MITOGENOME ANNOUNCEMENT

Complete genome sequence of chloroplast DNA (cpDNA) of *Chlorella sorokiniana*

Massimiliano Orsini[1], Roberto Cusano[1], Cristina Costelli[2], Veronica Malavasi[2], Alessandro Concas[1], Andrea Angius[1,3], and Giacomo Cao[1,2,4]

[1]Center for Advanced Studies, Research and Development in Sardinia (CRS4), Loc. Piscina Manna, Pula, Province of Cagliari, Italy, [2]Research Unit of the National Interuniversity Consortium "The Chemistry for the Environment" and Interdepartmental Center of Environmental Science and Engineering (CINSA), University of Cagliari, Cagliari, Italy, [3]Institute of Genetic and Biomedical Research (IRGB), National Research Council (CNR), Monserrato, Province of Cagliari, Italy, and [4]Department of Mechanical, Chemical and Materials Engineering, University of Cagliari, Piazza D'armi, Cagliari, Italy

Abstract

The complete chloroplast genome sequence of *Chlorella sorokiniana* strain (SAG 111–8 k) is presented in this study. The genome consists of circular chromosomes of 109,811 bp, which encode a total of 109 genes, including 74 proteins, 3 rRNAs and 31 tRNAs. Moreover, introns are not detected and all genes are present in single copy. The overall AT contents of the *C. sorokiniana* cpDNA is 65.9%, the coding sequence is 59.1% and a large inverted repeat (IR) is not observed.

The last few decades have seen a rising interest in chloroplast genome characterization and, as a result, a growing number of algae chloroplasts have been sequenced to better understand their evolution. The chloroplast genome of algae belonging to the Chlorophyta, displays extraordinary variability in terms of quadripartite structure, global gene organization and intron composition (de Cambiaire et al., 2006). Within the Chlorophyte-clade *sensu* Lewis & McCourt (2004), the Trebouxiophyceae group consists mainly of unicellular and colonial green algae, which are found mostly in subaerial habitats and as phycobionts in lichens. This group includes the Chlorellaceae family (Brunnthaler, 1915) which is characterized by the well-known genus *Chlorella* Beijerinck. One member of this genus is *Chlorella sorokiniana* which is a non-motile unicellular alga which can grow phototrophically by exploiting wastewaters and flue gases as costless sources of inorganic nutrients and CO_2, respectively. When cultivated under specific operating conditions, this strain is capable to accumulate significant amounts of fatty acids to be viably exploited for the production of biofuels (Lizzul et al., 2014). These characteristics make this strain, as well as most of the species belonging to the *Chlorella* genus, particularly promising in the biotechnological, environmental and energy sectors (Concas et al., 2010, 2012, 2014). In this study, the complete chloroplast genome of the *C. sorokiniana* (SAG 111–8 k), is presented. Unialgal culture was maintained under axenic conditions in shaken flasks illuminated by a photon

flux of 90–100 μE m^{-2} s^{-1} and containing modified WARIS-H culture medium (McFadden & Melkonian, 1986) without soil extract. Whole-genome sequencing of the *C. sorokiniana* strain was performed using Illumina HiSeq2000 platform (Pula (CA), Italy). The cpDNA sequence of *C. sorokiniana* (GenBank accession no. KJ397925) assembles as a circular map of 109,811 bp which encodes a total of 109 genes (Figure 1). The circular map was obtained using the CGView software (Stothard & Wishart, 2005). This chloroplast gene repertoire includes 74 proteins, 3 rRNAs and 31 tRNAs. The 74 proteins accounts for, in turn, 18 ribosomal proteins, an elongation factor and components of an RNA polymerase (rpo). In the genome, which doesn't show introns, all genes are present in single copy. Moreover, as well as for other *Chlorella* strains, the genome is lacking large inverted repeat (IR). *Chlorella sorokiniana* belongs to the group that displays the standard architecture of the rRNA operon of Chlrphyta structure (16S rDNA – tRNA Ile – tRNA Ala – 23S rDNA). The overall AT content of the *C. sorokiniana* cpDNA is 65.9% and the coding sequence is 59.1% (54,279 bp). The sequence analysis, here reported, highlights that the plastid genome of *C. sorokiniana*, does not display a quadripartite structure. Multiple sequence alignment compared to *C. variabilis* and *C. vulgaris* cpDNAs (NC_015359 and NC_001865, respectively) reveals a closer relationship to the former species with no particular rearrangements in the cpDNA sequence. On the contrary, a modest overall identity and large rearrangements are observed in the obtained chloroplast genome with respect to the *C. vulgaris* cpDNAs. On the other hand, the comparison among chloroplast encoded proteins showed a relevant similarity except for a small difference in two subunits of the *C. vulgaris* RNA polymerase.

