



OrganelleIntel

Distribution and Nomenclature of Protein-coding Genes in 12 Sequenced Chloroplast Genomes

BETTINA STOEBE^{1,*}, WILLIAM MARTIN² and
KLAUS V. KOWALLIK¹

¹Botanisches Institut, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, D-40225 Düsseldorf, Germany; ²Institut für Genetik, Technische Universität Braunschweig, Spielmannstr. 7, D-38023 Braunschweig, Germany

Abbreviations: bind., binding; Chl, *Chlorella vulgaris*; Cpa, *Cyanophora paradoxa*; Epi, *Epifagus virginiana*; Eug, *Euglena gracilis*; hom., homologue; Mar, *Marchantia polymorpha*; Nic, *Nicotiana tabacum*; Odo, *Odontella sinensis*; Ory, *Oryza sativa*; Pin, *Pinus thunbergii*; Pla, *Plasmodium falciparum*; Por, *Porphyra purpurea*; prot., protein; RT, reverse transcriptase; sim., similar to; SU, subunit; Syn, *Synechocystis* sp. PCC6803; Zea, *Zea mays*

A number of chloroplast genomes have been sequenced and many more are in the pipeline. We have investigated the information contained in completely sequenced chloroplast genomes for the purpose of studying plant evolution and the process of chloroplast genome migration to the nucleus (Martin et al., 1998). In the course of our work, we have noticed that a very practical limitation exists with regard to using sequenced chloroplast genomes for phylogenetic (and related) purposes: one has to find out which proteins are encoded on which genomes. This is much more tedious than it might seem at first sight.

We have tabulated all of the protein-coding genes in sequenced chloroplast genomes that we could identify using BLAST searches which are (i) encoded in at least two sequenced chloroplast genomes or (ii) encoded in at least one sequenced chloroplast genome and in the genome of *Synechocystis* sp. PCC6803 (Kaneko et al., 1996). We have updated the nomenclature set down by Reardon and Price (1995) by expanding the list of hypothetical chloroplast open reading frames (*ycfs*) and assigning (tentative) gene names to *ycfs* when similarity to prokaryotic sequences of known function seemed sufficiently robust to justify it.

In the course of BLAST searches, we identified a number of nuclear-encoded homologues of chloroplast-encoded proteins (Martin et al., 1998).

* Author for correspondence. e-mail: stoebe@uni-duesseldorf.de

We did not search EST databases for this purpose, due to a large number of apparently chloroplast-derived transcripts in the EST databases, a circumstance that made meaningful evaluation of the BLAST results currently impossible.

Table 1 summarizes the protein-coding gene content in the completely sequenced chloroplast genomes of *Zea mays* (X86563; Maier et al., 1995), *Oryza sativa* (X15901; Hiratsuka et al., 1989), *Nicotiana tabacum* (S54304; Shinozaki et al., 1986; Z00044; Wakasugi et al., 1998), *Pinus thunbergii* (D17510; Wakasugi et al., 1994), *Marchantia polymorpha* (X04465; Ohyama et al., 1986), *Euglena gracilis* (Z11874; Hallick et al., 1993), *Porphyra purpurea* (U38804; Reith and Munholland, 1995), *Odontella sinensis* (Z67753; Kowallik et al., 1995), *Chlorella vulgaris* (AB001684; Wakasugi et al., 1997), *Epifagus virginiana* (M81884; Wolfe et al., 1992), *Plasmodium falciparum* (X95275 and X95276; Wilson et al., 1996) and *Cyanophora paradoxa* (U30821; Stirewalt et al., 1995). The putative or known functions of the encoded products are mostly taken from the designations in Kaneko et al. (1996), in a few cases we have deviated slightly from those definitions. The Table is largely self-explanatory.

We would be grateful for suggestions or corrections that should be introduced in the next update of the Table.

