Plant Molecular Biology **19**: 149–168, 1992.
© 1992 *Kluwer Academic Publishers. Printed in Belgium.*

The chloroplast genome

Masahiro Sugiura
Center for Gene Research, Nagoya University, Chikusa, Nagoya, Japan 464-01

Key words: Cloroplast, DNA genome gene, intron, splicing, *trans*-splicing

Introduction

Chloroplasts are intracellular organelles in plants which contain the entire machinery necessary for the process of photosynthesis. They also participate in the biosynthesis of amino acids, nucleotides, lipids and starch. Mendel's law was rediscovered at the beginning of this century, and in 1909 Baur and Correns separately published the first reports of non-Mendelian inheritance based on studies of variegation in higher plants. Some of the green-and-white variegated leaves were shown to be caused by factors inherited in a non-Mendelian manner. Further analysis of variegation in higher plants revealed that the genetic determinants for these characters were associated with chloroplasts. However, the difficulty of obtaining specific chloroplast mutations has limited the study of non-Mendelian genetics in higher plants. Our knowledge of extranuclear genetics came primarily from studies using the unicellular alga *Chlamydomonas*.

The demonstration of a unique DNA species in chloroplasts [e.g. 94] has led to intensive studies of both the structure of chloroplast DNA and its expression. These studies have been accelerated by gene cloning and DNA sequencing techniques developed in the mid-1970s. The first physical map of chloroplast DNA was constructed for maize in 1976 [3] and the first chloroplast gene was cloned in 1977 [5]. These studies and others established a new field, 'chloroplast molecular biology,' and the organization and expression of chloroplast genomes were among the most extensively studied fields in plant molecular biology. After 10 years the entire sequence of the chloroplast DNA was determined in tobacco, liverwort and then in rice. Sequences for defined regions of many other chloroplast DNAs have also been completed, but the identification and expression analysis of many chloroplast genes have mostly been done with several representative higher plants and green algae.

The purpose of this paper is to review briefly the historical background as well as our latest knowledge of the chloroplast genome, emphasizing its structural aspect, followed by a short discussion of future research. The references will attempt to cite the first one or two reports, or reviews. Other aspects of the chloroplast genome have been presented in several recent reviews [10, 40, 82, 88, 128].

Genome structure

The presence of unique, double-stranded and high-molecular-weight DNA in chloroplasts was demonstrated by the distinct and characteristic buoyant density of these molecules in CsCl gradients [18, 94]. Since 1963, CsCl gradient centrifugation has been used widely to identify and isolate chloroplast DNA from plants [95]. However, difficulties were encountered early in isolating the chloroplast DNA of higher plants because their buoyant densities were similar to those of the nuclear DNAs. Chloroplast DNA is now prepared from highly purified intact chloroplasts. The circularity of chloroplast DNA was first reported by Manning *et al.* [75], who used an electron microscope and observed circular DNA molecules with a contour length of 44.5 μm in lysates

of *Euglena* chloroplasts. Circular chloroplast DNAs have since been observed in many other plants. Chloroplast DNA molecules appeared to have a uniform contour length within a given plant species. Convincing evidence for the homogeneity of chloroplast DNA molecules was provided by digestion with class II restriction endonucleases [4]. Furthermore, mapping the restriction endonuclease fragments always yielded a circular map. The first physical map of chloroplast DNA was thus determined from maize [3]. Restriction site mapping is routinely used to determine the size of chloroplast DNA, and is almost a prerequisite for further studies.

Almost all chloroplast DNAs fall into the size range of 120 to 160 kb [88]. Among chloroplast genomes for which an accurate size estimate exists, the siphonous green alga *Codium fragile* has the smallest chloroplast DNA known (85 kb) while the green alga *Chlamydomonas moewusii* has the largest (292 kb). The chloroplast genome of the giant green alga *Acetabularia* is more complex than those of other plants and its genome size appears to be 2000 kb. The population of chloroplast DNA in a plant species is generally homogeneous. However, the chloroplast genome of the brown alga *Pylaiella littoralis* has been shown to be composed of two different circular DNA molecules of 133 kb and 58 kb in size [74]. Though genetically homogeneous, chloroplast DNA often consists of two groups of molecules differing only in the relative orientation of the single-copy regions [87]. A small proportion of the molecules exists in dimer, trimer and tetramer forms [20].

One of the outstanding features of the chloroplast DNAs found in most plants is the presence of a large inverted repeat (IR) which ranges from 6 to 76 kb in length [88]. Most of the size variation among land plant chloroplast DNAs can be accounted for by changes in the length of the IR. The segments of the IR are separated by one large and one small single-copy region (LSC and SSC, respectively). Pea, broad bean, alfalfa and pine chloroplast DNAs are exceptions to this pattern and lack IRs [e.g. 63]. It has been suggested that the IR was present in the common ancestor of land plants and one segment of the IR was lost in some legumes and conifers during evolution [88]. However, our recent analysis indicates that the loss of the IR is partial at least in black pine as its genome retains a residual IR (unpublished). The chloroplast DNA from *Euglena gracilis* contains three tandem repeats, each of which contains an rRNA gene cluster. Thus, chloroplast DNAs can be classified into three groups: chloroplast DNAs lacking IRs (group I), chloroplast DNAs containing IRs (group II) and chloroplast DNAs with tandem repeats (group III).

The entire nucleotide sequence of the chloroplast DNA is now established for tobacco (155 844 bp) [103], liverwort (121 024 bp) [85] and rice (134 525 bp) [53]. The gene order present in tobacco (Fig. 1) is most representative of land plants, probably reflecting the ancestral gene order among higher plants. Most of the chloroplast DNA from maize, pea and *Euglena* have already been sequenced, and the determination of complete chloroplast DNA sequences from several other plants including *Arabidopsis* and pine is in progress. It was originally believed that the gene organization of chloroplast DNAs was relatively uniform from species to species. More recent analyses of chloroplast genomes from a variety of land plants and algae has revealed that this is not always the case. New genes not present in vascular plant chloroplast DNAs have been found one after another in chloroplast genomes from *Euglena*, marine algae and the cyanelles of *Cyanophora*. This gives impetus to diversify the range of plant species in which chloroplast genome analysis is undertaken. The accumulation of chloroplast DNA sequence data facilitates further analysis of chloroplast origin and evolution.

Gene structure

The average chloroplast genome contains about 120 kb of unique sequence, which is enough to encode 120 genes if one assumes that an average gene contains about 1 kb. The number of protein coding genes seems to be about 100 in addition to rRNA and tRNA genes [24].

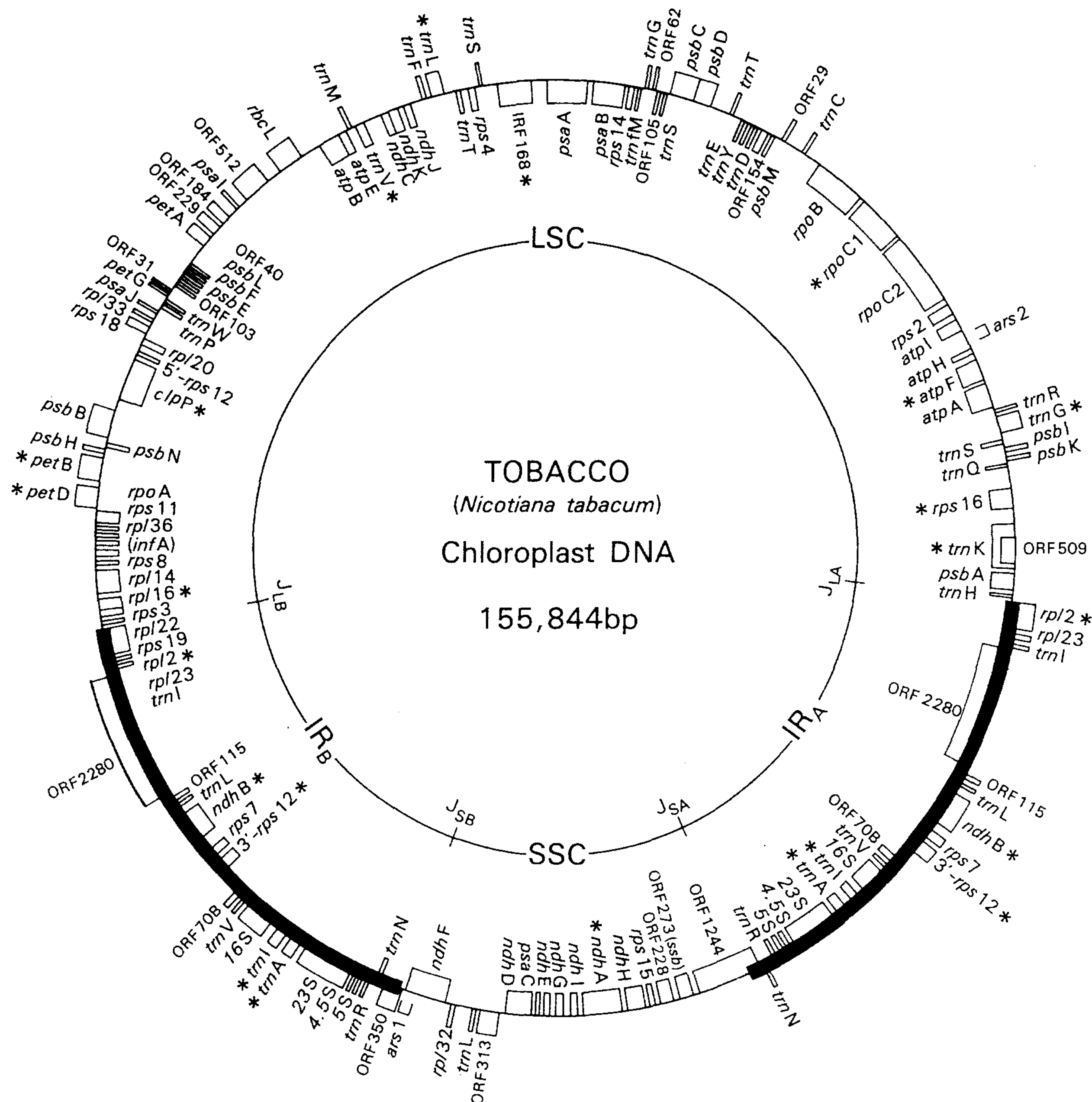

Fig. 1. Gene map of the tobacco chloroplast genome. Genes shown inside the circle are transcribed clockwise, and genes on the outside are transcribed counter-clockwise. Asterisks denote split genes. Major ORFs are included. IRF, intron-containing reading frame; IR, inverted repeat; LSC, large single-copy region; SSC, small single-copy region; J, junctions between IR and LSC/SSC. From the Research Grant Progress Report (1989) with minor revisions.

