Plant Molecular Biology Reporter
13 (4) 1995

pages 336–342

Genetic Resources

The Chloroplast Genome of a Chlorophyll *a+c*-Containing Alga, *Odontella sinensis*

Klaus V. Kowallik, Bettina Stoebe, Ina Schaffran,
Peter Kroth-Pancic, and Ulrich Freier

E-mail: kowallik@uni-duesseldorf.de
Institut für Botanik der Heinrich-Heine-Universität Düsseldorf, D 40225
Düsseldorf, Germany

Key Words: chloroplast genome, Chromophyta, diatom, *Odontella sinensis*

Abstract: The chloroplast genome of a marine centric diatom, *Odontella sinensis*, was cloned and sequenced. The circular genome is 119,704 bp in length (AC = Z67753;). It contains an inverted repeat sequence of 7,725 bp separating two single-copy regions of 38,908 and 65,346 bp, respectively, and 174 genes and open reading frames, of which nine are duplicated within the inverted repeat segments.

Chloroplast genomes from algae not belonging to the chlorophyll *a+b*-containing lineage have only recently been characterized by their entire nucleotide sequences. These include the plastid genome of *Porphyra purpurea* (191,028 bp; Reith & Munholland, 1995) and the cyanelle genome of the enigmatic unicellular flagellate *Cyanophora paradoxa* (135,599 bp; Stirewalt et al., 1995). To date, chloroplast genes from chlorophyll *a+c*-containing algae have been sequenced arbitrarily. They included genes that are known from land plant chloroplast genomes (*e.g., psaA; psbA, -C, -D; atpB, -E, -I, -H, -F, -A; rbcL; petB, -D;* cf. Kowallik, 1994), but also some that are novel or known to be nuclear in land plants (cf. Valentin et al., 1993).

In order to establish evolutionary traits among non-green plastids, the chloroplast genome of the centric diatom *Odontella sinensis* was cloned and sequenced. It contains 119,704 bp. An inverted repeat (IR) segment of 7,725 bp, containing the rRNA operon, separates a small single-copy (SSC) region of 38,908 bp from a large single-copy (LSC) region of 65,346

Abbreviations: IR, inverted repeat; LSC, large single-copy region; nt, nucleotide; SSC, small single-copy region.

