Microbiological Research 279 (2024) 127572

Contents lists available at ScienceDirect

Microbiological Research

journal homepage: www.elsevier.com/locate/micres

Characterization of Limnospira platensis PCC 9108 R-M and CRISPR-Cas systems

María Castillo [a,1], Govinda Guevara [b,1], Sara Baldanta [a,b], Patricia Suárez Rodríguez [b], Lucía Agudo [c], Juan Nogales [c], Asunción Díaz Carrasco [d], Fernando Arribas-Aguilar [e], Julián Pérez-Pérez [e], José Luis García [a], Beatriz Galán [a,*], Juana María Navarro Llorens [b]

[a] Microbial and Plant Biotechnology Department, Centro de Investigaciones Biológicas Margarita Salas-CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain
[b] Department of Biochemistry and Molecular Biology, Facultad de Ciencias Biológicas, Universidad Complutense de Madrid, Av. Complutense s/n, 28040 Madrid, Spain
[c] Department of Systems Biology, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CNB-CSIC), 28049 Madrid, Spain
[d] DNA Sequencing facility, Centro de Investigaciones Biológicas Margarita Salas-CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain
[e] SECUGEN SL, Centro de Investigaciones Biológicas Margarita Salas-CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Cyanobacteria
Methyltransferase
Limnospira platensis
Restriction-modification systems
CRISPR-Cas

ABSTRACT

The filamentous cyanobacterium Limnospira platensis, formerly known as Arthrospira platensis or spirulina, is one of the most commercially important species of microalgae. Due to its high nutritional value, pharmacological and industrial applications it is extensively cultivated on a large commercial scale. Despite its widespread use, its precise manipulation is still under development due to the lack of effective genetic protocols. Genetic transformation of Limnospira has been attempted but the methods reported have not been generally reproducible in other laboratories. Knowledge of the transformation defense mechanisms is essential for understanding its physiology and for broadening their applications. With the aim to understand more about the genetic defenses of L. platensis, in this work we have identified the restriction-modification and CRISPR-Cas systems and we have cloned and characterized thirteen methylases. In parallel, we have also characterized the methylome and orphan methyltransferases using genome-wide analysis of DNA methylation patterns and RNA-seq. The identification and characterization of these enzymes will be a valuable resource to know how this strain avoids being genetically manipulated and for further genomics studies.

1. Introduction

Arthrospira platensis, known commercially as spirulina, is an edible photosynthetic cyanobacterium widely used for food supply due to its high protein content (460–630 $g{\cdot}kg^{-1}$, dry matter) and for being a source of vitamins, essential amino acids, fatty acids and bioactive compounds with antioxidant or anti-inflammatory activities (Kose et al., 2017; Lupatini et al., 2017; AlFadhly et al., 2022). This species has also been considered for bioethanol and biodiesel production (Mohamadzadeh Shirazi et al., 2017; Esquivel-Hernández et al., 2021), antibody production (Jester et al., 2022; Saveria et al., 2022) or metal removal (Gunasundari and Senthil Kumar, 2017; Zinicovscaia et al., 2017; Yadav et al., 2022). A. platensis has been separated into the new genus Limnospira like other mass-produced Arthrospira species (Nowicka-Krawczyk et al., 2019).

Given the myriad possible biotechnological applications of L. platensis, its genome editing, or genetic manipulation is of paramount importance. The scarce availability of genetic tools for this species is still a drawback. In fact, up to now, there are few works where a genetic transformation has been successfully obtained in L. platensis strains (Kawata et al., 2004; Jeamton et al., 2017; Dehghani et al., 2018; Jester et al., 2022; Saveria et al., 2022; Tabakh et al., 2023). One of the problems that hinders transformation is the need to overcome the strong genetic barriers that L. platensis presents, which can differ from one strain to another. Genome comparisons revealed many genes that are considered major obstacles for stable transformation such as

* Corresponding author.
E-mail addresses: maria.castillo@cib.csic.es (M. Castillo), fguevara@ucm.es (G. Guevara), sabaldan@ucm.es (S. Baldanta), patriciasuarez@ucm.es (P.S. Rodríguez), l.algibe@gmail.com (L. Agudo), jnogales@cnb.csic.es (J. Nogales), a.diaz@cib.csic.es (A.D. Carrasco), f.arribas@secugen.es (F. Arribas-Aguilar), j.perez@secugen.es (J. Pérez-Pérez), jlgarcia@cib.csic.es (J.L. García), bgalan@cib.csic.es (B. Galán), jmnavarr@ucm.es (J.M. Navarro Llorens).
[1] María Castillo and Govinda Guevara authors contributed equally to this work.

https://doi.org/10.1016/j.micres.2023.127572
Received 25 September 2023; Received in revised form 5 December 2023; Accepted 6 December 2023
Available online 9 December 2023

restriction-modification systems (R-M) and CRISPR elements (Fujisawa et al., 2010; Cheevadhanarak et al., 2012; Xu et al., 2016; Suzuki et al., 2019).

R-M systems are key components of prokaryotic defense mechanism against invading DNA. They comprise two contrasting enzymatic activities: a restriction endonuclease (REase) and a methyltransferase (MTase), the REase recognizes and cleavages foreign DNA sequences at specific sites whereas MTase activity ensures discrimination between self and non-self-DNA by transferring a methyl group to the same specific sequences on self-genome (Vasu and Nagaraja, 2013). R-M systems are classified in four different types based on their subunit composition, sequence recognition, cleavage position, and substrate specificity (Roberts et al., 2003). Among them, Type II is the most widely studied and extensively used in genetic engineering. In addition, several genomes are also known to encode orphan or solitary MTases that are not associated to any known REase, as well as solitary REases that are not associated with any MTase (Blow et al., 2016).

DNA MTases specifically recognize a short palindromic sequence and catalyze the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to the target adenine or cytosine (Jeltsch, 2002). MTases are annotated into three groups based on the methylation products observed in prokaryotes: N^6-methyladenine, N^4-methylcytosine and N^5-methylcytosine (Wilson and Murray, 1991; Roberts et al., 2003). The marks introduced into DNA by methylases after DNA replication is not only important for defense against foreign DNA, but it is also involved in a variety of physiological processes including DNA repair, transcriptional regulation of essential processes, post-transcriptional gene regulation, and epigenetics (Anton and Roberts, 2021; Rolando et al., 2022).

