Annu. Rev. Genet. 1998. 32:437–59

EVOLUTION AND MECHANISM OF TRANSLATION IN CHLOROPLASTS

Masahiro Sugiura, Tetsuro Hirose, and Mamoru Sugita
Center for Gene Research, Nagoya University, Nagoya 464-8602, Japan;
e-mail: h44979a@nucc.cc.nagoya-u.ac.jp

KEY WORDS: chloroplast genome, ribosomal gene, ribosome-binding site, RNA editing, tRNA

ABSTRACT

The entire sequence (120 ∼ 190 kb) of chloroplast genomes has been determined from a dozen plant species. The genome contains from 87 to 183 known genes, of which half encode components involved in translation. These include a complete set of rRNAs and about 30 tRNAs, which are likely to be sufficient to support translation in chloroplasts. RNA editing (mostly C to U base changes) occurs in some chloroplast transcripts, creating start and stop codons and changing codons to retain conserved amino acids. Many components that constitute the chloroplast translational machinery are similar to those of *Escherichia coli*, whereas only one third of the chloroplast mRNAs contain Shine-Dalgarno–like sequences at the correct positions. Analyses conducted in vivo and in vitro have revealed the existence of multiple mechanisms for translational initiation in chloroplasts.

CONTENTS

0066-4197/98/1215-0437$08.00

INTRODUCTION

Chloroplasts are plant organelles that contain the entire machinery for the process of photosynthesis. In addition, chloroplasts possess their own genome, multiple copies of circular double-stranded DNA molecules, typically 150 kb in size, with over 100 different genes. According to the endosymbiotic theory, chloroplasts were derived from an ancestral photosynthetic prokaryote(s) related to cyanobacteria (72). This idea was supported at the molecular level by the early observations that nucleotide sequences of chloroplast 5S rRNAs and 16S rRNA genes are highly similar to those from the cyanobacterium *Anacystis nidulans* (*Synechococcus* PCC6301) (22, 111). Sequence analyses of chloroplast (plastid) genomes from a wide range of plant species and cyanobacterial genomes have provided data fundamental to estimating the origin and evolution of chloroplasts as well as the phylogenetic relationships among photosynthetic organisms. After inhabiting the eukaryotic cell lineage leading to plants, there was a massive transfer of genes from the ancestral organelle to the nucleus, and the chloroplast became dependent on the nucleus for most of their gene products and regulatory control (e.g. 36, 73).

This review focuses on the chloroplast translational machinery and discusses evolutionary relationships among chloroplast genes encoding its components (rRNAs, tRNAs, and ribosomal proteins) from a wide range of plants and related cyanobacterial genes. The alteration of genomic information by RNA editing and the mechanism of translational initiation in chloroplasts are also highlighted. This topic is very timely in light of the recent complete sequencing of several algal chloroplast genomes and a cyanobacterial genome and the substantial progress in the mechanistic analysis of RNA editing and translational initiation of chloroplast mRNAs. Other aspects of the chloroplast genome have been reviewed elsewhere (30–32, 38, 48, 75, 77, 101, 108).

OVERALL STRUCTURE OF THE CHLOROPLAST GENOME

The genome size of present-day cyanobacteria has been reported to range from 2700 kb (*Synechococcus* PCC 6301/7942 and PCC 7002) (19, 49) to 6400 kb (*Anabaena* PCC 7120) (2). The siphonous green alga *Codium fragile* has the smallest chloroplast (ct) DNA known (89 kb) (71), whereas the green alga *Chlamydomonas moewusii* has the largest (292 kb) ctDNA so far analyzed precisely (by constructing a physical restriction map) (113). However, there is a report that the ctDNA of the green alga *Acetabularia* could be as large as 400 kb as estimated by a rough method (restriction enzyme analysis) (63). Therefore, endosymbiont genome reduction to between 5% and 10% and massive gene transfer to the host nucleus are evident after inhabitation (e.g. 36, 83).

An outstanding feature of ctDNA is the presence of a large inverted repeat (IR) ranging from 5 kb to 76 kb in length (84). This arrangement results in duplication of the rRNA genes and other genes included within the IRs. Most of the size variation among ctDNAs in land plants can be accounted for by changes in the length of the IR. For example, geranium ctDNA is unusually large (217 kb), with most of the extra size due mainly to a 76 kb IR (85). The ctDNAs of some legumes, conifers, and algae are exceptions to this pattern and lack IRs (e.g. 84, 88). It has been suggested that the IRs were present in the common ancestor of land plants and that one segment of the IR was lost in some legumes and conifers during evolution (84). This hypothesis was supported, at least in part, by the observation that black pine ctDNA retains an apparent remnant (495 bp) of the large IR (112). No such IR has been observed in cyanobacterial genomes analyzed so far. The genomes of *Synechococcus* 6301 (49) and of *Synechocystis* PCC 6803 (50) contain two copies of rRNA gene clusters in the inverse orientation, but no additional genes are present in the repeats. Therefore, it is hypothesized that the IR was acquired during the reduction of the symbiont genome or that ancestral cyanobacteria-like organisms had the IR in their genomes.

Entire ctDNA sequences have been determined from eight land plants and six algae, and several more ctDNAs will soon be sequenced completely. These include dicots, *Nicotiana tabacum* (97), *Arabidopsis thaliana* (S Tabata, personal communication), and *Epifagus virginiana* (120); monocots, *Oryza sativa* (41) and *Zea mays* (69); gymnosperm, *Pinus thunbergii* (118); psilotopside, *Psilotum nudum* (117); bryophyte, *Marchantia polymorpha* (82); and algae, *Chlorella vulgaris* (116), *Euglena gracilis* (33), *Porphyra purpurea* (89), *Odontella sinensis* (57), *Cyanophora paradoxa* (66, 102), and *Cyanidioschyzon merolae* (80). Characterization of entire ctDNAs from diverse plant groups has helped in understanding the evolution of chloroplast genomes. "The chloroplast genome tree" has recently been proposed by analyzing a concatenate amino acid alignment from all protein-coding genes in 10 ctDNAs (73). This tree information is sufficient to resolve the evolution of these chloroplasts. Mapping of the presence and absence of multiple parallel gene losses are the underlying theme of ctDNA evolution, providing evidence for endosymbiotic gene transfer to the nucleus in plants.

The gene content and arrangement of ctDNAs in land plants are relatively uniform from species to species. However, this is not always the case for ctDNAs from a variety of algae. For example, the red alga *Porphyra* ctDNA contains over 70 new genes not found in the 8 sequenced ctDNAs from land plants (88, 89). Nonphotosynthetic and parasitic plants, such as *Epifagus*, are the exception and generally contain much reduced ctDNAs of 50 to 73 kb owing to the loss of many normal chloroplast genes (e.g. 84, 120).

Chloroplast genes fall into three main categories: genes involved in transcription/translation; those related to photosynthesis; and those encoding proteins

involved in the biosynthesis of amino acids, fatty acids, pigments, etc (51). The expression and structure/function relationships of photosynthesis-related genes have been extensively analyzed (e.g. 37) in parallel with the study of photosynthetic mechanisms. The presence of protein-encoding genes was deduced through a homology search (88), and little is know about their transcripts and translation products. In-depth discussion of their role is premature.

ORGANIZATION OF RIBOSOMAL GENES

Chloroplast genes encoding components of the translational machinery were initially found through homology with the corresponding genes of *Escherichia coli*, and these genes were later confirmed by analyzing their gene products. The translational machinery in chloroplasts was believed to be similar to that in *E. coli*. The chloroplast translational machinery consists of 70S ribosomes, tRNAs, aminoacyl tRNA synthetases, initiation factors, elongation factors, release factors, etc. Among these, rRNAs, tRNAs, ribosomal proteins, and several other factors are encoded by ctDNA (Table 1).