Initially, there were essentially three methods used to locate genes in the chloroplast genome. The first method was that of standard genetic analysis. Chloroplast genes coding for ribosomal proteins in *Chlamydomonas* and fraction I proteins in tobacco were analyzed by this method [95]. Analysis of the plastome mutants of *Antirrhinum* and *Pelargonium* suggested that some components of the thylakoid membrane are encoded in the chloroplast genome [46]. The sec-

ond method was RNA-DNA hybridization experiments which demonstrated the presence of rRNA and tRNA genes in the chloroplast genome [4]. Third, the study of proteins synthesized by isolated chloroplasts was a powerful technique used to detect chloroplast-encoded proteins [26]. The analysis of ribosome-deficient chloroplast mutants also provided information on proteins synthesized within the chloroplast [46]. Later, the application of gene cloning and DNA sequencing technologies gave us the primary structure of chloroplast genes and predicted the presence of new chloroplast genes.

Identification of chloroplast genes has been pursued extensively in maize, wheat, spinach, pea, tobacco, *Euglena* and *Chlamydomonas*. Chloroplast DNAs are now known to contain all the chloroplast rRNA genes (3–5 genes), 30–31 tRNA genes and about 100 genes encoding proteins, which fits the number of genes estimated from the size of the chloroplast DNA. The chloroplast genes that have been sequenced (including putative genes) are presented in Table 1. Gene nomenclature follows the proposal of Hallick [47]. The progress of chloroplast DNA analysis has been rapid and exceeds that of chloroplast protein analysis. At present there are many open reading frames (ORFs), which are potential polypeptide genes in the chloroplast genome. Homology searches in protein databases have predicted that some of the ORFs are protein genes. However, the final identification of chloroplast genes encoding polypeptides requires the analysis of the translation products. *In vitro* transcription-translation of specific DNA fragments followed by immunoassay and the determination of partial amino acid sequences of isolated chloroplast proteins are the two principal procedures used for this purpose.

Genetic system genes

Ribosomal RNA genes
Chloroplasts contain a 70S class of ribosomes which are distinct from the 80S ribosomes found in the cytoplasm. A 23S, a 5S and a 4.5S rRNA are all associated with the 50S subunit. The 4.5S rRNA has been found in higher-plant chloroplasts [e.g. 12] and it is homologous to the 3′ end of the 23S rRNA of prokaryotes. Maize chloroplast rRNA genes (*rDNA*s) were the first chloroplast genes cloned [5]. Sequencing of maize and tobacco chloroplast *rDNA*s revealed a gene order of 16S-23S-4.5S-5S and an interspersion of tRNA genes within this cluster [e.g. 97, 112]. In *Chlamydomonas reinhardtii* the *rDNA* cluster consists of 16S, 7S, 3S, 23S and 5S in this order [93]. The presence of two small *rDNA*s (3S and 7S) is unique in this alga. The *Chlamydomonas* 23S *rDNA* has one intron and was the first split gene found in the chloroplast genome. Moreover, the 23S *rDNA* from *C. eugametos* contains six introns and three short internal transcribed spacers that are post-transcriptionally excised to yield four rRNA species [115]. In *Euglena* (strain Z), there are three copies of the *rDNA* cluster arranged tandemly and an extra copy of the 16S *rDNA* [e.g. 38].

The chloroplast *rDNA*s described above are all arranged as in *Escherichia coli* (16S-23S-5S). However, the *rDNA* cluster of *Chlorella ellipsoidea* is split into two back-to-back operons; operon 1 comprising 16S rRNA, $tRNA^{Ile}$(GAU) and operon 2 comprising $tRNA^{Ala}$(UGC), 23S rRNA, 5S rRNA [121]. In *P. littoralis* chloroplasts, the large DNA contains two sets of *rDNA* in reverse orientation while the small DNA has a 16S rRNA pseudogene and a split 23S rRNA sequence separated by at least 4 kb [76]. The sequences of individual chloroplast *rDNA* are highly homologous to each other and to those from eubacteria, but *rDNA* operons differ significantly when land plants and algae are compared. They vary in the presence or absence of introns, additional small *rDNA* species and in *rDNA* organization. This has led to speculation that the origin and evolutionary history of chloroplast genomes are diverse.

Transfer RNA genes
Saturation hybridization of a total chloroplast tRNA fraction to chloroplast DNA indicated the presence of 20–40 tRNA genes on the chloroplast genome [e.g. 45]. Chloroplast genomes are thus

Table 1. Chloroplast genes.

Genes	Products	Remarks
Genes for the genetic system		
23S rDNA	23S rRNA	
16S rDNA	16S rRNA	
7S rDNA	7S rRNA	in *Chlamydomonas*
5S rDNA	5S rRNA	
4.5S rDNA	4.5S rRNA	in land plants
3S rDNA	3S rRNA	in *Chlamydomonas*
trnA-UGC	Ala-tRNA (UGC)	
trnR-ACG	Arg-tRNA (ACG)	
trnR-UCU	Arg-tRNA (UCU)	
trnR-CCG	Arg-tRNA (CCG)	in liverwort
trnN-GUU	Asn-tRNA (GUU)	
trnD-GUC	Asp-tRNA (GUC)	
trnC-GCA	Cys-tRNA (GCA)	
trnQ-UUG	Gln-tRNA (UUG)	
trnE-UUC	Glu-tRNA (UUC)	
trnG-GCC	Gly-tRNA (GCC)	
trnG-UCC	Gly-tRNA (UCC)	
trnH-GUG	His-tRNA (GUG)	
trnI-GAU	Ile-tRNA (GAU)	
trnI-CAU	Ile-tRNA (CAU)	
trnL-UAA	Leu-tRNA (UAA)	
trnL-CAA	Leu-tRNA (CAA)	
trnL-UAG	Leu-tRNA (UAG)	
trnK-UUU	Lys-tRNA (UUU)	
trnfM-CAU	fMet-tRNA (CAU)	
trnM-CAU	Met-tRNA (CAU)	
trnF-GAA	Phe-tRNA (GAA)	
trnP-UGG	Pro-tRNA (UGG)	
trnS-GGA	Ser-tRNA (GGA)	
trnS-UGA	Ser-tRNA (UGA)	
trnS-GCU	Ser-tRNA (GCU)	
trnT-GGU	Thr-tRNA (GGU)	
trnT-UGU	Thr-tRNA (UGU)	
trnW-CCA	Trp-tRNA (CCA)	
trnY-GUA	Tyr-tRNA (GUA)	
trnV-GAC	Val-tRNA (GAC)	
trnV-UAC	Val-tRNA (UAC)	
tscA	small RNA	in *Chlamydomonas*
rps2	30S r-protein CS2	
rps3	CS3	
rps4	CS4	
rps7	CS7	2 pieces in *Chlamydomonas*
rps8	CS8	
rps9	CS9	in *Cryptomonas*
rps10	CS10	in *Cryptomonas*
rps11	CS11	
rps12	CS12	
rps14	CS14	
rps15	CS15	
rps16	CS16	
rps18	CS18	
rps19	CS19	
rpl2	50S r-protein CL2	
rpl5	CL5	in *Euglena*
rpl13	CL13	in *Cryptomonas*
rpl14	CL14	
rpl16	CL16	
rpl20	CL20	
rpl21	CL21	in liverwort
rpl22	CL22	not in legumes

Table 1. (Continued)

Genes	Products	Remarks
rpl23	CL23	pseudogene in spinach
rpl32	CL32	
rpl33	CL33	
rpl36	CL36	
rpoA	RNA polymerase subunit α	
rpoB	subunit β	
rpoC1	subunit β'	
rpoC2	subunit β''	
tufA	elongation factor Tu	in algae
infA	initiation factor 1	pseudogene in tobacco
clpP	ATP-dependent protease, proteolytic subunit	
Genes for the photosynthetic system		
rbcL	Rubisco large subunit	
rbcS	small subunit	in red and brown algae
psaA	PSI P700 apoprotein A1	
psaB	P700 apoprotein A2	
psaC	9 kDa protein	
psaI	I-protein	
psaJ	J-protein	
psbA	PSII D1-protein	
psbB	47 kDa protein	
psbC	43 kDa protein	
psbD	D2-protein	
psbE	cytochrome b559 (8 kDa)	
psbF	cytochrome b559 (4 kDa)	
psbH	10 kDa phosphoprotein	
psbI	I-protein	
psbK	K-protein	
psbL	L-protein	
psbM	M-protein	
psbN	N-protein	
petA	*b/f* complex cytochrome *f*	
petB	cytochrome *b6*	
petD	subunit IV	
petG	subunit V	
atpA	H^+-ATPase subunit CF_1 α	
atpB	subunit CF_1 β	
atpE	subunit CF_1 ε	
atpF	subunit CF_0 I	
atpH	subunit CF_0 III	
atpI	subunit CF_0 IV	
ndhA	NADH dehydrogenase ND1	
ndhB	ND2	
ndhC	ND3	
ndhD	ND4	
ndhE	ND4L	
ndhF	ND5	
ndhG	ND6	
ndhH	49 kDa protein	
ndhI (*frxB*)	18 kDa protein	
ndhJ		ORF158–169 in LSC
ndhK (*psbG*)	27 kDa protein	
frxC	31 kDa protein	

believed to encode all the tRNA species used in chloroplast protein synthesis, although plant mitochondria take up some of their tRNAs from the cytoplasm. The tRNA gene (*trn*) maps were constructed by hybridizing purified individual tRNAs to chloroplast DNA fragments [e.g. 23] followed by the sequencing of tRNA genes [e.g. 35, 86]. All the tobacco chloroplast DNA fragments that hybridized to total chloroplast tRNAs have been sequenced, and thirty different tRNA genes were found [116]. A subsequent search of the complete sequence of tobacco chloroplast DNA yielded no new tRNA genes. Hence these 30 tRNA genes are probably all of the tRNA genes encoded in the chloroplast genome of tobacco and perhaps of other higher plants. The liverwort genome contains an additional $tRNA^{Arg}$(CCG) sequence.