CRISPR-Cas (that stand for Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins) is an adaptive immunity system that confers immune memory stored in the form of spacer sequences derived from foreign DNA inserted into the leader-repeat junction (Garrett et al., 2011; Koonin et al., 2017). Therefore, organisms possessing CRISPR-Cas systems can identify and degrade foreign DNA. Different CRISPR loci are highly present in cyanobacterial genomes (Shih et al., 2013; Hou et al., 2019), where they could be a barrier to transformation processes.

In addition, Limnospira has physical barriers that could prevent DNA uptake like the thick exopolysaccharide layer that may hinder DNA-cell and cell-cell contact while extracellular nucleases secreted by the host could degrade foreign DNA upon contact (Cao et al., 1996; Stucken et al., 2013).

This work deals with the recognition and characterization of some of the genetic barriers for transformation systems that operate in the strain L. platensis PCC 9108, concretely the R-M and CRISPR systems. Recently, several biotechnology-based techniques utilizing DNA methyltransferases have been used with several bacterial species other than Limnospira to improve transformation efficiency (Ren et al., 2022). Our data will also contribute to a better understanding of the biological role of DNA methylation in this cyanobacterium and will help to overcome the transformation step to develop an efficient gene transfer system in this organism.

2. Materials and methods

2.1. Bacterial strains, plasmids, and culture conditions

Strains and plasmids used in this study are listed in Table S1. L. platensis PCC 9108 was grown for 6–7 days and maintained in spirulina medium from University of Texas (USA) (here after referred as UTEX medium, https://utex.org/products/spirulina-medium) under controlled conditions on an illuminated shaker incubator (30 °C, 24 h light provided by cool, white lamps at a light intensity of 500 Lux and 120 rpm). For stock cultures, cells were transferred into new UTEX medium every week. Escherichia coli DH5α and E. coli C43 (DE3) were grown overnight either on solid or in liquid Lysogeny Broth medium at 37 °C and 250 rpm orbital shaking. Rhodococcus erythropolis CECT 3014 was grown at 30 °C in Lysogeny Broth medium for 2 days with shaking at 250 rpm and 15 μg/ml nalidixic acid. For plasmid selection, medium was supplemented with 50 μg/ml kanamycin, 100 μg/ml ampicillin or 34 μg/ml chloramphenicol (from Sigma-Aldrich) when appropriate.

2.2. Genome sequencing and analysis

L. platensis PCC 9108 genome was sequenced using Ion Torrent by ADM Lifesequencing (Paterna, Spain). An Ion Torrent Personal Genome Machine sequencer was used for single-end strategy with a coverture depth of 250X, resulting in 1293,308 reads with a median insert size of 343 bp. Newbler v2–7 (454 Life Sciences) software was used for assembling reads. The number of contigs delivered was 4157. To reduce the number of contigs we performed a second sequencing project using PacBio RSII technology by Macrogen (Seoul, South Korea) that rendered 103,730 reads with a median insert size of 11,603 bp. Genome data was processed with SMRT Analysis v2–3–0 software. The results from Ion Torrent and PacBio sequencing, were assembled de novo using Hierarchical Genome Assembly Process RS-HGAP-Assembly.3 (SMRT Pacific Biosciences v.2.3.0 (Chin et al., 2013)), SSPACE-LongRead (Boetzer and Pirovano, 2014) and LR_GapCloser (Xu et al., 2019). The final assembly contained 62 final contigs.

To further enhance the genome assembly, it was sequenced by Oxford Nanopore technology by SECUGEN S.L. (Madrid, Spain) and ran through the flow cell FLO-MIN-106D v R9 in a MinION equipment. The number of reads obtained was 63321 with a median read length of 3072 nt, a mean read length of 6059 nt, a read length N50 of 12708 nt, and a whole amount of sequence of 383.7 Mb. The longest read obtained had 130 kb. The assembly was performed filtering first the reads longer than 0.5 kb with the Filtlong tool (V0.2.0) (https://github.com/rrwick/Filtlong) and then using the Flye tool (flye (V2.9.1) (https://github.com/fenderglass/Flye) (Lin et al., 2016). The resulting assembly was manually curated and resulted in a single chromosome of 6763,969 bp. The genome project has been deposited at GenBank under the accession number CP066886.2.

The average nucleotide identity based on MUMmer (ANIm) was calculated using JSpeciesWS software (Richter et al., 2016). ANIm values were generated based on the NUCmer alignment for pairwise comparisons of closely related genomes.

2.3. Identification of putative MTase genes and CRISPR-Cas systems

RAST SERVER program (http://rast.nmpdr.org/) (Aziz et al., 2008) was used for searching R-M systems and orphan MTase genes in the L. platensis genome. Results were then contrasted using the BLASTP tool (non-putative and putative results) in the REBASE database (The Restriction Enzyme Database, http://rebase.neb.com) (Roberts et al., 2015) and BLASTP suite (Swiss-Prot and non-redundant database, excluding all the L. platensis entries) in the NCBI database, using a protein query.

In silico search of L. platensis genome for CRISPR-Cas systems was done with the CRISPRCasFinder program online of the Université Paris Sud (http://crispr.i2bc.paris-saclay.fr/Server/) (Grissa et al., 2007) with manual proofreading. When results comprised an evidence level more than 3 (as indicated in CRISPRfCasfinder) were considered as a positive locus. Secondary structures of direct repeat (DR) sequences were predicted on the RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). The search for similarities between each unique spacer was carried out by BLASTN against a database limited to RefSeq databases of Plasmids, Bacteria and Viruses, or the Env-nt Database at NCBI (https://www.ncbi.nlm.nih.gov/). The hits found within L. platensis CRISPR loci were removed. Hypothetical cas genes were found in close vicinity of these CRISPR putative sequences using RAST SERVER (http://rast.nmpdr.org/) (Aziz et al., 2008). The putative CRISPR-Cas

systems were classified according to (Makarova et al., 2015). The genes of interest were compared against the Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (Marchler-Bauer et al., 2005).

2.4. DNA extraction, restriction, digestion, and transformation processes

Chromosomal DNA extraction from L. platensis PCC 9108 was performed using the procedure previously published (Morin et al., 2010). DNA was suspended in water for following experiments.

