a] Hypothetical chloroplast open reading frames unique to *Chlamydomonas*. The *C. eugametos* ORF715 has been completely sequenced and found to code for a protein of 715 amino acids (our unpublished results), whereas the *C. eugametos* ORFA and the *C. reinhardtii* ORFB have been partly sequenced and found to encode proteins of at least 356 and 337 amino acids, respectively.

[b] Hypothetical chloroplast open reading frames conserved between land plants, or between land plants, *E. gracilis*, or algae. Their designation follows the nomenclature recommended by the Commission on Plant Gene Nomenclature of the International Society for Plant Molecular Biology (R. B. Hallick, personal communication), i.e. that used in the SwissProt database.

[c] Restriction fragments are designated by the enzymes that were used to generate them; nebulizer-generated fragments are designated by 'random', PCR-amplification fragments are designated by 'PCR', and the sequences of oligodeoxyribonucleotide probes are given in the 5′ to 3′ direction. Fragment sizes given in kb are approximations based on electrophoretic mobilities, while fragment sizes given in bp are based on sequence data.

[d] −, the gene probe starts at the indicated number of bp before the initiation codon; +, the gene probe starts or ends at the indicated number of bp following the initiation codon; + +, the gene probe ends at the indicated number of bp following the stop codon; nd, due to a lack of sequence data, it could not be determined where the probe starts or ends relative to the initiation or stop codons, respectively; [ ], the number indicates the size (bp) of the coding region entirely comprised within the fragment probe. For the tRNA probes, the numbering system employed by Sprinzl *et al.* [48] was used.

[e] Coordinates based on the corresponding *C. reinhardtii chlL* [24] and *rpoC2a* [13] sequences.

[f] Coordinates based on the corresponding *M. polymorpha* gene sequences [34].

[g] Our unpublished sequence data are indicated by an hyphen. We identified *psaJ*, *psbI* and ORFB in the DNA sequences from the references indicated.
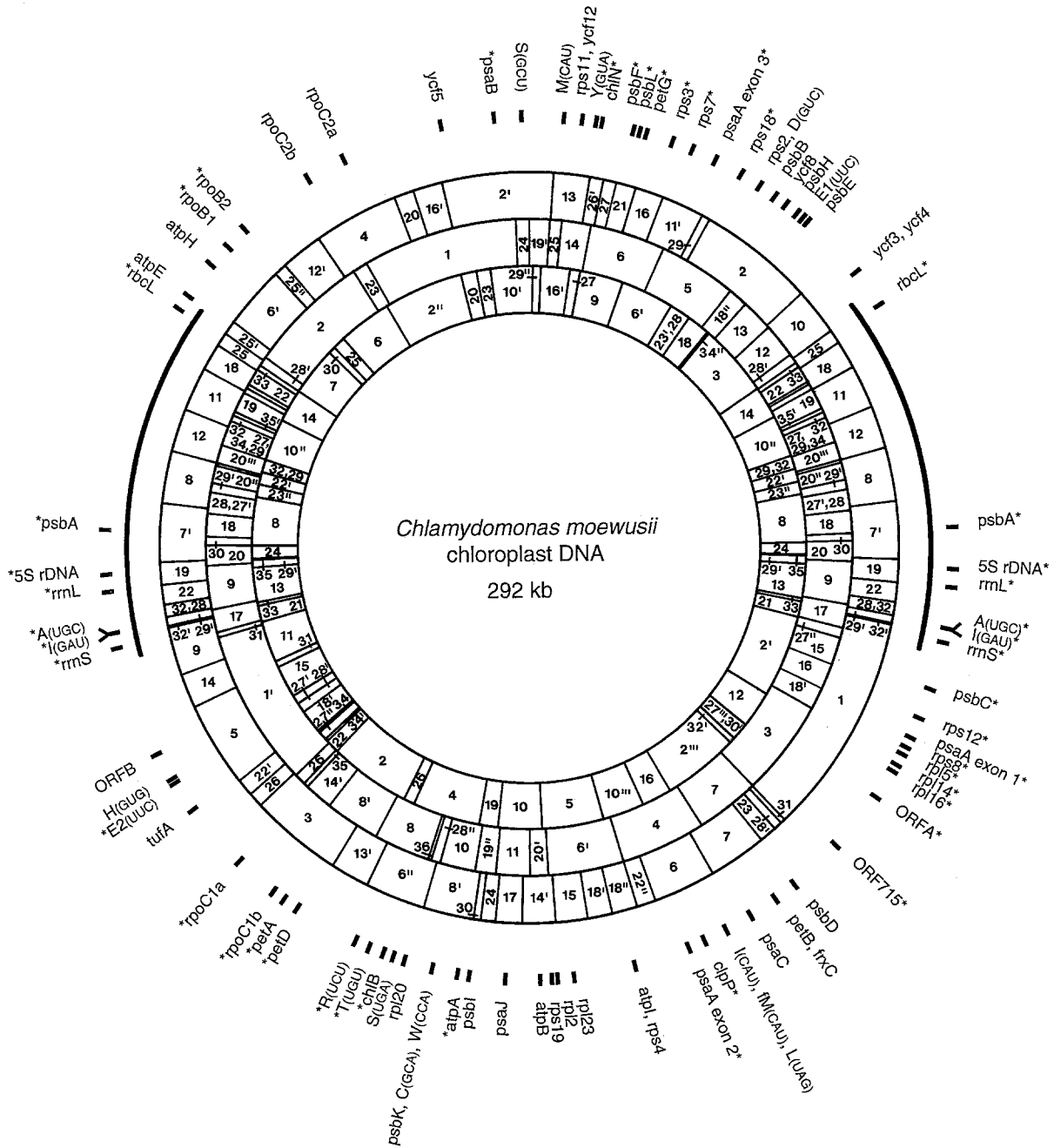
Fig. 1. Physical and gene maps of the *C. moewusii* cpDNA. The three circles from the inside to the outside represent the *Eco* RI, *Bst* EII and *Ava* I restriction maps, respectively [50]. The two curved lines outside these circles denote the minimal extent of the inverted repeat sequence, whereas the bars denote gene loci. The genes marked with an asterisk were unambiguously positioned on the basis of their partial or complete nucleotide sequence, whereas the others were located by Southern blot hybridizations [51, this study]. In the case of the latter genes, each of the bars coincides with the middle of the shortest DNA segment that revealed an hybridization signal. The *trnEI* probe hybridized to an *Eco* RI fragment of about 200 bp (34″), which was not previously mapped [50]. Hypothetical chloroplast open reading frames (*ycf*) that were shown to be conserved between land plants, or between land plants, *E. gracilis* or algae, were designated as recommended by the Commission on Plant Gene Nomenclature of the International Society for Plant Molecular Biology (R.B. Hallick, personal communication); this nomenclature is that used in the SwissProt

belled *C. reinhardtii* cpDNA under low-stringency conditions. The latter approach proved very effective in revealing genes, as all of the positive clones analyzed (55) disclosed coding sequences (representing 30 genes), including that of the ORF715 coding for a protein of 715 amino acids. These *C. eugametos* clones with sequence homology to the *C. reinhardtii* cpDNA represent only a small fraction (11%) of the recombinant clones screened (500), suggesting that a few genes are sufficiently conserved between the *C. eugametos* and *C. reinhardtii* cpDNAs to be detected by hybridization and/or that these genomes feature a substantial proportion of non-coding sequences.