Ribosomal RNA Gene Clusters

The chloroplast genome encodes all the rRNA species (16S, 23S, and 5S rRNAs), and their genes (*rrnS*) are clustered and transcribed as a unit (34). The genes for $tRNA^{Ile}$ (GAU) and $tRNA^{Ala}$ (UGC) are located in the spacer between 16S and 23S rRNA genes, so that the basic organization of chloroplast rRNA gene clusters is represented by 16S-$tRNA^{Ile}$ – $tRNA^{Ala}$ – 23S – 5S. Almost all

Table 1 Number of known genes on the entirely sequenced chloroplast genomes[a]

	Land plants		Algae				
	Photosynthetic[b]	*Epifagus*[c]	*Chlorella*	*Euglena*	*Odontella*	*Porphyra*	*Cyanelle*
Total	105 ~ 113	40	105	87	129	183	145
Genetic system							
rRNAs	4	4	3	3	3	3	3
tRNAs	30 ~ 32	17	31	27	25	35	33
r-proteins	20 ~ 21	15	21	21	44	47	37
Others	5–6	2	6	7	6	13	6
Photosynthesis							
Rubisco and thylakoids	31 ~ 32	0	32	27	42	44	41
NADH dehydrogenases	11	0	0	0	0	0	0
Phycobilisomes	0	0	0	0	0	10	7
Biosynthesis and miscellaneous	2 ~ 8	2	11	2	9	31	18

[a]Condensed from Table 6.1 in Kapoor & Sugiura (1998). *ycf*s and unique ORFs are not included.
[b]Including rice, maize, tobacco, black pine and liverwort.
[c]Nonphotosynthetic and parasitic flowering plant.

chloroplast rRNA gene clusters so far reported have this basic structure. Therefore, the ancestor of chloroplasts is suggested to be one that is related to *Synechococcus* 6301, in which its two *rrn* clusters have the same basic organization and are inversely oriented in its circular genome (49). This hypothesis is compatible with that obtained by sequence comparison of 5S rRNAs (22) and 16S rRNA genes (111) between land plant chloroplasts and *Synechococcus* 6301. *Synechocystis* 6803, a cyanobacterium distantly related to the above species (100), has two copies of *rrn* clusters with a different arrangement [16S – $tRNA^{Ile}$ – 23S – 5S, lacking the $tRNA^{Ala}$ gene (50)]. The *rrn* clusters similar to that of chloroplasts include *rrnA*, *rrnD*, and *rrnH* in *E. coli* (7); *rrnA* and *rrnO* in *Bacillus subtilis* (61); and *rrnA*, *rrnC*, and *rrnD* in *Haemophilus influenzae* Rd (25).

There are several variations from the basic organization among chloroplast *rrn* clusters (Figure 1). The large subunit rRNA is often split into two to four pieces. The 4.5S rRNA found in land plant chloroplasts is derived from the 3′ end of *E. coli*–type 23S rRNA (23), and the 7S/3S rRNAs found in *Chlamydomonas reinhardtii* chloroplasts from its 5′ end (92). The large subunit rRNA in *Chlamydomonas eugametos* chloroplasts is unusual and consists of four separate rRNA pieces [280(α), 52(β), 810(γ), and 1720(δ) nucleotides (nt) (114)]. Irrespective of how it is split, these rRNA pieces can assume a secondary structure almost identical to that proposed for the 23S rRNA of *E. coli*, confirming that the continuity of 23S rRNA is not essential for its function.

Ribosomal Protein Gene Clusters

Chloroplast 70S ribosomes contain about 60 different protein components (14, 34, 67, 103), several more protein species than those in *E. coli* ribosomes (21 proteins in the 30S subunit and 34 proteins in the 50S subunit). In land plant chloroplasts, one third of the ribosomal proteins are encoded by ctDNA and the remaining two thirds are encoded in the nuclear genome. For example, the tobacco chloroplast ribosome has from 58 to 62 proteins, with from 24 to 26 occurring in the 30S subunit and from 34 to 36 in the 50S subunit (14), among which 12 of the 30S subunit proteins and 9 of the 50S subunit proteins are encoded by tobacco ctDNA (109). Identity scores between chloroplast-encoded proteins and the counterparts in *E. coli* range from 74% (S12) to 18% (L32).

At least five novel proteins that are not homologous to any known *E. coli* ribosomal proteins have been reported in the higher plant chloroplast ribosome (103). These novel proteins are all encoded in the nucleus. ORFs similar to those encoding the novel components were not found in the entirely sequenced genomes from *E. coli* (7) and *Synechocystis* 6803 (50). These observations support the hypothesis that genes encoding the unique ribosomal proteins originally resided in the ancestral host genome and the gene products are exported into chloroplasts after acquiring transit-peptide sequences. There are still 20 or

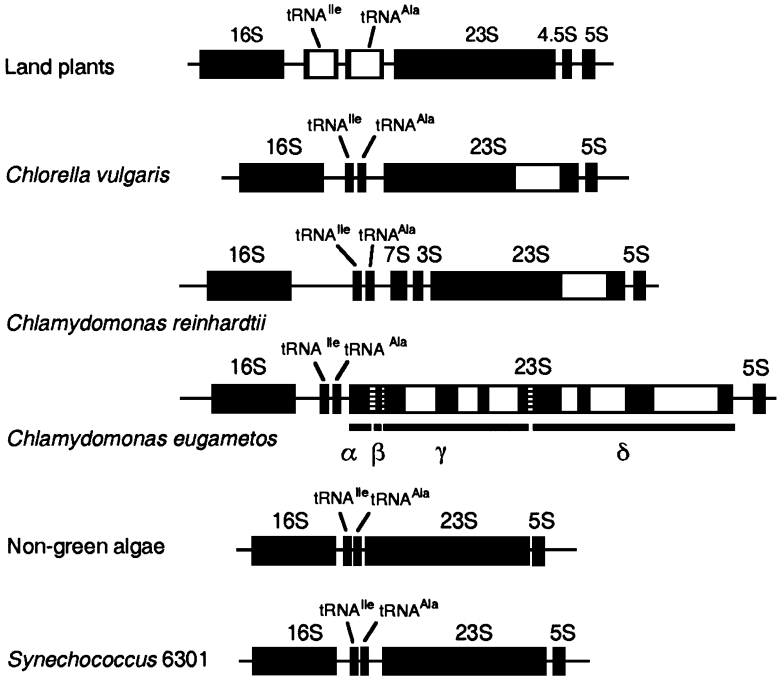

Figure 1 Arrangement of the rRNA gene clusters. Filled areas denote coding regions for mature RNA species and open boxes denote introns. Three internal transcribed spacer sequences (*hatched boxes*) and six introns interrupt the 23S gene of *Chlamydomonas eugametos* that is transcribed into mature four distinct rRNA pieces, termed α, β, γ, and δ (114). Non-green algae include *Euglena gracilis, Porphyra purpurea, Odontella sinensis, Cyanophora paradoxa,* and *Cyanidioschyzon merolae. E. coli rrnA, rrnD,* and *rrnH, B. subtilis rrnA* and *rrnO,* and *H. influenzae rrnC* and *rrnD* are similar to the *Synechococcus* cluster.

more significant ORFs in the ctDNA of higher plants, and it cannot be excluded that some of the ORFs may encode further novel ribosomal proteins.

Of the 21 ribosomal proteins in tobacco chloroplasts, 10 are encoded as a unit, which constitutes the largest chloroplast gene cluster (ctL23 cluster) including further genes coding for initiation factor 1 (IF-1) and the α subunit of RNA polymerase (107) (Figure 2). The L2 and L16 genes each contain single introns (107). The organization of this gene cluster is conserved in land plant ctDNAs, and it appears to have retained many genes from an ancestral symbiont operon that is similar to the L3 cluster of *Synechococcus* 6301 (105), which corresponds to an apparent fusion cluster of three *E. coli* operons (S10, *spc*, and α). The genes for S10, S14, L30, and S4 in the *E. coli* operons are not present in the L3 cluster of *Synechococcus* 6301 or in the ctL23 cluster, suggesting that