Restriction digestions were performed according to the manufacturer's instructions: 10–15 U of the commercial enzyme was used to cleave up to 1 μg DNA in 20 μl total volume and incubated at the recommended temperature for 6 h for genomic DNA or 2 h for other samples. Enzymes used were: *Ava*III, *Eco*RV, *Esp*I, HgiCI, HgiDI, *Nsp*V, *Sac*II, *Tth*111I and *Pst*I from Takara Bio Inc.; *Ava*I, *Age*I, *Sna*BI, *Bsi*WI and *Nsp*I from New England Biolabs Inc. Digested DNAs were analyzed by 1% agarose (Condalab) gel electrophoresis on TAE buffer (40 mM Tris-acetate, 2 mM EDTA, pH 7.6) and further staining with gelRed (Biotium). The L. platensis PCC 9108 DNA cleavage patterns were compared to *Eco*RV-digestion control generated restriction pattern to determine digestion results as complete (c), partial (p) or resistant (r) to enzymatic digestions. *Eco*RV R-M system is not present in L. platensis PCC 9108 genome and then the *Eco*RV restriction enzyme cuts the DNA of this strain and for this work it has been taken as a positive-control digestion.

E. coli competent cells were prepared and transformed by standard protocols (Green and Sambrook, 2012) while Rhodococcus transformation procedure was done as previously described (Fernández de Las Heras et al., 2014). Other DNA techniques such as plasmid isolation, ligations and restriction analysis followed standard methods (Green and Sambrook, 2012).

2.5. MTase gene cloning

Putative MTase genes were amplified by PCR from L. platensis genomic DNA with Invitrogen™ Platinum™ SuperFi™ DNA polymerase and synthetic oligonucleotides provided by Fisher Scientific (Table S2). PCR amplicons were digested and ligated into the pET29a+ expression DNA vector (Novagen) using T4 DNA ligase (Takara Bio Inc.) (Fig. S1A). Ligation products were transformed in E. coli DH5α and recombinant plasmid DNAs were isolated from kanamycin-resistant transformants. The presence of inserts of the expected size was confirmed by restriction analysis. Plasmids were then transformed in E. coli C43(DE3) (Sigma), a strain that allows the overexpression of toxic and membrane proteins.

Some of the MTases were alternatively cloned in the pTIP-QC1 vector, an expression E. coli-Rhodococcus shuttle vector (Nakashima and Tamura, 2004). pET29a+ -MTase plasmid DNA containing the putative MTase genes were digested and ligated to restricted pTIP-QC1 plasmid DNA using T4-DNA-ligase. All methylase genes were then under the control of Thiostrepton inducible promoter (TipA) present in the pTIP-QC1 vector (Nakashima and Tamura, 2004) (Fig. S1B). Ligation products were transformed in E. coli DH5α and recombinant plasmid DNAs were isolated from ampicillin-resistant transformants. The presence of inserts of the expected size was confirmed by restriction analysis and sequencing. Then, plasmids were transformed in R. erythropolis CECT 3014 competent cells by electroporation to overexpress MTase proteins.

Alternatively, the MTase gene of the HgiCI R-M system was also cloned in pET29a+ or pTIP-QC1 using as host E. coli C43 (DE3) or R. erytropolis CECT3014 respectively and coupled to the P_{lac} promoter in pUC26 using as host Pseudomonas putida CECT324.

2.6. In vivo MTase assays

E. coli C43(DE3) carrying pET29a+ with the cloned MTase genes were grown on Lysogeny Broth media containing 50 μg/ml kanamycin at 37 °C and 250 rpm orbital shaking up to $OD_{600\ nm}$ 0.6 for induction with 1 mM of isopropyl-β-D-thiogalactopyranoside (IPTG). Induced cultures were grown at 37 °C for 2 h more, and then the E. coli cells were collected by centrifugation. Plasmid DNA was isolated from E. coli C43 (DE3) cells and the presence or absence of specific methylation was confirmed by incubation of the plasmid DNA with the corresponding restriction endonuclease for 2 h, separation by agarose gel electrophoresis and visualization by Green Safe (Nzytech) staining. Unmethylated pET29a+ served as a control for restriction. When no suitable restriction sites were present in the pET29 construct, restriction sites for the predicted cleavage pattern were introduced into the pGEMT plasmid (Promega) using a PCR fragment amplified from L. platensis genomic DNA (Fig. S1C). Then, we co-transformed host cells with two plasmids for in vivo MTase assay.

The in vivo MTase assay in R. erythropolis was done as follows. R. erythropolis CECT3014 carrying pTIP-QC1 with the cloned MTase genes were grown on Lysogeny Broth medium containing 34 μg/ml chloramphenicol at 30 °C and 250 rpm orbital shaking up to $OD_{600\ nm}$ 0.6 for induction with 1 μg/ml of Thiostrepton (Sigma). Induced cultures were grown at 30 °C for 24 h more, and then cells were collected by centrifugation. Genomic DNA was isolated from R. erythropolis CECT3014 cells as previously described (Fernández de Las Heras et al., 2011) and the presence or absence of specific methylation was determined by digesting the genomic DNA with the corresponding restriction enzyme overnight and further agarose gel electrophoresis. Unmethylated genomic DNA from empty-pTIP-QC1 R. erythropolis CECT3014 strain served as a control for restriction. Protection of DNA against cleavage indicates efficient methylation.

2.7. Transcriptomic analysis

Total RNA was extracted with Direct-zol™ RNA MiniPrep Zymo Research Kit (Zymo Research). Three biological replicates were used to sequence the total RNA. The de novo transcriptome sequencing was performed by Macrogen (Seoul, South Korea) NGS Service (Illumina TruSeq RNA library, 6 GB/sample sequencing coverage) and fragments of 151 bp paired-end reads were obtained. Raw reads were trimmed and cleaned with Trimmomatic 0.39 to remove Illumina adapters and low-quality bases (Bolger et al., 2014). Trimmed reads were aligned to the L. platensis PCC 9108 genome (CP066886.2) using the STAR 2.7.3a. program (Dobin et al., 2013), followed by a read count with the featuresCount function of the R Bioconductor package Rsubread (Liao et al., 2019), both splitting multireads equally and ignoring them. Counts were afterwards converted to TPM (transcripts per kilobase million) to make samples comparable. Transcriptomic data obtained in this work can be found at SRR24757814 (NCBI database).

2.8. Methylome analysis

To develop a map of the methylated sites in the genome of L. platensis, Oxford Nanopore reads, fast5 files, were analyzed by SECUGEN S. L (Madrid, Spain). with TOMBO (https://github.com/nanoporetech/tombo) (Rand et al., 2017). Geneious Prime 2022 (https://www.geneious.com) was used for methylation sequence comparisons and further analysis. Each recognition sequence was identified in its respective TOMBO fasta file (5mC or 6mC). Reads containing these sequences were selected and those with overlapping regions (mostly on opposite strands) were merged, to count each recognition sequence just once. This process was performed using Biostrings of the R software version 4.2.0 (R Core Team, https://bioconductor.org/packages/Biostrings).