Besides revealing additional genes on the *C. eugametos* and *C. moewusii* cpDNAs, our sequence analysis confirmed the presence of most coding regions previously mapped by heterologous hybridizations [50, 51]. One notable exception, however, concerns *atpF*, which was reported to lie on part of the *C. moewusii Eco* RI fragment 19 in the immediate vicinity of *atpA*, another gene encoded by this restriction fragment [51]. In the course of the present study, we have entirely sequenced the 1.7 kb region of the *C. moewusii Eco* RI fragment 19 to which Turmel *et al.* [50, 51] localized positive hybridization signals with spinach *atpA* and *atpF* probes under low-stringency conditions. This region disclosed the whole coding sequence of *atpA* (1506 bp), but no sequence similarity to *atpF* (unpublished results). The positive signal that was detected with the spinach *atpF* probe is thus attributed to unspecific hybridization, and for this reason, we have

not represented this gene on the *C. moewusii* chloroplast gene map shown in Fig. 1.

Gene maps of the C. moewusii *and* C. reinhardtii *cpDNAs*

Figures 1 and 2 show the updated gene maps of the *C. moewusii* and *C. reinhardtii* cpDNAs as derived from the partial sequencing of the closely related *C. eugametos* and *C. moewusii* cpDNAs and from Southern blot hybridizations with 44 *Chlamydomonas* chloroplast gene-specific probes representing 41 distinct genes. This collection of gene probes consisted of *Taq* I and *Sau* 3AI subfragments originating from selected *C. eugametos* or *C. moewusii* chloroplast *Eco* RI fragments, of nebulizer-generated fragments (200–1500 bp) from the *C. eugametos* cpDNA, as well as of oligodeoxyribonucleotides and PCR amplification products that were designed from *C. eugametos*, *C. moewusii* or *C. reinhardtii* chloroplast gene sequences (Table 2). Note that probes for two genes (*psaJ*, *psbI*) and one reading frame (ORFB) that were not previously reported were derived from published *C. reinhardtii* cpDNA sequences [20, 31, 41]. For *rps7*, 5′ and 3′ gene-specific probes were employed to test the hypothesis that the *C. reinhardtii rps7* coding region consists of two exons mapping to different single-copy regions [39]. All of the hybridizations we performed, with few exceptions, revealed positive signals to specific *Ava* I, *Bst* EII and *Eco* RI fragments of the *C. moewusii* cpDNA (Table 3) or to *C. reinhardtii*

---

database. Reading frames unique to the *Chlamydomonas* chloroplasts were designated as ORFs (see Table 2). Note that the cpDNA of *C. eugametos* (243 kb) shares essentially the same gene organization as that of *C. moewusii* (292 kb) [50]. The difference in size between these cpDNAs is mainly accounted for by the presence of two extra sequences in *C. moewusii*: a 21 kb sequence spanning the region between *rbcL* and *psbA* and a 6 kb sequence in the interval between *petD* and *trnR* [2, 50]. Nucleotide sequences were reported for all of the *C. eugametos* or *C. moewusii* genes in the inverted repeat [12, 52, 53, 59], for *chlB* and *trnT* [38], for *petD* [2, 3], *trnR* (UCU) [2] as well as for *psaB* and *psbC* [55]. The partial or complete nucleotide sequences of the remaining genes were derived during the course of the present study [our unpublished results]. The *C. eugametos* ORF715 has been completely sequenced and found to code for a protein of 715 amino acids, whereas the *C. eugametos* ORFA has been partly sequenced and found to encode a protein of at least 356 amino acids with sequence similarity to the protein product of the *C. reinhardtii* ORF1 [58]. Note that it has not been demonstrated that the *C. moewusii rpoB* and *rpoC2* feature a split organization; however, the coding regions corresponding to the separate loci identified in *C. reinhardtii* have been designated as in this alga [13]. This figure shows only one of the two possible isomers of the chloroplast genome.

592

Fig. 2. Physical and gene maps of the *C. reinhardtii* cpDNA. The three circles from the inside to the outside represent the *Pst* I, *Bam* HI and *Eco* RI restriction maps, respectively [21]. The two curved lines outside these circles denote the minimal extent of the inverted repeat sequence, whereas the bars denote gene loci. The genes marked with an asterisk were unambiguously positioned on the basis of sequence data, while the others were located by Southern blot hybridization [21, this study]. In the case of the latter genes, each of the bars coincides with the middle of the shortest DNA segment that revealed an hybridization signal. The *rps3* gene is the ORF712 previously reported by Fong and Surzycki [14]. Hypothetical chloroplast open reading frames (*ycf*) that were shown to be conserved between land plants, or between land plants, *E. gracilis* or algae, were designated as recommended by the Commission on Plant Gene Nomenclature of the International Society for Plant Molecular Biology (R.B. Hallick, personal communication); this nomenclature is that used in the SwissProt database. Reading frames unique to the *Chlamydomonas* chloroplasts were designated as ORFs (see Table 2). References for all available gene sequences, with 14 exceptions, were cited by Harris [21]. The *psbH* and *ycf8* sequences were recently reported [25, 33]. We identified *psaJ* [31], *psbI* [30], *rpoC2b* [56], *trnD* [4], *trnfM*, *trnH* and ORFB [41] in published DNA sequences containing other genes or genetic elements. The *rps11* and *clpP* genes

Table 3. Summary of gene mapping hybridizations to the *C. moewusii* cpDNA.

Gene probe	*C. moewusii* hybridizing fragments		
	Ava I	*Bst* EII	*Eco* RI
chlL	7	7	2'''
ORF715	1	3	2'''
ORFB	5	1'	15
psaC	7	4	2'''
psaJ	17	11	10
psbH	2	18''	18
psbI	30	19''	19
psbK	8'	10	4
rpl20	6''	28''	4
rpl23	15	6'	5
*rpoC2*a	4	1	2''
*rpoC2*b	4	23	6
rps2	2	5	23'
rps4	18''	6'	5
rps11	13	14	16'
rps19	14	20'	5
trnC-GCA	8'	10	6
trnD-GUC	2	5	23'
trnE1-UUC	2	18''	34''a
trnE2-UUC	5	1'	28'
trnH-GUG	5	1'	28'
trnI-CAU	6	4	16
trnL-UAG	6	4	16
trnfM-CAU	6	4	16
trnM-CAU	13	25	16'
trnS-GCU	2'	24	10'
trnS-UGA	6''	8	4
trnW-CCA	8'	10	4
trnY-GUA	13	14	27
ycf3	2	13	3
ycf4	2	13	3
ycf5	2'	1	20
ycf8	2	5	18
ycf12	13	14	16'

[a] This *Eco* RI fragment of 200 bp was not previously mapped [50]; it is located between the *Eco* RI fragments 3 and 18 on the physical map presented in Fig. 1.

chloroplast *Bam* HI, *Eco* RI and *Pst* I fragments (Table 4). The *C. reinhardtii* probes specific to

Table 4. Summary of gene mapping hybridizations to the *C. reinhardtii* cpDNA.