Table 1. Protein-coding genes in sequenced chloroplast genomes

	Z	O	N	P	M	E	C	O	P	C	P	E	S	
	e	r	i	i	a	u	h	d	o	p	l	p	y	
	a	y	c	n	r	g	l	o	r	a	a	i	n	Probable or known function, Comments
<i>accA</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Acetyl-CoA carboxylase α SU
<i>accB</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Acetyl-CoA carboxylase biotin carrier SU
<i>accD</i>	-	-	+	+	+	-	+	-	+	-	-	+	+	Acetyl-CoA carboxylase carboxytransferase β SU (<i>ycf11</i> , <i>zfpA</i>)
<i>acpP</i>	-	-	-	-	-	-	-	-	-	+	+	+	-	+ Acyl carrier protein
<i>apcA</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Allophycocyanin α SU
<i>apcB</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Allophycocyanin β SU
<i>apcD</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Allophycocyanin γ SU
<i>apcE</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Phycobilisome core linker prot.
<i>apcF</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Allophycocyanin 18 kDa β SU
<i>argB</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Acetylglutamate kinase
<i>atpA</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	ATP synthase CF ₁ α SU
<i>atpB</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	ATP synthase CF ₁ β SU
<i>atpD</i>	-	-	-	-	-	-	-	-	-	+	+	+	-	+ ATP synthase CF ₁ δ SU
<i>atpE</i>	+	+	+	+	+	+	+	+	+	+	+	-	-	+ ATP synthase CF ₁ ϵ SU
<i>atpF</i>	+	+	+	+	+	+	+	+	+	+	+	-	-	+ ATP synthase CF ₀ SU I
<i>atpG</i>	-	-	-	-	-	-	-	-	-	+	+	+	-	+ ATP synthase CF ₀ SU II
<i>atpH</i>	+	+	+	+	+	+	+	+	+	+	+	-	-	+ ATP synthase CF ₀ SU III
<i>atpI</i>	+	+	+	+	+	+	+	+	+	+	+	-	-	+ ATP synthase CF ₀ SU IV
<i>bas1</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Thiol-specific antioxidant prot. (<i>ycf42</i>)
<i>bioY</i>	-	-	-	-	-	-	-	-	-	+	-	-	+	<i>Cpa</i> ORF188, <i>Syn</i> slr1365 biotin synthase
<i>carA</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Carbamoyl phosphate synthetase small SU
<i>ccbX</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	Orthologues in red type Calvin cycle operons of proteobacteria (<i>cfxQ</i> , <i>cbxX</i>) ¹
<i>ccsA</i>	+	+	+	+	+	-	+	+	+	+	+	-	-	+ Heme attachment to plastid cyt <i>c</i> (<i>ycf5</i>)
<i>cemA</i>	+	+	+	+	+	-	+	-	+	-	-	-	+	Envelope membrane prot. (<i>ycf10</i> , <i>hbp</i>)
<i>chlB</i>	-	-	-	+	+	-	+	-	+	+	-	-	+	Protochlorophyllide reductase ChlB SU
<i>chlI</i>	-	-	-	-	-	+	+	+	+	+	+	-	-	+ Protochlorophyllide reductase magnesium chelatase SU
<i>chlL</i>	-	-	-	+	+	-	+	-	+	+	+	-	-	+ Protochlorophyllide reductase ATP-bind. SU (<i>frxC</i>)
<i>chlN</i>	-	-	-	+	+	-	+	-	+	+	-	-	+	Protochlorophyllide reductase ChlN SU
<i>clpC</i>	-	-	-	-	-	-	-	-	-	+	+	-	+	Clp protease ATP-bind. SU
<i>clpP</i>	+	+	+	+	+	-	+	-	-	+	-	+	+	Clp protease proteolytic SU ²
<i>cpcA</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Phycocyanin α SU
<i>cpcB</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Phycocyanin β SU

Table 1. Continued

	Z	O	N	P	M	E	C	O	P	C	P	E	S	
	e	r	i	i	a	u	h	d	o	p	l	p	y	
	a	y	c	n	r	g	l	o	r	a	a	i	n	Probable or known function, Comments
<i>cpcG</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Phycobilisome rod-core linker polypeptide
<i>cpeA</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Phycoerythrin α SU
<i>cpeB</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Phycoerythrin β SU
<i>crtE</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Geranylgeranyl pyrophosphate synthetase hom.
<i>cysA</i>	-	-	-	-	+	-	+	-	-	-	-	-	-	+ Probable transport prot. (<i>mbpX</i>)
<i>cysT</i>	-	-	-	-	+	-	+	-	-	-	-	-	-	+ Probable transport prot. (<i>mbpY</i>)
<i>dfr</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ <i>Syn</i> sl10698 drug sensory prot. A (<i>ycf26</i>)
<i>dnaB</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	+ DNA-replication helicase SU
<i>dnaK</i>	-	-	-	-	-	-	-	-	-	+	+	+	-	+ Hsp70-type chaperone
<i>dsbD</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ <i>Por</i> ORF240, <i>Syn</i> thiol:disulfide interchange prot. DsbD sl10621
<i>fabH</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ β -ketoacyl-acyl carrier prot. synthase III
<i>fdx</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ <i>Por</i> ORF75a, <i>Syn</i> ssr3184, 2[4Fe-4S] ferredoxin
<i>ftrB</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Ferredoxin-thioredoxin reductase β SU
<i>ftsH</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Cell division prot., ATPase, protease (<i>ycf25</i>)
<i>ftsW</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Putative cell/organelle division prot.
<i>glnB</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Nitrogen regulatory prot. PII
<i>gltB</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Glutamate synthase (GOGAT)
<i>groEL</i>	-	-	-	-	-	-	-	-	-	+	+	+	-	+ Chaperonin, 60 kDa
<i>groES</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Chaperonin, 10 kDa
<i>hemA</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ 5-Aminolevulinic acid synthase
<i>hisH</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Histidinol-phosphate aminotransferase
<i>I-CvuI</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	- Intron encoded DNA endonuclease, similar to <i>Chlamydomonas reinhardtii</i> <i>I-CreI</i> (P05725)
<i>ilvB</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Acetohydroxyacid synthase large SU
<i>ilvH</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Acetohydroxyacid synthase small SU
<i>infA</i>	+	+	+	+	+	-	+	-	-	-	-	-	+	+ Translational initiation factor 1
<i>infB</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Translational initiation factor 2
<i>infC</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Translational initiation factor 3
<i>matK</i>	+	+	+	+	+	-	-	-	-	-	-	-	+	- Probable intron maturase (<i>ycf14</i>)
<i>minD</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Septum-site determining prot.
<i>minE</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	- 22% identity to <i>E. coli minE</i> , no hom. in <i>Syn</i> or cpDNA, tentative designation
<i>mntA</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ <i>Cpa</i> ORF244, manganese transport system ATP-bind. prot.
<i>mntB</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+ <i>Cpa</i> ORF299, manganese transport system