3. Results

3.1. Genome of L. platensis PCC 9108

L. platensis PCC 9108 (referred as PCC 9108 from now on in the text) was selected for this study given that among other strains tested in the laboratory, this strain was axenic and showed robust growth in minimal culture media. Then, the genome of PCC 9108 was sequenced and annotated (CP066886.2) yielding a single contig of 6763,969 bp and 6287 coding genes with a 44.2% GC content. This strain does not appear to contain plasmids.

There are 10 genomes sequenced of L. platensis in the NCIB GenBank-databases: GCA_025200965.1 and GCA_000307915.1 strains C1; GCA_000175415.3 strain Paraca; GCA_001611905.1 strain YZ; GCA_000210375.1 strain NIES-39; GCA_009176225.1 strain NIES-46; GCA_014698675.1 strain FACHB-439; GCA_014698385.1 strain FACHB-971; GCA_014698815.1 strain FACHB-835 and FO818640.1 strain PCC 8005). The genomes of these strains have a medium size of 6.14 Mb, a median GC content of 44% and around an average of 5196 coding genes with no trace of plasmid sequences. The genome of PCC 9108 displays a higher similarity with NIES-39 strain (6.78 Mb and 5924 coding genes) than with the rest of L. platensis genomes sequenced so far.

3.2. CRISPR-Cas systems in L. platensis PCC 9108

CRISPR-Cas systems have been found in most of sequenced cyanobacterial genomes, including L. platensis strains (Cai et al., 2013; Hou et al., 2019; Jungblut et al., 2021). In silico analysis using the CRISPRCasFinder program showed up to six CRISPR arrays and five cas clusters on the PCC 9108 genome (Fig. 1, Table S3). Although this number of CRISPR arrays is similar to that found in other L. platensis strains (i.e., 5–10 clusters) (Lefort et al., 2014; Xu et al., 2016; Silas et al., 2017; Suzuki et al., 2019), other 29 possible CRISPR sites, with only 1 or 2 spacers were found widely distributed in the genome of PCC 9108 (Table S4).

We have identified 6 direct repeats (DRs) sequences (Table S3) in the PCC 9108 genome, consisting of partially palindromic sequences of 35–37 nt that could form putative RNA secondary structures (Fig. S2). The characterization of CRISPR arrays containing different types of DRs is shown in Table S3. The size of these CRISPR loci varied from ~0.40 kb (CRISPR1) to ~3.22 kb (CRISPR5), corresponding to 5–40 unique sequences (spacers). Some CRISPR loci with similar DR are located next to each other (e.g., CRISPR1 and CRISPR2) or in different parts of the genome (e.g., CRISPR3 and CRISPR4), configuring larger genome CRISPR arrays (Table S3). On the other hand, although the number and sequence of spacers is different, we observed that the DR from CRISPR4