Gene probe	*C. reinhardtii* hybridizing fragments [a]		
	Bam HI	*Eco* RI	*Pst* I
chlB	7	1	24
ORF715	1	23	3
ORFA	10	33	3
psbH	6	19	7
*rpoC1*a	1	1	9
*rpoC1*b	1	6	19
*rpoC2*b	8	10	2
rps2	6	5	4
rps7 5'	4	7	8
rps7 3'	4	7	8
trnD-GUC	6	5	14
trnM-CAU	3	2	4
trnS-GCU	4	24	8
trnY-GUA	8	10	2
ycf3	3	30	4
ycf4	3	30	4
ycf5	13	7	7
ycf12	4	24	8

[a] The numbering system for the *Bam* HI and *Eco* RI fragments is that of Grant *et al.* [18].

tscA, trnG and *trnR*-ACG failed to hybridize to the *C. moewusii* cpDNA, while the *C. moewusii trnR*-UCU and *trnT* probes did not hybridize to the *C. reinhardtii* cpDNA.

Hybridizations of the *C. reinhardtii* cpDNA with 5' and 3' *rps7*-specific fragments from *C. eugametos* were found to be consistent with our sequence analysis of the *C. eugametos rps7* in indicating that the 5' and 3' coding regions of this gene are tightly linked. Probes derived from the 5' and 3' ends of the *C. eugametos rps7* indeed recognized a unique locus on the *C. reinhardtii* cpDNA; i.e., the *Eco* RI fragment 7 which was previously shown by sequencing to encode the 3' end of the gene [39]. It thus appears that the positive signal that was detected on the opposite

were sequenced in the laboratories of S.J. Surzycki (personal communication) and X.-Q. Liu (personal communication), respectively, whereas *ycf4, rps18* and *trnI* (CAU) were identified in our laboratory [our unpublished data]. Note that it has not been demonstrated that the *C. reinhardtii rpoC1* features two ORFs; however, the regions corresponding to the separate loci identified in *C. moewusii* have been designated as in this alga (see Fig. 1). This figure shows only one of the two possible isomers of the chloroplast genome.

single-copy region (*Eco* RI fragment 27) with an *Euglena gracilis* probe specific to the 5' end of *rps7* [44] is due to unspecific hybridization.

Discussion

Gene content and structure of individual genes in the C. moewusii and C. reinhardtii cpDNAs

Previous heterologous hybridizations and sequencing studies have led to the positioning of 20 genes on the *C. moewusii* cpDNA [50, 51] and of 56 genes on the divergent *C. reinhardtii* cpDNA [21]. In the present study, we have extended the number of mapped genes on each of these green algal cpDNAs to 74 and 75, respectively, by partial sequencing of the *C. eugametos* and *C. moewusii* cpDNAs and by hybridizations with *Chlamydomonas* chloroplast gene-specific sequences (see Figs. 1 and 2). The total length of cpDNA sequences determined so far in these two interfertile algae (about 140 bp) accounts for approximately half of the size of their chloroplast genome, and the number of genes identified represents about half of the gene content of land plant cpDNAs. The divergent *C. moewusii* and *C. reinhardtii* cpDNAs appear to share a similar gene complement as most of the chloroplast genes initially identified in *C. moewusii* were mapped in *C. reinhardtii* and vice versa. The few genes (*tscA* and four tRNA genes) that were positioned in only one of these two green algal cpDNAs are probably not sufficiently conserved in sequence to be localized by heterologous hybridizations. With the exception of *atpH*, the presence of all of the genes mapped so far on the *Chlamydomonas* chloroplast genome has been confirmed by partial or complete sequencing of the *C. eugametos*/ *C. moewusii* or *C. reinhardtii* coding regions. Of these genes, *rpl5*, *tscA*, the ORF715, ORFA and ORFB have not been reported in any land plant cpDNAs. Note that the *Chlamydomonas* *rpoB1* and *rpoB2* [13], *rpoC1a* and *rpoC1b* [our unpublished results], *rpoC2a* [13] and *rpoC2b* [our unpublished results] correspond to the land plant chloroplast *rpoB*, *rpoC1*, and *rpoC2* genes, respec-

tively. Four of the *Chlamydomonas* genes (*chlB*, *chlL*, *chlN* and *ycf12*) have been identified in *M. polymorpha* [34], but not in tobacco [45] and rice [23]. Of the genes encoded by land plant cpDNAs, those specifying subunits of the chlororespiratory NADH dehydrogenase (*ndhA* through *ndhK*) [1] may be absent from the *Chlamydomonas* chloroplast genome, as the large amount of chloroplast sequences characterized so far in both *C. eugametos*/*C. moewusii* and *C. reinhardtii* have not disclosed any similarity to such genes. It thus appears that a very small fraction of the increased size of these green algal DNAs relative to their land plant counterparts is explained by the presence of additional genes. Of the five genes that fall into this category, *rpl5* is the only one that is part of the larger set of chloroplast genes found in some early-diverging photosynthetic eukaryotes. It has been observed in the euglenophytes *E. gracilis* [10] and *Astasia longa* [47], in the red alga *Porphyra purpurea* [37] as well as in the protist *Cyanophora paradoxa* [7], and may thus be considered as an ancestral gene that has been retained by *Chlamydomonas*.

Sequence analyses suggest that the extra size of the *Chlamydomonas* chloroplast genome relative to its land plant counterpart is mainly accounted for by the presence of enlarged spacers between coding regions and also by the presence of unusually long genes featuring introns or other sequence elements that, in most cases, do not disrupt coding regions. For example, the spacer between the *C. eugametos* *rpl14* and *rpl16* genes is 1270 bp long [our unpublished results], whereas the corresponding spacer in *M. polymorpha* comprises only 98 bp [34]. As discussed below, the insertion of repeated sequences in spacers of ancestral green algal cpDNAs may have been a determinant factor in promoting the disruption of the multicistronic transcription units present in these DNAs. To date, a single gene (*psaA*) has been found to be interrupted by group II introns in the *C. eugametos*/*C. moewusii* [our unpublished results] and *C. reinhardtii* [see 40] cpDNAs, whereas five genes (*rrnS*, *rrnL* and the strongly transcribed protein-coding genes *psbA*, *psaB* and *psbC*) have been reported to contain optional

group I introns [see 55], some of which feature long ORFs encoding site-specific DNA endonucleases or potential proteins of unknown function [see 55]. As no similar introns reside in the corresponding genes of land plants, earlier-diverging eukaryotes and cyanobacteria, it appears that these elements have been inserted relatively late during the evolution of the green algal chloroplast genome. A third group of sequence elements has been identified exclusively in some chloroplast genes of *C. reinhardtii* (*rps3* [14], *rpoC2a* [13] and *clpP* [see 57]) and also appear to be of recent origin. These elements, which have not been characterized at the RNA and protein levels, code for potential proteins of 100–550 amino acids with no homology to any known proteins. As they are juxtaposed in-frame with coding regions of known identity and do not share any similarity with introns, chimeric proteins with domains of unknown identity are expected to result from translation of the transcripts derived from these DNA sequences. For the *C. reinhardtii clpP*, it has been suggested that the foreign coding region is spliced at the protein level [57].