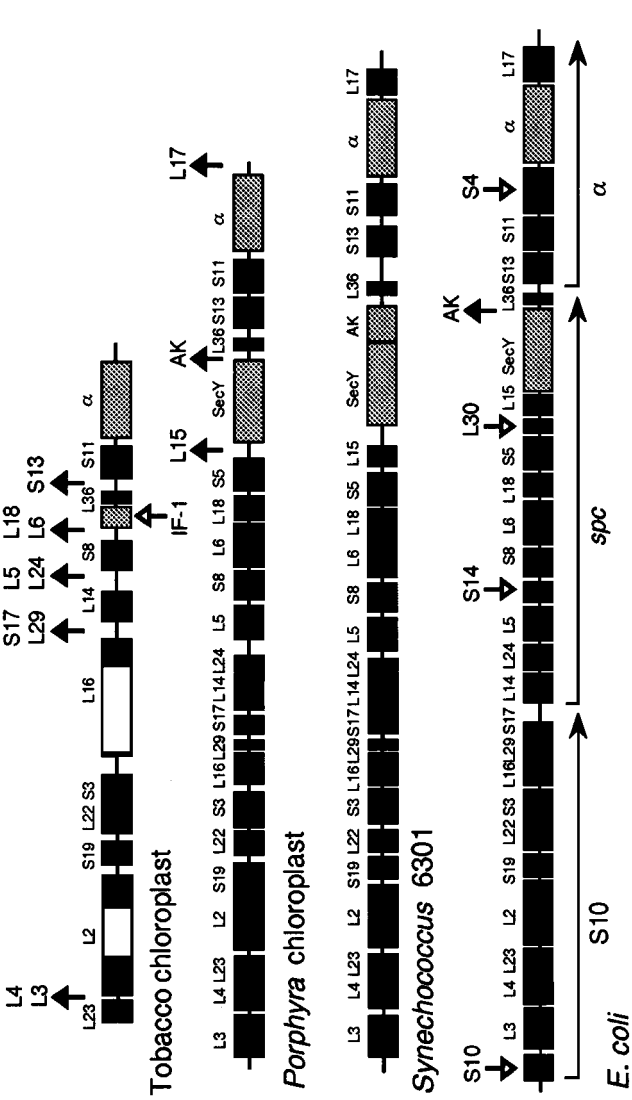

Figure 2 Arrangement of the ribosomal protein gene clusters. *Filled arrows* indicate genes that have been lost from the *Synechococcus* cluster (L3 cluster) but that have been translocated into the nuclear genome in tobacco (ctL23 cluster) and *Porphyra*. *Open arrows* indicate *E. coli* genes that have been inserted into the operons from the other loci with respect to *Synechococcus*. AK, α, and IF-1 show genes encoding adenylate kinase, α subunit of RNA polymerase, and initiation factor 1, respectively.

the divergence of *E. coli*–like and *Synechococcus* 6301–like ancestors occurred before the endosymbiosis.

The additional ribosomal protein genes present in the L3 cluster of *Synechococcus* 6301 are missing in land plant ctDNAs and have either been relocated within ctDNA (S14, S4) or transferred to the nuclear genome (34). In contrast, some algal ctDNAs contain an expanded ctL23 cluster; for example, 20 ribosomal protein genes are present in the corresponding cluster of *Porphyra purpurea* and *Cyanidioschyzon merolae*, which is close to the L3 cluster of *Synechococcus* 6301. Therefore, of the chloroplast genomes sequenced to date, that of *Porphyra* appears to be the most "primitive" (88). As shown in Figure 2, the IF-1 gene is present only in the ctL23 cluster of land plants. *Synechocystis* 6803 shares the same gene order as in *Synechococcus* 6301 except that the IF-1 gene is located between the AK and L36 gene within the cluster. Therefore, the evolution of chloroplast genomes might have involved not only gene transfer but also gene relocation (34, 103); alternatively, polyphyletic origins of chloroplasts have been suggested.

ORGANIZATION OF TRANSFER RNA GENES

Sequencing of entire ctDNAs revealed the presence of from 27 to 35 potential tRNA genes (*trnS*), except in *Epifagus*. In addition, one or more pseudo tRNA genes are also present in some ctDNAs (96). These pseudogenes found in monocotyledonous plants are located at or near the inversion endpoints, and the involvement of tRNA genes in genome rearrangements during evolution has been proposed (41, 47, 96). Six tRNA genes from land plants harbor long single introns (from 0.5 to 2.5 kb), while none or only one exists in algal chloroplasts (51). The chloroplast gene coding for $tRNA^{Leu}$ (UAA) contains a group I-type intron (13) and is found not only in land plants but also is *Chlorella* (116) and the cyanelle of *Cyanophora* (102). This split $tRNA^{Leu}$ gene has also been found in *Anabaena* (121) and *Synechococcus* (60, 104). Interestingly, the pre-$tRNA^{Leu}$ from the cyanobacteria self-splices in vitro (126), whereas no self-splicing of chloroplast pre-$tRNA^{Leu}$ molecules has been observed so far.

Transfer RNA Species and Codon Recognition

Tobacco ctDNA encodes 30 tRNA species. The location of most tRNA genes was consistent with that based on hybridization with isolated chloroplast tRNAs (5), and expression of all the 30 tRNA genes was confirmed by Northern blot analysis using total tRNA preparation from tobacco chloroplasts (110). All 61 possible codons are used in the chloroplast genes encoding 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.

No tRNA that recognizes codons CUU/C (Leu), CCU/C (Pro), GCU/C (Ala), or CGC/G (Arg) according to normal wobble base-pairing is not encoded by the tobacco genome. If the two-out-of-three mechanism operates in the chloroplast, three single tRNA species, $tRNA^{Pro}$(UGG), $tRNA^{Ala}$(UGC), and $tRNA^{Arg}$(ACG), can read all four Pro, Ala, and Arg codons, respectively (these tRNAs form only GC pairs in their first and second codon-anticodon interactions). If "U:N wobble" (U at the first position of anticodons must be unmodified) is assumed (U:N pairing between the first position of anticodons and the third positions of codons), CUU/C (Leu), GUU/C (Val), UCU/C (Ser), ACU/C (Thr) codons can be recognized by a single relevant tRNA species. Thus, the 30 tRNAs are probably sufficient to read all 61 codons (110). These hypotheses have been substantiated by a study in which some of these chloroplast tRNAs were purified from bean and their decoding properties analyzed in a tRNA-dependent wheat germ protein-synthesizing system (86).

The *Synechocystis* 6803 genome encodes 41 tRNA species that are sufficient to recognize all codons by normal wobble base-pairing (50), whereas no ctDNAs have so far been reported to contain a full set of tRNA genes, suggesting that one or more tRNA genes have been lost during chloroplast evolution. The number of anticodons not recognized by chloroplast-encoded tRNAs according to normal wobble base-pairing is dependent on the plant species. *Epifagus* ctDNA has only 17 tRNA genes, and the absence of 13 tRNA genes present in tobacco indirectly raises another possibility of translocation of tRNA genes to the nuclear genome and of import of nuclear-encoded tRNA species to effect translation (120). Both the completely sequenced genomes of *Mycoplasma pneumoniae* (816 kb) and *M. genitalium* (580 kb), the smallest prokaryotic genomes, contain an identical set of 33 tRNA genes that encode tRNAs necessary to recognize all codons (26, 40). Taken together, chloroplasts with only 27–35 tRNA genes may import one or more tRNA species so that supporting translation without using the expanded wobble base-pairing, although there is no direct evidence to support RNA transport into chloroplasts.

Codon Alteration by RNA Editing

Complete coding information was initially thought to be written in a DNA sequence. However, DNA information is often modified at the level of transcripts by RNA editing and by RNA splicing in chloroplasts. The first evidence for RNA editing in chloroplasts came from the maize *rpl2* transcript in which an ACG codon changes to a start codon AUG (46). A similar creation of initiation codons was also reported in other mRNAs (9, 59, 62, 78). Editing of certain transcripts is partial or differentially regulated by developmental and external conditions, indicating that editing is a critical determinant for gene expression as unedited transcripts are unavailable for translation (9, 45).

RNA editing is not limited to initiation codons but has been observed at internal codons (e.g. 68, 69) and, rarely, in untranslated regions (8). Most edited codons restore amino acids that are conserved in the corresponding proteins of other plants. Therefore, RNA editing is thought to be functionally significant, and its role is likely as a genetic correction process of wrong genomic sequences at the RNA level (8, 29, 69).

RNA editing occasionally creates stop codons, either UAA from CAA or UGA from CGA, which alters the size of translation products and results in the conservation of protein sizes (79, 115, 124, 125). Striking cases reported that two editing events result in the creation of an initiation codon and a stop codon within the same transcripts, leading to the formation of new reading frames, or new protein-coding genes at the mRNA level. The *petL* reading frame in black pine (115) and the *accD* reading frame in bracken (79) are thus produced at the RNA level.

The number of RNA editing sites in chloroplast transcripts was exhaustively analyzed; so far, 26 have been found in maize (69; R Bock, personal communication), 32 in tobacco (43; our unpublished data), and 26 in black pine (115), as these three genomes have been completely sequenced. About 0.13% of C residues in the genomic codons is altered to U residues in mRNAs. The C to U conversion does not occur at random, but rather is strongly biased (Table 2). Among the 82 editing sites (the sum of the three species) that cause amino acid substitution, 40 (48%) are UCA (Ser) to UUA (Phe) transitions and 13 (16%) are CCA (Pro) to CUA (Leu) changes. The second base of codons are frequently edited (72 of the 82 sites, 87%), and 9 changes were found at the first base. Only one was observed at the third base (Ser, CUC to CUU in the tobacco *atpA* transcript), thus not leading to amino acid substitution (silent editing) (42). If RNA editing is a genetic correction mechanism, it only works on limited genetic errors.