Table 1. Continued

	Z	O	N	P	M	E	C	O	P	C	P	E	S	
	e	r	i	i	a	u	h	d	o	p	l	p	y	
	a	y	c	n	r	g	l	o	r	a	a	i	n	Probable or known function, Comments
														membrane prot.
<i>moeB</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	<i>Por</i> ORF382, molybdopterin biosynthesis prot., <i>Syn</i> sll1536
<i>nadA</i>	-	-	-	-	-	-	-	-	-	+	-	-	+	Quinolinate synthase
<i>nblA</i>	-	-	-	-	-	-	-	-	-	+	-	-	+	Phycobilisome degradation prot. (<i>ycf18</i>)
<i>ndhA</i>	+	+	+	-	+	-	-	-	-	-	-	-	+	NADH-plastoquinone oxidoreductase SU 1
<i>ndhB</i>	+	+	+	-	+	-	-	-	-	-	-	-	+	NADH-plastoquinone oxidoreductase SU 2
<i>ndhC</i>	+	+	+	-	+	-	-	-	-	-	-	-	+	NADH-plastoquinone oxidoreductase SU 3
<i>ndhD</i>	+	+	+	-	+	-	-	-	-	-	-	-	+	NADH-plastoquinone oxidoreductase SU 4
<i>ndhE</i>	+	+	+	-	+	-	-	-	-	-	-	-	+	NADH-plastoquinone oxidoreductase SU 4L
<i>ndhF</i>	+	+	+	-	+	-	-	-	-	-	-	-	+	NADH-plastoquinone oxidoreductase SU 5
<i>ndhG</i>	+	+	+	-	+	-	-	-	-	-	-	-	+	NADH-plastoquinone oxidoreductase SU 6
<i>ndhH</i>	+	+	+	-	+	-	-	-	-	-	-	-	+	NADH-plastoquinone oxidoreductase 49 kDa SU
<i>ndhI</i>	+	+	+	-	+	-	-	-	-	-	-	-	+	NADH-plastoquinone oxidoreductase SU I
<i>ndhJ</i>	+	+	+	-	+	-	-	-	-	-	-	-	+	NADH-plastoquinone oxidoreductase SU J
<i>ndhK</i>	+	+	+	-	+	-	-	-	-	-	-	-	+	NADH-ubiquinone oxidoreductase SU K (<i>psbG</i>)
<i>ntcA</i>	-	-	-	-	-	-	-	-	-	+	-	-	+	Global nitrogen transcriptional regulator (<i>ycf28</i>)
<i>odpA</i>	-	-	-	-	-	-	-	-	-	+	-	-	+	Pyruvate dehydrogenase E1 component, A SU
<i>odpB</i>	-	-	-	-	-	-	-	-	-	+	-	-	+	Pyruvate dehydrogenase E1 component, B SU
<i>pbsA</i>	-	-	-	-	-	-	-	-	-	+	-	-	+	Heme oxygenase
<i>petA</i>	+	+	+	+	+	-	+	+	+	+	-	-	+	Apocytochrome <i>f</i>
<i>petB</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	Cytochrome <i>b₆/f</i> apoprotein
<i>petD</i>	+	+	+	+	+	-	+	+	+	+	-	-	+	Cytochrome <i>b₆/f</i> complex SU IV
<i>petF</i>	-	-	-	-	-	-	+	+	+	-	-	-	+	Ferredoxin
<i>petG</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	Cytochrome <i>b₆/f</i> complex SU V
<i>petJ</i>	-	-	-	-	-	-	+	-	-	-	-	-	+	Cytochrome <i>c₅₅₃</i>
<i>petL</i>	+	+	+	+	+	-	+	+	+	+	-	-	(+)	<i>Syn</i> sll1845, cyt. <i>b₆/f</i> complex SU VI (<i>ycf7</i>) ³
<i>petM</i>	-	-	-	-	-	-	+	+	+	-	-	-	+	Cytochrome <i>b₆/f</i> complex SU VII (<i>ycf31</i>)
<i>pgmA</i>	-	-	-	-	-	-	+	-	-	-	-	-	+	Phosphoglycerate mutase ⁴
<i>preA</i>	-	-	-	-	-	-	+	+	-	-	-	-	+	Prenyl transferase, related to <i>crtE</i>
<i>psaA</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSI P700 apoprotein A1
<i>psaB</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSI P700 apoprotein A2
<i>psaC</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSI iron-sulphur center FA/FB containing SU VII
<i>psaD</i>	-	-	-	-	-	-	+	+	-	-	-	-	+	PSI ferredoxin-bind. prot. SU II
<i>psaE</i>	-	-	-	-	-	-	+	+	+	-	-	-	+	PSI SU IV, 18–20 kDa
<i>psaF</i>	-	-	-	-	-	-	+	+	+	-	-	-	+	PSI SU III, plastocyanin-bind.