Considering that the *C. reinhardtii* chloroplast *rpoB*, *rpoC2* and *rps3* display unusual structures and produce undetectable levels of transcripts (at least under the growth conditions examined) [13, 14], it is possible that some of these genes and others in the *Chlamydomonas* chloroplast genome are not functional. Site-directed mutagenesis of these three genes and analyses of their products will be necessary to determine if they encode essential proteins. In the case of *rpoC2*, functional studies have already been reported by Goldschmidt-Clermont [15], who disrupted a presumed *Chlamydomonas*-specific ORF (ORF472) corresponding to the 3' end of the *C. reinhardtii rpoC2* sequence determined by Fong and Surzycki [13] and designated here as *rpoC2a*. This region was found to be essential for cell viability even on acetate medium [15], a result which is consistent with recent studies indicating that photosynthetic chloroplast genes are transcribed by the cpDNA-encoded RNA polymerase, whereas non-photosynthetic genes are transcribed by an enzyme encoded in the nucleus [22]. However, because the

C. reinhardtii rpoC2a sequence does not encode one of the conserved regions (region VI) found in all archaebacterial, bacterial and land plant RNA polymerases and features at its extreme 3' end a non-coding region that bears no similarity to intron consensus sequences, it is not clear how this gene is expressed in a functional protein. Our preliminary data suggest that the *Chlamydomonas rpoC2* resembles *rpoB* in being divided into two genes that encode separate polypeptides [our unpublished data]. Indeed, during the course of this work, a *C. eugametos* cpDNA clone showing sequence similarity to the region VI of land plant chloroplast *rpoC2* and also to the *C. reinhardtii* cpDNA region containing the ARS element 04 [56] was found to map, in both *C. moewusii* and *C. reinhardtii*, to sequences downstream of *rpoC2a*. It will be interesting to carry out detailed sequence and functional analyses of this new coding region designated as *rpoC2b*. With regard to *rpoC1*, our sequence analysis has revealed that, in *C. eugametos*, this gene resembles *rpoB* and *rpoC2* in featuring two separate ORFs that are closely linked (our unpublished results).

Although some *Chlamydomonas* chloroplast genes may be inactive, this does not appear to be the case for *tscA*, the ORF715 and ORFB, which have been identified uniquely in the *Chlamydomonas* chloroplast genome. Site-directed gene disruption experiments have clearly demonstrated that the small RNA encoded by *tscA* is required *in trans* for the *trans*-splicing of *psaA* exons 1 and 2 [see 40]. This RNA probably base-pairs with the separate exon 1 and exon 2 transcripts to form the characteristic structure of group II introns [40]. On the other hand, the ORF715 and ORFB are believed to be functional because they were found to be transcriptionally active and easily detectable by heterologous hybridization in all nine *Chlamydomonas* taxa we have examined thus far (our unpublished results). It will be interesting to construct disruption mutants of these genes in order to see if they are required for a vital function.

In both *C. moewusii* and *C. reinhardtii*, the coding region of *psaA* is made up of three exons mapping to widely scattered regions of the

chloroplast genome (see Figs. 1 and 2). Although the *psaA* exons of these two divergent taxa reside in very different gene contexts, they show exactly the same coding information [40, our unpublished results], indicating that the most recent common ancestor of *Chlamydomonas* also possessed this uncommon structure. As reported for *C. reinhardtii* [40], separate precursor RNAs are undoubtedly produced from the *C. moewusii psaA* exons and assembled *in trans* to yield the mature *psaA* transcript. In all other cpDNAs investigated, with the notable exception of the *E. gracilis* cpDNA [see 20], *psaA* is unsplit and uninterrupted by introns. In the land plant chloroplast genome, only *rps12* has been found to feature a split organization in which exons are interspersed with other genes [40].
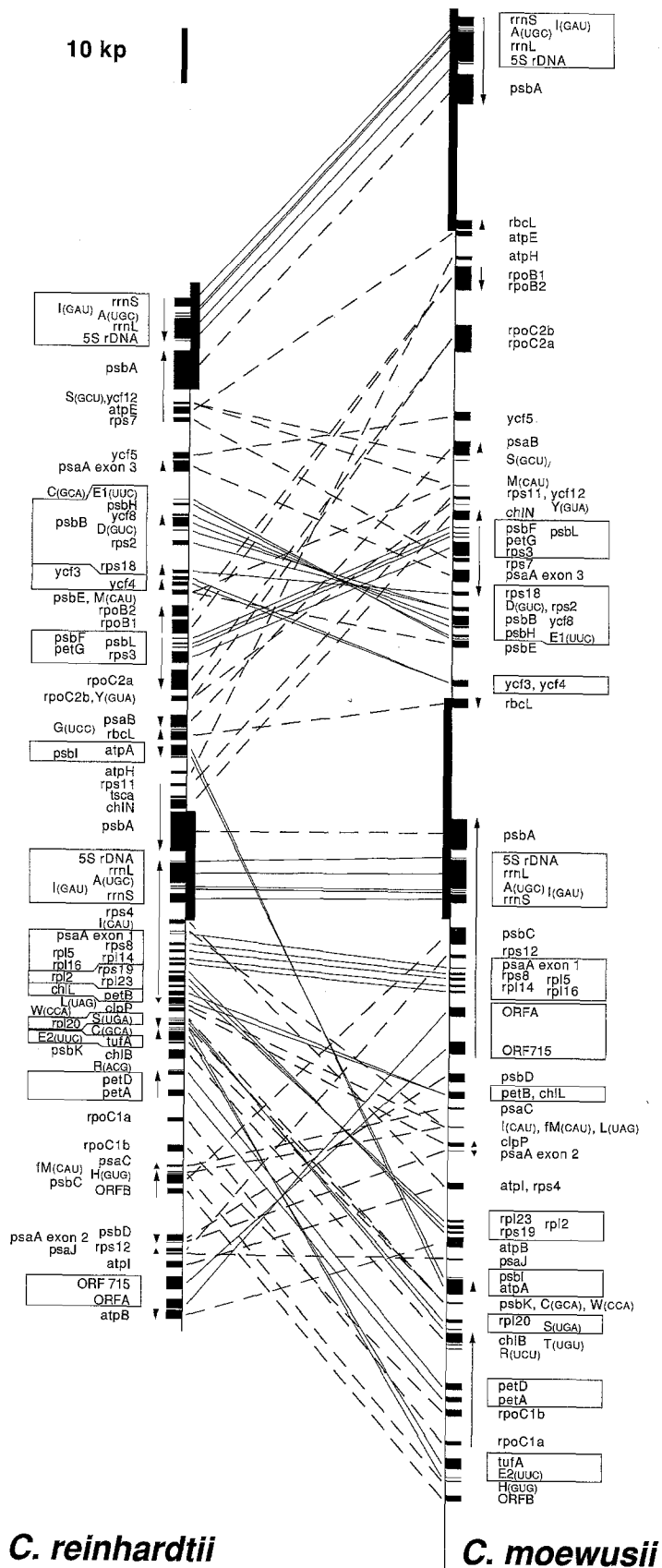
Comparative organization of the C. moewusii *and* C. reinhardtii *cpDNAs*

As shown in Fig. 3, the *C. moewusii* and *C. reinhardtii* cpDNAs are so extensively rearranged that the series of mutational events responsible for their scrambled gene order cannot be discerned. A single mutation, involving the inversion of the segment containing *rps18-rps2-trnD-psbB-ycf8-psbH-trnE1*, can be possibly distinguished. All gene rearrangements between the two green algal cpDNAs, with the exception of those accounting for the relocations of *rbcL* and the *atpA-psbI* pair, occurred within corresponding regions of the genome.