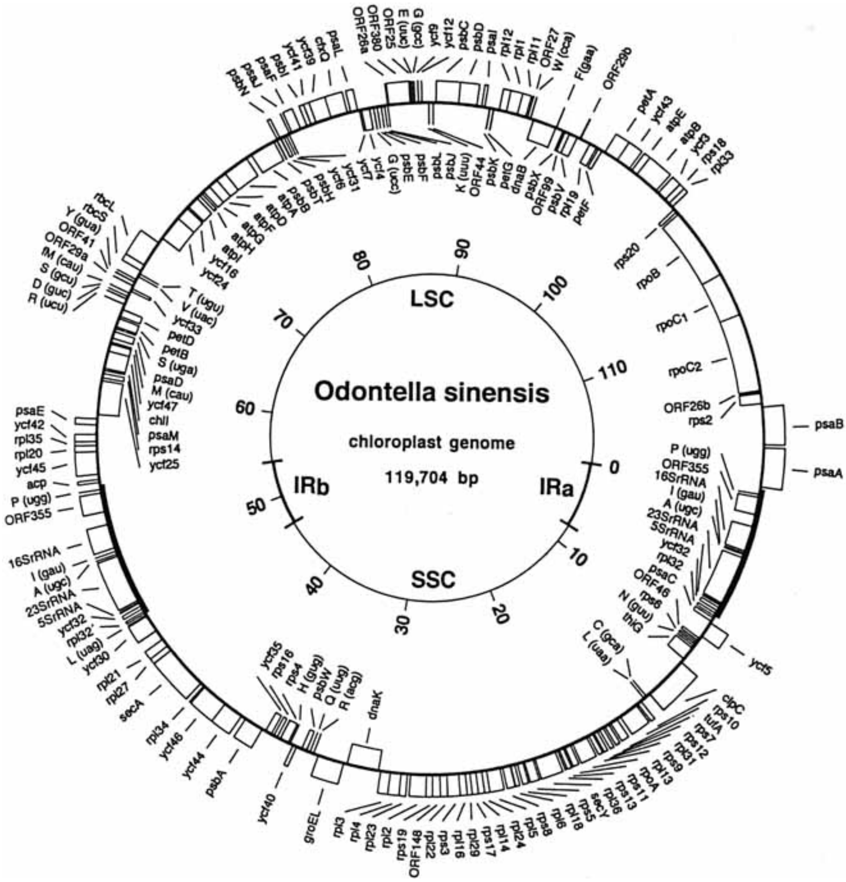

Fig. 1. Gene map of the circular chloroplast genome of *Odontella sinensis*. Genes transcribed clockwise are shown inside the circle, those transcribed counter-clockwise, are shown outside the circle. The inverted repeats are indicated by bold-type segments. Counting of the sequence starts at the right-hand inverted repeat (IRa) and continues clockwise.

bp (Fig. 1). The *Odontella* chloroplast genome reveals an overall coding capacity of 84.3 percent and has a (G+C)-content of 31.8 percent, which is equivalent to calculations based on the density of purified plastid DNA by equilibrium density sedimentation in gradients of CsCl (Kowallik, 1989).

To date, 174 genes and open reading frames (minimum of 25 amino acid residues) have been determined, of which nine are duplicated within the IR (Table I). No intron sequences or *ndh* genes have been found. Four pairs of genes show a 3′-5′ overlap (*rpl4/rpl23*, 8 nt; *ycf24/ycf16*, 1 nt; *atpF/atpD*, 4 nt; *psbD/psbC*, 53 nt). All three stop codons are

Table I. List of genes and ORFs of the *Odontella sinensis* chloroplast genome. The gene list proceeds clockwise from the LSC boundary of the right-hand inverted repeat (IRa). Genes are listed together with their putative gene products. ORFs with a (G+C)-content less than 30% (ORFs 25, 26a, 26b, 27, 29a, 29b, 41, 44, 46) are designated (A+T)-rich and may be non-coding. Numerical positions include the first nucleotide of the translation initiation codon as well as the last nucleotide of the stop codon. Transcription initiation and termination sites of tRNAs and rRNAs were determined according to multiple alignments and secondary structure models. Genes marked "(–)" are on the complementary strand.

Gene	Gene Product	Coding Sites	
	start IRa		1
trnP	tRNA-Pro(ugg)	267	340
ORF355		478	1545
rrn16	16S ribosomal RNA	2210	3694
trnI	tRNA-Ile(gau)	3761	3835
trnA	tRNA-Ala(ugc)	3838	3910
rrn23	23S ribosomal RNA	3976	6866
rrn5	5S ribosomal RNA	6938	7063
ycf32	ORF36	7318	7428
rpl32	50S ribosomal protein L32	7563	7739
	end IRa	7725	
psaC	PSI, Fe-S polypeptide SU VII, 9 kDa	7825	8073
ycf5	ORF312 (–)	8139	9077
ORF46	(A+T)-rich	9287	9427
rps6	30S ribosomal protein S6	9482	9772
trnN	tRNA-Asn(guu)	9852	9922
thiG	involved in thiamine biosynthesis	9944	10729
clpC	Clp protease, caseinolytic-like (–)	10894	13551
trnC	tRNA-Cys(gca)	13780	13850
trnL	tRNA-Leu(uaa)	13921	14006
rps10	30S ribosomal protein S10 (–)	14207	14518
tufA	elongation factor Tu (–)	14531	15760
rps7	30S ribosomal protein S7 (–)	15831	16307
rps12	30S ribosomal protein S12 (–)	16341	16721
rpl31	50S ribosomal protein L31 (–)	16773	16991
rps9	30S ribosomal protein S9 (–)	17005	17421
rpl13	50S ribosomal protein L13 (–)	17445	17864
rpoA	RNA polymerase α-chain (–)	17893	18831
rps11	30S ribosomal protein S11 (–)	18883	19275
rps13	30S ribosomal protein S13 (–)	19320	19691
rpl36	50S ribosomal protein L36 (–)	19713	19826
secY	preprotein-translocase subunit Y (–)	19866	21143
rps5	30S ribosomal protein S5 (–)	21166	21723
rpl18	50S ribosomal protein L18 (–)	21765	22172
rpl6	50S ribosomal protein L6 (–)	22219	22758
rps8	30S ribosomal protein S8 (–)	22777	23175