No chloroplast genes have been found which code for a tRNA with a 3′-CCA end. All the tRNAs deduced from the DNA sequences can form the clover leaf structure, and none exhibits an abnormal form. The presence of introns in chloroplast tRNA genes was first demonstrated in maize *trnI* and *trnA* located in the long spacer separating the 16S and 23S *rDNA*s [61]. Six chloroplast tRNA genes from land plants are now known to harbor long single introns (0.5–2.5 kb). The *trnG*-UCC gene contains an intron in the D-stem region, a feature unique to chloroplasts [21]. In contrast, no split tRNA genes have been found in algal chloroplasts.

In land plants, the tRNA genes are scattered over the chloroplast genome, while in *Euglena* most of the tRNA genes are clustered [49]. The *Euglena* genome contains an additional pseudo $tRNA^{Ile}$ gene in the *16S rDNA* leader region, the first pseudogene found in the chloroplast genome [86]. In monocot chloroplast genomes at least five pseudo tRNA genes have been found. These are located near the inversion endpoints, and the involvement of tRNA genes in genome inversions during evolution has been proposed [e.g. 100].

All 61 possible codons are used in chloroplast genes which encode polypeptides. The minimum number of tRNA species required for translation of all 61 codons is 32 if normal wobble base-pairing occurs in codon-anticodon recognition. As shown in Table 1, no tRNAs which recognize codons CUU/C(Leu), CCU/C(Pro), GCU/C(Ala) or CGC/A/G(Arg) [or CGC/A in liverwort] according to normal wobble base-pairing have been found. If the 'two-out-of-three' mechanism operates in the chloroplast, four single tRNA species, $tRNA^{Pro}$(UGG), $tRNA^{Ala}$(UGC) and $tRNA^{Arg}$(ACG) species can read all four Pro, Ala and Arg codons, respectively (note that these tRNAs form only GC pairs in their first and second codon-anticodon interactions). There is a tRNA gene in which the $tRNA^{Leu}$ anticodon is UAG and if this tRNA has an unmodified U in the first position of the anticodon, it can read all four Leu codons (CUN) by 'U:N wobble.' The bean, spinach and soybean $tRNAs^{Leu}$(UAG) have unmodified Us in their anticodons ($UA^{m7}G$) [90]. Thus, 30 tRNAs are probably sufficient to read all 64 codons [103]. These hypotheses have been supported by a recent study in which these four chloroplast tRNAs were purified from bean and their decoding properties were analyzed in a tRNA-dependent wheat germ protein synthesizing system [89].

Ribosomal protein genes

Chloroplast ribosomes contain about 60 different protein components, one-third of which are thought to be encoded by chloroplast DNA according to protein synthesis studies in isolated chloroplasts [e.g. 27]. Genes encoding chloroplast ribosomal proteins have been deduced through their homology with *E. coli* ribosomal protein genes [e.g. 110]. Twenty-one different ORFs potentially coding for polypeptides homologous to *E. coli* ribosomal proteins have been found in the tobacco, liverwort and rice chloroplast genomes [111]. The tobacco and rice genomes lack *rpl21* but contain *rps16* which has an intron. The black pine genome apparently lacks *rps16* (unpublished observations). Many of their translation products have been identified in spinach, pea and tobacco through partial amino acid sequencing [114]. Several nuclear-encoded chloroplast ribosomal proteins and their genes have also been analyzed. Among them two have no similarity to any bacterial ribosomal proteins, in-

dicating the uniqueness of chloroplast ribosomes [32, 58].

The *rpl23*, *rpl2*, *rps19*, *rpl22*, *rps3*, *rpl16*, *rpl14*, *rps8*, *infA*, *rpl36*, *rps11* and *rpoA* genes are clustered in this order in the *rpl23* operon and the arrangement corresponds to that of the homologous genes in the *E. coli* S10, *spc* and α operons [e.g. 113] (Fig. 2). This raises the possibility that the genes for ribosomal proteins of higher plant chloroplasts and *E. coli* may have evolved from a common ancestral gene set. However, *rpl22* was not found in legume chloroplast DNAs [105]. *rpl23* is split into two overlapping reading frames in spinach and several other higher plants [128], and its translation product has been identified and shown to be functional in tobacco [123]. The *rpl23* operon in *Euglena* is also similar to the gene arrangement in the *E. coli* operons [17] (Fig. 2). *Euglena rpl5* is a new gene not reported for any land plant chloroplast genomes to date, as are *rpl13*, *rps9* and *rps10* in the marine alga *Cryptomonas* [22].

An intron within a potential chloroplast ribosomal protein gene was first found in *Nicotiana debneyi rpl2*, but this intron is absent in spinach and some related dicots [128]. The higher-plant *rps16* and *rpl16* sequences also contain single introns while *Euglena rpl16*, *rpl23*, *rps2*, *rps3*, *rps8*, *rps11* and *rps19* contain multiple introns. More unusual is *rps12* in land plants, which consists of three exons and requires *trans*-splicing for expression (see the next section). The putative *rps7* gene of *C. reinhardtii* is split into two segments. The 5′ and 3′ portions of *rps7* are separately transcribed and no common transcript has been detected. A Shine-Dalgarno sequence occurs upstream of the 3′ portion, but the consensus sequences for chloroplast introns are absent. These results suggest that *trans*-splicing is probably not involved in *rps7* expression, but rather that the protein may be made in two pieces [33].

Translational factors

A sequence similar to the *E. coli* EF-Tu gene (*tufA*) has been found upstream of *rps12*/*rps7* in *Euglena* [81] and several other algae, but not in any land plant chloroplast DNAs sequenced to date. A putative gene for the initiation factor IF-1 (*infA*) was found in the *rpl23* gene cluster between *rps8* and *rpl36* in spinach [104]. Among the *infA*s so far sequenced, a tobacco *infA* homologue does not constitute an ORF but is transcribed along with neighboring genes. Recently a cDNA potentially coding for chloroplast IF-1 has been isolated, suggesting that *infA* has transferred into the nucleus in tobacco (T. Wakasugi *et al.*, unpublished). It is curious that only one of the *rpl23* gene cluster (containing 12 genes) is duplicated

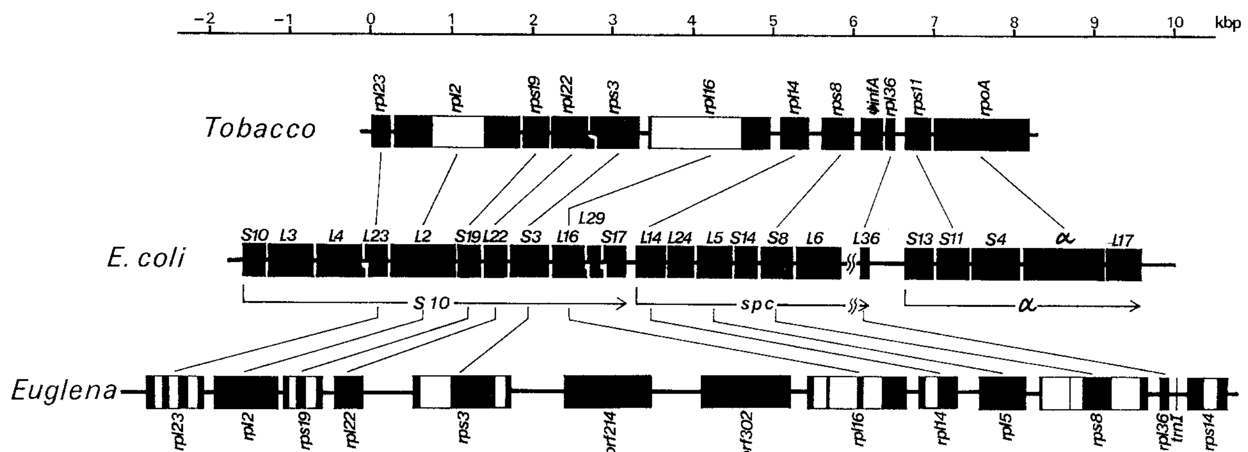

Fig. 2. Comparison of tobacco and *Euglena* chloroplast *rpl23* operons with the *E. coli* S10, *spc* and α operons. Exons are shown by filled boxes and introns by open boxes.

and transferred to the nucleus. A putative wheat chloroplast gene (*clpP*) encoding the proteolytic subunit of an ATP-dependent protease has been reported [37]. This enzyme degrades incomplete polypeptides and unassembled proteins in chloroplasts.

RNA polymerase subunit genes

It had been suggested that higher-plant chloroplast RNA polymerase is encoded in the nuclear genome [e.g. 71]. However, chloroplast DNA sequences hybridizing with the *E. coli* RNA polymerase genes were reported in *Chlamydomonas* [117]. Subsequent sequence analysis revealed that chloroplast DNA regions potentially coding for polypeptides similar to *E. coli* RNA polymerase α (*rpoA*), β (*rpoB*) and β' (*rpoC*) subunits were found in land plants [e.g. 84, 104]. An *E. coli rpoC* homologue is split into two parts, *rpoC1* and *rpoC2*, of which only *rpoC1* contains an intron [e.g. 85]. However in rice and maize, *rpoC1* is a continuous gene and *rpoC2* contains an extra sequence (380–450 bp) in the middle of it [57, 99]. No sequences similar to a bacterial *rpoD* have been found. The amino-terminal sequences of the 38 kDa, 120 kDa, 78 kDa and 180 kDa subunits of maize chloroplast RNA polymerase have recently been determined and found to correspond precisely to the sequences deduced from the maize *rpoA*, *rpoB*, *rpoC1* and *rpoC2* genes, respectively [55, 56]. These findings indicate that chloroplasts are the site of synthesis of some if not all of the chloroplast RNA polymerase subunits. We cannot rule out the possibility that a distinct species of RNA polymerase is imported from the cytoplasm (see the next section).

Photosynthetic system genes

Rubisco subunit genes

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) is the major stromal protein in chloroplasts and is composed of eight identical large subunits (LS) of 55 kDa and eight identical small subunits (SS) of 12 kDa. LS is encoded by the chloroplast DNA and SS is encoded by the nuclear DNA in higher plants and green algae. In contrast, the SS gene (*rbcS*) has been found in the chloroplast DNA from brown and red algae [e.g. 92]. The maize chloroplast gene for LS (*rbcL*) was the first chloroplast protein gene cloned and sequenced [78]. *rbcL* has become the most widely sequenced gene, enabling comparison for the determination of phylogenetic relationships among plant species [128]. The *rbcL* genes of higher plants and *Chlamydomonas* contain no introns while nine introns have been found in the *Euglena rbcL* gene [64]. In the chloroplast genomes which contain it, *rbcS* is located downstream from *rbcL* and constitutes an operon with *rbcL* as has been reported in cyanobacteria and cyanelles. No intron has been found in chloroplast *rbcS* genes while the nuclear *rbcS* genes have one to three introns.