Although the *C. moewusii* and *C. reinhardtii* cpDNAs differ by complex sequence rearrange-ments, 38 genes scattered throughout the genome define 12 conserved clusters of closely linked loci: (1) *rrnS-trnI-trnA-rrnL*-5S rDNA; (2) *psbF-psbL-petG-rps3*; (3) *rps18-rps2-trnD-psbB-ycf8-psbH-trnE1*; (4) *ycf3-ycf4*; (5) *rpl16-rpl14-rpl5-rps8-psaA* exon 1; (6) ORF715-ORFA; (7) *petB-chlL*; (8) *rpl23-rpl2-rps19*; (9) *atpA-psbI*; (10) *trnS-rpl20*; (11) *petA-petD*; (12) *tufA-trnE2* (Fig. 3). For all of these clusters, with the exception of those consisting of *rps18-rps2-trnD-psbB-ycf8-psbH-trnE1*, *ycf3-ycf4* and *petB-chlL*, partial or complete sequence analysis of the genes from both green algae or from *C. reinhardtii* alone (see Figs. 1 and 2 and accompanying legends) has indicated that these genes are not only tightly linked, but that they are also encoded on the same DNA strand (see Fig. 3). It is therefore very likely that these nine gene clusters are colinear. All nine clusters were most probably present in the cpDNA of the most recent common ancestor of *C. moewusii* and *C. reinhardtii*, and it is also most probable that some of them formed multicistronic transcriptional units. Consistent with the latter hypothesis, four of the nine clusters show similarities to ancestral multicistronic operons found in land plant and algal genomes (see below). Comparative sequence analysis of the remaining three chloroplast gene clusters (*rps18-rps2-trnD-psbB-ycf8-psbH-trnE1*, *ycf3-ycf4* and *petB-chlL*) from *C. moewusii* and *C. reinhardtii* together with gene mapping in other *Chlamydomonas* taxa will be necessary to determine whether they represent continuous stretches of DNA sequences that were present in a common ancestor of *Chlamydomonas* taxa. Given the limited sequence data available, it is possible that the genes contained in these

Fig. 3. Comparative gene organization of the *C. moewusii* and *C. reinhardtii* cpDNAs. DNAs are drawn to scale and are linearized at one of the junctions of the inverted repeat (denoted by thick lines) and the single-copy region bordering the *rrnS* genes. Gene loci are denoted by dark areas, with their size reflecting the length of coding regions as determined from *Chlamydomonas* gene sequences or from their counterparts in *M. polymorpha*. Note, however, that the coding regions of *rrnL* and *psbA* from both green algae as well as those of the *C. moewusii rrnS*, *psaB* and *psbC* are oversized in this figure, as the intron sequences interrupting them were not discriminated. All corresponding *C. moewusii* and *C. reinhardtii* gene loci are connected by lines: those that are part of conserved clusters (framed areas) are linked by solid lines, whereas the remaining genes are connected by dashed lines. For all genes that are indicated on only one of the two green algal cpDNAs (either the *C. moewusii* or *C. reinhardtii* cpDNA), our heterologous hybridizations failed to identify their counterparts in the compared DNA. For each gene, the polarity of the DNA strand containing the coding region was denoted by an arrow when this information was available. Note that contiguous genes with the same polarity were assigned a common arrow. References for all *Chlamydomonas* chloroplast gene sequences are cited in the legends of Figs. 1 and 2.

10 kp

C. reinhardtii gene labels (left):

rrnS
I(GAU) A(UGC)
rrnL
5S rDNA

psbA

S(GCU),ycf12
atpE
rps7

ycf5
psaA exon 3

C(GCA)/E1(UUC)
psbH
psbB ycf8
D(GUC)
rps2

ycf3 rps18
ycf4
psbE, M(CAU)
rpoB2
rpoB1

psbF petG psbL
rps3

rpoC2a
rpoC2b, Y(GUA)

G(UCC) psaB
rbcL
psbI atpA

atpH
rps11
tsca
chlN

psbA

5S rDNA
rrnL
A(UGC)
I(GAU) rrnS

rps4
I(CAU)
psaA exon 1 rps8
rpl5 rpl14
rpl16 rps19
rpl2 rpl23
chlL
L(UAG) petB
W(CCA) clpP
rpl20 S(UGA)
E2(UUC) C(GCA)
psbK tufA
chlB
R(ACG)

petD
petA

rpoC1a

rpoC1b
psaC
fM(CAU) H(GUG)
psbC ORFB

psaA exon 2 psbD
psaJ rps12
atpI

ORF715
ORFA
atpB

C. moewusii gene labels (right):

rrnS I(GAU)
A(UGC)
rrnL
5S rDNA

psbA

rbcL
atpE
atpH
rpoB1
rpoB2

rpoC2b
rpoC2a

ycf5
psaB
S(GCU)
M(CAU)
rps11, ycf12
chlN Y(GUA)
psbF psbL
petG
rps3
rps7
psaA exon 3
rps18
D(GUC), rps2
psbB ycf8
psbH E1(UUC)
psbE

ycf3, ycf4
rbcL

psbA

5S rDNA
rrnL
A(UGC) I(GAU)
rrnS

psbC
rps12
psaA exon 1
rps8 rpl5
rpl14 rpl16

ORFA

ORF715

psbD
petB, chlL
psaC
I(CAU), fM(CAU), L(UAG)
clpP
psaA exon 2

atpI, rps4

rpl23 rpl2
rps19
atpB
psaJ
psbI
atpA
psbK, C(GCA), W(CCA)
rpl20 S(UGA)
chlB T(UGU)
R(UCU)

petD
petA
rpoC1b

rpoC1a
tufA
E2(UUC)
H(GUG)
ORFB

C. reinhardtii   **C. moewusii**

598

clusters are closely linked in both *C. moewusii* and *C. reinhardtii*, but show differences in polarity or gene order.

Considering that a fraction of the genes encoded by the *C. moewusii* and *C. reinhardtii* cpDNAs have not yet been mapped and that the relative orders of some genes could not be determined in the present study, it is possible that there exist additional conserved clusters between *Chlamydomonas* cpDNAs and that those already identified encompass longer segments of the chloroplast genome. For example, if the *C. moewusii clpP* and *trnL* as well as *psaC* and *trnfM* were found to be contiguous and in the same orientation(s) as in *C. reinhardtii*, these genes would define two new conserved clusters. Moreover, the *ycf3-ycf4* cluster could be enlarged to include *psbE* if these three genes were found to display the same arrangement in the two algae. Similarly, the *rpl20-trnS* cluster could be enlarged by adding *trnC*. Obviously, the precise identification of the number and extent of colinear segments between the *C. moewusii* and *C. reinhardtii* cpDNAs will require complete sequencing of these DNAs.