rpl5	50S ribosomal protein L5 (–)	23209	23925
rpl24	50S ribosomal protein L24 (–)	23957	24190
rpl14	50S ribosomal protein L14 (–)	24191	24556
rps17	30S ribosomal protein S17 (–)	24569	24823
rpl29	50S ribosomal protein L29 (–)	24868	25137
rpl16	50S ribosomal protein L16 (–)	25144	25557
rps3	30S ribosomal protein S3 (–)	25609	26253
rpl22	50S ribosomal protein L22 (–)	26295	26642
ORF148	(–)	26691	27137
rps19	30S ribosomal protein S19 (–)	27149	27427
rpl2	50S ribosomal protein L2 (–)	27472	28299
rpl23	50S ribosomal protein L23 (–)	28324	28632
rpl4	50S ribosomal protein L4 (–)	28625	29272
rpl3	50S ribosomal protein L3 (–)	29320	29910
dnaK	Hsp70-type chaperone	30213	32057
groEL	chaperonin, 60 kDa (–)	32125	33711
trnR	tRNA-Arg(acg)	33891	33964
trnQ	tRNA-Gln(uug)	34020	34089
psbW	PSII, protein W, 13 kDa	34326	34673
ycf40	ORF73, homologous to *Porphyra* ORF71 (–)	34997	35218
trnH	tRNA-His(gug)	35294	35368
rps4	30S ribosomal protein S4	35435	35955
rps16	30S ribosomal protein S16	36238	36477
ycf35	ORF128	36607	36993
psbA	PSII, D1 reaction-center protein (–)	37339	38421
ycf44	ORF382, homologous to *Porphyra* ORF437 (–)	38642	39790
ycf46	ORF497, homologous to *Porphyra* ORF491 (–)	39799	41292
rpl34	50S ribosomal protein L34 (–)	41404	41550
secA	preprotein-translocase subunit α (–)	41583	44249
rpl27	50S ribosomal protein L27 (–)	44272	44523
rpl21	50S ribosomal protein L21 (–)	44546	44863
ycf30	former *trsE*(*rbcR* homolog) (–)	45509	46438
trnL	tRNA-Leu(uag) (–)	46484	46564
	end IRb	46634	
rpl32'	50S ribosomal protein L32' (–)	46632	46796
ycf32	ORF36 (–)	46931	47041
rrn5	5S ribosomal RNA (–)	47296	47421
rrn23	23S ribosomal RNA (–)	47493	50383
trnA	tRNA-Ala(ugc) (–)	50449	50521
trnI	tRNA-Ile(gau) (–)	50524	50598
rrn16	16S ribosomal RNA (–)	50665	52149
ORF355	(–)	52814	53881
trnP	tRNA-Pro(ugg) (–)	54019	54092
	start IRb		54358
acp	acyl carrier protein (–)	54453	54695
ycf45	ORF455, homologous to *Porphyra* ORF565 (–)	54953	56320
rpl20	50S ribosomal protein L20 (–)	56561	56905