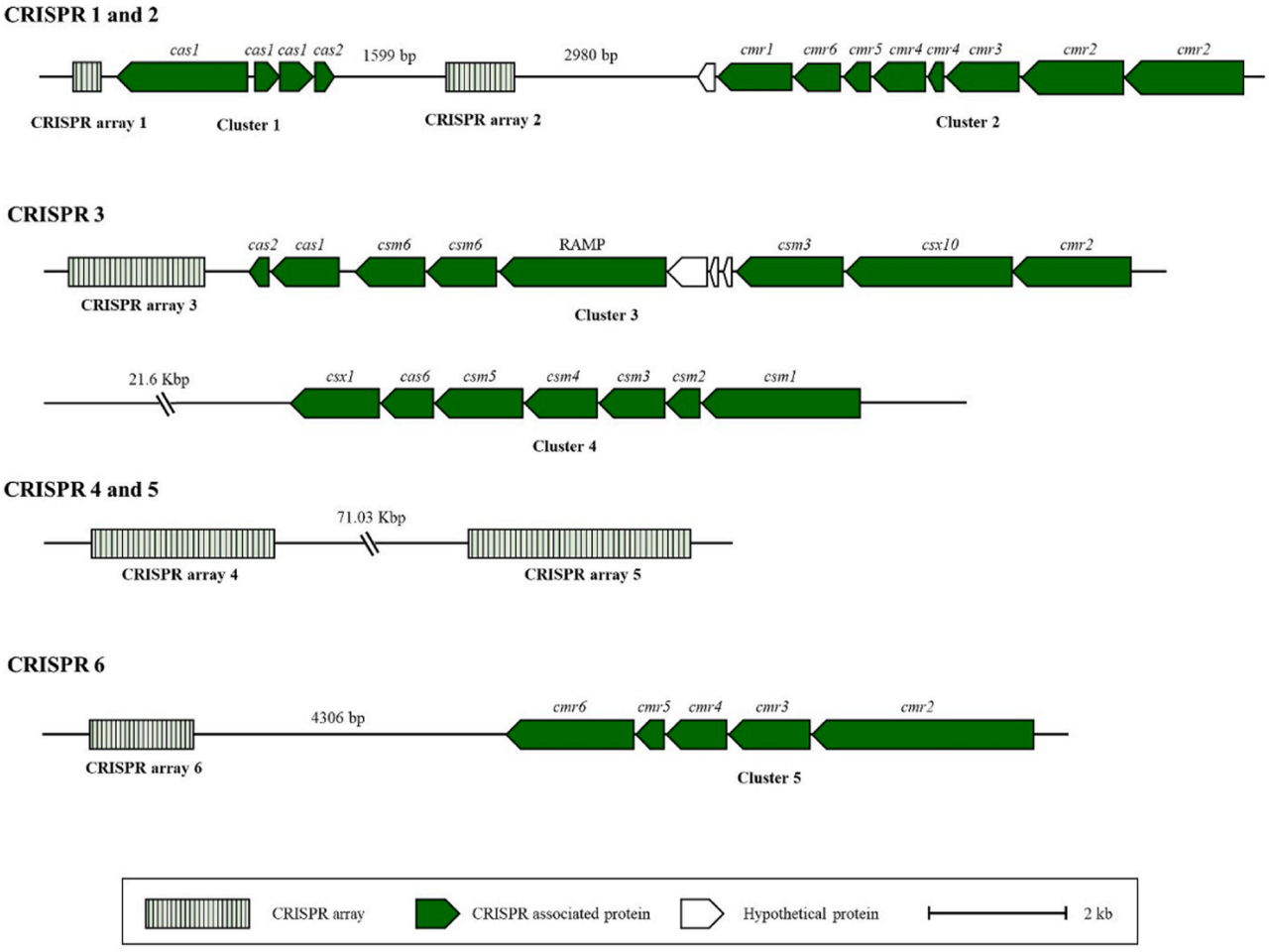

Fig. 1. Structures of CRISPR-Cas systems identified in L. platensis PCC 9108 genome. CRISPR-Cas systems consist of a CRISPR array (two and several hundred repeating sequences of DNA of 25–35 bases separated by unique spacers genetic memories of past invaders of 30–40 bases) and a cluster of cas genes that are organized in one or more operons (Shmakov et al., 2020). In silico analysis of L. platensis PCC 9108 genome showed the presence of six CRISPR arrays and five sets of cas gene clusters. Cas, CRISPR-associated protein; Cmr, effector complex effector of the Type III CRISPR system that binds to target RNA; RAMP, repeat-associated mysterious proteins involved in the processing of CRISPR loci encoded crRNA and silencing of invading RNA molecules; Csm, a type III-A CRISPR-Cas interference complex; Csx, a protein containing the CRISPR-associated Rossmann fold (CARF) domain gene for which no clear link to a particular CRISPR subtype has been established.

is the complementary sequence to DR from CRISPR5. These interruptions could be caused due to the insertion of a long repetitive sequence that results in a poor assembly. Some of these repetitive sequences are IS (insertion sequence) elements that are reported to be associated to CRISPR arrays (Horvath et al., 2009; Kuno et al., 2012) and have been detected in PCC 9108. Even though the few mismatches between some of DRs (e.g., from CRISPR1 and CRISPR2, or CRISPR3 and CRISPR4), the structure shows some differences regarding the length of the stem-loop (Fig. S2). The stem-loop in DR highlights its importance for the functionality of CRISPR-Cas systems, but only DR from CRISPR5 contains a conserved GAAAC sequence located at 3′, that has been reported to possibly act as a binding site for Cas proteins (Kunin et al., 2007). Additionally, a total of 137 unique spacers (from 32 to 54 nt) were identified in all CRISPR loci of the PCC 9108 genome (Table S5).

Subsequently, we searched for genes encoding CRISPR-associated proteins (Cas) in the vicinity of each CRISPR locus using RAST SERVER and checked the presence of conserved domains using the Conserved Domain Database (CDD). Table 1 shows the 36 cas genes identified, 30 of them grouped in 4 cas gene clusters. CRISPR1, CRISPR2, CRISPR3 and CRISPR6 have cas genes nearby; however, an incomplete CRISPR-Cas system was observed in CRISPR4 and CRISPR5 (Fig. 1). A set of cas1 and cas2 was identified between CRISPR1 and CRISPR2, and a cluster cmr2-cmr3-cmr4-cmr5-cmr6 3 kb away from CRISPR2. We have also found a gene cluster composed by csm1-csm2-csm3-csm4-csm5-cas6-csx1 which do not have a CRISPR array close by, similarly to L. platensis YZ, NIES-39, C1 and PCC 8005 strains (Xu et al., 2016). However, in PCC 9108 at 21.6 kb away, CRISPR3 is surrounded by a group of cas genes: a RAMP (Repeat-Associated Mystery Protein) ORF (open reading frame), two csm6, one cas6, one cas1 and two cas2. Finally, a cluster cmr2-cmr3-cmr4-cmr5-cmr6-cas1 seems to be associated with CRISPR6 (Fig. 1). PCC 9108 presents four cas1 genes and two cas genes, similarly to L. platensis NIES-39, although L. platensis CS328, Paraca and PCC 8005 strains present three cas1 genes and two cas genes (Cai et al., 2013).

We have found clusters of subtypes III-A (Csm module) and III-B (Cmr module). Subtype III-A often possess cas1, cas2 and cas6 and most of the III-B systems lack these genes (Makarova and Koonin, 2015). Cluster 2, subtype III-B, (Table 1) lacks genes cas1, cas2 and cas6, and seems to depend on other CRISPR-Cas systems present nearby. In the other clusters, cas1, cas2 or cas6 genes are present. Moreover, a gene encoding a cas6 without a CRISPR array or other cas was detected in PCC 9108 genome (Table 1). These results show the complexity of the organization of CRISPR-Cas systems.

3.3. R-M systems in L. platensis PCC 9108

R-M systems are found in most of the sequenced bacterial genomes (Loenen and Raleigh, 2014; Roberts et al., 2015) and could represent a significant barrier to horizontal gene transfer. Depth analysis of PCC 9108 genome using REBASE database showed the presence of several R-M systems. We have identified four complete sets of type I R-M systems (hsdMSR) (Table 2). We have found two solitary type I methylase

Table 1
CRISPR/Cas system-associated proteins found in the genome of L. platensis PCC 9108.