RNA editing has been identified in chloroplast transcripts from all major lineages of land plants (27) but has not been reported in those from algae and cyanobacteria. Recently, extensive C to U and novel U to C editing was reported in the *rbcL* and *atpB* transcripts from a bryophyte, the hornwort *Anthoceros formosae*; 20 and 29 sites in *rbcL* and *atpB* transcripts, respectively (124, 125). If these numbers are extrapolated to the whole genome, editing frequency would be comparable to that of mitochondria in higher plants [$\sim$1000 sites (29)]. However, no editing in either organelle was found in another bryophyte, *Marchantia polymorpha* (29). One explanation is that this bryophyte originally had this capacity but lost it at some point after *M. polymorpha* diverged from other bryophytes (39). Phylogenetic analysis of RNA editing in chloroplast as well as plant mitochondria suggests that editing arose in the bryophyte, although

Table 2 Codon transition caused by editing of chloroplast transcripts from maize, tobacco, and black pine

First base	Second base U	Second base C	Second base A	Second base G	Third base
U	Phe ←3	Ser	Tyr ←	Cys ←	U
	Phe ←3	Ser ←1	Tyr	Cys	C
	Leu ←40	Ser	Stop ←	Stop	A
	Leu ←5	Ser	Stop	Trp ←	G
C	Leu	Pro ←1	His ←1	Arg ←1	U
	Leu ←1	Pro ←1	His	Arg	C
	Leu ←13	Pro ←1	Gln ←2	Arg	A
	Leu ←1	Pro ←1	Gln	Arg ←1	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met ←5	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val ←1	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Arrows show the direction of codon transitions. Numbers (82 sites in total) indicate the observed frequencies of the respective transitions.

neither editing frequency nor editing pattern of a specific transcript correlates with the phylogenetic tree (27). The mechanism of RNA editing in chloroplasts is being studied (e.g. 10, 11, 15, 16, 52, 70, 98) and is not discussed here.

UNIQUE MECHANISMS OF TRANSLATION

Certain components constituting the chloroplast translational machinery, e.g. tRNAs, rRNA, ribosomal proteins, IF-1∼3, EF-Tu, and EF-G, were identified through their similarity with those of the translational machinery in *E. coli*. The discovery of chloroplast-specific ribosomal proteins (see above) indicates the existence of additional components unique to chloroplasts that cannot be identified via homology, although the extent of such components/factors is not yet known. These novel components/factors so far found include chloroplast-specific ribosomal proteins and mRNA-specific regulatory factors which are all nuclear-encoded (see below).

Functional mRNA Species

Mature and functional mRNAs in chloroplasts have no cap-structures nor poly(A) tails, as found in cytoplasmic mRNAs. Poly(A)-rich sequences found in certain chloroplast mRNAs are not the poly(A) tail in cytoplasmic mRNAs but these are the target for rapid exonucleolytic decay of the mRNAs (58, 64, 65). The primary transcript harbors 5′ triphosphates, and its 5′-UTR is often processed endonucleolytically at a specific site(s). The 3′-end trimming generally occurs toward a hair-pin structure in the 3′-UTR (e.g. 31).

Among 70 protein-coding genes and 9 *ycfS* (conserved ORFs) present in the tobacco chloroplast genome, five (*ndhF, psbA, psbM, psbN*, and *rbcL*) are transcribed monocistronically (106). A methyl jasmonate-induced change in the length of the 5′-UTR was reported to impair translation of the barley *rbcL* transcript (87), suggesting that specific 5′-processing (5′ ends having monophosphates) is essential for certain chloroplast mRNAs to be functional. However, at least some other primary transcripts seem to be functional because transcripts with 5′-triphosphates from tobacco *psbA* are highly active in translation in vitro (44). In *Chlamydomonas*, nearly all chloroplast genes appear to be transcribed and translated as monocistronic mRNAs (90, 91).

Many chloroplast genes in higher plants are transcribed as polycistronic pre-mRNAs, which are then endonucleolytically processed into mostly monocistronic mRNAs (119, 122). Though the pre-mRNA from the maize *psbB-psbH-petB-petD* operon was translatable (3), processed monocistronic mRNAs were suggested to be more efficient in translation than their precursor (4). The tobacco *psaC-ndhD* gene cluster is transcribed as a dicistronic pre-mRNA that is then cleaved into monocistronic mRNAs. In vitro translation assays showed that the dicistronic mRNA is not functional and that the intercistronic cleavage is a prerequisite for both *psaC* and *ndhD* translation (45). An intramolecular interaction between an 8-nt sequence in the *psaC* coding region and its complementary 8-nt sequence in the 5′ *ndhD* UTR inhibited translation of the dicistronic pre-mRNA. Therefore, functional mRNAs in chloroplasts are, in principle, monocistronic, unlike those in *E. coli* and related bacteria.

Ribosome-Binding Sites in mRNAs

In *E. coli*, almost all mRNAs have the Shine-Dalgarno (SD) sequence (typically GGAGG), which can bind to the 3′ end of 16S rRNA in the 30S ribosomal subunit and initiate accurate translation (76). The SD sequence is located 7 ± 2 nt upstream from the initiation codon and this position is critical for its function. Figure 3 shows the 5′-UTRs of 10 ribosomal protein genes present in the tobacco ctL23 cluster and their counterparts in *Synechococcus* 6301 and *E. coli*. All the *E. coli* genes have SD sequences at a conserved position and ATG codons at the initiation site, which is reasonable to produce equimolar

Tobacco chloroplast

	-30	-20	-10		+10
rpl23	ACTATAACTC	ATAG**GGA**AGA	AAATCGATTT	ATG	GCGATAC
rpl2	AAAAGAACTT	AAAAAAAAAT	ACTTAATAGC	ATG	GCGATAC
rps19	TACAAAAAAA	AAAAAAATA**G GAG**TAACGTT		GTG	ACACGTT
rpl22	CTTATGATTC	AGTAGTA**GGA GG**CAAACCTT		ATG	CTAAAGA
rps3	AATAGTGGGG	ATGTATAATA	GTGGG**GGA**TT	ATG	GGACAAA
rpl16	AAATTCGATT	GACCATTTGA	TATAATTGCT	ATG	CTTAGTC
rpl14	CCCTTTGCAT	TCAAAGAAC**G GA**TTAAAAAA		ATG	ATTCAAC
rps8	ATTT**GGAGG**C	CCCAATAATT	TTAGTTCATC	ATG	GGTAGGG
rpl36	GATTCAGAAT	TAAGGTAAGG	AAT**GAG**AAAT	ATG	AAAATAA
rps11	CTCTGACTCA	TATTTAC**GAG**	ATGATAAAAT	ATG	GCAAAAG

Synechococcus PCC6301

	-30	-20	-10		+10
rpl23	CAGCGCTGGA	GAAGATTCA**G GAGG**TCTACG		GTG	GCTGAAG
rpl2	AGTCGTCGCA	AGTA**GGA**AAT	CACCAAGATT	ATG	GGAATCC
rps19	GCTTCTGCTA	TTCACGGAC**G GAG**TTGAATT		ATG	GCTCGCT
rpl22	GCCGGTCGTT	GAGATTTCT**G AG**ATAGATCC		ATG	GCTGTAG
rps3	GTCTAGGACG	CGAACAGA**GA GG**GAATACCT		GTG	GGACAGA
rpl16	AGATCGCTCT	AACGAAGGTT	AA**GAG**TCACC	ATG	CTCAGTC
rpl14	CGAAAGCGTC	GGTGCCT**GAG G**AAGCCATCC		ATG	ATTCAAC
rps8	TAAGAGGCAC	TTCCAAGCGA	A**GAGG**AAATT	ATG	GCGGTCA
rpl36	CAAAACTTT**G GAG**TCTTGCC		GAATACTGGC	ATG	AAAGTAA
rps11	GCATTTTGTC	TTGCTCAGAC	**GAGG**TTTACT	ATG	GCTCGAC

E. coli

	-30	-20	-10		+10
rpl23	TGCTGTTAAG	CAAGTT**GAGG AG**ATGCTGGC		ATG	ATTCGTG
rpl2	CGGCGCTGAG	TAAGTC**GGAG G**AGTAATACA		ATG	GCAGTTG
rps19	GCCGTAGCAA	ATAATTTTA**G AGG**ATAAGCC		ATG	CCACGTT
rpl22	AGCGAAGAAG	AAATAAGGTA	**GGAGG**AAGAG	ATG	GAAACTA
rps3	GTCCGATCGC	TGAGACTCT**G GAG**ACTAGCA		ATG	GGTCAGA
rpl16	CGTAAAGGCC	GTAAATAA**GG AG**CGTCGCTG		ATG	TTACAAC
rpl14	CAGTAGTAGT	TGACATTAGC	**GGAG**CCTAAA	ATG	ATCCAAG
rps8	CACCAATTGA	ATCACG**GGAG GT**AAAGACAG		ATG	AGCTAGC
rpl36	TGGTCGCCCG	AGAAGTTAC**G GAG**AGTAAAA		ATG	AAAGTTC
rps11	CGATCAAGAA	ATAATC**GGGG**	TGATTGAATA	ATG	GCAAAGG

Figure 3 Comparison of the 5′-UTRs of 10 ribosomal protein genes in the ctL23 cluster from tobacco chloroplasts and from the corresponding *Synechococcus* 6301 and *E. coli* genes (see *Figure 2*). The SD-like sequences are shown by bold characters with underlining.

ribosomal proteins. This is not the case for chloroplasts. SD-like sequences are found but their positions are highly variable, and *rpl2* and *rpl16* do not even have such sequences. The SD-like sequences of *rps19*, *rpl22*, *rpl14*, and *rps11* are likely to be functional but the others seem not to be active based on the spacing. In addition, the initiation codon of *rps19* is GTG whose translational efficiency is roughly one tenth that of ATG (our unpublished result). The existence of an additional and unique mechanism(s) is hypothesized to produce a compatible level of ribosomal proteins in chloroplasts.