Most of the ancestral operons that characterize the plastid genome organization of land plants and earlier-diverging photosynthetic eukaryotes appear to have been disrupted before the emergence of *Chlamydomonas*. In both *C. moewusii* and *C. reinhardtii*, only 16 chloroplast genes are organized similarly to such operons and remarkably, all of them map to colinear segments between these two green algal cpDNAs. These *Chlamydomonas* genes featuring an ancestral organization (underlined) and the conserved clusters to which they are associated are as follows: (1) *rrnS-trnI-trnA-rrnL-5SrDNA*; (2) *psbF-psbL-petG-rps3*; (3) *rps18-rps2-trnD-psbB-ycf8-psbH-trnE1*; (4) *rpl23-rpl2-rps19*; (5) *rpl16-rpl14-rpl5-rps8-psaA* exon 1. Of the four corresponding operons in the land plant chloroplast genome, only the rRNA operon (*rrnS-trnI-trnA-rrnL*-5S rDNA) shows exactly the same gene content as the equivalent *Chlamydomonas* cluster. The remaining three operons (*psbE-psbF-psbL-psbJ*, *psbB-ycf8-psbH-petB-petD*, *rpl23-rpl2-rps19-rpl22-rps3-rpl16-rpl14-*

rps8-infA) feature a larger set of genes, indicating that remnants of these ancestral operons are present in *Chlamydomonas* cpDNAs. Since in two cases such remnant sequences map to conserved clusters containing additional *Chlamydomonas* genes with identical polarity in their downstream region (*psbF-psbL-petG-rps3* and *rpl16-rpl14-rpl5-rps8-psaA* exon 1), the gene rearrangements that led to the partial destruction of the corresponding ancestral operons might have involved not only the relocation of coding regions elsewhere on the cpDNA, but also the creation of novel transcription units by fusion of coding regions to the 3' portions of fragmented ancestral operons. Whether these derived gene organizations constitute transcription units in *Chlamydomonas* cpDNAs remains to be established. To date, a single study has been reported on the transcriptional organization of *Chlamydomonas* chloroplast gene clusters showing similarities with ancestral operons. This study, bearing on the *C. reinhardtii psbB-ycf8-psbH* cluster, revealed that *psbB* and *ycf8* are cotranscribed [33].

Of the multiple sequence rearrangements that marked the evolution of the *Chlamydomonas* chloroplast genome, one seems to have led to disruption of the ancestral region containing *rpl23*, *rpl2*, *rps19*, *rpl16*, *rpl14*, *rpl5*, *rps8* and the *psaA* exon 1. This gene cluster, which differs from the corresponding land plant operon by the absence of *rpl22*, *rps3* and *infA* and also by the presence of *rpl5* and the *psaA* exon 1, is found in *C. reinhardtii*, while it is divided into two separate fragments, *rpl23-rpl2-rps19* and *rpl16-rpl14-rpl5-rps8-psa* A exon 1, in *C. moewusii* (Fig. 3). These fragments differ in their polarity and are separated by approximately 42 kb in the single-copy region bordering the *rrnS* gene. If splitting of this ancestral region occurred only once, we expect that all green algae clustering in the lineage represented by *C. reinhardtii* will display the unsplit ribosomal protein gene cluster, whereas the others belonging to the lineage represented by *C. moewusii* will feature both the split and unsplit versions of this region or only the split version, depending upon the time when this gene rearrangement occurred relative to the divergence

time of these two major *Chlamydomonas* lineages. It may be significant that the ancestral green algal gene cluster was fragmented in the region corresponding to the *rps3* and *rpl22* loci in land plant cpDNAs. Given that the equivalent ribosomal protein operon of *E. gracilis* contains an additional reading frame (ORF516) between *rps3* and *rpl16* [20], it is possible that this particular region is prone to gene rearrangement simply because close linkage of *rps19* and *rpl16* is not a critical factor for gene expression as opposed to the other ribosomal protein-coding genes in the conserved *Chlamydomonas* gene cluster(s), which may be constrained to remain together for such an expression. However, these constraints may be inexistent and random events may have led to disruption of the ribosomal protein-coding gene cluster in *C. moewusii*. Two independent observations support this idea: first, the positions of most gene loci in relation to others in the *Chlamydomonas* chloroplast genome do not appear to be important for gene expression and, second, in the case of the two closely linked genes that are encoded by the same DNA strand in the conserved *petA-petD* cluster, recent studies have shown that they are transcribed independently in *C. reinhardtii* [42]. Also a few genes mapping to dispersed loci of the *C. moewusii* cpDNA (*psbD* and *psaA* exon 2 [40], *rps7* and *atpE* [39], *psaJ* and *rps12* [31, our unpublished data]) are known to be tightly linked and cotranscribed in the *C. reinhardtii* cpDNA.

Interestingly, the chloroplast *petA* and *petD* genes are also adjacent to each other and encoded by the same DNA strand in *Scenedesmus obliquus* [26, 27], a green alga (Chlorophyceae) having no alliance with the polyphyletic *Chlamydomonas* genus [9]. This observation strongly supports the notion that these genes were present not only in the most recent common ancestor of *C. moewusii* and *C. reinhardtii*, but also in the most recent common ancestor of *Chlamydomonas* and *Scenedesmus*. This gene cluster from the latter ancestor might have also included *trnR*-UCU, a gene found immediately downstream of *petD* in the *C. eugametos* cpDNA [our unpublished results], but separated from the latter locus, in the

C. moewusii cpDNA, by an extra sequence of 6 kb [2] (see Fig. 1). Like its *C. eugametos* homologue, the *Scenedesmus trnR*-UCU is closely linked to *petD* [26], and in these two algae as well as in *C. moewusii*, all three genes share the same polarity [2, 26, and our unpublished results]. Our heterologous hybridization with a *C. moewusii trnR*-UCU probe did not allow us to map this gene on the *C. reinhardtii* cpDNA; however, the region of this DNA situated downstream of *petD* was reported to contain *trnR*-ACG [60], a locus that could not be located by heterologous hybridization on the *C. moewusii* cpDNA. Assuming that the *petA-petD-trnR*-UCU cluster was present in the most recent common ancestor of *Chlamydomonas* and *Scenedesmus*, this observation implies that *trnR*-UCU was transferred to another locus after the divergence of the major *Chlamydomonas* lineages. Identification of the *petA-petD-trnR*-UCU cluster in other *Chlamydomonas* and non-*Chlamydomonas* taxa will be necessary to demonstrate without any ambiguity this hypothesis.

