Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: The existence of genes possibly involved in chloroplast division

(DNA sequence/evolution/gene map/*minD*/*rpoA*)

Tatsuya Wakasugi*†, Toshiyuki Nagai*, Meenu Kapoor*, Mamoru Sugita*, Mari Ito‡, Shiho Ito‡, Junko Tsudzuki‡, Keiko Nakashima‡, Takahiko Tsudzuki§, Yasuhiko Suzuki¶, Akira Hamada¶, Tutomu Ohta¶, Atsushi Inamura¶, Koichi Yoshinaga¶, and Masahiro Sugiura*∥

*Center for Gene Research, Nagoya University, Nagoya 464-01, Japan; ‡School of Life Studies, Sugiyama Jogakuen University, Nagoya 464, Japan; §Data Processing Center, Aichi-Gakuin University, Nisshin 470-01, Japan; and ¶Faculty of Science, Shizuoka University, Shizuoka 422, Japan

ABSTRACT   The complete nucleotide sequence of the chloroplast genome (150,613 bp) from the unicellular green alga *Chlorella vulgaris* C-27 has been determined. The genome contains no large inverted repeat and has one copy of rRNA gene cluster consisting of 16S, 23S, and 5S rRNA genes. It contains 31 tRNA genes, of which the tRNA^{Leu}(GAG) gene has not been found in land plant chloroplast DNAs analyzed so far. Sixty-nine protein genes and eight ORFs conserved with those found in land plant chloroplasts have also been found. The most striking is the existence of two adjacent genes homologous to bacterial genes involved in cell division, *minD* and *minE*, which are arranged in the same order in *Escherichia coli*. This finding suggests that the mechanism of chloroplast division is similar to bacterial division. Other than *minD* and *minE* homologues, genes encoding ribosomal proteins L5, L12, L19, and S9 (*rpl5, rpl12, rpl19,* and *rps9*); a chlorophyll biosynthesis Mg chelating subunit (*chlI*); and elongation factor EF-Tu (*tufA*), which have not been reported from land plant chloroplast DNAs, are present in this genome. However, many of the new chloroplast genes recently found in red and brown algae have not been found in *C. vulgaris*. Furthermore, this algal species possesses two long ORFs related to *ycf1* and *ycf2* that are exclusively found in land plants. These observations suggest that *C. vulgaris* is closer to land plants than to red and brown algae.

Structure and expression of the chloroplast genome have been studied in a number of plants. Gene content and the sequence of many genes in chloroplast DNA are relatively conserved among land plants and the Euglenophyta *Euglena gracilis* (1–3). However, recent analyses of the entire chloroplast genome from nongreen algae (4, 5) and the cyanelle genome of *Cyanophora paradoxa* (6) have revealed that this is not always the case. For example, the chloroplast genome of the red alga *Porphyra purpurea* contains over 70 new genes not found in land plant and *Euglena* chloroplasts (4, 7). It is generally believed that land plants evolved from green algae (8) and that during this evolution, extensive rearrangements occurred within the chloroplast genomes. To understand the process of chloroplast genome evolution, information on repeated sequences, intergenic regions, and pseudogenes in chloroplast DNA is extremely helpful. Therefore, entire nucleotide sequences of green algal chloroplast genomes have been awaited.

The unicellular green alga *Chlorella vulgaris* C-27 is the organism for which the synchronous culture was first developed (9), and this synchronous system has long been used for studies of the cell cycle from physiological and biochemical aspects (10). In addition, the *Chlorella* cell contains a single large chloroplast. Thus, analysis of the *Chlorella* chloroplast DNA provides a unique opportunity to investigate not only the evolution but also the cell cycle-dependent expression of chloroplast genes. Though interesting, information on chloroplast DNA from the genus *Chlorella* was fragmentary. With the aim of fulfilling the gap in our understanding of this green alga, we undertook the sequencing of the chloroplast genome of *C. vulgaris*. Here we report the complete nucleotide sequence of the *Chlorella* chloroplast genome and its entire gene organization. The most notable feature of the genome is the existence of two genes homologous to bacterial genes (*minD* and *minE*) related to cell division. This observation suggests the conservation of division mechanisms during evolution of chloroplasts from their ancestral prokaryotes.

MATERIALS AND METHODS

Chlorella vulgaris C-27 (formerly designated *Chlorella ellipsoidea* C-27 Tamiya's strain) was grown in M-4NA medium (9). Chloroplast DNA was prepared from chloroplast-rich fractions (11) and purified by CsCl equilibrium centrifugation (12). Chloroplast DNA was partially digested with either *Bgl*II or *Eco*RI and cloned into the Charomid 9–28 vector as described (13). A *Pst*I fragment (10.5 kb) containing *rbcL* (11) was used as a probe for initial screening of Charomid clones containing the chloroplast DNA region. Nineteen overlapping Charomid clones covering the total chloroplast genome were selected by linking with restriction and hybridization analyses. Subfragments from these inserts were cloned into Bluescript SK/KS, and DNA sequences were determined by the dideoxynucleotide chain-termination method manually and by using automated DNA sequencers. DNA sequences were analyzed using the GENETYX program (Software Development Co., Tokyo) on personal computers and an HP 9000-845 computer (Hewlett–Packard).