Among the 79 protein-coding genes (including 9 *ycfs*) in tobacco chloroplasts, 30 contain no SD-like sequences within 20 nt upstream from the start codon, and the remaining 49 have SD-like sequences but not in a conserved position. Mutagenesis and in vitro translation assays revealed that three SD-like sequences, GGA at positions −14 to −12 in *rps14* (43), GGAG at −18 to −15 in *atpE*, and GGAGG at −10 to −6 in *rbcL* (our unpublished observation) are required for translational initiation. Assuming that SD-like sequences present between −18 and −6 are functional in tobacco chloroplasts, 38 of the 49 genes are likely to have functional SD-like sequences. The distribution of SD-like sequences is similar to that from other plant species (6, 12, 28, 94, 101). Similarity is also observed in *Synechococcus* 6301 (see Figure 3), again supporting the cyanobacterial origin of chloroplasts. Clearly, the mechanism of translational initiation in chloroplasts is not always similar to that found in *E. coli*.

The establishment of chloroplast transformation for *Chlamydomonas reinhardtii* has allowed examination in vivo of the effect of site-directed mutations introduced into the chloroplast genome. Deletion of an SD-like sequence from the *psbA* mRNA (GGAG, −31 to −28) abolished translation (74), whereas mutation of that from the *petD* mRNA (GGA, −12 to −10) did not affect translation (95). Replacement mutagenesis of the SD-like sequences had no effect on the expression of the *atpB* mRNA (GGAGG −115 to −111), *atpE* mRNA (GAAG, −22 to −19), *rps4* mRNA (GAAG, −149 to −146), and *rps7* mRNA (GGA, −116 to −114) (24). Furthermore, insertion of GGAGG at positions −9 to −5 in the four mRNAs did not enhance expression of these mRNAs. Based on the observations, a unique translational initiation system was suggested to be present in *C. reinhardtii*. However, all SD-like sequences examined seem to be too far from start codons, because the location of SD-sequences (7 ± 2 nt) is critical for ribosome-binding in *E. coli* (76).

In *Euglena*, replacement mutations of the SD-like sequence from the *atpH* mRNA (GGAGUU, −10 to −5) showed little effect on the formation of translational initiation complexes in vitro (6). Similarly, introduction of GGGAG at positions −11 to −7 into the *rbcL* mRNA resulted in only a modest increase in the complex formation (56). This situation is similar to that observed in *Chlamydomonas*.

To study the translation start site selection in chloroplasts, ribosome-binding sites on mRNAs were investigated using primer extension inhibition analysis (toeprinting). The barley *rbcL* mRNA has a GGAGG at positions −10 to −6 and is associated with ribosomes at the translation initiation site (54). This result suggests that the SD sequence is functional in chloroplast translation. Though the *psbA* mRNA lacks an SD-like sequence within −12 region, toeprint analysis revealed that ribosomes bind to its initiation region, suggesting the existence of an additional mechanism to select correct initiation sites (54).

Using an in vitro translation system from tobacco chloroplasts, *cis*-elements have been identified that are essential for tobacco *psbA* mRNA translation (44). Two are complementary to the 3′ terminus of chloroplast 16S rRNA and probably necessary for the association of the 30S ribosomal subunit (AAG at −11 to −9 termed as RBS1 and UGAU at −25 to −22 as RBS2). The other is an AU-rich sequence (UAAAUAAA at −21 to −14) located between RBS1 and RBS2; this is another critical element for translation, termed the AU-box. In vitro mRNA competition experiments revealed the existence of a *trans*-acting translation factor(s), possibly interacting with the AU-box (44). As an important development, this in vitro system holds promise for elucidating *cis*/*trans*-elements involved in the translation of chloroplast mRNAs. An SD-like sequence (GGAG) located at −36 to −33 was found not to affect translation, due probably to its location far upstream (44). Unique *cis*-elements and/or *trans*-acting factors have been sought for mRNAs without SD-like sequences in their 5′-UTRs. The 5′-UTR of tobacco *atpB* mRNA is rich in U residues, and in vitro translation analysis indicated that the U-rich sequence is essential for translation, possibly together with a *trans*-acting factor(s) (our unpublished result). Based on these in vitro studies, tobacco chloroplasts have at least three different mechanisms for translational initiation.

Evidence for the involvement of nuclear-encoded *trans*-factors in translation come mostly from genetic and molecular analyses of nuclear mutants of *Chlamydomonas* (20, 21, 28, 35, 53, 75, 90, 91, 93, 99, 123, 127, 128), whereas protein factors that interact with the 5′-UTR of several chloroplast mRNAs have also been reported from higher plants (1, 44, 55). Many of these factors seem to be mRNA-specific, suggesting the presence of unexpectedly larger numbers of regulatory proteins unique to chloroplast translation.

Initiation Codons

A critical phase of translation initiation is the selection of the correct initiation codon. The initiation codon of chloroplast mRNAs from land plants is AUG and rarely GUG. UUG has been assigned as the potential initiation codon for the IF-1 gene from *Chlorella* chloroplasts (116), but there is no experimental

evidence to confirm whether this gene is expressed. Among the 79 protein-coding genes in tobacco chloroplasts, only two mRNAs have GUG as initiation codons (*rps19, psbC*). The start codon of two mRNAs (*psbK, ycf15*) has not been defined (either AUG or GUG), and the remaining mRNAs possess AUG codons in which two are created from ACG codons by RNA editing (*psbL, ndhD*) (see below).

To study translational initiation in vivo, initiation codons were mutated by chloroplast transformation in *Chlamydomonas*. Change of the *petD* initiation codon from AUG to either AUC or AUU caused temperature-sensitive growth, probably because of inefficient translation of *petD* mRNA (17). Then, the start codon AUG of *petA* was altered to AUU, ACG, ACC, ACU, and UUC. The translation product (cytochrome f) accumulated to detectable levels in the four mutants but not in the one with UUC. Only one mutant with an AUU codon grew well. These results suggest that an AUG codon is not required to specify the site of initiation but that the efficiency of initiation depends on the identity of the initiation codon in *Chlamydomonas* (18).

Phenotypic changes and accumulation of translation products can be detected by the in vivo analysis described above. In contrast, in vitro translation assays allow us to define the initiation site by measuring the size of translation products, and also to eliminate possible effects of coding regions and 3′-UTRs on translational initiation by replacing them with a foreign sequence (e.g. *lacZ* mRNA) (44). Using the tobacco *psbA* mRNA, changing the initiation codon AUG to ACG abolished translation in vitro (44). The replacement of AUG by GUG decreased translation to about 10% of the wild type in several tobacco mRNAs (our unpublished data).

There are often multiple possible initiation codons in protein-coding genes and it is not easy to determine which is the real start codon. Determining the N termini of translation products is straightforward, albeit laborious. In land plants, RNA editing creates AUG from ACG in *rpl2, psbL, ndhD, petL,* and *accD* mRNAs (9, 46, 59, 62, 78, 79, 115). The *ndhD* mRNA from tobacco contains potential start codons AUG and GUG, but the downstream ACG is edited to produce AUG (78). In vitro translation assays clearly indicated that only the edited AUG acts as the initiation codon (45). These observations indicate that chloroplast protein structures or exact protein-coding regions cannot always be predicted from their genomic sequences. This proviso should be kept in mind when comparing gene sequences for evolutionary studies.

PROBLEMS AND PROSPECTS

Entire sequences of chloroplast genomes have been determined from a wide variety of plants and of a cyanobacterial genome, providing an enormous

amount of evolutionary information. Close comparison not only of gene sequences but also of gene arrangements gives strong support to the endosymbiotic theory of chloroplast origin.