From our comparative analysis of the *C. moewusii* and *C. reinhardtii* cpDNAs and from already published studies of cpDNAs from other green algal groups, it is clear that green algal genomes are very plastic in their gene organization and appear to be much less constrained than their land plant counterparts to retain a compact gene organization; however, the molecular mechanisms underlying this great diversity remain unknown. The proliferation of repeated sequences in intergenic spacers of green algal cpDNAs may be related to their great plasticity [6]. Such elements have been proposed to play an active role in chloroplast gene rearrangements, as they are present in increased amounts in the most highly rearranged land plant cpDNAs and are located close to endpoints of inversions in some species [see 36]. It has been speculated that they act as transposable elements and/or, alternatively, that they mediate rearrangements by facilitating homologous intra- and inter-molecular recombination events [see 36]. Assuming that they promote gene rearrangements in green algal cpDNAs, specific sequences seem to be unnecessary as the

C. moewusii and *C. reinhardtii* cpDNAs do not apparently share the same families of dispersed repeats as evidenced by hybridization [6] and sequence data. In addition to short dispersed repeats, tRNA genes have been implicated in gene reshuffling in land plant cpDNAs via intermolecular recombination events [36]. To demonstrate clearly the role of these two types of sequence elements in the tremendous rearrangements that occurred during the evolution of the *Chlamydomonas* chloroplast genome, it will be essential to analyze the endpoints of rearranged cpDNA segments from very closely related taxa showing specific mutations.

In future studies, it will also be important to examine green algal lineages that are basal relative to *Chlamydomonas* in order to establish whether the derived gene clusters and peculiar gene organizations observed in the *C. moewusii* and *C. reinhardtii* cpDNAs are features specific to this polyphyletic genus or of a larger taxonomic group. In addition, such studies should reveal whether some green algal lineages have retained a large number of evolutionarily primitive cotranscribed genes and are evolving under the same constraints as their land plant counterparts.

References

1. Arizmendi JM, Runswick MJ, Skehel JM, Walker JE: NADH:ubiquinone oxidoreductase from bovine heart mitochondria. A fourth nuclear encoded subunit with a homologue encoded in chloroplast genomes. FEBS Lett 301: 237–242 (1992).
2. Bergeron A: Analyse structurale d'un ADN linéaire de 6 kilopaires de bases chez *Chlamydomonas moewusii*. M.Sc. thesis, Université Larval (1990).
3. Bergeron A, Boulanger J, Turmel M: Nucleotide sequence of the chloroplast *petD* gene of *Chlamydomonas eugametos*. Nucl Acids Res 17: 3593 (1989).
4. Berry-Lowe SL, Johnson CH, Schmidt GW: Nucleotide sequence of the *psbB* gene of *Chlamydomonas reinhardtii* chloroplasts. Plant Physiol 98: 1541–1543 (1992).
5. Birnboim HC, Doly J: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl Acids Res 7: 1513–1523 (1979).
6. Boynton JE, Gillham NW, Newman SM, Harris EH: Organelle genetics and transformation of *Chlamydomonas*. In: Hermann RG (ed) Plant Gene Research, vol VI, pp. 3–64, Springer-Verlag, Vienna (1992).
7. Bryant DA, Stirewalt VT: The cyanelle genome of *Cyanophora paradoxa* encodes ribosomal proteins encoded by the chloroplast genome of higher plants. FEBS Lett 259: 273–280 (1990).
8. Buchheim MA, Turmel M, Zimmer EA, Chapman RL: Phylogeny of *Chlamydomonas* (Chlorophyta) based on cladistic analysis of nuclear 18S rRNA sequence data. J Phycol 26: 689–699 (1990).
9. Buchheim MA, Chapman RL: Phylogeny of *Carteria* (Chlorophyceae) inferred from molecular and organismal data. J Phycol 28: 362–374 (1992).
10. Christopher DA, Hallick RB: *Euglena gracilis* chloroplast ribosomal protein operon: A new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. Nucl Acids Res 17: 7591–7608 (1989).
11. Devereux J, Haeberli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12: 387–395 (1984).
12. 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 15: 277–282 (1989).
13. Fong SE, Surzycki SJ: Chloroplast RNA polymerase genes of *Chlamydomonas reinhardtii* exhibit an unusual structure and arrangement. Curr Genet 21: 485–497 (1992).
14. Fong SE, Surzycki SJ: Organization and structure of

plastome *psbF*, *psbL*, *petG* and ORF712 genes in *Chlamydomonas reinhardtii*. Curr Genet 21: 527–530 (1992).

15. Goldschmidt-Clermont M: Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker for site-specific directed transformation of *Chlamydomonas*. Nucl Acids Res 19: 4083–4089 (1991).

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

17. Gowans CS: Genetics of *Chlamydomonas moewusii* and *Chlamydomonas eugametos*. In: Lewin RA (ed) The Genetics of Algae, Botanical Monographs, vol. 12, pp. 145–173. Blackwell Scientific Publications, Oxford (1976).

18. Grant DM, Gillham NW, Boynton JE: Inheritance of chloroplast DNA in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA 77: 6067–6071 (1980).

19. Gray MW: The endosymbiont hypothesis revisited. Int Rev Cytol 141: 233–357 (1992).

20. Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E: Complete sequence of *Euglena gracilis* chloroplast DNA. Nucl Acids Res 21: 3537–3544 (1993).

21. Harris EH: *Chlamydomonas reinhardtii* chloroplast genome. In: O'Brien SJ (ed) Genetic Maps, 6th ed., Book 2, pp. 165–168. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1993).

22. Hess WR, Prombona A, Fieder B, Subramanian AR, Börner T: Chloroplast *rps15* and the *rpoB/C1/C2* gene cluster are strongly transcribed in ribosome-deficient plastids: evidence for a functioning non-chloroplast-encoded RNA polymerase. EMBO J 12: 563–571 (1993).

23. Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C-R, Meng B-Y, Li Y-Q, Kanno A, Nishizawa Y, Harai A, Shinozaki K, Sugiura M: The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217: 185–194 (1989).

24. Huang C, Liu X-Q: Nucleotide sequence of the *frxC*, *petB* and *trnL* genes in the chloroplast genome of *Chlamydomonas reinhardtii*. Plant Mol Biol 18: 985–988 (1992).

25. Johnson CH, Schmidt GW: Nucleotide sequence of the *psbH* gene of *Chlamydomonas reinhardtii*. Accession number Z15133.

26. Kück U: The intron of a plastid gene from a green alga contains an open reading frame for a reverse transcriptase-like enzyme. Mol Gen Genet 218: 257–267 (1989).

27. Kück U, Godehardt I, Schmidt U: A self-splicing group II intron in the mitochondrial large subunit rRNA (LSUrRNA) gene of the eukaryotic alga *Scenedesmus obliquus*. Nucl Acids Res 18: 2691–2697 (1990).

28. Lemieux B, Lemieux C: Extensive sequence rearrangements in the chloroplast genomes of the green algae *Chlamydomonas eugametos* and *Chlamydomonas reinhardtii*. Curr Genet 10: 213–219 (1985).

29. Lemieux C, Turmel M, Seligy VL, Lee RW: The large subunit of ribulose-1,5-bisphosphate carboxylase-oxygenase is encoded in the inverted repeat sequence of the *Chlamydomonas eugametos* chloroplast genome. Curr Genet 9: 139–145 (1985).