Table 1. Continued

	Z	O	N	P	M	E	C	O	P	C	P	E	S	
	e	r	i	i	a	u	h	d	o	p	l	p	y	
	a	y	c	n	r	g	l	o	r	a	a	i	n	Probable or known function, Comments
<i>psaI</i>	+	+	+	+	+	-	+	+	+	+	-	-	+	PSI SU VIII
<i>psaJ</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSI SU IX
<i>psaK</i>	-	-	-	-	-	-	-	-	+	-	-	-	+	PSI SU X
<i>psaL</i>	-	-	-	-	-	-	-	+	+	-	-	-	+	PSI SU XI, reaction center
<i>psaM</i>	-	-	-	+	+	+	+	+	+	+	-	-	+	PSI M-polypeptide
<i>psbA</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII D1 reaction center prot.
<i>psbB</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII CP47 chlorophyll apoprotein
<i>psbC</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII CP43 chlorophyll apoprotein
<i>psbD</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII D2 reaction center prot.
<i>psbE</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII cytochrome <i>b</i> ₅₅₉ α SU
<i>psbF</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII cytochrome <i>b</i> ₅₅₉ β SU
<i>psbH</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII 10 kDa phosphoprotein
<i>psbI</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII I polypeptide
<i>psbJ</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII J protein
<i>psbK</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII K protein
<i>psbL</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII L protein
<i>psbM</i>	+	+	+	+	+	-	+	-	-	+	-	-	+	PSII M protein
<i>psbN</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII N protein
<i>psbT</i>	+	+	+	+	+	+	+	+	+	+	-	-	+	PSII T protein (<i>ycf8</i>)
<i>psbV</i>	-	-	-	-	-	-	-	+	+	+	-	-	+	PSII cytochrome <i>c</i> ₅₅₀ (<i>petK</i>)
<i>psbW</i>														Gene nomenclature problem: see <i>ycf79</i> and footnote ⁵
<i>psbX</i>	-	-	-	-	-	-	-	+	+	+	-	-	+	PSII X 4.1 kDa protein
<i>rbcLg</i>	+	+	+	+	+	+	+	-	-	+	-	-	+	Rubisco large SU, green type
<i>rbcLr</i>	-	-	-	-	-	-	-	+	+	-	-	-	(+)	Rubisco large SU, red type
<i>rbcR</i>	-	-	-	-	-	-	-	+	+	+	-	-	+	<i>Syn</i> sll0998 LysR family of transcriptional regulators (<i>ycf30</i>) ⁶
<i>rbcSg</i>	-	-	-	-	-	-	-	-	-	+	-	-	+	Rubisco small SU, green type
<i>rbcSr</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	(+) Rubisco small SU, red type
<i>rdpO</i>	-	-	-	-	-	-	-	-	-	-	-	-	(+)	<i>Eug</i> ORF506 S34497, <i>Scenedesmus obliquus</i> RT <i>rdpO</i> S05341, <i>Calothrix</i> PCC7601 S40013
<i>rne</i>	-	-	-	-	-	-	-	-	-	+	-	-	+	RNA component of RNaseE, RNA processing
<i>rpl1</i>	-	-	-	-	-	-	-	-	+	+	+	-	+	Ribosomal protein L1
<i>rpl2</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	Ribosomal protein L2
<i>rpl3</i>	-	-	-	-	-	-	-	-	+	+	+	-	+	Ribosomal protein L3
<i>rpl4</i>	-	-	-	-	-	-	-	-	+	+	-	+	+	Ribosomal protein L4
<i>rpl5</i>	-	-	-	-	-	-	-	-	+	+	+	+	+	Ribosomal protein L5

