

Genetic Resource

Nucleotide Sequence of the Cyanelle Genome from *Cyanophora paradoxa*

Veronica L. Stirewalt, Christine B. Michalowski,
Wolfgang Löffelhardt, Hans J. Bohnert, and Donald A. Bryant

E-mail: dab14@psuvm.psu.edu; bohnert@biosci.arizona.edu
(VLS, DAB) Department of Biochemistry and Molecular Biology, The
Pennsylvania State University, University Park, PA, 16802, U.S.A.
(CBM, HJB) Departments of Biochemistry, Molecular and Cellular Biology,
and Plant Sciences, The University of Arizona, Tucson, AZ 85721, U.S.A.
(WL) Institut für Biochemie und Molekulare Zellbiologie und
Ludwig-Boltzmann-Forschungsstelle für Biochemie, Universität Wien,
A-1030, Vienna, Austria.

Key Words: *Cyanophora paradoxa*, cyanelle, chloroplast, plastid, nucleotide sequence.

Abstract: The complete nucleotide sequence of the cyanelle genome of *Cyanophora paradoxa* Pringsheim strain LB 555 was determined (accession number U30821). The circular molecule is 135,599 base pairs in length. The physical map of this DNA molecule is shown along with identified genes and open reading frames.

C *yanophora paradoxa* (*Glaucocystophyceae*) is a bi-flagellated protist that contains cyanobacterium-like plastids termed cyanelles. The cyanelles of *Cyanophora paradoxa* are conspicuous because they are surrounded by a lysozyme-sensitive peptidoglycan wall that is typical of those associated with cyanobacteria but that is not found in other plastid types. Thus, this organism has frequently been considered a “living fossil” and a paradigm for the invasion of a eukaryotic cell by a cyanobacterium (for a review, see Löffelhardt and Bohnert, 1994). The complete nucleotide sequence, 135,599 bp, of the cyanelle DNA from the

Abbreviations: LSC, large single-copy region; SSC, small single-copy region.

unicellular alga *Cyanophora paradoxa* Pringsheim strain LB555 has been determined (GeneBank accession number: U30821). The chromosome (Fig. 1) has a G+C content of 30.4 percent and is characterized by two inverted-repeat segments of 11,285 bp each (IR_A and IR_B , respectively), separated by a large single-copy region (LSC) of 94,946 bp and a small single-copy region (SSC) of 18,083 bp. Approximately 192 identified genes and open reading frames (of which ten are duplicated in the inverted repeats) have been identified thus far. Although the genome is smaller than that of tobacco, it encodes about 30 percent more genes than does the tobacco chloroplast genome. Genes encoded in the chloroplast genomes of higher plants but not found in the cyanelle genome include *ndhA-K*, *rps15*, *rpl23*, *infA*, *atpI*, *accD*, and *matK*.

To catalog genes and reading frames on the cyanelle DNA, the molecule has arbitrarily been linearized at the beginning of the inverted repeat 5' to the end of one of the 16S rRNA genes. Counting proceeds through IR_A , LSC, IR_B , and SSC (Fig. 1). Identified genes are given names as recommended (Hallick and Bottomley, 1983; Hallick, 1989); unknown reading frames larger than 25 codons are labeled "ORF" followed by the number of codons. Unknown reading frames that have been found in other plastid DNAs are labeled *ycf* (Hallick and Bairoch, 1993). Genes for tRNAs are identified by the amino acid with which they are charged in the single-letter code.

While a full discussion of the sequence characteristics will be presented elsewhere, we should note the special features of cyanelle DNA:

- a single type-I intron located in *trnL*(UUA) in a position that is conserved in many cyanobacteria and all plastid DNAs (Kuhnel et al., 1990; Xu et al., 1990);
- a full complement of tRNA genes (36 in total) and the gene *rnpB* (= RNA subunit of RNaseP);
- one set of rRNA genes (16 S, 23 S, and 5 S, with *trnA* and *trnI* genes in the 16 S-23 S spacers) in each IR;
- 37 ribosomal protein genes (18 small-subunit proteins; 19 large-subunit proteins);
- four genes encoding RNA polymerase subunit proteins, and *tufA* encoding translation factor Tu;
- seven genes encoding phycobiliproteins (including all chromophorylated subunits of the phycobilisome);

Fig. 1: Physical map and gene map of the cyanelle genome of *Cyanophora paradoxa* (opposite page). See table of gene symbols on p. 324.

- 25 genes encoding subunit proteins of the photosystem I complex (8 proteins) and photosystem II complex (17 subunits);
- seven genes encoding ATPase subunit proteins;
- six genes encoding subunits of the cytochrome *b₆f* complex;
- the *petF* gene encoding type I [2Fe-2S] ferredoxin;
- genes for the molecular chaperones GroEL, GroES, and DnaK located within IR segments; in addition gene *secY* is present, encoding a subunit of the preprotein translocation machinery;
- two different genes (*clpP1* and *clpP2*) encoding ClpP subunits of the Clp protease;
- one gene with homology to *ftsW*, possibly involved in cyanelle division or cell wall biosynthesis;
- 13 identified genes encoding proteins for functions in metabolism;
- cyanobacterial L_8S_8 RuBisCO encoded by an *rbcL-rbcS* operon;
- three ORFs with possible functions in the transcriptional regulation of gene expression (two OmpR family members; one LysR family member);
- 15 ORF and 21 *ycf* (ranging from 27 to 333 codons);
- five ORF/*ycf* with possible functions as membrane transporters;
- an ORF (*ycf17*) encoding a protein of 49 amino acids with strong homology to all members of the CAB/ELIP/HLIP protein superfamily;
- no *ndh* genes or pseudo-*ndh* reading frames (genes for NADH dehydrogenase).