Locus	Cas	Description	Accession (CDD)	Length (kb)	Cas Type
Cluster 1, 4 genes					
AP9108_13660	Cas1	CRISPR associated protein Cas1	pfam01867	1.800	I-D
AP9108_34470	Cas1	CRISPR-associated endonuclease Cas1	-	0.360	
AP9108_34475	Cas1	CRISPR/Cas system-associated protein Cas1	cl00656	0.483	
AP9108_13670	Cas2	CRISPR/Cas system-associated protein Cas2	cd09725	0.282	
Cluster 2, 8 genes					
AP9108_34490	Cmr1	CRISPR/Cas system-associated RAMP superfamily protein.	cl43403	1.080	III-B
AP9108_34495	Cmr6	CRISPR/Cas system-associated RAMP superfamily protein	cd09726	0.657	
AP9108_13720	Cmr5	-	-	0.396	
AP9108_13725	Cmr4	CRISPR type III-B/RAMP module RAMP protein Cmr4 (TIGR02580)	cl21471	0.789	
AP9108_13730	Cmr4	CRISPR/Cas system-associated RAMP superfamily protein	cd09726	0.225	
AP9108_13735	Cmr3	CRISPR/Cas system-associated RAMP superfamily protein Cmr3 III-B (cd09748)	cl21471	1.080	
AP9108_34505	Cmr2	CRISPR-associated protein	pfam12469	1.467	
AP9108_34500	Cmr2	CRISPR/Cas system-associated protein Cas10	cl21742	1.722	
Cluster 3, 8 genes					
AP9108_17030	Cas2	CRISPR/Cas system-associated protein Cas2	cd09725	0.285	III-A+ III-B
AP9108_17035	Cas1	CRISPR associated protein Cas1	pfam01867	0.975	
AP9108_17040	Csm6	CRISPR/Cas system-associated protein Csm6_III-A	cd09742	1.101	
AP9108_17045	Csm6	CRISPR/Cas system-associated protein Csm6_III-A	cd09742	1.125	
AP9108_17050	RAMP	CRISPR/Cas system-associated RAMP superfamily protein (TIGR03986)	cl21471	2.406	
AP9108_17070	Csm3	CRISPR/Cas system-associated RAMP superfamily protein Csm3 III-A (cd09683)	cl21471	1.551	
AP9108_17075	Csx10	CRISPR-associated RAMP protein, Csx10 family (TIGR02674). Restricted to cyanobacteria	cl21471	2.427	
AP9108_17080	Cmr2	CRISPR/Cas system-associated protein Cas10 (COG1353)	cl34236	1.722	
Cluster 4, 7 genes					
AP9108_17170	Csx1	CRISPR-associated (Cas) DxTHG family, Csx1	pfam09455	1.278	III-A
AP9108_17175	Cas6	Class 1 type III CRISPR-associated endoribonuclease Cas6	cd21141	0.759	
AP9108_17180	Csm5	CRISPR/Cas system-associated RAMP superfamily protein Csm5 III-A (cd09662)	cl21471	1.269	
AP9108_17185	Csm4	CRISPR/Cas system-associated RAMP superfamily protein Csm4 III-A Superfamily	cl23831	1.041	
AP9108_17190	Csm3	CRISPR type III-A/MTUBE-associated RAMP protein Csm3	TIGR02582	0.945	
AP9108_17195	Csm2	Csm2 Type III-A	pfam03750	0.408	
AP9108_17200	Csm1	Type III-A CRISPR-associated protein Cas10/Csm1	-	2.310	
Cluster 5, 6 genes					
AP9108_24945	Cmr6	CRISPR type III-B/RAMP module RAMP protein Cmr6 (TIGR01898)	cl21471	1.863	III-B
AP9108_24950	Cmr5	-	-	0.396	
AP9108_24955	Cmr4	CRISPR type III-B/RAMP module RAMP protein Cmr4 (TIGR02580)	cl21471	0.876	
AP9108_24960	Cmr3	CRISPR/Cas system-associated RAMP superfamily protein Cmr3 III-B (cd09748)	cl21471	1.179	
AP9108_24965	Cmr2	CRISPR-associated protein Cas10/Cmr2, subtype III-B (TIGR02577)	cl30290	3.213	
Other cas					
AP9108_24120	Csx3	CRISPR-associated protein Cas csx3	pfam09620	0.930	III
AP9108_20955	Cas6	Class 1 type III CRISPR-associated endoribonuclease Cas6	cd21141	1.128	III

The genes of interest were compared against the Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (Marchler-Bauer et al., 2005). - = Conserved Domain not found.

Table 2
Type I R-M systems found in L. platensis PCC 9108.

	Gene name	Homology in REBASE	Hypothetical recognition sequence	Modification base	Homology NCBI Blastp Database non-redundant (excluding L. platensis) or swissprot	Locus tag Protein accession number	Presence in other L. platensis strains[1]
Complete system: endonuclease and methylase							
System I	**hsdM**	M.TdeDI (676 aa, 64%) (11310)	GCAYNNNNNNCATC	N6A (u.b.)	Type I restriction-modification system subunit M [Oscillatoria acuminata PCC 6304] AFY80819.1; 679 aa; 656/681(96%)	AP9108_02990 QQW29826.1 (681 aa)	All strains
	hsdS	S.CspFWC2II (447 aa, 32%) (618739)			Restriction endonuclease subunit S [Limnospira fusiformis KN01] ULB45797.1; 422 aa; 253/343(74%)	AP9108_02985 QQW29825.1 (392 aa)	NIES-39, YZ, Paraca
	hsdR	SplZ (1013 aa, 100%) (11612)	Unknown (c.s.r.)		Helicase, type I site-specific restriction-modification system restriction subunit [Oscillatoria acuminata PCC 6304] AFY80823.1; 1013 aa; 971/1013(96%)	AP9108_02970 QQW29823.1 (1013 aa)	All strains
System II	**hsdM**	M.Bgl188I (518 aa, 86%) (253836)	C**m6**ACNNNNNNTAYG	N6A	Type I restriction enzyme EcoprrI methylase subunit [Escherichia coli] Q47163.1; 520 aa; 422/513(82%)	AP9108_19130 QQW27368.1 (513 aa)	All strains
	hsdS1	S.Kox9149II (409 aa, 49%) (381603)			Type I restriction enzyme EcoR124II specificity subunit [Escherichia coli] P10485.2; 404 aa; 56/173(32%)	AP9108_19120 QQW27366.1 (252 aa)	NIES-39
	hsdS2	S.Vmi33654I (421 aa, 31%) (138749)			Type I restriction enzyme EcoR124II specificity subunit [Escherichia coli] P10485.2; 404 aa; 101/250(40%)	AP9108_19125 QQW27367.1 (166 aa)	None
	hsdR	NgoAV-1 (1032 aa, 41%) (10879) EcoR124II (1033 aa, 32%) (989)	GCANNNNNNNNTGC (c.s.r.) GAANNNNNNNRTCG (c.s.r.)		Type I restriction enzyme EcoR124II endonuclease subunit [Escherichia coli] P10486.1; 1033 aa; 333/442(75%) P10486.1; 1033 aa; 445/561(79%)	AP9108_35190 WAK73726.1 (559 aa) AP9108_35185 WAK73725.1 (448 aa)	NIES-39
System III	**hsdM**	M.CspNS6III (482 aa; 65%) (127803)	CRA**m6**ANNNNNNNGTCY	N6A	Type I restriction enzyme EcoEI methylase subunit [Escherichia coli] Q47282.1; 490 aa; 163/506(32%)	AP9108_15455 QQW31720.1 (504 aa)	C1, YZ
	hsdS	S.Hin2866IV (437 aa, 17%) (28524)			Type I restriction enzyme EcoDI specificity subunit [Escherichia coli] P06991.1; 444 aa; 77/184(42%)	AP9108_15460 QQW32057.2 (395 aa)	YZ
	hsdR	FtnUII (782 aa, 35%) (13980)	CYYANNNNNNNCTC (c.s.r.)		Type I restriction enzyme EcoEI endonuclease subunit [Escherichia coli] Q47281.1; 813 aa; 265/794(33%)	AP9108_15465 QQW31721.1 (774 aa)	C1, YZ
System IV	**hsdM**	M.Lsp6406VII (809 aa, 72%) (53023)	GG**m6**AGNNNNNCTC	N6A	Type I restriction enzyme EcoR124II methylase subunit [Escherichia coli] P10484.1; 520 aa; 165/504(33%)	AP9108_33250 WAK74298.1 (814 aa)	NIES-39, Paraca
	hsdS	S.AspNIH1IV (442aa, 41%) (232213)			Hypothetical protein [Nitrosomonas europaea] MBV6388509.1; 416 aa; 268/404(66%)	AP9108_04995 QQW30145.1 (417 aa)	YZ
	Mtase	M.TspPSB1IX (412 aa, 32%) (428420)	CCTNAGG	N4	DNA adenine methylase [Arthrospira sp. SH-MAG29] MBS0016857.1; 397aa; 391/397(98%)	AP9108_05000 (402 aa)	NIES-39, YZ, Paraca
	hsdR	EcoO127I (1031 aa, 46%) (19328)	Undetermined (c.s.r.)		HsdR family type I site-specific deoxyribonuclease [Arthrospira sp. PLM2. Bin9] TVU53307.1; 1027 aa; 671/700(96%)	AP9108_05015 WAK74660.1 (704 aa)	All strains
Incomplete system: only methylase							
	hsdM	M.TspPSB1VII (525 aa, 24%) (90496)	GTG**m6**ANNNNNNTGC	N6A	Type I restriction enzyme MjaIX methylase subunit [Methanocaldococcus jannaschii DSM 2661] Q60297.1; 558 aa; 72/186(39%)	n.a-RS (214 aa)	NIES-39, Paraca
	hsdM	M.Hsp6II (571 aa, 24%) (243140)	GC**m6**AANNNNNTCC	N6A	Type I restriction enzyme HindI methylase subunit [Haemophilus influenzae Rd KW20] Q57168.1; 443 aa; 141/319(44%)	AP9108_33145 WAK74657.1 (43 aa) AP9108_33150 WAK74277.1 (110 aa) AP9108_33155	NIES-39