RESULTS AND DISCUSSION

Overall Genome Organization. The circular chloroplast DNA of *Chlorella vulgaris* C-27 is 150,613 bp long, and its gene

map is shown in Fig. 1. The sequence is numbered counter-clockwise from the 5′-protruding end (A) of the *Eco*RI-cleavage site in *ycf10*. Overall GC content of the chloroplast DNA is 31.6%, which is closer to those of *Odontella* (31.8%; ref. 5), cyanelle (30.4%; ref. 6), and *Porphyra* (33%; ref. 4) than to angiosperms (38–39%; refs. 14–16). *C. vulgaris* chloroplast DNA contains no large inverted repeat, commonly found in many chloroplast DNAs (1), and hence possesses one copy of rRNA gene cluster. This situation is different from that of *C. ellipsoidea* C-87, in which rRNA genes are located in the inverted repeat of 22.5 kbp and are split into two operons: operon 1, *rrn16–trnI*, and operon 2, *trnA–rrn23–rrn5* (17, 18). In spite of lacking the inverted repeat and other significant repeated sequences, the genome size (151 kbp) of *C. vulgaris* chloroplasts is similar to that of angiosperm chloroplasts (1, 14–16), indicating that this genome has 20–30 kbp additional unique sequences with respect to those of angiosperms.

Identified genes and conserved ORFs (*ycf*s) are listed in Table 1. So far, 103 gene species and 8 *ycf*s, among which only two tRNA genes are duplicated, have been identified. The *Chlorella* genome has 10 genes not found in land plant chloroplasts, whereas it lacks all 11 *ndh* genes found in photosynthetic land plants except black pine (19). Therefore, the number of known genes is similar to that of land plants (14–16, 19, 20) and also of *Euglena* (21). Many new genes not found in land plant chloroplasts have been reported in red and brown algal chloroplasts and cyanelles (4–6). A limited num-

ber of such genes are present in the *Chlorella* genome (see Table 1).

Unlike *Euglena* chloroplast DNA, which has at least 155 introns (22), *C. vulgaris* possesses only three split genes, each containing a single group I intron (see below). Only one intron is present in the cyanelle genome (in *trnL*-UAA; ref. 6), whereas no introns have been found in *Porphyra* (4) and *Odontella* (5) chloroplast genomes. Therefore, intron contents are highly diverse in algal species, although they are conserved in land plants (≈20 introns; ref. 2).

Extensive rearrangements of the *Chlorella* chloroplast genome are evident relative to tobacco as well as to *Euglena*. However, several clusters are highly conserved in the entirely sequenced genomes (*rrn* and *rpl23* operons, *rps2-atpIHFA*, *atpBE*, *rpoBC₁C₂*, *psaAB*, *psbBTNH*, *petBD*, *psbEFLJ*, *psbDC*, and *rps12/7*). Therefore, the genome shuffling occurred mostly as blocks, which might have been necessary for genome function.