Photosystem II genes

The thylakoid membranes have four distinct complexes [e.g. 36]: photosystems I and II (PSI, PSII), the cytochrome *b*/*f* complex and ATP synthase. The genes encoding thylakoid proteins have usually been isolated and identified through protein analysis, in which a protein component is purified and its antibody is prepared. A cloned DNA fragment is then placed in a transcription-translation system, or an isolated mRNA (or hybrid-selected mRNA) is translated in a cell-free system. The protein product is then identified using the specific antibody [e.g. 52]. After chloroplast DNA sequences became available, some genes were identified by western blotting analysis using antibodies against synthetic oligopeptides deduced from the DNA sequences [e.g. 106] and by comparing partial amino acid sequences of isolated proteins with the amino acid sequences derived from the DNA sequences [e.g. 28].

At least 12 components of PSII are encoded in the chloroplast genome. The gene for the 32 kDa protein Q_B or D1 protein (*psbA*) was the first PS gene sequenced in spinach and *Nicotiana debneyi* [126]. The 32 kDa protein binds to herbicides such as atrazine and DCMU. Therefore *psbA* is agronomically important and is another widely analyzed gene. The *psbA* genes isolated from

herbicide-resistant mutants have point mutations at codon 264 of the protein which result in substitution of glycine or alanine for serine [e.g. 54]. In land plants all the PSII genes are continuous while some of the algal *psb* genes are split by one to six introns [e.g. 60]. In higher plants the *psbD* gene overlaps *psbC* by about 50 bp [e.g. 1], suggesting that chloroplasts must have a specific mechanism for producing a proper amount of each component of a given complex.

Photosystem I genes

Five components of PSI are encoded in the chloroplast genome. The genes for subunits A1 and A2 of the P700 chlorophyll *a* apoprotein (*psaA* and *psaB*) were first sequenced from maize [29]. The *psaA* and *psaB* genes in higher plants contain no introns, are situated tandemly and are about 45% homologous at the amino acid level. The predicted A1 and A2 products contain a leucine-zipper motif, which is probably involved in dimerization of these subunits [65, 118]. In *Chlamydomonas* the *psaA* gene is divided into three exons scattered around the chloroplast genome, while *psaB* is uninterrupted [68]. The three distantly separated exons of *psaA* produce a functional mRNA by *trans*-splicing (see the next section).

Cytochrome b/f *complex genes*

The cytochrome *b/f* complex consists of six components, four of which are encoded by the chloroplast genome [e.g. 51, 120]. The *petB* and *petD* genes are clustered with *psbB* and *psbH* in higher plants and they constitute a transcription unit [e.g. 52]. In higher plants both *petB* and *petD* contain single introns with short first exons (6–8 bp). In the green alga KS3/2 *petD* contains a 3.5 kb intron, the largest chloroplast intron reported to date, which has an ORF (608 codons) showing significant homology with reverse transcriptase genes [67].

ATP synthase genes

ATP synthase consists of two parts, CF_1 and CF_0. CF_1 is composed of five different subunits and CF_0 is composed of four different subunits. The genes for six subunits are present in the chloroplast genome. Genes for the β and ε subunits (*atpB* and *atpE*) were first sequenced from maize and spinach [66, 127]. The *atpB* and *atpE* genes in most higher plants overlap by 4 bp, so that the first two bases of the TGA stop codon of *atpB* and an A residue preceding it form the ATG initiation codon of *atpE*. The genes for the three CF_0 subunits (*atpI*, *atpH*, *atpF*) are clustered just before *atpA* [e.g. 8]. The deduced amino acid sequences of these six subunits show homology with their counterparts in *E. coli*.

ndh *genes*

Eleven chloroplast DNA sequences (*ndh*) whose predicted amino acid sequences resemble those of components of the respiratory-chain NADH dehydrogenase from mitochondria have been found in a variety of plants [e.g. 85, 103]. The *ndhA* and *ndhB* genes contain single introns. As most of these sequences are actively transcribed and the *ndhA* and *B* transcripts are spliced rapidly, they are likely to be the genes for components of a chloroplast NADH dehydrogenase [e.g. 77]. These observations suggest the existence of a respiratory-chain in chloroplasts, although it remains to be determined whether or not all of these transcripts are translated into functional proteins.

Gene expression

Transcription and promoters

RNA polymerase

Chloroplasts contain at least two different RNA polymerase activities, a soluble enzyme and a transcriptionally active chromosome (TAC) [40]. A soluble DNA-dependent RNA polymerase was highly purified from maize chloroplasts and its subunit composition was analyzed (see the previous section). A TAC was first isolated from *Euglena* chloroplasts [50]. RNA polymerase associated with TAC is tightly bound to chloroplast DNA and preferentially transcribes the rRNA genes.

For the precise initiation of transcription, RNA

polymerase requires additional factors. Such factors have been isolated from maize (S-factor [59], BF fraction [125]), from spinach (a σ-like polypeptide [70]) and from mustard [14]. In parallel with the characterization of RNA polymerases and their accessory factors, *in vitro* transcription systems have been developed to identify chloroplast promoters [e.g. 42, 73]. Chloroplast primary transcripts are known to harbor 5′-triphosphates which can be specifically labeled with [^{32}P]GTP and guanylyltransferase (*in vitro* capping). The *in vitro* capping assay can therefore identify the initiation site of transcription *in vivo*.

Promoter sequences

The upstream regions of many initiation sites determined by *in vitro* capping contain DNA sequences similar to the '– 10' and '– 35' *E. coli* promoter motifs. The identification and characterization of chloroplast promoters have been done using deleted and mutated genes and *in vitro* transcription systems [43, 44, 73]. These analyses have confirmed that '– 35'- and '– 10'-like sequences are required for proper transcription (Fig. 3). However, a class of chloroplast tRNA genes has been identified which do not require their 5′upstream regions for transcription *in vitro* [41]. Relevant to this, one of the four primary transcripts for spinach *atpB* completely lacks an untranslated leader; the transcription start site is at the translation start codon [6]. The *psbA* genes in higher plants contain both prokaryotic-type '– 35' and '– 10' regions and between them a sequence motif similar to the nuclear TATA box. Mutation experiments have shown that the TATA box-like region is also critical for correct *psbA* transcription *in vitro* [25]. Thus at present the chloroplast genome is known to contain at least three structurally distinct upstream regions: regions containing the '– 35'/'– 10' promoter motifs, the TATA box-like promoter and no upstream promoter. Furthermore, they imply that there are multiple RNA polymerase species and/or multiple σ-like factors in chloroplasts. This is consistent with nuclear and chloroplast origins of chloroplast RNA polymerases (see 'Genetic system genes,' above).

Transcription from deleted or mutated genes has been studied *in vitro* as described above, but with recently developed systems for stably introducing foreign genes into chloroplast genomes it has become possible to do so *in vivo*. This method will permit testing of the conclusions derived from *in vitro* studies [e.g. 9, 13].

Transcript processing and introns

The chloroplast genome contains over 120 genes and about 50 transcription units, suggesting that

		"–35" "TATA" "–10"
Mustard	*psbA*	TTGGTTGACATGGCTATATAAGTCATGTTATACTGTTCAAT
Spinach	*psbA*	TTGGTTGACACGGGCATATAAGGCATGTTATACTGTTGAAT
	rbcL	TGGGTTGCGCCATATATATGAAAGAGTATACAATAATGATG
	atpB	TCTTGACAGTGGTATATGTTGTATATGTATATCCTAGATGT
	atpB (5)	ATTTTTGCAAAAAATTTCGACATACTTTACTATATATTATG (met)
	trnM	TTATATTGCTTATATATAATATTTGATTTATAATCAATCTA

Fig. 3. Chloroplast promoter regions identified by using deleted/mutated genes and *in vitro* transcription systems. Mustard *psbA* [25, 73] spinach *psbA* to *trnM* [43, 44] and spinach *atpB*(5) (the promoter for a fifth transcript starting at the translation initiation codon [6]) are shown.

chloroplast genes are generally cotranscribed. The detection of primary transcripts has actually shown that most of the chloroplast genes are transcribed polycistronically. Multiple transcripts are observed for most chloroplast gene clusters and these are mainly the results of multiple RNA processing of the primary transcripts. Processing of rRNA and tRNA precursors and of precursors from split genes is required to form functional RNA molecules. A couple of the chloroplast operons consist of functionally distinct genes; e.g. the *psbDC* operon contains two PSII genes and a tRNA gene [7], and the *psaA* operon has two PSI genes and a ribosomal protein gene [79]. Some chloroplast operons are known to be constitutively transcribed. These findings suggest that posttranscriptional processing of primary transcripts represents an important step in the control of chloroplast gene expression [40].

RNA cutting

Chloroplast polycistronic transcripts are generally processed into many overlapping shorter RNA species. Some of the shorter RNAs are monocistronic but others are not. Several transcript sets contain multiple 5′ ends, which result from the cutting of precursor RNAs and from multiple transcriptional initiation. Detailed analysis of polycistronic transcripts have been made for the rRNA gene cluster [e.g. 19], the tRNA gene cluster [39], the *psbB* operon [119] and the *psbDC* operon [e.g. 31]. For example, the RNA pattern of the spinach *psbB* operon (*psbB-psbH-petB-petD*) is complex and resolves into 18 major RNA species [119]. All RNA species arise from the cutting of 5.6 kb primary transcript rather than from multiple transcription initiation and termination events. Processing results ultimately in the formation of monocistronic mRNAs for each of the two PSII proteins and a dicistronic mRNA for both *pet* subunits. These mono- and dicistronic mRNAs are thought to be major translatable mRNAs. Almost all of the transcripts from the maize *psbB* operon cosediment with polysomes, suggesting that they are translated. Intercistronic cutting is not always required for translation of these RNAs [2].

Most chloroplast transcription units contain short inverted repeats at their 3′ ends, which were originally thought to function as transcription terminators. The role of such inverted repeats has been examined using an *in vitro* transcription system. It was found that these inverted repeats are ineffective as transcription terminators *in vitro* but serve as accurate and efficient RNA-processing signals [107]. The stability of RNAs containing inverted repeats at their 3′ ends is greatly enhanced.