30. Leu S, Schlesinger J, Micheals A, Shavit N: Complete DNA sequence of the *Chlamydomonas reinhardtii* chloroplast *atpA* gene. Plant Mol Biol 18: 613–616 (1992).

31. Liu X-Q, Gillham NW, Boynton JE: Chloroplast ribosomal protein gene *rps12* of *Chlamydomonas reinhardtii*: wild-type sequence, mutation to streptomycin resistance and dependence, and function in *Escherichia coli*. J Biol Chem 264: 16100–16108 (1989).

32. Mattox K, Stewart K: Classification of the green algae: a concept based on comparative cytology. In: Irvine DEG, John DM (eds) Systematics of the Green Algae, pp. 29–72, Academic Press, London (1984).

33. Monod C, Goldschmidt-Clermont M, Rochaix J-D: Accumulation of chloroplast *psbB* RNA requires a nuclear factor in *Chlamydomonas reinhardtii*. Mol Gen Genet 231: 449–459 (1992).

34. Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H: Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha*. Nature 322: 572–574 (1986).

35. O'Neill GP, Schön A, Chow H, Chen M-W, Kim Y-C, Söll D: Sequence of tRNA^{Glu} and its genes from the chloroplast genome of *Chlamydomonas reinhardtii*. Nucl Acids Res 18: 5893 (1990).

36. Palmer JD: Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil IK (eds) Cell Culture and Somatic Cell Genetics of Plants, vol 7a: The Molecular Biology of Plastids, pp. 5–53, Academic Press, San Diego (1991).

37. Reith M, Munholland J: A high-resolution gene map of the chloroplast genome of the red alga *Porphyra purpurea*. Plant Cell 5: 465–475 (1993).

38. Richard M, Bellemare G: Nucleotide sequence of *Chlamydomonas moewusii* chloroplastic tRNA-Thr. Nucl Acids Res 18: 3061 (1990).

39. Robertson D, Boynton JE, Gillham NW: Cotranscription of the wild-type chloroplast *atpE* gene encoding the CF_1/CF_0 epsilon subunit with the 3' half of the *rps7* gene in *Chlamydomonas reinhardtii* and characterization of frameshift mutations in *atpE*. Mol Gen Genet 221: 155–163 (1990).

40. Rochaix J-D: Post-transcriptional steps in the expression of chloroplast genes. Annu Rev Cell Biol 8: 1–28 (1992).

41. Rochaix J-D, Kuchka M, Mayfield S, Schirmer-Rahire M, Girard-Bascou J, Bennoun P: Nuclear and chloroplast mutations affect the synthesis or stability of

the chloroplast *psbC* gene product in *Chlamydomonas reinhardtii*. EMBO J 8: 1013–1021 (1989).

42. Sakamoto W, Kindle KL, Stern DB: *In vivo* analysis of *Chlamydomonas* chloroplast *petD* gene expression using stable transformation of β-glucuronidase translational fusions. Proc Natl Acad Sci USA 90: 497–501 (1993).

43. Schlösser UG: Species-specific sporangium autolysins (cell-wall-dissolving enzymes) in the genus *Chlamydomonas*. In: Irvine DEG, John DM (eds) Systematics of the Green Algae, pp. 409–418. Academic Press, London (1984).

44. Schmidt RJ, Hosler JP, Gillham NW, Boynton JE: Biogenesis and evolution of chloroplast ribosomes: cooperation of nuclear and chloroplast genes. In: Steinbeck KE, Bonitz S, Arntzen CJ, Bogorad L (eds) Molecular Biology of the Photosynthetic Apparatus, pp. 417–427, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1985).

45. Shinozaki N, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M: The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5: 2043–2049 (1986).

46. Silk GW, Dela Cruz F, Wu M: Nucleotide sequence of the chloroplast gene for the 4 kD K polypeptide of photosystem II (*psbK*) and the *psbK-tufA* intergenic region of *Chlamydomonas reinhardtii*. Nucl Acids Res 18: 4930 (1990).

47. Siemeister DA, Buchholz C, Hachtel W: Genes for the ribosomal proteins are retained on the 73 kb DNA from *Astasia longa* that resembles *Euglena* chloroplast DNA. Curr Genet 18: 457–464 (1990).

48. Sprinzl M, Hartmann T, Weber J, Blank J, Zeidler R: Compilation of tRNA sequences and sequences of tRNA genes. Nucl Acids Res 17 (suppl): r1–r172 (1989).

49. Takahashi Y, Goldschmidt-Clermont M, Soen S-Y, Franzén LG, Rochaix J-D: Directed chloroplast transformation in *Chlamydomonas reinhardtii*: insertional inactivation of the *psaC* gene encoding the iron sulfur protein destabilizes photosystem I. EMBO J 10: 2033–2040 (1991).

50. Turmel M, Bellemare G, Lemieux C: Physical mapping of differences between the chloroplast DNAs of the interfertile algae *Chlamydomonas eugametos* and *Chlamydomonas moewusii*. Curr Genet 11: 543–552 (1987).

51. Turmel M, Lemieux B, Lemieux C: The chloroplast genome of the green alga *Chlamydomonas moewusii*: localization of protein-coding genes and transcriptionally active regions. Mol Gen Genet 214: 412–419 (1988).

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

53. 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 218: 293–311 (1991).

54. 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 232, 446–467 (1993).

55. Turmel M, Mercier J-P, Côté M-J: Group I introns interrupt the chloroplast *psaB* and *psbC* and the mitochondrial *rrnL* gene in *Chlamydomonas*. Nucl Acids Res 21: 5242–5250 (1993).

56. Vallet J-M, Rahire M, Rochaix J-D: Localization and sequence analysis of chloroplast DNA sequences of *Chlamydomonas reinhardtii* that promote autonomous replication in yeast. EMBO J 3: 415–421 (1984).

57. Weeks DP: *Chlamydomonas*: an increasingly powerful model plant cell system. Plant Cell 4: 871–878 (1992).

58. Woessner JP, Gilham NW, Boynton JE: The sequence of the chloroplast *atpB* gene and its flanking regions in *Chlamydomonas reinhardtii*. Gene 44: 17–28 (1986).

59. Yan RCA, Dove M, Seligy VL, Lemieux C, Turmel M, Narang SA: Complete nucleotide sequence and mRNA-mapping of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) from *Chlamydomonas moewusii*. Gene 50: 259–270 (1986).

60. Yu W, Spreitzer RJ: Sequences of the *trnR*-ACG and *petD* that contain a tRNA-like element within the chloroplast genome of *Chlamydomonas reinhardtii*. Nucl Acids Res 19: 957 (1992).

61. Yu W, Zhang D, Spreitzer RJ: Sequences of the *Chlamydomonas reinhardtii* chloroplast genes encoding tRNA^{Ser} and ribosomal protein L20. Plant Physiol 100: 1079–1080 (1992).

62. Zhang D, Spreitzer RJ: Nucleotide sequences of the *Chlamydomonas reinhardtii* chloroplast genes for tryptophan and glycine transfer RNAs. Nucl Acids Res 17: 8873 (1989).