Genes Potentially Involved in Chloroplast Division. The most notable feature of the *Chlorella* chloroplast genome is the existence of genes related to those for *Escherichia coli* cell division. As shown in Fig. 2, the amino acid sequences deduced from ORF282 and the following ORF127 shows 39% and 22% identities with the predicted products from *E. coli minD* and *minE* genes, respectively (23). The *E. coli* minicell locus consists of three genes, *minC* (231 codons), *minD* (269 codons), and *minE* (88 codons), in this order, whose coordinate expression is required for the placement of a division
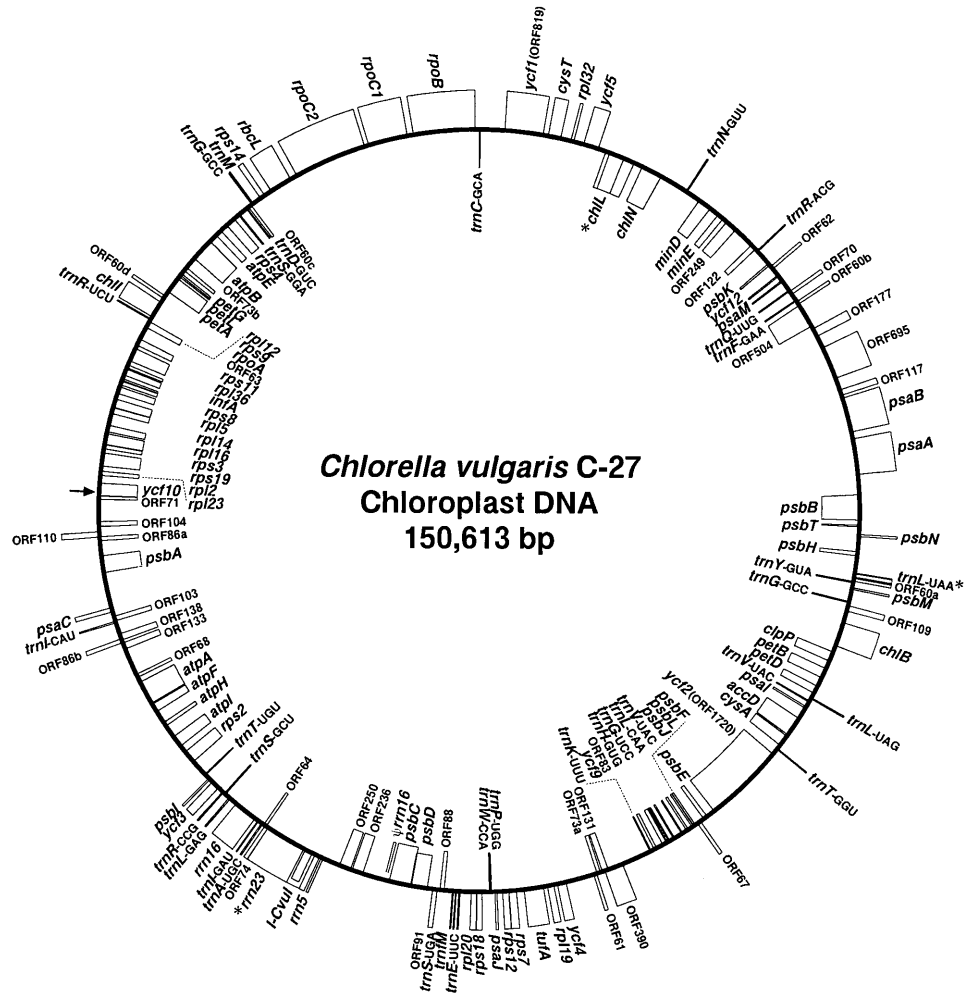

FIG. 1.   Gene map of the *C. vulgaris* C-27 chloroplast genome. Genes shown on the inside of the circle are transcribed clockwise, and genes on the outside are transcribed counterclockwise. ORFs of ≥60 codons are included. Asterisks denote split genes. Nucleotide positions are numbered counterclockwise from the arrow (position 1) in *ycf10*.