Table 1. Continued

	Z	O	N	P	M	E	C	O	P	C	P	E	S				
	e	r	i	i	a	u	h	d	o	p	l	p	y				
	a	y	c	n	r	g	l	o	r	a	a	i	n	Probable or known function, Comments			
<i>rpl6</i>	-	-	-	-	-	-	-	-	-	+	+	+	+	-	+	Ribosomal protein L6	
<i>rpl9</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	Ribosomal protein L9	
<i>rpl11</i>	-	-	-	-	-	-	-	-	-	+	+	+	-	-	+	Ribosomal protein L11	
<i>rpl12</i>	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-	+	Ribosomal protein L12 ⁷
<i>rpl13</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	-	+	Ribosomal protein L13	
<i>rpl14</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	Ribosomal protein L14
<i>rpl16</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Ribosomal protein L16
<i>rpl18</i>	-	-	-	-	-	-	-	-	-	+	+	+	-	-	+	Ribosomal protein L18	
<i>rpl19</i>	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	+	Ribosomal protein L19
<i>rpl20</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	Ribosomal protein L20
<i>rpl21</i>	-	-	-	-	+	-	-	-	-	+	+	+	-	-	+	Ribosomal protein L21	
<i>rpl22</i>	+	+	+	+	+	+	+	+	+	-	+	+	+	-	-	+	Ribosomal protein L22
<i>rpl23</i>	+	+	+	+	+	+	+	+	+	+	-	+	-	+	+	Ribosomal protein L23	
<i>rpl24</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	-	+	Ribosomal protein L24	
<i>rpl27</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	-	+	Ribosomal protein L27	
<i>rpl28</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	-	+	Ribosomal protein L28	
<i>rpl29</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	-	+	Ribosomal protein L29	
<i>rpl31</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	-	+	Ribosomal protein L31	
<i>rpl32</i>	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	Ribosomal protein L32	
<i>rpl33</i>	+	+	+	+	+	+	-	-	+	+	+	-	+	+	+	Ribosomal protein L33	
<i>rpl34</i>	-	-	-	-	-	-	-	-	-	+	+	+	-	-	+	Ribosomal protein L34	
<i>rpl35</i>	-	-	-	-	-	-	-	-	-	+	+	+	-	-	+	Ribosomal protein L35	
<i>rpl36</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Ribosomal protein L36	
<i>rpoA</i>	+	+	+	+	+	-	+	+	+	+	+	-	-	+	+	RNA polymerase α SU	
<i>rpoB</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	RNA polymerase β SU	
<i>rpoC1</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	RNA polymerase β' SU	
<i>rpoC2</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	RNA polymerase β'' SU ⁸	
<i>rps1</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	Ribosomal protein S1	
<i>rps2</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Ribosomal protein S2	
<i>rps3</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Ribosomal protein S3	
<i>rps4</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Ribosomal protein S4	
<i>rps5</i>	-	-	-	-	-	-	-	-	-	+	+	+	+	-	+	Ribosomal protein S5	
<i>rps6</i>	-	-	-	-	-	-	-	-	-	+	+	+	-	-	+	Ribosomal protein S6	
<i>rps7</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Ribosomal protein S7	
<i>rps8</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Ribosomal protein S8	
<i>rps9</i>	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	+	Ribosomal protein S9

Table 1. Continued

	Z	O	N	P	M	E	C	O	P	C	P	E	S		
	e	r	i	i	a	u	h	d	o	p	l	p	y	Probable or known function, Comments	
	a	y	c	n	r	g	l	o	r	a	a	i	n		
<i>rps10</i>	-	-	-	-	-	-	-	-	+	+	+	-	-	+	Ribosomal protein S10
<i>rps11</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Ribosomal protein S11
<i>rps12</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Ribosomal protein S12
<i>rps13</i>	-	-	-	-	-	-	-	-	+	+	+	-	-	+	Ribosomal protein S13
<i>rps14</i>	+	+	+	+	+	+	+	+	+	+	+	-	+	+	Ribosomal protein S14
<i>rps15</i>	+	+	+	+	+	-	-	-	-	-	-	-	-	+	Ribosomal protein S15
<i>rps16</i>	+	+	+	-	-	-	-	-	+	+	+	-	-	+	Ribosomal protein S16
<i>rps17</i>	-	-	-	-	-	-	-	-	+	+	+	+	-	+	Ribosomal protein S17
<i>rps18</i>	+	+	+	+	+	+	+	+	+	+	+	-	+	+	Ribosomal protein S18
<i>rps19</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Ribosomal protein S19
<i>rps20</i>	-	-	-	-	-	-	-	-	+	+	+	-	-	+	Ribosomal protein S20
<i>secA</i>	-	-	-	-	-	-	-	-	+	+	-	-	-	+	Preprotein-translocase SU A
<i>secY</i>	-	-	-	-	-	-	-	-	+	+	+	-	-	+	Preprotein-translocase SU Y
<i>syfB</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+	Phenylalanine tRNA synthetase
<i>syh</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+	Histidine tRNA synthetase
<i>thiG</i>	-	-	-	-	-	-	-	-	+	+	-	-	-	+	ThiG protein, thiamine biosynthesis
<i>trpA</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+	Tryptophane synthase α SU
<i>trpG</i>	-	-	-	-	-	-	-	-	-	+	+	-	-	+	Anthranilate synthase component II, glutamine amidotransferase
<i>trxA</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+	Thioredoxin
<i>tsf</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+	Translational elongation factor Ts
<i>tufA</i>	-	-	-	-	-	+	+	+	+	+	+	+	-	+	Translational elongation factor Tu
<i>upp</i>	-	-	-	-	-	-	-	-	-	+	-	-	-	+	<i>Por</i> ORF198, uracil phosphoribosyltransferase sll1035
<i>ycf1</i>	-	-	+	+	+	-	-	-	-	-	-	-	-	+	* see also <i>ycf77</i> ⁹
<i>ycf2</i>	-	-	+	+	+	-	-	-	-	-	-	-	-	+	* sim. FtsH-like proteins, see also <i>ycf78</i> ¹⁰
<i>ycf3</i>	+	+	+	+	+	-	+	+	+	+	+	-	-	+	<i>Syn</i> slr0823
<i>ycf4</i>	+	+	+	+	+	+	+	+	+	+	+	-	-	+	<i>Syn</i> sll0226
<i>ycf6</i>	+	+	+	+	+	-	-	+	+	+	-	-	+		<i>Syn</i> sml0004
<i>ycf9</i>	+	+	+	+	+	+	+	+	+	+	+	-	-	+	<i>Syn</i> sll1281
<i>ycf12</i>	-	-	-	+	+	+	+	+	+	+	+	-	-	+	<i>Syn</i> sll0047
<i>ycf13</i>	-	-	-	-	-	+	-	-	-	-	-	-	-	(+)	<i>Anabaena</i> U13767 intron-encoded ORF (Bauer <i>et al.</i> , 1997), in <i>Astasia</i> P14761
<i>ycf15</i>	+	-	+	-	-	-	-	-	-	-	-	-	-	-	* <i>Zea</i> ORF99
<i>ycf16</i>	-	-	-	-	-	-	-	-	+	+	+	-	-	+	<i>Syn</i> slr0075 abc-transporter