The stability of chloroplast mRNAs and the protein interaction with their 3′ -inverted repeats have been investigated in spinach [e.g. 108], *Chlamydomonas* [e.g. 69, 109], barley [e.g. 30, 98] and mustard [83]. The observations suggest that nuclear-encoded proteins function in chloroplast mRNA maturation and differential mRNA stability, which are major control steps in chloroplast gene expression. Recently a 28 kDa RNA-binding protein which is responsible for processing the 3′ ends of chloroplast mRNA has been isolated [96] and related proteins containing RNA-binding domains have also been reported [72, 122].

Introns and splicing

Introns in chloroplast genes were first reported for the *23S rDNA* of *C. reinhardtii* [93]. Chloroplast genes which have been found to contain introns are listed in Table 2. Most genes possessing introns in higher plants contain single introns, while *Euglena* and *Chlamydomonas* polypeptide encoding genes have multiple introns [e.g. 48, 91]. Six chloroplast tRNA genes in higher plants have introns but none are known in algae. The presence of introns can be predicted by sequence homology with known genes (e.g. tRNA genes, ATP synthase genes, ribosomal protein genes) and by conserved intron boundary sequences (see below). However, experimental analyses are required to confirm the existence of introns and to determine the splice sites of pre-RNAs.

Introns found in chloroplast genes can be classified into four groups on the basis of the intron boundary sequences and possible secondary structures [16, 102]. Chloroplast group I introns

Table 2. Chloroplast genes containing introns.

Gene	Number of introns			Remarks
	higher plants	*Euglena*	*Chlamydomonas*	
23SrDNA	0	0	1	ORF, an intron in *Chlorella*
			6	in *C. eugametos*
16SrDNA	0	0	1	in *C. moewusii*
trnL-UAA	1	0	0	
trnI-GAU	1	0	0	
trnA-UGC	1	0	0	
trnV-UAC	1	0	0	
trnG-UCC	1	0	0	
trnK-UUU	1	0		ORF
rps2	0	4		
rps3	0	2	0	
rps8	0	3		
rps11	0	2	0	
rps12	3 exons	0	0	*trans*-splicing
rps14	0	1		
rps16	1			
rps19	0	2		
rpl2	1	0	0	no intron in spinach
rpl14	0	1	0	
rpl16	1	3	0	
rpl22	0	1		
rpl23	0	3		pseudogene in spinach
rpoB	0	8		
rpoC1	1	11		no intron in rice and maize
rpoC2	0	2		
tufA	no gene	3	0	
clpP	2			no intron in wheat and rice
rbcL	0	9	0	
psaA	0	3	3 exons	*trans*-splicing
psaB	0	6	0	
psbA	0	4	4	2 introns in *C. moewusii*
psbB	0	4		
psbC	0	1	0	ORF
psbE	0	2		
psbF	0	1		an-intron-within-an-intron
petB	1	1	0	no intron in *Chlorella*
petD	1		0	no intron in *Chlorella*
			1	3.5 kb intron in alga KS3/2
atpF	1	3		
ndhA	1			
ndhB	1			
IRF167–170	2			

ORF means the presence of an ORF in an intron. IRF indicates an intron-containing reading frame which is present in front of the *psaA* operon.

can be folded with a secondary structure typical of group I introns of fungal mitochondrial genes [11, 80]. The introns of *trnL*, the *23S rDNA* and *C. moewusii psbA* belong to this group. Introns of *trnI* and *trnA* in the 16S-23S spacer can be folded into a secondary structure which is similar to the postulated structure of group II introns in fungal mitochondrial genes, but their boundary sequences are different from those of chloroplast group III introns. Chloroplast group III introns have conserved boundary sequences GTGYGRY at the 5′ ends and RYCNAYY(Y)YNAY at the 3′ ends, and include introns in protein-encoding genes, *trnV*-UAC, *trnG*-UCC and *trnK*-UUU from higher plants. Their postulated secondary structures are similar to those of group II introns (therefore groups II and III are sometimes combined). It should be noted that introns in protein-encoding genes and some tRNA genes in higher plant chloroplasts show common features. This is not true of nuclear genes which are split. A fourth intron group has been described for *Euglena rpl14*, *rpl16*, *rpl23*, *rps2*, *rps3*, *rps8*, *rps11*, *rps14*, *rps19* and *tufA* [e.g. 16]. These introns are uniform in size (95–109 bp), share common features with each other and are distinct from chloroplast group I – III introns.

The tobacco gene for a ribosomal protein, CS12, is divided into one copy of *5′-rps12* and two copies of *3′-rps12*. *5′-rps12* contains exon 1 consisting of 38 codons and *3′-rps12* consists of exon 2 (78 codons), a 536 bp intron and exon 3 (7 codons). This gene structure was designated as a 'divided' gene [103]. The *5′*- and *3′-rps12* segments are separated by 28 kb and are transcribed independently. These two transcripts are spliced in *trans* to produce a mature mRNA for CS12 [e.g. 124] (Fig. 4). The 3′ -flanking sequence of exon 1 and the 5′ -flanking sequence of exon 2 fit the conserved boundary sequences of chloroplast group III introns. It is noteworthy that the tobacco *rps12* gene requires both *cis* and *trans* splicing in order to produce the mature mRNA. Liverwort *rps12* is divided into two parts; the *5′*- and *3′-rps12* segments are present in single copies and are located on opposing DNA strands [85]. The mRNA is also produced by *trans*-splicing [62].

The *Chlamydomonas psaA* gene is also divided into three parts [68]. The first exon of 30 codons is 50 kb away from the second exon (60 codons), which is itself 90 kb away from exon 3 (661 codons). All exons are flanked by the consensus intron boundary sequences. The three exons are transcribed independently as precursors, and the synthesis of mature *psaA* mRNA involves *trans* assembly of these three separate transcripts [e.g. 15]. Interestingly, exon 2 is cotranscribed with the upstream *psbD* gene, and *psaB* is uninterrupted as are *psaA* and *psaB* of higher plants (Fig. 4). At least one additional chloroplast locus (*tscA*) is required for *trans*-splicing of exons 1 and 2 and produces a small RNA of about 430 bases [34].

Chloroplast introns can be classified into three to four groups, suggesting the presence of multiple splicing pathways. Clear self-splicing of pre-RNAs from split chloroplast genes has not been demonstrated *in vitro*. The group III conserved intron boundary sequence is similar to that found in nuclear protein-encoding genes, suggesting that at least one group of chloroplast intron sequences is removed by a mechanism similar to that operating in the nucleus. Splicing of nuclear pre-mRNAs is catalyzed by protein-RNA complexes. RNA molecules are not thought to be imported into chloroplasts from the cytoplasm, which suggests that the RNA components of these complexes, if there are any, should be encoded in the chloroplast genome. Aside from *Chlamydomonas tscA* RNA, the tobacco chloroplast genome has been found to encode a small RNA species which is not tRNA or rRNA (unpublished observations). No *in vitro* splicing systems in chloroplasts are currently available. This makes it difficult to analyze individual steps in splicing and to detect factors involved in splicing in chloroplasts.

Conclusions

Thirty-eight different genes for RNA components and 74 different genes for polypeptides (including putative genes) have now been reported (see Table 1). Most of the putative genes (some *rps*/*rpl*,

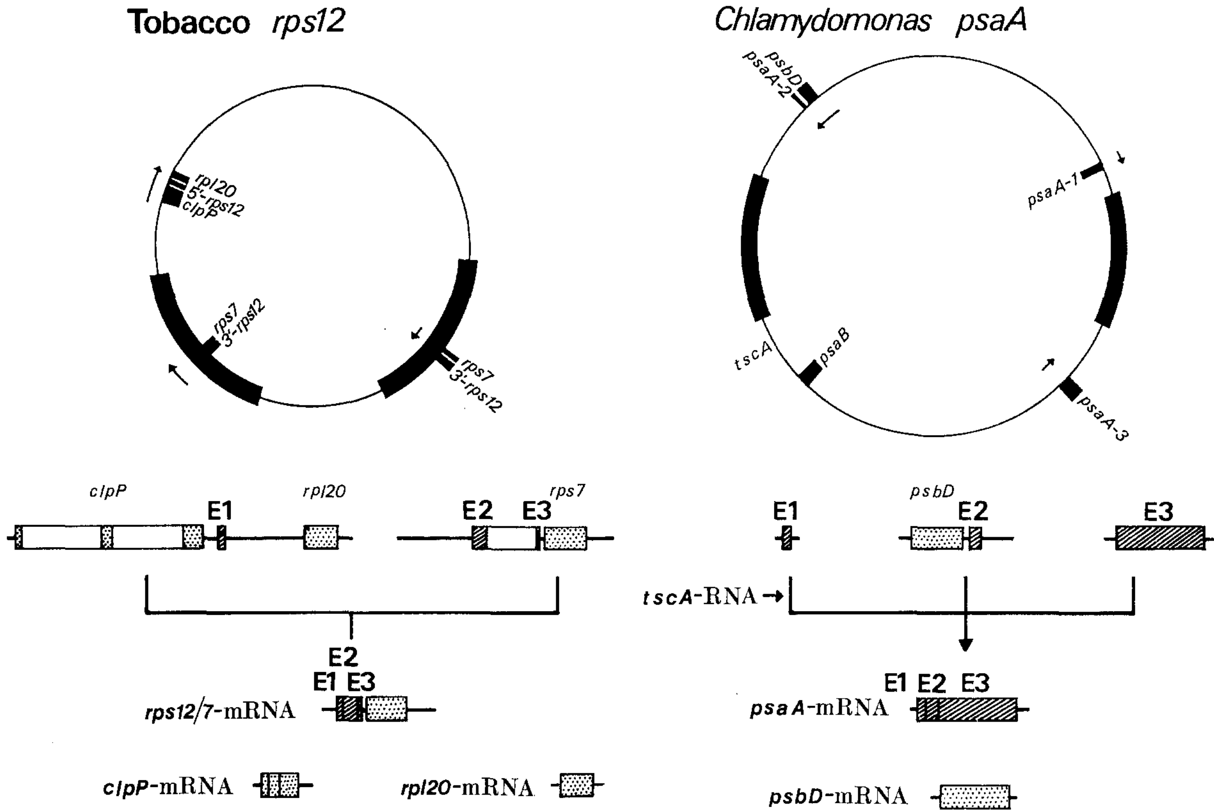

Fig. 4. Scheme for *rps12* and *psaA* mRNA maturation from separate pre-mRNAs [15, 124]. Upper circles show the location of genes and the direction of transcription. Bold lines, IR. In tobacco, *5'-rps12* contains exon 1 (E1), and *3'-rps12* has exon 2 (E2)-intron-exon 3 (E3). In *Chlamydomonas*, *psaA*-1, 2 and 3 contain exons 1, 2 and 3 (E1, E2, E3), respectively. The pathway is shown in the lower part. Slashed/dotted boxes indicate exons and open boxes, introns.

tuf, *inf*, *clp*, *ndh* and *frx*), which have been identified through homology with the corresponding genes of other organisms, are transcribed in chloroplasts, but their translation products remain to be isolated and characterized. There are still 26–34 ORFs (each over 29 codons in length) and twelve of them are conserved in size and sequence among the chloroplast genomes which have been completely sequenced [101]. Further efforts must be made to determine whether these ORFs (including putative genes) are functional chloroplast genes coding for polypeptides. Isolation of translation products followed by amino acid sequencing appears to be the best way to do so. There remain several long sequenced regions that have been assigned no genes and contain no significant ORFs. The *tscA* RNA was found in such a region from *Chlamydomonas* chloroplast DNA and further small RNA species might be encoded there.