(*continued on next page*)

Table 2 (*continued*)

	Gene name	Homology in REBASE	Hypothetical recognition sequence	Modification base	Homology NCBI Blastp Database non-redundant (excluding L. platensis) or swissprot	Locus tag Protein accession number	Presence in other L. platensis strains[1]
Complete system: endonuclease and methylase							
						WAK74278.1 (181 aa)	
Incomplete system: only endonuclease							
	hsdR	Mcr10I (1067 aa; 16%) (555904)	GGCANNNNNNTCC (c.s.r.)		Type I restriction enzyme HindI endonuclease subunit [Haemophilus influenzae Rd KW20] O05052.1; 1055 aa; 130/258(50%)	AP9108_19165 QQW32109.2 (285 aa)	NIES-39, Paraca

1 Compared to the L. platensis Paraca, NIES-39, C1 and YZ strains with the largest number of R-M systems (Xu et al., 2016)
c.s.r.: cleavage sites are random; N6A: N6-methyladenosine; C5: 5-methylcytosine; N4: N4-methylcytosine; u.b.: unknown base; n.a-RS: not annotated but identified in RAST-SERVER.

subunits (hsdM) and one restriction subunit (hsdR) as well (Table 2).

Regarding, type II R-M systems L. platensis PCC 9108 genome contains at least 13 hypothetical genes (Table 3A) that are widely distributed among cyanobacteria and other prokaryotes (Houmard and de Marsac, 1988; Roberts et al., 2015). One of the genes was not found by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP), but it was identified as a methyltransferase by RAST-SERVER, and we manually proposed it as M.*Nsp*V (pb 6513,369 -> 6511,675 from CP066886.2). In addition to R-M systems, 8 MTases with non-identifiable endonucleases were found in PCC 9108 and 6 R-M systems formed by one single enzyme having putatively both capabilities (Table 3B).

The arrangement of the hypothetical type II R–M systems genes found in PCC 9108 genome is shown in Fig. 2. The 13 systems are composed of a MTase and a REase except for two systems (system *Sna*BI and system *Nsp*I) that present an extra ORF. This gene codes a specialized transcription regulatory protein called C (controller) protein, that is described to have a regulatory role in other R-M systems (Heidmann et al., 1989; Ives et al., 1992; Rimšeliené et al., 1995; Anton et al., 1997; Vijesurier et al., 2000; Cesnaviciene et al., 2003; Zheleznaya et al., 2003).

3.4. MTase activity in L. platensis PCC 9108

Once the type II R-M systems were analyzed in silico, we verified the functionality of the MTases of the 13 R-M systems. To probe DNA protection against these restriction systems, genomic DNA from L. platensis PCC 9108 was treated with commercial restriction endonucleases equivalent to the type II REases found in the genome, and further analyzed by electrophoresis. Digestion with commercial REase *Eco*RV, that is not found within the PCC 9108 R-M systems, was used as a digestion control. We observed that PCC 9108 genome was completely protected from digestion by the probed enzymes with the only exception of *Ava*I (Fig. 3A) indicating that DNA was not fully methylated and sufficiently protected. On the other hand, L. platensis PCC 9108 genome was completely protected against the other enzymes used (Fig. 3A).

To verify the MTase activity of each potential MTase encoding gene, the genes were cloned into the expression vector pET29a+ to carry out an in vivo MTase assay. The methylation activity was determined by digesting the plasmid, isolated from cells containing the relevant MTase gene, with the corresponding commercial endonuclease (Fig. 3B). Six out of the twelve cloned MTases were functional in E. coli, showing DNA protection from restriction digestion by the specific REase that corresponds to each particular MTase: M.*Ava*III, M.*Esp*I, M.*Tth*111I, M.*Sna*BI, M.*Bsi*WI and M.*Pst*I. On the other hand, DNA methyltransferases M.*Ava*I, M.HgiDI, M.*Nsp*I, M.*Nsp*V, M.*Age*I and M.*Sac*II expressed on plasmid pET29a+ were not fully functional (Fig. 3B).

Non-active MTases were then expressed in Rhodococcus to probe its activity using pTIP-QC1 as an expression vector. After the MTases overexpression in R. erythropolis CECT3014 the gDNA was isolated to test the restriction enzymes. The results confirmed the positive activity of all cloned MTases except M.*Ava*I (Fig. 3C). The smear observed for MTase *Ava*I-like from L. platensis PCC 9108 suggests that this enzyme could have a low activity and therefore the DNA could not be fully methylated yielding a partial digestion. Further experiments should be done to verify the function of this R-M system.

The results show that twelve out of thirteen functional methyltransferases were expressed in heterologous hosts. Only MTase of HgiCI system was not functional in any of the 3 different expression systems (pET29a+ in E. coli C43, pTIP-QC1 in R. erytropolis CECT3014 and pUC26 in Pseudomonas putida CECT324) (data not shown).