Table 1. Continued

	Z	O	N	P	M	E	C	O	P	C	P	E	S	
	e	r	i	i	a	u	h	d	o	p	l	p	y	
	a	y	c	n	r	g	l	o	r	a	a	i	n	Probable or known function, Comments
ycf17	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn ssl1633, slr0839, ssr2595, ssl2542, ssr1789 ELIP superfamily (Grimm <i>et al.</i> , 1989)
ycf19	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Syn ssr2142
ycf20	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Syn sll1509
ycf21	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn sll1797
ycf22	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Syn sll0751, sll1002
ycf23	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn slr2032
ycf24	-	-	-	-	-	-	-	-	-	+	+	+	+	- Syn slr0074 abc-transporter
ycf27	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn slr0115, slr0947 regulatory component of sensory transduction system
ycf29	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn slr1783 regulatory component of sensory transduction system
ycf32	-	-	-	-	-	-	-	-	-	+	+	+	-	+ Syn sml0007
ycf33	-	-	-	-	-	-	-	-	-	+	+	+	-	+ Syn ssl1417
ycf34	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn ssr1425
ycf35	-	-	-	-	-	-	-	-	-	+	+	+	-	+ Syn sll0661
ycf36	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn sll0584
ycf37	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn slr0171
ycf38	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn sll0760
ycf39	-	-	-	-	-	-	-	-	-	+	+	+	-	+ Syn sll1218, slr0399
ycf40	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn ssr0102
ycf41	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn slr1034
ycf43	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn sll0194
ycf44	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn slr2087
ycf45	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn slr0692
ycf46	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn slr0480, slr0374 sim. SU 4 of 26S proteasome ¹¹
ycf47	-	-	-	-	-	-	-	-	-	+	+	-	-	+ Syn ssr3307
ycf48	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Cpa ORF333, Syn slr2034
ycf49	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Cpa ORF102, Syn sll0608
ycf50	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Cpa ORF108, Syn slr2073
ycf51	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Cpa ORF163, Syn sll1702
ycf52	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Por ORF174, Syn sll0286
ycf53	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Por ORF238, Syn sll0558
ycf54	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Por ORF108, Syn slr1780
ycf55	-	-	-	-	-	-	-	-	-	+	-	-	-	+ Por ORF320, Syn sll1879