rpl35	50S ribosomal protein L35 (–)	56915	57109
ycf42	ORF204, homologous to *Porphyra* ORF199 (–)	57112	57726
psaE	PSI, subunit IV (–)	57835	58050
ycf25	ORF 644, homologous to *E. coli* cell-division protein FtsH	58339	60273
rps14	30S ribosomal protein S14	60428	60730
psaM	PSI, protein M	60884	60976
chlI	chlorophyll biosynthesis, probable magnesium-chelatase subunit	61054	62115
ycf47	ORF74, homologous to *Porphyra* ORF71	62143	62367
trnM	tRNA-Met(cau)	62586	62659
psaD	PSI, ferredoxin-binding protein II	62744	63163
trnS	tRNA-Ser(uga)	63199	63286
petB	cytochrome b6	63341	63988
petD	cytochrome b6/f complex, subunit IV	64035	64517
trnR	tRNA-Arg(ucu) (–)	64774	64845
trnD	tRNA-Asp(guc) (–)	65093	65167
trnS	tRNA-Ser(gcu) (–)	65253	65340
trnfM	tRNA-fMet(cau) (–)	65380	65452
ycf33	ORF64	65614	65808
ORF29a	(A+T)-rich (–)	65920	66009
ORF41	(A+T)-rich (–)	66057	66182
trnY	tRNA-Tyr(gua) (–)	66350	66431
trnV	tRNA-Val(uac)	66555	66626
trnT	tRNA-Thr(ugu)	66647	66717
rbcS	Rubisco, small subunit (–)	66864	67283
rbcL	Rubisco, large subunit (–)	67323	68795
ycf24	ORF486	69086	70546
ycf16	ORF251	70546	71301
atpI	ATP synthase CF_O subunit IV	71429	72157
atpH	ATP synthase CF_O subunit III	72227	72475
atpG	ATP synthase CF_O subunit II	72587	73057
atpF	ATP synthase CF_O subunit I	73121	73660
atpD	ATP synthase CF_1 subunit δ	73657	74220
atpA	ATP synthase CF_1 subunit α	74268	75779
psbB	PSII, CP47 chlorophyll apoprotein	76048	77577
psbT	PSII, protein T, 3 kDa(*ycf8*)	77629	77727
psbN	PSII, protein N (–)	77752	77883
psbH	PSII, phosphoprotein, 10 kDa	77966	78169
ycf6	ORF29	78255	78344
ycf31	ORF42	78384	78512
psaJ	PSI, subunit IX, 5 kDa (–)	78575	78700
psaF	PSI, subunit III, plastocyanin-binding (–)	78732	79289
psbI	PSII, protein I, 4.8 kDa (–)	79654	79770
ycf41	ORF113, homologous to *Porphyra* ORF111 (–)	79855	80196
ycf39	ORF319 (–)	80212	81171
cfxQ	involved in Rubisco-expression (–)	81201	82076
psaL	PSI, subunit XI (–)	82171	82623
ycf7	ORF31	82800	82895

ycf4	ORF181	82917	83462
trnG	tRNA-Gly(ucc)	83508	83578
psbE	cytochrome b_{559} α-chain	83634	83888
psbF	cytochrome b_{559} β-chain	83901	84032
psbL	PSII, protein L	84068	84184
psbJ	PSII, protein J	84209	84328
ORF26a	(A+T)-rich (–)	84426	84506
ORF380		84538	85680
ORF25	(A+T)-rich (–)	85745	85822
trnE	tRNA-Glu(uuc) (–)	85882	85953
trnG	tRNA-Gly(gcc) (–)	85995	86066
ycf9	ORF61 (–)	86186	86371
ycf12	ORF34 (–)	86419	86523
trnK	tRNA-Lys(uuu)	86730	86801
ORF44	(A+T)-rich	86932	87066
psbC	PSII, CP43 chlorophyll apoprotein (–)	87163	88578
psbD	PSII, D2 reaction-center protein (–)	88526	89581
psaI	PSI, subunit VIII (–)	89824	89940
psbK	PSII, protein K	90146	90280
petG	cytochrome b6/f complex, subunit V	90417	90530
rpl12	50S ribosomal protein L12 (–)	90778	91161
rpl1	50S ribosomal protein L1 (–)	91234	91926
rpl11	50S ribosomal protein L11 (–)	91949	92374
ORF27	(A+T)-rich (–)	92456	92539
trnW	tRNA-Trp(cca) (–)	92660	92732
dnaB	DNA-replication helicase	92815	94182
trnF	tRNA-Phe(gaa) (–)	94184	94256
psbX	PSII, protein X, 4.1 kDa	94593	94709
ORF99		94783	95082
psbV	PSII, cytochrome c_{550} (*petK*)	95139	95630
ORF29b	(A+T)-rich (–)	95659	95748
rpl19	50S ribosomal protein L19	96389	96751
petF	ferredoxin	96895	97194
petA	apocytochrome f (–)	97278	98222
ycf43	ORF263, homologous to *Porphyra* ORF254 (–)	98256	99047
atpE	ATP synthase CF$_1$ subunit ε (–)	99151	99552
atpB	ATP synthase CF$_1$ subunit β (–)	99566	100993
ycf3	ORF179 (–)	101162	101701
rps18	30S ribosomal protein S18 (–)	101792	102010
rpl33	50S ribosomal protein L33 (–)	102014	102208
rps20	30S ribosomal protein S20	102420	102701
rpoB	RNA polymerase β-chain	102865	107004
rpoC1	RNA polymerase β'-chain	107028	109559
rpoC2	RNA polymerase β''-chain	109599	114044
ORF26b	(A+T)-rich	114058	114138
rps2	30S ribosomal protein S2	114159	114848
psaB	PSI, P700 apoprotein A2 (–)	114970	117171
psaA	PSI, P700 apoprotein A1 (–)	117351	119609