Correspondence: Giacomo Cao, Department of Mechanical, Chemical and Materials Engineering, University of Cagliari, Piazza d'Armi, Cagliari 09123, Italy. Tel: +39 0706755058. Fax: +39 0706755057. E-mail: giacomo.cao@dimcm.unica.it
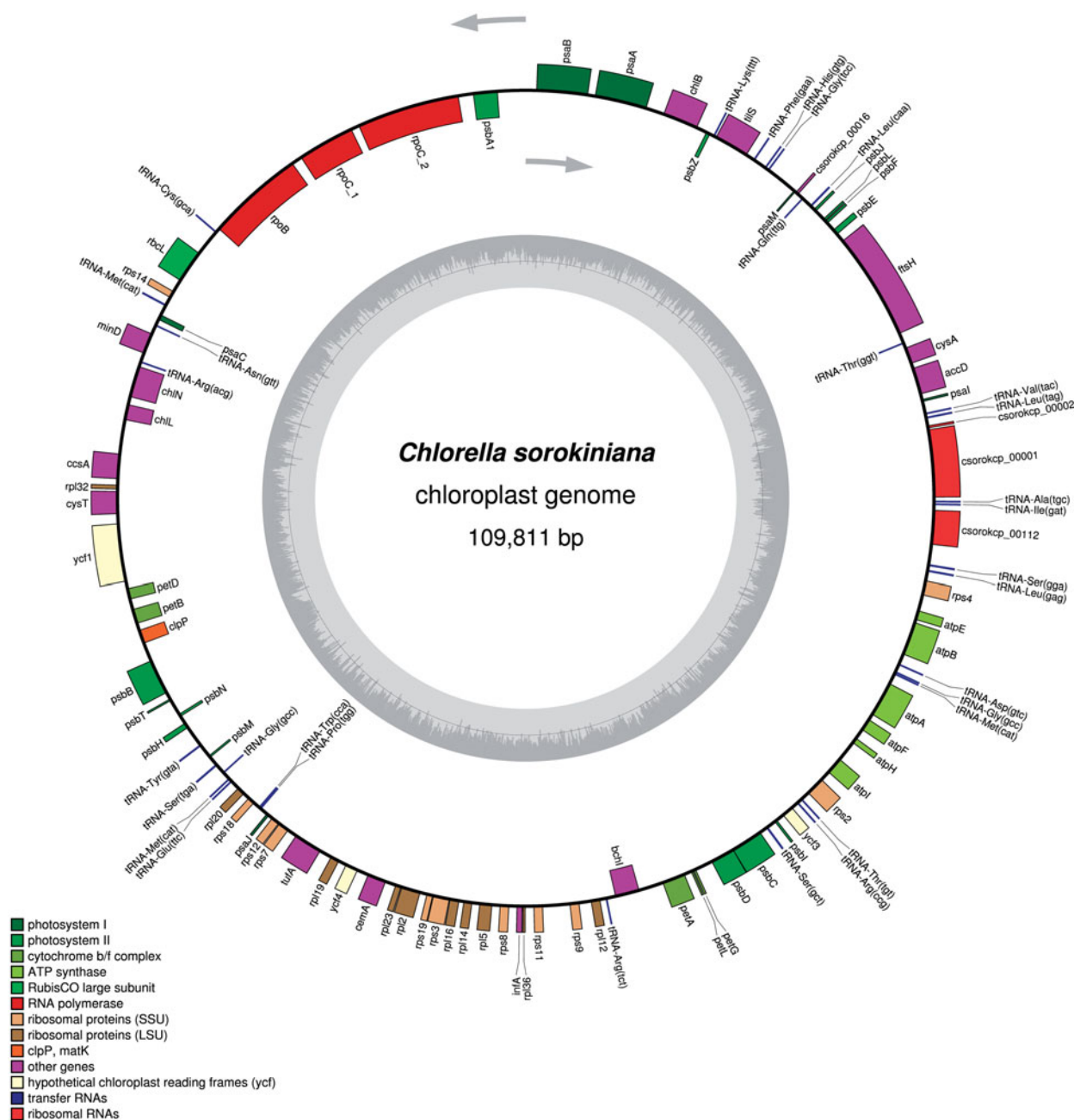
RIGHTSLINK

Figure 1. Gene map of the chloroplast genome of *C. sorokiniana*.

Declaration of interest

References

Brunnthaler J. (1915). Chlorophyceae 2. In: Pascher A, editors. Süßwasserflora Deutschlands, Österreichs und der Schweiz. Jena: Fischer. p 183–6.

Concas A, Pisu M, Cao G. (2010). Novel simulation model of the solar collector of BIOCOIL photobioreactors for CO_2 sequestration with microalgae. Chem Eng J 157:297–303.

Concas A, Lutzu GA, Pisu M, Cao G. (2012). Experimental analysis and novel modeling of semi-batch photobioreactors operated with *Chlorella vulgaris* and fed with 100% (v/v) CO_2. Chem Eng J 213:203–13.

Concas A, Steriti A, Pisu M, Cao G. (2014). Comprehensive modeling and investigation of the effect of iron on the growth rate and lipid accumulation of *Chlorella vulgaris* cultured in batch photobioreactors. Bioresource Technol 153:340–50.

de Cambiaire JC, Otis C, Lemieux C, Turmel M. (2006). The complete chloroplast genome sequence of the chlorophycean green alga *Scenedesmus obliquus* reveals a compact gene organization and a biased distribution of genes on the two DNA strands. BMC Evol Biol 6:37.

Lewis LA, McCourt RM. (2004). Green algae and the origin of land plants. Am J Bot 91:1535–56.

Lizzul AM, Hellier P, Purton S, Baganz F, Ladommatos N, Campos L. (2014). Combined remediation and lipid production using *Chlorella sorokiniana* grown on wastewater and exhaust gases. Bioresource Technol 151:11–18.

McFadden GI, Melkonian M. (1986). Culture Medium Waris-H. Phycologia 25:551–7.

Stothard P, Wishart DS. (2005). Circular genome visualization and exploration using CGView. Bioinformatics 21:537–9.