Table 1. Gene list of *C. vulgaris* C-27 chloroplast DNA

<table>
<tr><td colspan="5" align="center">RNA genes</td></tr>
<tr><td colspan="5">Ribosomal RNAs</td></tr>
<tr><td>23SrDNA*</td><td>16SrDNA</td><td>5SrDNA</td><td></td><td></td></tr>
<tr><td colspan="5">Transfer RNAs</td></tr>
<tr><td>trnA(UGC)</td><td>trnC(GCA)</td><td>trnD(GUC)</td><td>trnE(UUC)</td><td>trnF(GAA)</td></tr>
<tr><td>trnG(GCC) ×2</td><td>trnG(UCC)</td><td>trnH(GUG)</td><td>trnI(CAU)</td><td>trnI(GAU)</td></tr>
<tr><td>trnK(UUU)</td><td>trnL(CAA)</td><td>trnL(GAG)†</td><td>trnL(UAA)*</td><td>trnL(UAG)</td></tr>
<tr><td>trnfM(CAU)</td><td>trnM(CAU)</td><td>trnN(GUU)</td><td>trnP(UGG)</td><td>trnQ(UUG)</td></tr>
<tr><td>trnR(ACG)</td><td>trnR(CCG)</td><td>trnR(UCU)</td><td>trnS(GGA)</td><td>trnS(GCU)</td></tr>
<tr><td>trnS(UGA)</td><td>trnT(GGU)</td><td>trnT(UGU)</td><td>trnV(UAC) ×2</td><td>trnW(CCA)</td></tr>
<tr><td>trnY(GUA)</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="5" align="center">Protein genes</td></tr>
<tr><td colspan="5">Photosynthesis</td></tr>
<tr><td colspan="5">Photosystem I</td></tr>
<tr><td>psaA</td><td>psaB</td><td>psaC</td><td>psaI</td><td>psaM</td></tr>
<tr><td colspan="5">Photosystem II</td></tr>
<tr><td>psbA</td><td>psbB</td><td>psbC</td><td>psbD</td><td>psbE</td></tr>
<tr><td>psbF</td><td>psbH</td><td>psbI</td><td>psbJ</td><td>psbK</td></tr>
<tr><td>psbL</td><td>psbM</td><td>psbN</td><td>psbT</td><td></td></tr>
<tr><td colspan="5">Cytochrome</td></tr>
<tr><td>petA</td><td>petB</td><td>petD</td><td>petG</td><td>petL</td></tr>
<tr><td colspan="5">ATP synthase</td></tr>
<tr><td>atpA</td><td>atpB</td><td>atpE</td><td>atpF</td><td>atpH</td></tr>
<tr><td>atpI</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="5">Chlorophyll biosynthesis</td></tr>
<tr><td>chlB</td><td>chlI†</td><td>chlL*</td><td>chlN</td><td></td></tr>
<tr><td colspan="5">Rubisco</td></tr>
<tr><td>rbcL</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="5">Ribosomal proteins</td></tr>
<tr><td colspan="5">Large subunits</td></tr>
<tr><td>rpl2</td><td>rpl5†</td><td>rpl12†</td><td>rpl14</td><td>rpl16</td></tr>
<tr><td>rpl19†</td><td>rpl20</td><td>rpl23</td><td>rpl32</td><td>rpl36</td></tr>
<tr><td colspan="5">Small subunits</td></tr>
<tr><td>rps2</td><td>rps3</td><td>rps4</td><td>rps7</td><td>rps8</td></tr>
<tr><td>rps9†</td><td>rps11</td><td>rps12</td><td>rps14</td><td>rps18</td></tr>
<tr><td>rps19</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="5">Transcription/translation</td></tr>
<tr><td colspan="5">RNA polymerase</td></tr>
<tr><td>rpoA</td><td>rpoB</td><td>rpoC1</td><td>rpoC2</td><td></td></tr>
<tr><td colspan="5">Translation factors</td></tr>
<tr><td>infA</td><td>tufA†</td><td></td><td></td><td></td></tr>
<tr><td colspan="5">Division</td></tr>
<tr><td>minD†</td><td>(minE)†</td><td></td><td></td><td></td></tr>
<tr><td colspan="5">Miscellaneous proteins</td></tr>
<tr><td>accD</td><td>clpP</td><td>cysA</td><td>cysT</td><td>I-CvuI†</td></tr>
<tr><td colspan="5">Conserved ORFs</td></tr>
<tr><td>(ycf1)</td><td>(ycf2)</td><td>ycf3</td><td>ycf4</td><td>ycf5</td></tr>
<tr><td>ycf9</td><td>ycf10</td><td>ycf12</td><td></td><td></td></tr>
</table>

Parentheses indicate genes/ORFs with limited homology. ×2, Two copies.
*Genes containing introns.
†Genes not found in land plants.

septum at the mid-point of the rod-shaped cell (23). The MinC and MinD proteins seem to act in concert to form a division inhibitor capable of blocking nonspecific septation, and the MinE protein gives topological specificity to the division inhibitor so that septation is limited to the proper division site at midcell (23). Overexpression of *Chlorella* MinD in *E. coli* caused aberrant division of the cell (unpublished data), strongly suggesting that the *Chlorella minD* is involved in chloroplast division.

In the chloroplast genome of *Porphyra* and *Odontella*, a sequence consisting of 628 or 622 codons (*ycf25*) homologous to *E. coli ftsH* (644 codons; ref. 24) has been reported recently (4, 5). The C-terminal half of the protein deduced from ORF1720 in *Chlorella* shows partial homology with *ftsH* products of *E. coli* and *ycf25* products of *Porphyra* and *Odontella*; regions corresponding to the three sequences characteristic to

FtsH, the motifs A and B of a putative ATP-binding sequence and the DPAL motif, are especially well conserved. ORF1720 is the largest ORF in *Chlorella* and can be classified, though of limited homology, as *ycf2* (land plant ORF2280; ref. 25), whose deduced protein shows partial similarity with FtsH in *E. coli* (26). A large number of cell division mutants in *E. coli* have been identified; *fts* mutants (filamentation temperature sensitive) fail to septate when grown at high temperatures (27). Many Fts proteins are involved in septation, of which FtsZ forms a ring-like structure at the leading edge of the septum and appears to be the target of the MinC-MinD inhibitor. A cDNA encoding the chloroplast-localized FtsZ homologue has recently been identified in *Arabidopsis* (28). ORF with homology to *E. coli ftsW* has been reported in the cyanelle genome (6). Taken together, the existence of homologues of *E. coli* genes involved in cell division implies that chloroplast division
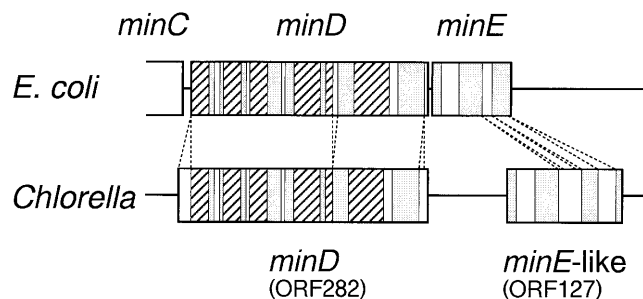
FIG. 2.   Schematic representation of the similarity between *Chlorella* ORF282/ORF127 and *E. coli minD/minE*. The gene arrangement and a comparison of their deduced amino acid sequences are shown. The hatched, shaded, and open areas represent identities of >50%, 30–50%, and <30%, respectively.