used, with a preference for TAA. Four genes start with GTG instead of ATG (*rps3, rps13, rpl23, rbcS*).

Identified genes are named according to Hallick and Bottomley (1983) and Hallick (1989). Open reading frames shared by homologous sequences of other chloroplast genomes are designated as *ycf* (Hallick and Bairoch, 1994), genes for tRNAs are labelled using the one-letter code of their respective amino acids. Open reading frames unique to *Odontella* are designated as *ORF*, followed by the number of encoded amino acid residues. Genes shared with the *Porphyra purpurea* chloroplast genome and the cyanelle genome of *Cyanophora paradoxa*, but unknown from land-plant chloroplast genomes, include *acp, atpD, atpG, dnaK, groEL, petF, psaE, psaF, psbV, psbW, psbX, rbcS, tufA, secY*, and *ycf16, -24, -30, -31, -32, -33, -35, -39*. In addition, several genes for ribosomal subunits (*rpl1, -3, -6, -11, -12, -18, -34, -35*, and *rps5, -6, -10, -13, -17, 20*) appear to be unique among plastid genomes of non-green algae. Whereas eight open reading frames (*ycf40-47*) have counterparts in the *Porphyra* chloroplast genome, none of these ORFs were found in the cyanelle genome.

A detailed description of the sequence and its phylogenetic implications will appear elsewhere.

Acknowledgment: This work was supported by the Deutsche Forschungsgemeinschaft (**Molekulare Evolution der Pflanzen, grants Ko 439/6-1 and 439/6-2**).

References

Hallick, R.B. 1989. Proposals for the naming of chloroplast genes. II. Update to the nomenclature of genes for thylakoid membrane polypeptides. Plant Mol. Biol. Reptr. 7:266-275.

Hallick, R.B. and A. Bairoch. 1994. Proposals for the naming of chloroplast genes. III. Nomenclature for open reading frames encoded in chloroplast genomes. Plant Mol. Biol. Reptr. 12:29-30.

Hallick, R.B. and W. Bottomley. 1983. Proposals for the naming of chloroplast genes. Plant Mol. Biol. Reptr. 1:38-43.

Kowallik, K.V. 1989. Molecular aspects and phylogenetic implications of plastid genomes of certain chromophytes, pp.101-124 in *The Chromophyte Algae: Problems and Perspectives* (Green J.C., B.S.C. Leadbeater, and W.L. Diver, eds.). The Systematics Association Special Volume 38. Clarendon Press, Oxford.

Kowallik, K.V. 1993. Origin and evolution of plastids from chlorophyll *a*+*c*-containing algae: Suggested ancestral relationship to red and green algal plastids, pp. 223-263 in *Origins of Plastids* (Lewin, R.A., ed.). Chapman and Hall, New York and London.

Reith, M., and J. Munholland. 1995. Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. Plant Mol. Biol. Reptr. 13:338-340.

Stirewalt, V.S., C. B. Michalowski, W. Löffelhardt, H. J. Bohnert, and D. A. Bryant. 1995. Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. Plant Mol. Biol. Reptr. 13:327-332.

Valentin, K., R.A. Cattolico and K. Zetsche. 1993. Phylogenetic origin of the plastids, pp. 193-212 in *Origins of Plastids* (Lewin, R.A., ed.). Chapman and Hall, New York and London.