Table 1. Continued

	Z	O	N	P	M	E	C	O	P	C	P	E	S	
	e	r	i	i	a	u	h	d	o	p	l	p	y	
	a	y	c	n	r	g	l	o	r	a	a	i	n	Probable or known function, Comments
<i>yef56</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ Por ORF263a, <i>Syn</i> slr0050
<i>yef57</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ Por ORF114, <i>Syn</i> slr1417 sim. prot. in <i>nifU</i> region
<i>yef58</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ Por ORF149, <i>Synechococcus</i> M95288, <i>Syn</i> slr2049
<i>yef59</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ Por ORF349, <i>Syn</i> sll1214 leucine zipper protein
<i>yef60</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ Por ORF203, <i>Syn</i> sll1737
<i>yef61</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ Por ORF75b, <i>Syn</i> ssl2982, hom. in <i>Guillardia theta</i> cpDNA (S. Douglas, pers. commun.)
<i>yef62</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ Por ORF327, <i>Syn</i> slr1278
<i>yef63</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ Por ORF263b, <i>Syn</i> slr1045
<i>yef64</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ Por ORF107, <i>Syn</i> slr1846 ¹²
<i>yef65</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ Por ORF99, <i>Synechococcus</i> U62737, <i>Syn</i> slr0923, hom. in <i>Guillardia theta</i> cpDNA (S. Douglas, pers. commun.)
<i>yef66</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ <i>Odo</i> ORF99, <i>Syn</i> slr0503, <i>Mar</i> ORF135 A05009
<i>yef67</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ * <i>Eug</i> ORF161 P48337, <i>Astasia longa</i> ORF170 P34778
<i>yef68</i>	+	+	-	+	-	-	-	-	-	-	-	-	-	+ * ORF in <i>trnI</i> intron <i>Pin</i> ORF75a P52807, <i>Zea</i> ORF134 P03938, <i>Ory</i> ORF133b P12173
<i>yef69</i>	+	+	+	-	-	-	-	-	-	-	-	-	-	+ * <i>Nic</i> ORF131 A05210, <i>Ory</i> ORF72 JQ0277, <i>Zea</i> ORF58 S58628
<i>yef70</i>	+	+	-	-	-	-	-	-	-	-	-	-	-	+ * <i>Ory</i> ORF91 JQ0209, <i>Zea</i> ORF69 S58539
<i>yef71</i>	+	+	-	-	-	-	-	-	-	-	-	-	-	+ * <i>Ory</i> ORF82 JQ0268, <i>Zea</i> ORF75 S58641
<i>yef72</i>	+	+	-	-	-	-	-	-	-	-	-	-	-	+ * <i>Ory</i> ORF137 JQ0269, <i>Zea</i> ORF137 S58640
<i>yef73</i>	+	+	-	-	-	-	-	-	-	-	-	-	-	+ * <i>Ory</i> ORF249 JQ0274, <i>Zea</i> ORF173 S58632
<i>yef74</i>	+	+	-	-	-	-	-	-	-	-	-	-	-	+ * ORF in <i>trnA</i> intron <i>Ory</i> ORF109 JQ0280, <i>Zea</i> ORF49 S58608
<i>yef75</i>	+	+	-	-	-	-	-	-	-	-	-	-	-	+ * <i>Ory</i> ORF63a JQ0283, <i>Zea</i> ORF63 S58623
<i>yef76</i>	+	+	-	-	-	-	-	-	-	-	-	-	-	+ * <i>Ory</i> ORF85b JQ0278, <i>Zea</i> ORF85 S58627
<i>yef77</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ <i>Chl</i> ORF1720, in positions 950–1150 residual similarity to <i>yef46</i> and <i>Syn fth3</i> P73437 ⁹
<i>yef78</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ * <i>Chl</i> ORF819, hom. in <i>Chlamydomonas reinhardtii</i> cpDNA X92726 ¹⁰
<i>yef79</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ <i>Syn</i> sll1398, annotated as 13 kDa protein in PSII ⁵
<i>yef80</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+ <i>Por</i> ORF450, hom. in <i>Guillardia theta</i> cpDNA (S. Douglas, pers. commun.)

Plus and minus signs indicate presence and absence, respectively, of the gene in the given chloroplast genome. A plus sign in parentheses (+) indicates that the gene is not found in BLAST searches in the *Synechocystis* genome, but was scored for specific reasons as being cyanobacterial in origin and present in the ancestral plastid genome. Synonyms are indicated in parentheses, except in the case of ORF designations. Accession numbers refer to PIR, GenBank, or SwissProt data bases. Reading frames designated as pseudogenes in *Epifagus* (Wolfe et al., 1992) and black pine cpDNA (Wakasugi et al., 1994) were scored as absent here. The last revision of *ycf* nomenclature ended with *ycf47* (Reardon and Price, 1995).

* No similarity to prokaryotic genes was detected in BLAST searches.

¹ *cbxX* is located in proteobacterial *cbx* operons (*rbcLr-rbcSr-cbxX*), not to be confused with *rbcX* in cyanobacterial *cbx* operons (*rbcLg-rbcX-rbcSg*; Li and Tabita, 1997).

² *Cyanophora* cpDNA encodes two distinct *clpP* genes that are 36% identical, whereby *Synechocystis* also possesses two distinct *clpP* genes that are however 74% identical.

³ *petL*, a 31 amino acid long component of the cytochrome *b₆/f* complex, is too short to detect its *Synechocystis* homologue (gi1653694) among the highest BLAST scores, but it was counted as being of cyanobacterial origin.

⁴ *pgmA* has related but paralogous nuclear homologues for cytosolic phosphoglycerate mutase in many higher plants (e.g. *Nic* P35494).

⁵ There is a severe problem with *psbW* nomenclature. The gene designated in previous studies of chloroplast genomes (including Martin et al., 1998) as *psbW* shares no similarity with the nuclear-encoded gene for the psbW protein functionally characterized from spinach chloroplasts (Lorkovic et al., 1995). The spinach gene designation has priority. The chloroplast-encoded protein previously designated as PsbW is similar to *Synechocystis* sll1398 (GenBank 1651690), annotated as a 13 kDa PSII protein. Since a function for the chloroplast gene previously described as *psbW* has not been reported, we designate it here as *ycf79*.

⁶ Closely related to the Calvin cycle transcriptional regulator *cbxR* (Qian and Tabita, 1996).