The sequence analysis of algal chloroplast DNA and cyanelle DNA of *Cyanophora* has revealed the presence of new genes not found in land plant chloroplast DNAs as well as significant differences in genome organization in comparison both with each other and with land plants. Further analysis of chloroplast genomes from distantly related plant species as well as cyanelle and cyanobacterial genomes will provide the fundamental data needed to estimate the origins and process of chloroplast genome evolution as well as the phylogenetic relationships among plant species. Relevant to this, portions of chloroplast genomes are found in both nuclear and mitochondrial genomes (called 'promiscuous sequences'). The process of chloroplast DNA sequence transfer and the functional significance within the nucleus and the mitochondrion are interesting subjects for future study.

The molecular mechanism of chloroplast gene expression and its control are very interesting subjects. The number of RNA polymerase species in chloroplasts still remains to be answered. The structure of chloroplast genes and their modes of expression as presently understood exhibit both prokaryotic and eukaryotic features. This implies the presence of multiple RNA polymerase species. The control of chloroplast gene expression operates during several steps, transcription, post-transcription, translation and post-translation during chloroplast gene expression. It is also affected by environmental factors such as light. Recent studies suggest that the role of transcription in controlling gene expression is rather limited and that gene expression is more tightly controlled at the post-transcriptional level. This includes, RNA processing (cutting), RNA splicing and RNA stabilization. Chloroplast mRNA stability is currently being studied in detail.

The mechanism of chloroplast pre-RNA splicing is also very intriguing. For example, the tobacco chloroplast genome contains 18 distinct split genes and a comparable number of split genes have been found to date in *Euglena*. The introns found in chloroplast genes can be classified into three to four groups based on their structures. The largest group includes introns in some protein-encoding genes and in tRNA genes. Two genes, land plant *rps12* and *Chlamydomonas psaA* genes, require *trans*-splicing for expression. Some introns contain ORFs the predicted polypeptides of which show homology with maturases, reverse transcriptases and DNA endonucleases. Splicing of chloroplast pre-RNA is a most promising target for future study. For study of this process it is essential to develop *in vitro* splicing systems for chloroplasts. Recently chloroplast transformation systems have been developed in *Chlamydomonas* and tobacco and this method appears to be promising not only for *in vivo* the analysis of transcription but also for analysis of splicing pathways. At this time the method needs improvement before it can be used routinely.

The chloroplast contains a compact genetic system with circular DNAs of about 150 kb. This system has many interesting features as described in this paper. Many researchers have taken advantage of the relative simplicity of the chloroplast genetic system compared to nuclear and prokaryotic systems to produce new findings and hypotheses which are not only restricted to the plant world but apply to other organisms as well. It can be expected that further research on the chloroplast genome as a model system will continue to yield such exciting results.

References

1. Alt J, Morris J, Westhoff P, Herrmann RG: Nucleotide sequence of the clustered genes for the 44 kd chlorophyll *a* apoprotein and the '32 kd'-like protein of the photosystem II reaction center in the spinach plastid chromosome. Curr Genet 8: 597–606 (1984).
2. Barkan A: Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic mRNAs. EMBO J 7: 2637–2644 (1988).
3. Bedbrook JR, Bogorad L: Endonuclease recognition sites mapped on *Zea mays* chloroplast DNA. Proc Natl Acad Sci USA 73: 4309–4319 (1976).
4. Bedbrook JR, Kolodner R: The structure of chloroplast DNA. Annu Rev Plant Physiol 30: 593–620 (1979).
5. Bedbrook JR, Kolodner R, Bogorad L: *Zea mays* chloroplast ribosomal RNA genes are part of a 22,000 base pair inverted repeat. Cell 11: 739–749 (1977).
6. Bennett DC, Rogers SA, Chen LJ, Orozco Jr EM: A primary transcript in spinach chloroplasts that completely lacks a 5′ untranslated leader region. Plant Mol Biol 15: 111–119 (1990).
7. Berends T, Gamble PE, Mullet JE: Characterization of the barley chloroplast transcription units containing psaA-psaB and psbD-psbC. Nucl Acids Res 15: 5217–5240 (1987).
8. Bird CR, Koller B, Auffret AD, Huttly AK, Howe CJ, Dyer TA, Gray JC: The wheat chloroplast gene for CF_0 subunit I of ATP synthase contains a large intron. EMBO J 4: 1381–1388 (1985).
9. Blowers AD, Ellmore GS, Klein U, Bogorad L: Transcriptional analysis of Endogenous and foreign genes in chloroplast transformants of *Chlamydomonas*. Plant Cell 2: 1059–1070 (1990).
10. Bogorad L, Vasil IK (eds): The Molecular Biology of Plastids. Academic Press, San Diego (1991).
11. Bonnard G, Michel F, Weil JH, Steinmetz A: Nucleotide sequence of the split tRNA-Leu (UAA) gene from *Vicia faba* chloroplasts: evidence for structural homologies of the chloroplast tRNA-Leu intron with the intron from the autosplicable *Tetrahymena* ribosomal RNA precursor. Mol Gen Genet 194: 330–336 (1984).

12. Bowman CM, Dyer TA: 4.5S ribonucleic acid, a novel ribosome component in the chloroplasts of flowering plants. Biochem J 183: 605–613 (1979).
13. Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM, Jones AR, Randolph-Anderson BL, Robertson D, Klein TM, Shark KB, Sanford JC: Chloroplast transformation in *Chlamydomonas* with high velocity microprojectills. Science 240: 1534–1538 (1988).
14. Bülow S, Link G: Sigma-like activity from mustard (*Sinapis alba* L.) chloroplasts conferring DNA-binding and transcription specificity to *E. coli* core RNA polymerase. Plant Mol Biol 10: 349–357 (1988).
15. Choquet Y, Goldschmidt-Clermont M, Girard-Bascou J, Kück U, Bennoun P, Rochaix JD: Mutant phenotypes support a *trans*-splicing mechanism for the expression of the tripartite *psaA* gene in the *C. reinhardtii* chloroplast. Cell 52: 903–913 (1988).
16. 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: 7591–7607 (1989).
17. Christopher DA, Hallick RB: Complex RNA maturation pathway for a chloroplast ribosomal protein operon with an internal tRNA cistron. Plant Cell 2: 659–671 (1990).
18. Chun EHL, Vaugham MH, Rich A: The isolation and characterization of DNA associated with chloroplast preparations. J Mol Biol 7: 130–141 (1963).
19. Delp G, Igloi GL, Kössel H: Identification of *in vivo* processing intermediates and of splice junctions of tRNAs from maize chloroplasts by amplification with the polymerase chain reaction. Nucl Acids Res 19: 713–716 (1991).
20. Deng XW, Wing RA, Gruissem W: The chloroplast genome exists in multimeric forms. Proc Natl Acad Sci USA 86: 4156–4160 (1989).
21. Deno H and Sugiura M: Chloroplast tRNA-Gly gene contains a long intron in the D stem: nucleotide sequences of tobacco chloroplast genes for tRNA-Gly (UCC) and tRNA-Arg (UCU). Proc Natl Acad Sci USA 81: 405–408 (1984).
22. Douglas SE (1991): Unusual organization of a ribosomal protein operon in the plastid genome of *Cryptomonas* Φ: evolutionary considerations. Curr Genet 19: 289–294 (1991).
23. Driesel AJ, Crouse EJ, Gordon K, Bohnert HJ, Herrmann RG, Steinmetz A, Mubumbila M, Keller M, Burkard G, Weil JH: Fractionation and identification of spinach chloroplast transfer RNAs and mapping of their genes on the restriction map of chloroplast DNA. Gene 6: 285–306 (1979).
24. Dyer TA: The chloroplast genome: its nature and role in development. In: Baker NR, Barber J (eds) Chloroplast Biogenesis, pp. 23–69. Elsevier, Amsterdam (1984).
25. Eisermann A, Tiller K, Link G: *In vitro* transcription and DNA binding characteristics of chloroplast and etioplast extracts from mustard (*Sinapis alba*) indicate differential usage of the *psbA* promoter. EMBO J 9: 3981–3987 (1990).
26. Ellis RJ: Chloroplast proteins: synthesis, transport and assembly. Annu Rev Plant Physiol 32: 111–137 (1981).
27. Eneas-Filho J, Hartley MR, Mache R: Pea chloroplast ribosomal proteins: characterization and site of synthesis. Mol Gen Genet 184: 484–488 (1981).
28. Fish LE, Bogorad L: Identification and analysis of the maize P700 chlorophyll a proteins of PSI-A1 and PSI-A2 by high pressure liquid chromatography analysis and partial sequence determination. J Biol Chem 261: 8134–8139 (1986).
29. Fish LE, Kück U, Bogorad L: Two partially homologous adjacent light-inducible maize chloroplast genes encoding polypeptides of the P700 chlorophyll a-protein complex of photosystem I. J Biol Chem 260: 1413–1421 (1985).
30. Gamble PE, Mullet JE: Blue light regulates the accumulation of two *psbD-psbC* transcripts in barley chloroplasts. EMBO J 8: 2785–2794 (1989).
31. Gamble PE, Sexton TB, Mullet J: Light-dependent changes in *psbD* and *psbC* transcripts of barley chloroplasts: accumulation of two transcripts maintains *psbD* and *psbC* translation capability in mature chloroplasts. EMBO J 7: 1289–1297 (1988).
32. Gantt JS: Nucleotide sequences of cDNAs encoding four complete nuclear-encoded plastid ribosomal proteins. Curr Genet 14: 519–528 (1988).
33. Gillham NW, Harris EH, Randolph-Anderson BL, Boynton JE, Hauser CR, McElwain KB, Newman SM: Molecular genetics of chloroplast ribosomes in *Chlamydomonas*. In: Mache R, Stutz E, Subramanian AR (eds) The Translational Apparatus of Photosynthetic Organelles, pp. 127–144. Springer-Verlag, Berlin (1991).
34. Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD: A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. Cell 65: 135–143 (1991).
35. Graf L, Kössel H, Stutz E: Sequencing of 16S-23S spacer in a ribosomal RNA operon of *Euglena gracilis* chloroplast DNA reveals two tRNA genes. Nature 286: 908–910 (1980).
36. Gray JC: Genetics and synthesis of chloroplast membrane proteins in Photosynthesis. In: Amesz J (ed) Photosynthesis, pp. 319–342. Elsevier, Amsterdam (1987).
37. Gray JC, Hird SM, Dyer TA: Nucleotide sequence of a wheat chloroplast gene encoding the proteolytic subunit of an ATP-dependent protease. Plant Mol Biol 15: 947–950 (1990).
38. Gray PW, Hallick RB: Physical mapping of the *Euglena gracilis* chloroplast DNA and ribosomal RNA gene region. Biochem 17: 284–289 (1978).
39. Greenberg BM, Gruissem W, Hallick RB: Accurate pro-