The sequences of four algal plastid DNAs are or will soon be available (Table I): those of the cyanelle genome of *Cyanophora paradoxa* described here, the 191,028-bp plastid DNA from the red alga *Porphyra purpurea* (Reith, 1995; Reith & Munholland, 1995), the 119,704-bp plastid genome of the diatom *Odontella sinensis* (class *Bacillariophyceae*) (Kowallik et al., 1995), and the 143,170-bp plastid genome of *Euglena gracilis* (Hallick et al., 1993). These sequences complement those of chloroplast DNAs from six higher plant species: the liverwort *Marchantia polymorpha* (Ohyama et al., 1986); the gymnosperm *Pinus thunbergii* (Wakasugi et al., 1994); the monocots *Oryza sativa* (Hiratsuka et al., 1989) and *Zea mays* (140,386 bp; Maier et al., 1995); and the dicots *Nicotiana tabacum* (Shinozaki et al., 1986) and *Epifagus virginiana*, a parasitic, nonphotosynthetic species (Wolfe et al., 1992).

Acknowledgments. We wish to thank Hans Kössel, Klaus Kowallik, and Michael Reith for sharing unpublished information. This work has primarily been supported by a grant from the U.S. Department of Agriculture (National Research Initiative—Plant Genome Program) to H.J.B. and D.A.B. Additional support was

Table I. Fully sequenced plastid genomes.

Plant Species	Genome Size (bp)	Reference
Algae		
<i>Cyanophora paradoxa</i>	135,599	this article
<i>Porphyra purpurea</i>	119,704	Reith, 1995; Reith & Munholland, 1995
<i>Odontella sinensis</i>	119,680	Kowallik et al., 1995
<i>Euglena gracilis</i>	143,170	Hallick et al., 1993
<i>Chlorella ellipsoidea</i>	155,000	M. Sugiura, personal communication
Higher Plants		
<i>Marchantia polymorpha</i>	121,024	Ohyama et al., 1986
<i>Pinus thunbergii</i>	119,707	Wakasugi et al., 1994
<i>Nicotiana tabacum</i>	155,844	Shinozaki et al., 1986
<i>Epifagus virginiana</i>	70,028	Wolfe et al., 1992
<i>Oryza sativa</i>	134,525	Hiratsuka et al., 1989
<i>Zea mays</i>	140,386	Maier et al., 1995

provided by the National Science Foundation (DMB-8818997; D.A.B.), the Arizona Agricultural Experiment Station (H.J.B.), and the Austrian Research Council (W.L.).

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. Repr.* 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. Repr.* 12:S29-30.
- Hallick, R.B. and W. Bottomley. 1983. Proposals for the naming of chloroplast genes. *Plant Mol. Biol. Repr.* 1:38-43.
- Hallick, R.B., L. Hong, R.G. Drager, M.R. Favreau, A. Montfort, B. Orsat, A. Spielmann, and E. Stutz. 1993. Complete sequence of the *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* 21:3537-3544.
- Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, C. Sun, B. Meng, Y. Li, A. Kanno, Y. Nishizawa, A. Hirai, K. Shinozaki and M. Sugiura. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes account for a major plastid inversion during the evolution of cereals. *Mol. Gen. Genet.* 217:185-194.
- Kowallik, K.V., B. Stoebe, I. Schaffran, and U. Freier. 1995. The chloroplast genome of a chlorophyll *a+c*-containing alga, *Odontella sinensis*. *Plant Mol. Biol. Repr.* 13:226-342.
- Kuhnel, M.G., R. Strickland and J.D. Palmer. 1990. An ancient group I intron shared by eubacteria and chloroplasts. *Science* 250:1570-1573.

- Löffelhardt, W. and H.J. Bohnert. 1994. Molecular biology of cyanelles. In: *The Molecular Biology of Cyanobacteria*, (ed., D.A. Bryant), pp. 65-89, Kluwer Academic Publ., Dordrecht.
- Maier, R.M., K. Neckermann, G.I. Igloi and H. Kössel. 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.
- Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S.-i. Aota, H. Inokuchi, and H. Ozeki. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha*. *Nature* 322:572-574.
- Reith, M. 1995. Molecular biology of rhodophyte and chromophyte plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 46:549-575.
- Reith, M. and J. Munholland. 1995. Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol. Biol. Repr.* 13:333-345.
- Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B.Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada and M. Sugiura. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *EMBO J.* 5:2043-2049.
- Wakasugi, T., J. Tsudzuki, S. Ito, K. Nakashima, T. Tsudzuki and M. Sugiura. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc. Natl. Acad. Sci. U.S.A.* 91:9794-9798.
- Wolfe, K.H., C.W. Morden and J.D. Palmer. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc. Natl. Acad. Sci. U.S.A.* 89:10648-10652.
- Xu, M.Q., S.D. Kathe, H. Goodrich-Blair, S.A. Nierzwicki-Bauer and D.A. Shub. 1990. Bacterial origin of a chloroplast intron: Conserved self-splicing group I introns in cyanobacteria. *Science* 250:1566-1570.