is similar in mechanism to that of bacterial cells. This hypothesis is compatible with the cytological observations showing that chloroplast division occurs at the mid-point of oval plastids and is mediated by a ring-like structure, the plastid-division ring (29).

Genes for Polypeptides of the Transcription-Translation System. A total of 21 ribosomal protein genes have been identified: 11 proteins of the 30S subunit and 10 proteins of the 50S subunit. Among them, *rpl5*, *rpl12*, *rpl19*, and *rps9* are absent in land plant chloroplasts (30, 31) but present in other algal chloroplasts (refs. 4–6 and 21; *rpl19* is absent in *Euglena*). *rps15*, *rps16*, and *rpl22*, which are present in many land plants, are missing in *Chlorella*.

The *Chlorella* genome contains both *infA* and *tufA*, the former has so far been found in land plants, whereas the latter has been found only in algae (3). In this respect, the *Chlorella* chloroplast genome seems to be closer to land plants than the other algae so far analyzed. It should be noted that the initiation codon of *infA* is UUG, which has not been reported in any chloroplast mRNA genes.

The genes encoding subunits of *E. coli*-type RNA polymerase, *rpoA*, *rpoB*, *rpoC1*, and *rpoC2*, are present. *Chlorella rpoA* (278 codons) is shorter than those of *E. coli* (329 codons) and of land plant chloroplasts (≈340 codons; ref. 32) mainly because of absence of the 3′-terminal portion (Fig. 3). As compared with the α subunit of *E. coli* (33), the *Chlorella* α subunit deduced from a DNA sequence possesses an extra sequence of 41 aa in the middle, whereas it lacks a region corresponding to the C-terminal domain of ≈100 aa. The C-terminal domain (or contact domain) is known to interact with class I transcription factors (34) and to bind the upstream element rich in AT located ahead of the "−35/−10" motif in some promoters, such as *rrnB* P1 in *E. coli* (35). Lack of the C-terminal domain-like region in the putative α subunit suggests the absence of class I-like transcription factors in *Chlorella* chloroplasts. *rpoA* is absent from the *Euglena* and *Epifagus* genomes (21, 36), and the *Chlamydomonas rpoA* homologue consists of only 131 codons (37). Therefore, the possibility that *rpoA* in *Chlorella* does not produce a functional subunit of RNA polymerase cannot be ruled out.

Genes for the Photosynthetic Apparatus. The *Chlorella* chloroplast DNA contains at least 35 genes related to the photosynthesis apparatus. Unlike *Euglena*, this alga contains *psaI*, *psbM*, *petA*, *petD*, and *petL*. Similar to the cyanelle and *Porphyra*, the *Chlorella* contains all the four genes (*chlB*, *chlI*, *chlL*, and *chlN*), known as chloroplast genes, related to chlorophyll biosynthesis. The *chlL* gene contains a 951-bp group I intron, the first example of a split *chlL* gene (38). This intron is able to efficiently self-splice *in vitro* in the presence of GTP. The *Chlorella* chloroplast genome lacks all 11 *ndh* genes found in most land plants (3) and their pseudogenes as found in black pine (19). No *ndh* genes have been found in parasitic plants (36,
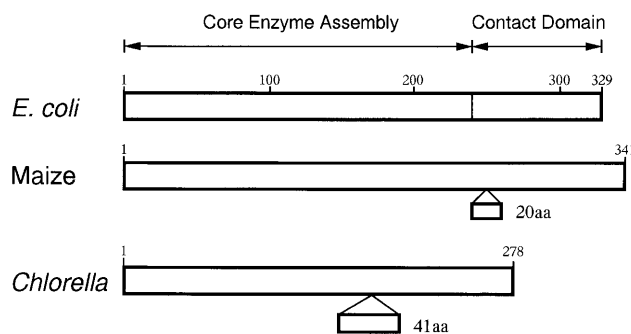

FIG. 3.   Structures of RNA polymerase α subunits deduced from *rpoA* genes of *E. coli* (33, 34), maize chloroplasts (32), and *Chlorella* chloroplasts. Functional domains of the *E. coli* α subunit are indicated at the top. Numbers above the boxes represent numbers of amino acids residues. The α subunits of maize and *Chlorella* have extra sequences with respect to that of *E. coli*, and these are shown below with the number of amino acid residues (aa). Corresponding regions (without the extra sequences) between *E. coli* and maize and between *E. coli* and *Chlorella* show 30% and 28% identities, respectively.