⁷ Chloroplast *rpl12* has in some cases also been designated as *rpl7*.

⁸ *rpoC2* is sometimes also designated as *rpoD* in cpDNA, but the gene designated *rpoD* in *Syn* is a sigma factor, the designation *rpoC2* for the chloroplast gene is less ambiguous and hence preferable.

⁹ We found no significant similarity between cpDNA genes previously designated as *ycf1* (Reardon and Price, 1995) and the gene tentatively designated as *ycf1* in the *Chlorella* genome (Wakasugi et al., 1997), that is hence designated as *ycf77* here.

¹⁰ We found no significant similarity between cpDNA genes previously designated as *ycf2* (Reardon and Price, 1995) and the gene tentatively designated as *ycf2* in the *Chlorella* genome (Wakasugi et al., 1997), that is hence designated as *ycf78* here.

¹¹ *ycf46* has a related but paralogous nuclear homologue in *Ory* D17789 (Suzuka et al., 1994).

¹² *ycf64* has similarity to glutaredoxin-like proteins, nuclear homologues are found in yeast and *Plasmodium*, a distant homologue is nuclear in *Arabidopsis* (e327479 in Z97339).

References

- Bauer CC, Ramaswamy KS, Endley S, Scappino LA, Golden JW and Haselkorn R (1997) Suppression of heterocyst differentiation in *Anabaena* PCC7120 by a cosmid carrying wild-type genes encoding enzymes for fatty acid synthesis. *FEMS Microbiol Lett* 151: 23–30.
- Grimm B, Kruse E and Kloppstech K (1989) Transiently expressed early light-inducible thylakoid proteins share transmembrane domains with light-harvesting chlorophyll binding proteins. *Plant Mol Biol* 13: 583–593.
- Hallick RB, Hong L, Drager RG, Favreau M, Monfort A, Orsat B, Spielmann A and Stutz E (1993) Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res* 21: 3537–3544.
- Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C-R, Meng B-Y, Li Y-Q, Kanno A, Nishizawa Y, Hirai A, Shinozaki K and Sugiura M (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: 185–194.
- Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirose M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M and Tabata S (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: 109–136.
- Kowallik KV, Stoebe B, Schaffran I, Kroth-Pancic P and Freier U (1995) The chloroplast genome of a chlorophyll *a+c*-containing alga, *Odontella sinensis*. *Plant Mol Biol Repr* 13: 336–342.
- Lorkovic ZJ, Schroeder WP, Pakrasi HB, Irrgang KD, Herrmann RG and Oelmueller R (1995) Molecular characterization of PsbW, a nuclear-encoded component of the photosystem II reaction center complex in spinach. *Proc Natl Acad Sci USA* 92: 8930–8934.
- Li L-A and Tabita FR (1997) Maximum activity of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase of *Anabaena* sp. strain CA requires the product of the *rbcX* gene. *J Bacteriol* 197: 3793–3796.
- Maier RM, Neckermann K, Igloi GL and Kössel H (1995) Complete sequence of maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol* 251: 614–628.
- Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M and Kowallik KV (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393: 162–165.
- Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H and Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322: 572–574.
- Qian Y and Tabita FR (1996) A global signal transduction system regulates aerobic and anaerobic CO₂ fixation in *Rhodobacter sphaeroides*. *J Bacteriol* 178: 12–18.
- Reardon EM and Price CA (1995) Plastid genomes of three non-green algae are sequenced. *Plant Mol Biol Repr* 13: 320–326.
- Reith M and Munholland J (1995) Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol Biol Repr* 13: 333–335.
- Shinozaki K, Ohme M, Tanaka M, Wakasugi M, Hayashida M, Matsubayashi T, Zaita M, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY,

- Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H and Sugiura M (1986) The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5: 2043–2049.
- Stirewalt VL, Michalowski CB, Löffelhardt W, Bohnert HJ and Bryant DB (1995) Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Mol Biol Repr* 13: 327–332.
- Suzuka I, Koga-Ban Y, Sasaki T, Minobe Y and Hashimoto J (1994) Identification of cDNA clones for rice homologs of the human immunodeficiency virus-1 Tat binding protein and subunit 4 of human 26S protease (proteasome). *Plant Sci* 103: 33–40.
- Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T and Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of black pine *Pinus thunbergii*. *Proc Natl Acad Sci USA* 91: 9794–9798.
- Wakasugi T, Nagai T, Kapoor M, Sugita M, Ito M, Ito S, Tsudzuki J, Nakashima K, Tsudzuki T, Suzuki Y, Hamada A, Ohta T, Inamura A, Yoshinaga K and Sugiura M (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: 5967–5972.
- Wakasugi T, Sugita M, Tsudzuki T and Sugiura M (1998) Updated gene map of tobacco chloroplast DNA. *Plant Mol Biol Repr* 16: 231–241.
- Wilson RJM, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW and Williamson DH (1996) Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J Mol Biol* 261: 155–172.
- Wolfe KH, Morden CW and Palmer JD (1992) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc Natl Acad Sci USA* 89: 10648–10652.