Eighty-two different genes were described when the complete nucleotide sequence of tobacco chloroplast genome was reported in 1986. During the past decade, 23 additional genes have been identified although a comparable number of significant ORFs remains to be identified. This indicates that the identification of novel genes is an extremely difficult task, requiring creative experimental approaches. In addition, several long regions still have not been assigned any genes nor any significant ORFs. The *sprA* gene encoding a small RNA was found in such a region and further small RNA species might be encoded there. Thus final conclusions cannot be drawn, even about the full coding potential of the chloroplast genome.

The process of chloroplast gene expression was initially thought to be similar to that of *E. coli*, owing to the strong resemblance of certain chloroplast components involved in transcription and translation to those of *E. coli*. However, only sequences homologous to the *E. coli* genes encoding components of its RNA polymerase and translational machinery were identified as the chloroplast counterparts.

Recent studies have shown that chloroplast gene expression is much more complex than previously thought; there are multiple classes of promoters and of RNA polymerases, multiple RNA processing steps (RNA cleavage/trimming, *cis*/*trans*-splicing, RNA editing, and RNA stability), and multiple mechanisms for translational initiation. Corresponding to these findings, chloroplast-specific components/factors are being identified by biochemical methods. Results from genetic studies using *Chlamydomonas* suggest that from five to ten specific nuclear-encoded factors are required for the expression of specific chloroplast genes.

Why chloroplasts have such complex processes remains to be determined. It has been postulated that these posttranscriptional steps are remnants of the old RNA world for transferring proper genetic messages into proteins, and that present-day nuclear genomes acquired accurate genetic information in their DNA sequences, co-linear with protein sequences, over the course of evolution. This also seems to be the case for transcription and translation of present-day prokaryotes, which have adopted and polished single systems from several trial pieces. Thus the chloroplast genetic system may offer a window on evolutionary antiquity.

ACKNOWLEDGMENTS

We thank T Wakasugi for discussions on this chapter and apologize to those whose work was not directly cited owing to space limitations.

Literature Cited

1. Alexander C, Faber N, Klaff P. 1998. Characterization of protein-binding to the spinach chloroplast *psbA* mRNA 5′ untranslated region. *Nucleic Acids Res.* 26:2265–72
2. Bancroft I, Wolk CP, Oren EV. 1989. Physical and genetic maps of the genome of the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* 171:5940–48
3. Barkan A. 1988. Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic mRNAs. *EMBO J.* 7: 2637–44
4. Barkan A, Walker M, Nolasco M, Johnson D. 1994. A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. *EMBO J.* 13:3170–81
5. Bergmann P, Seyer P, Burkard G, Weil JH. 1984. Mapping of transfer RNA genes on tobacco chloroplast DNA. *Plant Mol. Biol.* 3:29–36
6. Betts L, Spremulli LL. 1994. Analysis of the role of the Shine-Dalgarno sequence and mRNA secondary structure on the efficiency of translational initiation in the *Euglena gracilis* chloroplast *atpH* mRNA. *J. Biol. Chem.* 269:26456–63
7. Blattner FR, Plunkett G III, Bloch CA, Perna NT, Burland V, et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–62
8. Bock R, Albertazzi F, Freyer R, Fuchs M, Ruf S, et al. 1997. Transcript editing in chloroplasts of higher plants. See Ref. 95a, pp. 123–37
9. Bock R, Hagemann R, Kössel H, Kudla J. 1993. Tissue- and stage-specific modulation of RNA editing of the *psbF* and *psbL* transcript from spinach plastids—a new regulatory mechanism. *Mol. Gen. Genet.* 240:238–44
10. Bock R, Hermann M, Kössel H. 1996. In vivo dissection of *cis*-acting determinants for plastid RNA editing. *EMBO J.* 15:5052–59
11. Bock R, Kössel H, Maliga P. 1994. Introduction of a heterologous editing site into the tobacco plastid genome: the lack of RNA editing leads to a mutant phenotype. *EMBO J.* 13:4623–28
12. Bonham-Smith PC, Bourque DP. 1989. Translation of chloroplast-encoded mRNA: potential initiation and termination signals. *Nucleic Acids Res.* 17:2057–80
13. Bonnard G, Michel F, Weil JH, Steinmetz A. 1984. 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.* 194:330–36
14. Capel MS, Bourque DP. 1982. Characterization of *Nicotiana tabacum* chloroplast and cytoplasmic ribosomal proteins. *J. Biol. Chem.* 257:7746–55
15. Chaudhuri S, Carrer H, Maliga P. 1995. Site-specific factor involved in the editing of the *psbL* mRNA in tobacco plastids. *EMBO J.* 14:2951–57
16. Chaudhuri S, Maliga P. 1997. New insights into plastid RNA editing. *Trends Plant Sci.* 2:5–6
17. Chen X, Kindle K, Stern D. 1993. Initiation codon mutations in the *Chlamydomonas* chloroplast *petD* gene result in temperature-sensitive photosynthetic growth. *EMBO J.* 12:3627–35
18. Chen X, Kindle KL, Stern DB. 1995. The initiation codon determines the efficiency but not the site of translation initiation in *Chlamydomonas* chloroplasts. *Plant Cell* 7:1295–305
19. Chen X, Widger WR. 1993. Physical genome map of the unicellular cyanobacterium *Synechococcus* sp. strain PCC 7002. *J. Bacteriol.* 175:5106–16
20. Danon A. 1997. Translational regulation in the chloroplast. *Plant Physiol.* 115: 1293–98
21. Danon A, Mayfield SP. 1991. Light regulated translational activators:identification of chloroplast gene specific mRNA binding proteins. *EMBO J.* 10:3993–4001
22. Dyer TA, Bowman CM. 1979. Nucleotide sequences of chloroplast 5S ribosomal ribonucleic acid in flowering plants. *Biochem. J.* 183:595–604
23. Edwards K, Bedbrook J, Dyer T, Kössel H. 1981. 4.5S rRNA from *Zea mays*

chloroplasts shows structural homology with the 3′end of prokaryotic 23S rRNA. *Biochem. Int.* 2:533–38
24. Fargo DC, Zhang M, Gillham NW, Boynton JE. 1998. Shine-Dalgarno-like sequences are not required for translation of chloroplast mRNAs in *Chlamydomonas reinhardtii* chloroplasts or in *Escherichia coli*. *Mol. Gen. Genet.* 257:271–82
25. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496–512
26. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, et al. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* 270:397–403
27. Freyer R, Kiefer-Meyer MC, Kössel H. 1997. Occurrence of plastid RNA editing in all major lineages of land plants. *Proc. Natl. Acad. Sci. USA* 94:6285–90
28. Gillham NW, Boynton JE, Hauser CR. 1994. Translational regulation of gene expression in chloroplasts and mitochondria. *Annu. Rev. Genet.* 28:71–93
29. Gray MW. 1996. RNA editing in plant organelles: a fertile field. *Proc. Natl. Acad. Sci. USA* 93:8157–59
30. Gruissem W. 1989. Chloroplast gene expression: how plants turn their plastids on. *Cell* 56:161–70
31. Gruissem W, Tonkyn JC. 1993. Control mechanisms of plastid gene expression. *Crit. Rev. Plant Sci.* 12:19–55
32. Hagemann R. 1992. Plastid genetics in higher plants. See Ref. 35a, pp. 65–96
33. Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, et al. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* 21:3537–44
34. Harris EH, Boynton JE, Gillham NW. 1994. Chloroplast ribosomes and protein synthesis. *Microbiol. Rev.* 58:700–54
35. Hauser CR, Gillham NW, Boynton JE. 1996. Translational regulation of chloroplast genes. *J. Biol. Chem.* 271:1486–97
35a. Herrmann RG, ed. 1992. *Cell Organelles*. Wien: Springer
36. Herrmann RG. 1997. Eukaryotism, towards a new interpretation. See Ref. 95a, pp. 73–118
37. Herrmann RG, Oelmüller R, Bichler J, Schneiderbauer A, Steppuhn J, et al. 1991. The thylakoid membrane of higher plants: genes, their expression and interaction. In *Plant Molecular Biology*, ed. RG Herrmann, BA Larkins, 2:411–27. New York: Plenum
38. Herrmann RG, Westhoff P, Link G. 1992. Biogenesis of plastids in higher plants. See Ref. 35a, pp. 275–349
39. Hiesel R, Combettes B, Brennicke A. 1994. Evidence for RNA editing in mitochondria of all major groups of land plants except the Bryophyta. *Proc. Natl. Acad. Sci. USA* 91:629–33
40. Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, et al. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* 24:4420–49
41. Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, et al. 1989. 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.* 217:185–94
42. Hirose T, Fan H, Suzuki JY, Wakasugi T, Tsudzuki T, et al. 1996. Occurrence of silent RNA editing in chloroplasts: its species specificity and the influence of environmental and developmental conditions. *Plant Mol. Biol.* 30:667–72
43. Hirose T, Kusumegi T, Sugiura M. 1998. Translation of tobacco chloroplast *rps14* mRNA depends on a Shine-Dalgarno-like sequence in the 5′-untranslated region but not internal RNA editing in the coding region. *FEBS Lett.* 430:257–60
44. Hirose T, Sugiura M. 1996. *Cis*-acting elements and *trans*-acting factors for accurate translation of chloroplast *psbA* mRNAs: development of an *in vitro* translation system from tobacco chloroplasts. *EMBO J.* 15:1687–95
45. Hirose T, Sugiura M. 1997. Both RNA editing and RNA cleavage are required for translation of tobacco chloroplast *ndhD* mRNA: a possible regulatory mechanism for expression of a chloroplast operon consisting of functionally unrelated genes. *EMBO J.* 16:6804–11
46. Hoch B, Maier RM, Appel K, Igloi GL, Kössel H. 1991. Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353:178–80
47. Howe CJ, Barker RF, Bowman CM, Dyer TA. 1988. Common features of three inversions in wheat chloroplast DNA. *Curr. Genet.* 13:343–49
48. Igloi GL, Kössel H. 1992. The transcriptional apparatus of chloroplasts. *Crit. Rev. Plant Sci.* 10:525–58
49. Kaneko T, Matsubayashi T, Sugita M, Sugiura M. 1996. Physical and gene maps of the unicellular cyanobacterium