39) and in any analyzed chloroplast genomes from algae including *Chlamydomonas reinhardtii*, in which chlororespiration was reported to occur (40), and several possible *ndh* products have been detected (41). Therefore, it has been suggested that all the *ndh* genes have been transferred to the nuclear genome in algae.

Genes Encoding Stable RNAs. The rRNA gene cluster is arranged in the order *rrn16-trnI-trnA-rrn23-rrn5*, among which *rrn23* possesses an intron of 815 bp. This intron is of the group I type, inserted at the same site as the *rrn23* intron (888 bp) of *Chlamydomonas reinhardtii* chloroplasts, and contains an internal ORF (ORF161) whose predicted polypeptide (161 aa) is similar to the I-*Cre*I DNA endonuclease (163 aa) encoded by the *Chlamydomonas* intron (42) and to the I-*Ceu*I DNA endonuclease (218 aa) encoded by the group I intron of *Chlamydomonas eugametos rrn23* (43). ORF161 is, therefore, tentatively designated as I-*Cvu*I. The *Chlorella* 23S pre-rRNA is capable of self-splicing *in vitro* (44) as in the case of *Chlamydomonas* pre-rRNA (45). The *C. vulgaris rrn23* intron is considerably different in sequence and site of insertion from the intron (243 bp) found in *rrn23* of *C. ellipsoidea* (46).

Thirty-one different tRNA genes were identified based on homology with those from other chloroplasts and prokaryotes. Twenty-nine of them correspond to those identified in angiosperm chloroplasts (3). *trnV*(GAC), generally located in front of *rrn* clusters in land plant chloroplasts, is absent, whereas two additional tRNA genes, *trnR*(CCG) and *trnL*(GAG), are present; these two genes have not been found in angiosperms and land plants, respectively (3). No pseudo-tRNA genes have been found so far. *trnL*(UAA) is the only tRNA gene containing an intron (218 bp), which is of the group I type, like those that have been reported in land plant chloroplasts (47). No self-splicing of the pre-tRNA^{Leu}(UAA) *in vitro* has so far been observed.
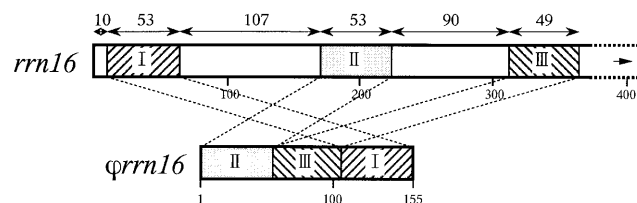

FIG. 4.   Structures of the 5′ part of *rrn16* and the entire *rrn16* in *Chlorella*. Numbers above *rrn16* represent sizes of homologous regions (bp) and those below indicate nucleotide positions from 5′ ends. Sequence identities of regions I, II, and III are 94%, 92%, and 84%, respectively.

As mentioned earlier, *trnG*(GCC) and *trnV*(UAC) are present in two copies each. Black pine chloroplasts have four duplicated tRNA genes, and a substantial portion of the flanking sequences are identical or near identical (48), whereas identity is restricted to the coding regions with up to seven flanking nucleotides in the two duplicated *Chlorella* tRNA genes. These observations suggest that the duplication of tRNA genes during evolution occurred by a different mechanism. Unlike tRNA genes from land plant chloroplasts, four tRNA genes, *trnI*(GAU), *trnA*(UGC), *trnG*(UCC), and *trnM*, encode the 3′ CCA end. It is noteworthy that *trnW*(CCA) directly follows *trnP*(UGG) with no intergenic spacer.

All 61 sense codons are used in the chloroplast DNA sequences encoding proteins, and the codon usage is not strikingly deviated from that in land plant chloroplasts (49). No deduced tRNAs with anticodons complementary to codons GUU/C (Val), CCU/C (Pro), and GCU/C (Ala) have been found. However, the 31 tRNAs encoded in the *Chlorella* genome are likely to be sufficient to decode all the protein genes using the "two-out-of-three" and the "U:N wobble" recognition mechanisms described for the tobacco chloroplast translation system (14).

Miscellaneous Genes and ORFs. The *matK* gene present in the long intron of *trnK* in land plants (3, 50) is absent in the *Chlorella*, as observed in all the other algae whose *trnK* genes are continuous genes. The *Chlorella* genome contains two other genes, *clpP* and *accD*, which are present in land plants (except no *accD* in monocots) but are absent in *Euglena* and *Odontella* (also no *accD* in the cyanelle and no *clpP* in *Porphyra*). *Chlorella* ORF236 and ORF266 show substantial homology (34% and 42% amino acid identity) with *E. coli* genes, *cysA* (365 codons) and *cysT* (277 codons), encoding sulfate transport ATP-binding protein and permease protein, respectively (51). We adopted here *cysA/T* for these ORFs. Their homologous sequences have been reported in *Marchantia* chloroplasts (*mbpX* and *mbpY*; ref. 20).