cessing and pseudouridylation of chloroplast transfer RNA in a chloroplast transcription system. Plant Mol Biol 3: 97–109 (1984).
40. Gruissem W: Chloroplast RNA: transcription and processing. In Marcus A (ed) Biochemistry of Plants, pp. 151–191. Academic Press, San Diego (1989).
41. Gruissem W, Elsner-Menzel C, Latshaw S, Narita JO, Schaffer MA, Zurawski G: A subpopulation of spinach chloroplast tRNA genes does not require upstream promoter elements for transcription. Nucl Acids Res 14: 7541–7556 (1986).
42. Gruissem W, Greenberg BM, Zurawski G, Prescott DM, Hallick RB: Biosynthesis of chloroplast transfer RNA in a spinach chloroplast transcription system. Cell 35: 815–828 (1983).
43. Gruissem W, Zurawski G: Identification and mutational analysis of the promoter for a spinach chloroplast transfer RNA gene. EMBO J 4: 1637–1644 (1985).
44. Gruissem W, Zurawski G: Analysis of promoter regions for the spinach chloroplast *rbc* L, *atp* B and *psb*A genes. EMBO J 4: 3375–3383 (1985).
45. Haff LA, Bogorad L: Hybridization of maize chloroplast DNA with transfer ribonucleic acids. Biochem 15: 4105–4109 (1976).
46. Hagemann R: Genetics and molecular biology of plastids of higher plants. In: Stadler Symposium vol. 11, pp. 91–116. University of Missouri, Columbia (1979).
47. Hallick RB: Proposals for the naming of chloroplast genes. II Update to the nomenclature of genes for thylakoid membrane polypeptides. Plant Mol Biol Rep 7: 266–275 (1989).
48. Hallick RB, Gingrich JC, Johanningmeier U, Passavant CW: Introns in *Euglena* and *Nicotiana* chloroplast protein genes. In: van Vloten-Doting L, Groot GSP, Hall TC (eds) Molecular Form and Function of the Plant Genome, pp, 211–231. Plenum Press, New York (1985).
49. Hallick RB, Hollingsworth MJ, Nickoloff JA: Transfer RNA genes of *Euglena gracilis* chloroplast DNA. Plant Mol Biol 3: 169–175 (1984).
50. Hallick RB, Lipper C, Richards OD, Rutter WJ: Isolation of a transcriptionally active chromosome from chloroplasts of *Euglena gracilis*. Biochem 15: 3039–3045 (1976).
51. Heinemeyer W, Alt J, Herrmann RG: Nucleotide sequence of the clustered genes for apocytochrome b6 and subunit 4 of the cytochrome *b*/*f* complex in the spinach plastid chromosome. Curr Genet 8: 543–549 (1984).
52. Herrmann RG, Westhoff P, Alt J, Tittgen J and Nelson N: Thylakoid membrane protein and their genes. In: van Vloten-Doting L, Groot GSP, Hall TC (eds) Molecular Form and Function of the Plant Genome, pp. 233–256. Plenum Press, New York (1985).
53. Hiratsuka J, Shimada H, Whittier RF, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li Y, Kanno A, Nishizawa Y, Hirai 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).
54. Hirschberg J, McIntosh L: Molecular basis of herbicide resistance in *Amaranthus hybridus*. Science 222: 1346–1349 (1983).
55. Hu J, Bogorad L: Maize chloroplast RNA polymerase: the 180-, 120-, and 38-kilodalton polypeptides are encoded in chloroplast genes. Proc Natl Acad Sci USA 87: 1531–1535 (1990).
56. Hu J, Troxler RF, Bogorad L: Maize chloroplast RNA polymerase: the 78-kilodalton polypeptide is encoded by the plastid *rpoC1* gene. Nucl Acids Res 19: 3431–3434 (1991).
57. Igloi GL, Meinke A, Döry I, Kössel H: Nucleotide sequence of the maize chloroplast rpo $B/C_1/C_2$ operon: comparison between the derived protein primary structures from various organisms with respect to functional domains. Mol Gen Genet 221: 379–394 (1990).
58. Johnson CH, Kruft V, Subramanian AR: Identification of a plastid-specific ribosomal protein in the 30 S subunit of chloroplast ribosomes and isolation of the cDNA clone encoding its cytoplasmic precursor. J Biol Chem 265: 12790–12795 (1990).
59. Jolly SO, Bogorad L: Preferential transcription of cloned maize chloroplast DNA sequences by maize chloroplast RNA polymerase. Proc Natl Acad Sci USA 77: 822–826 (1990).
60. Karabin GD, Farley M, Hallick RB: Chloroplast gene for M_r 32 000 polypeptide of photosystem II in *Euglena gracilis* is interrupted by four introns with conserved boundary sequences. Nucl Acids Res 12: 5801–5812 (1984).
61. Koch W, Edwards K, Kössel H: Sequencing of the 16S-23S Spacer in a ribosomal RNA operon of zea mays chloroplast DNA reveals two split tRNA genes. Cell 25: 203–213 (1981).
62. Kohchi T, Umesono K, Ogura Y, Komine Y, Nakahigashi K, Komano T, Yamada Y, Ozeki H, Ohyama K: A nicked group II intron and *trans*-splicing in liverwort, *Marchantia polymorpha*, chloroplasts. Nucl Acids Res 16: 10025–10036 (1988).
63. Koller B, Delius H: *Vicia faba* chloroplast DNA has only one set of ribosomal RNA genes as shown by partial denaturation mapping and R-loop analysis. Mol Gen Genet 178: 261–269 (1980).
64. Koller B, Gingrich JC, Stiegler GL, Farley MA, Delius H, Hallick RB: Nine introns with conserved boundary sequences in the *Euglena gracilis* chloroplast ribulose-1,5-bisphosphate carboxylase gene. Cell 36: 545–553 (1984).
65. Kössel H, Döry I, Igloi G, Maier R: A leucine-zipper motif in photosystem I. Plant Mol Biol 15: 497–499 (1990).
66. Krebbers ET, Larrinua IM, McIntosh L, Bogorad L:

The maize chloroplast genes for the β and ε subunits of the photosynthetic coupling factor CF1 are fused. Nucl Acids Res 10: 4985–5002 (1982).
67. 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: 257–265 (1989).
68. Kück U, Choquet Y, Schneider M, Dron M, Bennoun P: Structural and transcriptional analysis of two homologous genes for the P700 chlorophyll *a*-apoproteins in *Chlamydomonas reinhardtii*: evidence for *in vivo trans*-splicing. EMBO J 6: 2185–2195 (1987).
69. Kuchka MR, Goldschmidt-Clermont M, van Dillewijn J, Rochaix JD: Mutation at the *Chlamydomonas* nuclear *NAC2* locus specifically affects stability of the chloroplast *psbD* transcript encoding polypeptide D2 of PS II. Cell 58: 869–876 (1989).
70. Lerbs S, Bräutigam E, Mache R: DNA-dependent RNA polymerase of spinach chloroplasts: characterization of α-like and σ-like polypeptides. Mol Gen Genet 211: 459–464 (1988).
71. Lerbs S, Bräutigam E, Parthier B: Polypeptides of DNA-dependent RNA polymerase of spinach chloroplast: characterization by antibody-linked polymerase assay and determination of sites of synthesis. EMBO J 4: 1661–1666 (1985).
72. Li Y, Sugiura M: Three distinct ribonucleoproteins from tobacco chloroplasts: each contains a unique amino terminal acidic domain and two ribonucleoprotein consensus motifs. EMBO J 9: 3059–3066 (1990).
73. Link G: DNA sequence requirements for the accurate transcription of a protein-coded plastid gene in a plastid *in vitro* system from mustard (*Sinapis alba* L.). EMBO J 3: 1697–1704 (1984).
74. Loiseaux-de Goër S, Markowicz Y, Dalmon J, Audren H: Physical maps of the two circular plastid DNA molecules of the brown alga *Pylaiella littoralis* (L.) Kjellm. Curr Genet 14: 155–162 (1988).
75. Manning JE, Wolstenholme DR, Ryan RS, Hunter JA, Richards OC: Circular chloroplast DNA from *Euglena gracilis*. Proc Natl Acad Sci USA 68: 1169–1173 (1971).
76. Markowicz Y, Loiseaux-de Goër S, Mache R: Presence of a 16S rRNA *pseudogene* in the bi-molecular plastid genome of the primitive brown alga *Pylaiella littoralis*. Evolutionary implication. Curr Genet 14: 599–608 (1988).
77. Matsubayashi T, Wakasugi T, Shinozaki K, Shinozaki KY, Zaita N, Hidaka T, Meng BY, Ohto C, Tanaka M, Kato A, Maruyama T, Sugiura M: Six chloroplast genes (*ndh A* to *F*) homologous to human mitochondrial genes encoding components of the respiratory-chain NADH dehydrogenase are actively expressed: determination of the splice-sites in *ndh A* and *ndh B* pre- mRNAs. Mol Gen Genet 210: 385–393 (1987).
78. McIntosh L, Poulsen C, Bogorad L: Chloroplast gene sequence for the large subunit of ribulose bisphosphate carboxylase of maize. Nature 288: 556–560 (1980).
79. Meng BY, Tanaka M, Wakasugi T, Ohme M, Shinozaki K, Sugiura M: Cotranscription of the genes encoding two P700 chlorophyll *a* apoproteins with the gene for ribosomal protein CS14: determination of the transcriptional initiation site by *in vitro* capping. Curr Genet 14: 395–400 (1988).
80. Michel F, Dujon B: Conservation of RNA secondary structure in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. EMBO J 2: 33–38 (1983).
81. Montandon PE, Stutz E: The genes for the ribosomal proteins S12 and S7 are clustered with the gene for the EF-Tu protein on the chloroplast genome of *Euglena gracilis*. Nucl Acids Res 12: 2851–2859 (1984).
82. Mullet JE: Chloroplast development and gene expression. Annu Rev Plant Physiol 39: 475–502 (1988).
83. Nickelsen J, Link G: Interaction of a 3′ RNA region of the mustard *trnK* gene with chloroplast proteins. Nucl Acids Res 17: 9637–9648 (1989).
84. Ohme M, Tanaka M, Chunwongse J, Shinozaki K, Sugiura M: A tobacco chloroplast DNA sequence possibly coding for a polypeptide similar to *E. coli* RNA polymerase β subunit. FEBS Lett 200: 87–90 (1986).
85. 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* chloroplast DNA. Nature 322: 572–574 (1986).
86. Orozco EM Jr, Rushlow KE, Dodd JR, Hallick RB: *Euglena gracilis* chloroplast ribosomal RNA transcription units II. Nucleotide sequence homology between the 16S-23S ribosomal RNA spacer and the 16S ribosomal RNA leader regions. J Biol Chem 255: 10997–11003 (1980).
87. Palmer JD: Chloroplast DNA exists in two orientations. Nature 301: 92–93 (1983).
88. Palmer JD: Comparative organization of chloroplast genomes. Annu Rev Genet 19: 325–354 (1985).
89. Pfitzinger H, Weil JH, Pillay DTN, Guillemaut P: Codon recognition mechanisms in plant chloroplasts. Plant Mol Biol 14: 804–814 (1990).
90. Pillay DTN, Guillemaut G, Weil JH: Nucleotide sequences of three soybean chloroplast tRNAs-Leu and re-examination of bean chloroplast $tRNA_2$-Leu sequence. Nucl Acids Res 12: 2997–3001 (1984).
91. Plant AL, Gray JC: Introns in chloroplast protein-coding genes of land plants. Photosynth Res 16: 23–39 (1988).
92. Reith M, Cattolico RA: Inverted repeat of *Olisthodiscus luteus* chloroplast DNA contains genes for both subunits of ribulose-1,5-bisphosphate carboxylase and the 32 000-dalton Q_B protein: phylogenetic implications. Proc Natl Acad Sci USA 83: 8599–8603 (1986).
93. Rochaix JD, Malnoe P: Anatomy of the chloroplast ribosomal DNA of *Chlamydomonas reinhardtii*. Cell 15: 661–670 (1978).

94. Sager R, Ishida MR: Chloroplast DNA in *Chlamydomonas*. Proc Natl Acad Sci USA 50: 725–730 (1963).
95. Sager R, Schlanger G: Chloroplast DNA: physical and genetic studies. In: King RC (ed) Handbook of Genetics, vol. 5, pp. 371–423. Plenum Press, New York (1976).
96. Schuster G, Gruissem W: Chloroplast mRNA 3′ end processing requires a nuclear-encoded RNA-binding protein. EMBO J 10: 1493–1502 (1991).
97. Schwarz Z, Kössel H: The primary structure of 16S rDNA from *Zea mays* chloroplast is homologous to *E. coli* 10S rRNA. Nature 283: 739–742 (1980).
98. Sexton TB, Christopher DA, Mullet JE: Light-induced switch in barley *psbD-psbC* promoter utilization: a novel mechanism regulating chloroplast gene expression. EMBO J 9: 4485–4494 (1990).
99. Shimada H, Fukuta M, Ishikawa M, Sugiura M: Rice chloroplast RNA polymerase genes: the absence of an intron in *rpoC1* and the presence of an extra sequence in *rpoC2*. Mol Gen Genet 221: 395–402 (1990).
100. Shimada H, Sugiura M: Pseudogenes and short repeated sequences in the rice chloroplast genome. Curr Genet 16: 293–301 (1989).
101. Shimada H, Sugiura M: Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucl Acids Res 19: 983–995 (1991).
102. Shinozaki K, Deno H, Sugita M, Kuramitsu S, Sugiura M: Intron in the gene for the ribosomal proteins S16 of tobacco chloroplast and its conserved boundary sequences. Mol Gen Genet 202: 1–5 (1986).
103. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Shinozaki, KY, 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).
104. Sijben-Müller G, Hallick R, Alt J, Westhoff P, Herrmann RG: Spinach plastid genes coding for initiation factor IF-1, ribosomal protein S11 and RNA polymerase α-subunit. Nucl Acids Res 14: 1029–1044 (1986).
105. Spielmann A, Roux E, von Allmen JM, Stutz E: The soybean chloroplast genome: complete sequence of the *rps*19 gene, including flanking parts containing exon 2 of *rpl* 2 (upstream), but lacking *rpl* 22 (downstream). Nucl Acids Res 16: 1199 (1988).
106. Steinmetz AA, Castroviejo M, Sayre RT, Bogorad L: Protein PSII-G: an additional component of photosystem II identified through its plastid gene in maize. J Biol Chem 261: 2485–2488 (1986).
107. Stern DB, Gruissem W: Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell 51: 1145–1157 (1987).
108. Stern DB, Jones H, Gruissem W: Function of plastid mRNA 3′ inverted repeats: RNA stabilization and gene-specific protein binding. J Biol Chem 264: 18742–18750 (1989).
109. Stern DB, Radwanski ER, Kindle KL: A 3′ stem/loop structure of the *Chlamydomonas* chloroplast *atpB* gene regulates mRNA accumulation *in vivo*. Plant Cell 3: 285–297 (1991).
110. Sugita M, Sugiura M: Putative gene of tobacco chloroplast coding for ribosomal protein similar to *E. coli* ribosomal protein S19. Nucl Acids Res 11: 1913–1918 (1983).
111. Sugiura M, Torazawa K, Wakasugi T: Chloroplast genes coding for ribosomal proteins in land plants. In: Mache R, Stutz E, Subramanian AR (eds) The Translational Apparatus of Photosynthetic Organelles, pp. 59–69. Springer-Verlag, Berlin (1991).
112. Takaiwa F, Sugiura M: The nucleotide sequence of 4.5S and 5S ribosomal RNA genes from tobacco chloroplasts. Mol Gen Genet 180: 1–4 (1980).
113. Tanaka M, Wakasugi T, Sugita M, Shinozaki K, Sugiura M: Genes for the eight ribosomal proteins are clustered on the chloroplast genome of tobacco (*Nicotiana tabacum*): similarity to the S10 and *spc* operons of *Escherichia coli*. Proc Natl Acad Sci USA 83: 6030–6034 (1986).
114. Thomas F, Massenet O, Dorne AM, Briat JF, Mache R: Expression of the *rpl23*, *rpl2* and *rps19* genes in spinach chloroplasts. Nucl Acids Res 16: 2461–2472 (1988).
115. 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).
116. Wakasugi T, Ohme M, Shinozaki K, Sugiura M: Structures of tobacco chloroplast genes for tRNA-Ile (CAU), tRNA-Leu (CAA), tRNA-Cys (GCA), tRNA-Ser (UGA) and tRNA-Thr (GGU): a compilation of tRNA genes from tobacco chloroplasts. Plant Mol Biol 7: 385–392 (1986).
117. Watson JC, Surzycki SJ: Both the chloroplast and nuclear genomes of *Chlamydomonas reinhardtii* share homology with *Escherichia coli* genes for transcriptional and translational components. Curr Genet 7: 201–210 (1983).
118. Webber AN, Malkin R: Photosystem I reaction-centre proteins contain leucine zipper motifs: a proposed role in dimer formation. FEBS Lett 264: 1–4 (1990).
119. Westhoff P, Herrmann RG: Complex RNA maturation in chloroplasts: the *psb*B operon from spinach. Eur J Biochem 171: 551–564 (1988).
120. Willey DL, Auffret AD, Gray JC: Structure and topology of cytochrome *f* in pea chloroplast membranes. Cell 36: 555–562 (1984).
121. Yamada T, Shimaji M: Peculiar feature of the organization of rRNA genes of the *Chlorella* chloroplast DNA. Nucl Acids Res 14: 3827–3839 (1986).
122. Ye L, Li Y, Fukami-Kobayashi K, Go M, Konishi T,

Watanabe A, Sugiura M: Diversity of a ribonucleoprotein family in tobacco chloroplasts: two new chloroplast ribonucleoproteins and a phylogenetic tree of ten chloroplast RNA-binding domains. Nucl Acids Res 19: 6485–6490 (1991).

123. Yokoi F, Tanaka M, Wakasugi T, Sugiura M: The chloroplast gene for ribosomal protein CL23 is functional in tobacco. FEBS Lett 281: 64–66 (1991).
124. Zaita N, Torazawa K, Shinozaki K, Sugiura M: *Trans* splicing *in vivo*: joining of transcripts from the 'divided' gene for ribosomal protein S12 in the chloroplasts of tobacco. FEBS Lett 210: 153–156 (1987).
125. Zaitlin D, Hu J, Bogorad L: Binding and transcription of relaxed DNA templates by fractions of maize chloroplast extracts. Proc Natl Acad Sci USA 86: 876–880 (1989).
126. Zurawski G, Bohnert HJ, Whitfeld PR, Bottomley W: Nucleotide sequence of the gene for the M_R 32,000 thylakoid membrane protein from *Spinacia oleracea* and *Nicotiana debneyi* predicts a totally conserved primary translation product of M_R 38,950. Proc Natl Acad Sci USA 79: 7699–7703 (1982).
127. Zurawski G, Bottomley W, Whitfeld PR: Structure of the genes for the β and ε subunit of spinach chloroplast ATPase indicates a dicistronic mRNA and an overlapping translation stop/start signal. Proc Natl Acad Sci USA 79: 6260–6264 (1982).
128. Zurawski G, Clegg MT: Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. Annu Rev Plant Physiol 38: 391–418 (1987).