Synechococcus sp. strain PCC6301 genome. *Plant Mol. Biol.* 31:193–201
50. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, et al. 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* 3:109–36
51. Kapoor S, Sugiura M. 1998. Expression and regulation of plastid genes. In *Photosynthesis: A Comprehensive Treatise*, ed. AS Raghavendra, pp. 72–86. Cambridge: Cambridge Univ. Press
52. Karcher D, Bock R. 1998. Site-selective inhibition of plastid RNA editing by heat shock and antibiotics:a role for plastid translation in RNA editing. *Nucleic Acids Res.* 26:1185–90
53. Kim J, Mayfield SP. 1997. Protein disulfide isomerase as a regulator of chloroplast translational activation. *Science* 278: 1954–57
54. Kim J, Mullet JE. 1994. Ribosome-binding sites on chloroplast *rbcL* and *psbA* mRNAs and light-induced initiation of D1 translation. *Plant Mol. Biol.* 25:437–48
55. Klaff P, Gruissem W. 1995. A 43 kD light-regulated chloroplast RNA-binding protein interacts with the *psbA* 5′ nontranslated leader RNA. *Photosynth. Res.* 46:235–48
56. Koo JS, Spremulli LL. 1994. Analysis of the translational initiation region on the *Euglena gracilis* chloroplast ribulose-bisphosphate carboxylase/oxygenase (*rbcL*) messenger RNA. *J. Biol. Chem.* 269: 7494–500
57. Kowallik KV, Stoebe B, Schaffran I, Kroth-Pancic P, Freier U. 1995. The chloroplast genome of a chlorophyll a+c-containing alga, *Odontella sinensis*. *Plant Mol. Biol. Rep.* 13:336–42
58. Kudla J, Hayes R, Gruissem W. 1996. Polyadenylation accelerates degradation of chloroplast mRNA. *EMBO J.* 15:7137–46
59. Kudla J, Igloi GL, Metzlaff M, Hagemann R, Kössel H. 1992. RNA editing in tobacco chloroplasts leads to the formation of a translatable *psbL* mRNA by a C to U substitution within the itnitiation codon. *EMBO J.* 11:1099–103
60. Kuhsel MG, Strickland R, Palmer JD. 1990. An ancient group I intron shared by eubacteria and chloroplasts. *Science* 250:1570–73
61. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390:249–56
62. Kuntz M, Camara B, Weil JH, Schantz R. 1992. The *psbL* gene from bell pepper (*Capsicum annuum*): plastid RNA editing also occurs in non-photosynthetic chromoplasts. *Plant Mol. Biol.* 20:1185–88
63. Leible MB, Berger S, Schweiger HG. 1989. The plastome of *Acetabularia mediterranea* and *Batophora oersteddii*: inter- and intraspecific variability and physical properties. *Curr. Genet.* 15:355–61
64. Lisitsky I, Klaff P, Schuster G. 1996. Addition of destabilizing poly(A)-rich sequences to endonuclease cleavage sites during the degradation of chloroplast mRNA. *Proc. Natl. Acad. Sci. USA* 93: 13398–403
65. Lisitsky I, Klaff P, Schuster G. 1997. Blocking polyadenylation of mRNA in the chloroplast inhibits its degradation. *Plant J.* 12:1173–78
66. Löffelhardt W, Bohnert HJ, Bryant DA. 1997. The cyanelles of *Cyanophora paradoxa*. *Crit. Rev. Plant Sci.* 16:393–413
67. Mache R. 1990. Chloroplast ribosomal proteins and their genes. *Plant Sci.* 72:1–12
68. Maier RM, Hoch B, Zeltz P, Kössel H. 1992. Internal editing of the maize chloroplast *ndhA* transcript restores codons for conserved amino acids. *Plant Cell* 4:609–16
69. Maier RM, Neckermann K, Igloi GL, Kössel H. 1995. Complete sequence of the maize chloroplast genome:gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.* 251:614–628
70. Maier RM, Zeltz P, Kössel H, Bonnard G, Gualberto JM, et al. 1996. RNA editing in plant mitochondria and chloroplasts. *Plant Mol. Biol.* 32:343–65
71. Manhart JR. Kelly K, Dudock BS, Palmer JD. 1989. Unusual characteristics of *Codium fragile* chloroplast DNA revealed by physical and gene mapping. *Mol. Gen. Genet.* 216:417–21
72. Margulis L. 1970. *Origin of Eukaryotic Cells*. New Haven: Yale Univ. Press
73. Martin W, Stöbe B, Goremykin V, Hansmann S, Hasegawa M, et al. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393:162–165
74. Mayfield SP, Cohen A, Danon A, Yohn CB. 1994. Translation of the *psbA* mRNA of *Chlamydomonas reinhardtii* requires a structured RNA element contained within