A sequence (155 bp) highly homologous to portions of *rrn16* is present ≈10 kbp downstream from the *rrn16* gene. This sequence (*φrrn16*) consists of three discrete fragments of ≈50 bp, as shown in Fig. 4. The process of *φrrn16* formation is intriguing; it could have originated by complex rearrangement within a limited rDNA region or by rejoining of pieces of reverse transcription products.

The *Chlorella* genome has eight *ycf*s, including two large ones related to *ycf1* and *ycf2*, though homology is limited. The *ycf1* and *ycf2* homologues have not been found in *Euglena*, *Porphyra*, *Odontella*, and the cyanelle, whereas they are present in land plants except monocot plants. On the other hand, additional *ycf* genes (*ycf16* through *ycf47*) conserved among *Porphyra*, *Odontella*, and the cyanelle, or two of them (52) were not found in *Chlorella*. These observations further suggest that green algae are closer to land plants than nongreen algal species.

No significant repeated sequences other than the two duplicated tRNA genes nor pseudogenes (or gene pieces) except *φrrn16* have been found in the *Chlorella* genome. Therefore, the 20- to 30-kbp extra single-copy sequences unique to *Chlorella* probably represent intergenic regions. However, it cannot be ruled out that these sequences may contain additional genes encoding polypeptides. There are 16 ORFs of more than 100 codons, 20 ORFs of 60–99 codons, and 60 ORFs of 40–59 codons. The chloroplast genome of the green alga *Nephroselmis olivacea* has been completely sequenced and found to contain many new ORFs, most of which, however, have no homology with those in *Chlorella* (M. Turmel and C. Lemieux, personal communication). Moreover, the detection of new structural RNA genes, such as *tscA* in *Chlamydomonas* (53), *sprA* in tobacco (54), and *rnpB* in *Porphyla* (4) and cyanelle (6), suggests that the intergenic sequence in *Chlorella* may encode additional structure RNAs.

1. Palmer, J. D. (1991) in *The Molecular Biology of Plastids*, eds. Bogorad, L. & Vasil, I. K. (Academic, San Diego), pp. 5–53.
2. Sugiura, M. (1992) *Plant Mol. Biol.* **19,** 149–168.
3. Sugiura, M. (1996) in *Molecular Genetics of Photosynthesis*, eds. Anderson, B., Salter, A. H. & Barber, J. (IRL, Oxford), pp. 58–74.
4. Reith, M. & Munholland, J. (1995) *Plant Mol. Biol. Rep.* **13,** 333–335.
5. Kowallik, K. V., Stoebe, B., Schaffran, I., Kroth-Pancic, P. & Freier, U. (1995) *Plant Mol. Biol. Rep.* **13,** 336–342.
6. Stirewalt, V. L., Michalowski, C. B., Löffelhardt, W., Bohnert, H. J. & Bryant, D. A. (1995) *Plant Mol. Biol. Rep.* **13,** 327–332.
7. Reith, M. & Munholland, J. (1993) *Plant Cell* **5,** 465–475.
8. Graham, L. E. (1996) *J. Plant Res.* **109,** 241–251.
9. Tamiya, H., Iwamura, T., Shibata, K., Hase, E. & Nihei, T. (1953) *Biochim. Biophys. Acta* **12,** 23–40.
10. Tamiya, H. (1966) *Annu. Rev. Plant Physiol.* **17,** 1–26.
11. Yoshinaga, K., Ohta, T., Suzuki, Y. & Sugiura, M. (1988) *Plant Mol. Biol.* **10,** 245–250.
12. Sager, R. & Ishida, M. R. (1963) *Proc. Natl. Acad. Sci. USA* **50,** 725–730.
13. Yoshinaga, K., Iinuma, H., Masuzawa, T. & Ueda, K. (1996) *Nucleic Acids Res.* **24,** 1008–1014.
14. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., *et al.* (1986) *EMBO J.* **5,** 2043–2049.
15. 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., Hirai, A., Shinozaki, K. & Sugiura, M. (1989) *Mol. Gen. Genet.* **217,** 185–194.
16. Maier, R. M., Neckermann, K., Igloi, G. L. & Kössel, H. (1995) *J. Mol. Biol.* **251,** 614–628.
17. Yamada, T. (1983) *Curr. Genet.* **7,** 481–487.
18. Yamada, T. & Shimaji, M. (1987) *Mol. Gen. Genet.* **208,** 377–383.
19. Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T. & Sugiura, M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9794–9798.
20. Ohyama, K., Fukuzawa, H., Kohchi, T., Sano, T., Sano, S., Shirai, H., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. & Ozeki, H. (1988) *J. Mol. Biol.* **203,** 281–298.
21. Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A. & Stutz, E. (1993) *Nucleic Acids Res.* **21,** 3537–3544.
22. Thompson, M. D., Copertino, D. W., Thompson, E., Favreau, M. R. & Hallick, R. B. (1995) *Nucleic Acids Res.* **23,** 4745–4752.
23. de Boer, P. A. J., Crossley, R. E. & Rothfield, L. I. (1989) *Cell* **56,** 641–649.
24. Tomoyasu, T., Yuki, T., Morimura, S., Mori, H., Yamanaka, K., Niki, H., Hiraga, S. & Ogura, T. (1993) *J. Bacteriol.* **175,** 1344–1351.
25. Downie, S. R., Katz-Downie, D. S., Wolfe, K. H., Calie, P. J. & Palmer, J. D. (1994) *Curr. Genet.* **25,** 367–378.
26. Wolfe, K. H. (1994) *Curr. Genet.* **25,** 379–383.
27. de Boer, P. A. J., Cook, W. R. & Rothfield, L. I. (1990) *Annu. Rev. Genet.* **24,** 249–274.
28. Osteryoung, K. W. & Vierling, E. (1995) *Nature (London)* **376,** 473–474.
29. Kuroiwa, T. (1991) *Int. Rev. Cytol.* **128,** 1–62.
30. Subramanian, A. R. (1993) *Trends Biochem. Sci.* **18,** 177–180.
31. Harris, E. H., Boynton, J. E. & Gillham, N. W. (1994) *Microbiol. Rev.* **58,** 700–754.
32. Hu, J. & Bogorad, L. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 1531–1535.
33. Bedwell, D., Davis, G., Gosink, M., Post, L., Nomura, M., Kestler, H., Zengel, J. M. & Lindahl, L. (1985) *Nucleic Acids Res.* **13,** 3891–3903.
34. Ishihama, A. (1992) *Mol. Microbiol.* **6,** 3283–3288.
35. Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. & Gourse, R. L. (1993) *Science* **262,** 1407–1413.
36. Wolfe, K. H., Morden, C. W. & Palmer, J. D. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10648–10652.

37. Surzycki, S. J., Hong, T. H. & Surzycki, J. A. (1986) in *Regulation of Chloroplast Differentiation* (Liss, New York), pp. 511–516.
38. Kapoor, M., Wakasugi, T., Yoshinaga, K. & Sugiura, M. (1996) *Mol. Gen. Genet.* **250,** 655–664.
39. Haberhausen, G. & Zetsche, K. (1994) *Plant Mol. Biol.* **24,** 217–222.
40. Peltier, G. & Schmidt, G. W. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 4791–4795.
41. Wu, M., Nie, Z. Q. & Yang, J. (1989) *Plant Cell* **1,** 551–557.
42. Dürrenberger, F. & Rochaix, J. D. (1991) *EMBO J.* **10,** 3495–3501.
43. Gauthier, A., Turmel, M. & Lemieux, C. (1991) *Curr. Genet.* **19,** 43–47.
44. Kapoor, M., Nagai, T., Wakasugi, T., Yoshinaga, K. & Sugiura, M. (1997) *Curr. Genet.*, in press.
45. Herrin, D. L., Chen, Y. F. & Schmidt, G. W. (1990) *J. Biol. Chem.* **265,** 21134–21140.
46. Yamada, T. & Shimaji, M. (1987) *Curr. Genet.* **11,** 347–352.
47. Bonnard, G., Michel, F., Weil, J. H. & Steinmetz, A. (1984) *Mol. Gen. Genet.* **194,** 330–336.
48. Tsudzuki, J., Ito, S., Tsudzuki, T., Wakasugi, T. & Sugiura, M. (1994) *Curr. Genet.* **26,** 153–158.
49. Shimada, H. & Sugiura, M. (1991) *Nucleic Acids Res.* **19,** 983–995.
50. Sugita, M., Shinozaki, K. & Sugiura, M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3557–3561.
51. Sirko, A., Hryniewicz, M., Hulanicka, D. & Böck, A. (1990) *J. Bacteriol.* **172,** 3351–3357.
52. Reardon, E. M. & Price, C. A. (1995) *Plant Mol. Biol. Rep.* **13,** 320–326.
53. Goldschmidt-Clermont, M., Choquet, Y., Girard-Bascou, J., Michel, F., Schirmer-Rahire, M. & Rochaix, J. D. (1991) *Cell* **65,** 135–143.
54. Vera, A. & Sugiura, M. (1994) *EMBO J.* **13,** 2211–2217.