the 5′ untranslated region. *J. Cell Biol.* 127:1537–45
75. Mayfield SP, Yohn CB, Cohen A, Danon A. 1995. Regulation of chloroplast gene expression. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 46:147–66
76. McCarthy JEG, Brimacombe R. 1994. Prokaryotic translation: the interactive pathway leading to initiation. *Trends Genet.* 10:402–7
77. Mullet JE. 1993. Dynamic regulation of chloroplast transcription. *Plant Physiol.* 103:309–13
78. Neckermann K, Zeltz P, Igloi GL, Kössel H, Maier RM. 1994. The role of RNA editing in conservation of start codons in chloroplast genomes. *Gene* 146:177–82
79. Nishikawa A, Wakasugi T, Shimada H, Yamada K. 1996. Both an initiation codon and a termination codon are created by RNA editing in *accD* transcripts of the fern (*Pteridium aquilium*) chloroplasts. *Plant Cell Physiol.* 37(Suppl.):s88
80. Ohta N, Sato N, Kuroiwa T. 1998. Analyses of the plastid genome of protoflorideophyceous algae *Cyanidioschzon merolae*. *Plant Cell Physiol.* 39(Suppl.):s54
81. Ohta N, Sato N, Nozaki H, Kuroiwa T. 1997. Analysis of the cluster of ribosomal protein genes in the plastid genome of a unicellular red alga *Cyanidioschyzon merolae*: translocation of the *str* cluster as an early event in the rhodophyte-chromophyte lineage of plastid evolution. *J. Mol. Evol.* 45:688–95
82. Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, et al. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322:572–74
83. Palmer JD. 1990. Contrasting modes and tempos of genome evolution in land plant organelles. *Trends Genet.* 6:115–20
84. Palmer JD. 1991. Plastid chromosomes: structure and evolution. In *The Molecular Biology of Plastids*, ed. L Bogorad, IK Vasil, pp. 5–53. San Diego: Academic
85. Palmer JD, Nugent JM, Herbon LA. 1987. Unusual structure of geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *Proc. Natl. Acad. Sci. USA* 84:769–73
86. Pfitzinger H, Weil JH, Pillay DTN, Guillemaut P. 1990. Codon recognition mechanisms in plant chloroplasts. *Plant Mol. Biol.* 14:805–14
87. Reinbothe S, Reinbothe C, Heintzen C, Seidenbecher C, Parthier B. 1993. A methyl jasmonate-induced shift in the length of the 5′untranslated region impairs translation of the plastid *rbcL* transcript in barley. *EMBO J.* 12:1505–12
88. Reith M, Munholland J. 1993. A high-resolution gene map of the chloroplast genome of the red alga *Porphyra purpurea*. *Plant Cell* 5:465–75
89. Reith M, Munholland J. 1995. Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol. Biol. Rep.* 13:333–35
90. Rochaix JD. 1992. Post-transcriptional steps in the expression of chloroplast genes. *Annu. Rev. Cell Biol.* 8:1–28
91. Rochaix JD. 1996. Post-transcriptional regulation of chloroplast gene expression in *Chlamydomonas reinhardtii*. *Plant Mol. Biol.* 32:327–41
92. Rochaix JD, Darlix JL. 1982. Composite structure of the chloroplast 23S ribosomal RNA genes of *Chlamydomonas reinhardtii*: evolutionary and functional implications. *J. Mol. Biol.* 159:383–95
93. Rochaix JD, Kuchka M, Mayfield S, Schirmer-Rahire M, Girard-Bascou J, et al. 1989. Nuclear and chloroplast mutations affect the synthesis or stability of the chloroplast *psbC* gene product in *Chlamydomonas reinhardtii*. *EMBO J.* 8:1013–21
94. Ruf M, Kössel H. 1988. Occurrence and spacing of ribosome recognition sites in mRNAs of chloroplasts from higher plants. *FEBS Lett.* 240:41–44
95. Sakamoto W, Chen X, Kindle KL, Stern DB. 1994. Function of the *Chlamydomonas reinhardtii petD* 5′ untranslated region in regulating the accumulation of subunit IV of the cytochrome b_6/f complex. *Plant J.* 6:503–12
95a. Schenk HEA, Herrmann RG, Jeon KW, Müller NE, Schwemmler W, eds. 1997. *Eukaryotism* and *Symbiosis*. Berlin: Springer
96. Shimada H, Sugiura M. 1989. Pseudogenes and short repeated sequences in the rice chloroplast genome. *Curr. Genet.* 16:293–301
97. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, et al. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* 5:2043–49
98. Smith HC, Gott JM, Hanson MR. 1997. A guide to RNA editing. *RNA* 3:1105–23
99. Stampacchia O, Girard-Bascou J, Zanasco JL, Zerges W, Bennoun P, et al. 1997. A nuclear-encoded function essential for translation of the chloroplast *psaB* mRNA in *Chlamydomonas*. *Plant Cell* 9:773–82
100. Stanier RY, Kunisawa R, Mandel M, Cohen-Bazire G. 1971. Purification and

properties of unicellular blue-green algae (order *Chroococcales*). *Bacteriol. Rev.* 35:171–205
101. Stern DB, Higgs DC, Yang J. 1997. Transcription and translation in chloroplasts. *Trends Plant Sci.* 2:308–315
102. Stirewalt VL, Michalowski CB, Löffelhardt W, Bohnert HJ, Bryant DA. 1995. Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Mol. Biol. Rep.* 13:327–32
103. Subramanian AR. 1993. Molecular genetics of chloroplast ribosomal proteins. *Trends Biochem. Sci.* 18:177–81
104. Sugita M, Luo L, Ohta M, Itadani H, Matsubayashi T, et al. 1995. Genes encoding the group I intron-containing $tRNA^{Leu}$ and subunit L of NADH dehydrogenase from the cyanobacterium *Synechococcus* PCC 6301. *DNA Res.* 2:71–76
105. Sugita M, Sugishita H, Fujishiro T, Tsuboi M, Sugita C, et al. 1997. Organization of a large gene cluster encoding ribosomal proteins in the cyanobacterium *Synechococcus* sp. strain PCC 6301: comparison of gene clusters among cyanobacteria, eubacteria and chloroplast genomes. *Gene* 195:73–79
106. Sugita M, Sugiura M. 1996. Regulation of gene expression in chloroplasts of higher plants. *Plant Mol. Biol.* 32:315–26
107. Sugiura M. 1992. The chloroplast genome. *Plant Mol. Biol.* 19:49–68
108. Sugiura M. 1996. Structure and replication of chloroplast DNA. In *Frontiers in Molecular Biology: Molecular Genetics of Photosynthesis*, ed. B Andersson, AH Salter, J Barber, pp. 58–74. Oxford: IRL
109. Sugiura M, Torazawa K, Wakasugi T. 1991. Chloroplast genes coding for ribosomal proteins in land plants. In *The Translational Apparatus of Photosynthetic Organelles*, ed. R Mache, E Stutz, AR Subramanian, pp. 59–69. Berlin: Springer
110. Sugiura M, Wakasugi T. 1989. Compilation and comparison of transfer RNA genes from tobacco chloroplasts. *Crit. Rev. Plant Sci.* 8:89–101
111. Tomioka N, Sugiura M. 1983. The complete nucleotide sequence of a l6S ribosomal RNA gene from a blue-green alga, *Anacystis nidulans*. *Mol. Gen. Genet.* l9l:46–50
112. Tsudzuki J, Nakashima K, Tsudzuki T, Hiratsuka J, Shibata M, et al. 1992. Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ*, *trnK*, *psbA*, *trnI* and *trnH* and the absence of *rps16*. *Mol. Gen. Genet.* 232:206–14
113. Turmel M, Bellemare G, Lemieux C. 1987. Physical mapping of differences between the chloroplast DNAs of the interfertile algae *Chlamydomonas eugametos* and *Chlamydomonas moewusii*. *Curr. Genet.* 11:543–52
114. Turmel M, Boulanger J, Schnare MN, Gray MW, Lemieux C. 1991. 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
115. Wakasugi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, et al. 1996. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: The pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. *Proc. Natl. Acad. Sci. USA* 93:8766–70
116. Wakasugi T, Nagai T, Kapoor M, Sugita M, Ito M, et al. 1997. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division. *Proc. Natl. Acad. Sci. USA* 94:5967–72
117. Wakasugi T, Nishikawa A, Yamada K, Sugiura M. 1998. Complete nucleotide sequence of the plastid genome from a fern, *Psilotum nudum*. *Endocytobiosis Cell Res.* 13(Suppl.):147
118. Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, et al. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc. Natl. Acad. Sci. USA* 91:9794–98
119. Westhoff P, Herrmann RG. 1988. Complex RNA maturation in chloroplasts:The *psbB* operon from spinach. *Eur. J. Biochem.* 171:551–64
120. Wolfe KH, Morden CW, Palmer JD. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc. Natl. Acad. Sci. USA* 89:10648–52
121. Xu MQ, Kathe SD, Goodrich-Blair H, Nierzwicki-Bauer SA, Shub DA. 1990. Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. *Science* 250:1566–70
122. Yao WB, Meng BY, Tanaka M, Sugiura M. 1989. An additional promoter within the protein-coding region of the *psbD-psbC* gene cluster in tobacco chloroplast DNA. *Nucleic Acids Res.* 17:9583–91
123. Yohn CB, Cohen A, Danon A, Mayfield SP. 1996. Altered mRNA binding activity

and decreased translation initiation in a nuclear mutant lacking translation of the chloroplast *psbA* mRNA. *Mol. Cell. Biol.* 16:3560–66

124. Yoshinaga K, Iinuma H, Masuzawa T, Ueda K. 1996. Extensive RNA editing of U to C in addition to C to U substitution in the *rbcL* transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. *Nucleic Acids Res.* 24:1008–14
125. Yoshinaga K, Kakehi T, Shima Y, Iinuma H, Masuzawa T, et al. 1997. Extensive RNA editing and possible double-stranded structures determining editing sites in the *atpB* transcripts of hornwort chloroplasts. *Nucleic Acids Res.* 25:4830–34
126. Zaug AJ, McEvoy MM, Cech TR. 1993. Self-splicing of the group I intron from *Anabaena* pre-tRNA: requirement for base-pairing of the exons in the anticodon stem. *Biochem.* 32:7946–53
127. Zerges W, Girard-Bascou J, Rochaix JD. 1997. Translation of the chloroplast *psbC* mRNA is controlled by interactions between its 5′ leader and the nuclear loci TBC1 and TBC3 in *Chlamydomonas reinhardtii*. *Mol. Cell. Biol.* 17:3440–48
128. Zerges W, Rochaix JD. 1994. The 5′ leader of a chloroplast mRNA mediates the translational requirements for two nucleus-encoded functions in *Chlamydomonas reinhardtii*. *Mol. Cell. Biol.* 14:5268–77