3.5. Whole-genome methylation pattern of L. platensis PCC 9108

Once the MTases of PCC 9108 were characterized, we proceeded to study its methylome by MinION sequencing to establish the global methylation pattern in this strain. This method allows the direct detection of N6-adenine and C5-cytosine methylation, that represents the genome methylation from 8 out of 13 type II R-M analyzed in PCC 9108 (see Table 4 and Fig. S2). As a result, in total, there are 63,824 mC sites from 1503,924 C sites, which represents a 4.2% of methylated cytosines, in the PCC 9108 genome. From these, the 44.5% are CG, 13.1% are CHG and 42.5% are CHH (in which H could be A, C or T). Similarly, the 3.5% of the adenosines are methylated (65,102 out of 1881,535), being 25.9% in the context AG, 9.8% AHG and 64.3% AHH. On the other hand, the 92.7% and the 94.2% of the gene bodies (from the start to stop codon) present 5mC and 6 mA respectively, more than the 84.4% and 81.2% of 5mC and 6 mA found in intergenic regions.

There are 1824 putative recognition sites for the 8 MTases of the R-M system type II analyzed in the genome of PCC 9108, and around 85% of them (1557 sites) are methylated. The methylation of those sites varies depending on the methylase, from 23.7% of M. *Nsp*I to 96.6% of M. HgiCI. The most common recognition site by each MTase and the number of methylated sites by each one is shown in Table 4. Of the total methylated sites, 48% are due to M. HgiCI/*Ban*I followed by M.*Age*I (21,8%) and M. AplI (10.8%).

3.6. Expression analysis of R-M and CRISPR-Cas systems

RNA-seq analysis was performed to evaluate the relative abundance of mRNA of the genes involved in the genetic barriers (mainly R-M and CRISPR-Cas systems) of PCC 9108. The relative expression profile of R-M system type I genes from PCC 9108 is depicted in Fig. 4A. As expected, the complete type I R-M systems were expressed: System I (from 5.3 to 8.1-fold), system II (from 0.7 to 15.8-fold), system III (from 8.0 to 40.3-fold) and system IV (from 1.6 to 4.1-fold). The hsdM gene was the most expressed (from 2.6-fold to 40.3-fold) and the hsdR gene the least expressed (from 0.7 to 10.9-fold). In contrast, RT-PCR of hsdM gene of system III was null according to Zhao et al. (2006), while the transcriptome showed high expression (around 40.3-fold) for this gene. The Mtase gene of the first incomplete system (Table 2) was the most

Table 3A
Complete type II R–M systems found in L. platensis PCC 9108.

System	Prototype in REBASE (Reference)	Homology in REBASE	Hypothetical recognition sequence	Modification base	Homology NCBI Blastp Database non-redundant (excluding L. platensis) or Swiss-Prot	Locus tag Protein accession number	Presence in other L. platensis strains[1]
Complete system: endonuclease and methylase							
I	M.*Ava*I (3287)	M.LinBM01 (486 aa, 18%) (591318)	CYCGRG	N4 (u.b.)	Type II methyltransferase M.*Ava*I [Nostoc sp. PCC 7120] P0A461.1; 482 aa; 36/77(47%)	AP9108_03600 QQW29926.1 (100 aa)	All strains
	*Ava*I (165)	LinBM01 (318 aa, 12%) (591317)	C^YCGRG		Type II restriction enzyme *Ava*I [Nostoc sp. PCC 7120] P0A457.1; 315 aa; 28/51(55%)	AP9108_03605 QQW29927.1 (65 aa)	All strains
II	M.*Ava*III (4673)	M.*Ava*III (295 aa, 73%)	ATGC**A**^{m6}T	N6A	DNA adenine methyltransferase YhdJ; M.EcoKII [Escherichia coli K-12] P28638.3; 294 aa; 157/269(58%)	AP9108_30655 QQW29081.1 (295 aa)	All strains
	*Ava*III (167)	*Ava*III (381 aa, 41%)	ATGCAT (u.c.)		Conserved hypothetical protein [Limnospira maxima CS-328] EDZ95504.1; 232 aa; 231/232(99%)	AP9108_30660 QQW29082.1 (232 aa)	NIES-39, C1
III	M.*Nsp*V (3475)	M.*Nsp*V (440 aa, 40%)	TTCGAA	N4 or N6A (u.b.)	Type II methyltransferase M.*Nsp*V [Nostoc sp. PCC 7524] P35678.2 (480 aa); 260/469(55%)	n.a-RS 6513,369 -> 6511,675 pb (564 aa)	All strains
	*Nsp*V (1407)	*Nsp*V (220 aa, 39%)	TT^CGAA		Type II restriction enzyme *Nsp*V [Nostoc sp. PCC 7524] P35677.2; 220 aa; 115/143(80%)	AP9108_31230 QQW32282.2 (145 aa)	NIES-39, C1
IV	M.*Sna*BI (3758)	M.*Sna*BI (228 aa, 69%)	TA**C**^{m4}GTA	N4	Hypothetical protein [Symploca sp. SIO2E9] NEP00966.1; 231 aa; 183/218(84%)	AP9108_33210 WAK74288.1 (230 aa)	NIES-39, Paraca
	*Sna*BI (1707)	*Sna*BI (223 aa, 57%)	TAC^GTA		Hypothetical protein [Arthrospira sp. SH-MAG29] MBS0016888.1; 221 aa; 221/221 (100%)	AP9108_04405 QQW30041.1 (221aa)	NIES-39, YZ
	C.*Sna*BIP (3816)	C.*Bcl*I (77 aa, 37%) (4590)			Control protein C.*Sma*I [Serratia marcescens] P14307.1; 84 aa; 24/65(37%)	AP9108_04410 QQW30042.1 (79 aa)	All strains
V	M.*Age*I (3276)	M. *Age*I (429 aa, 52%)	AC**C**^{m5}GGT	C5	Type II methyltransferase M.*Age*I [Thalassobius gelatinovorus] P94147.1; 429 aa; 239/428(56%)	AP9108_27710 QQW28652.1 (451 aa)	NIES-39, YZ
	*Age*I (42)	*Age*I (278 aa, 58%)	A^CCGGT		Type II restriction enzyme *Age*I [Thalassobius gelatinovorus] Q9KHV6.3; 278 aa; 167/275(61%)	AP9108_27705 QQW28651.1 (276 aa)	NIES-39, YZ, Paraca
VI	M.*Esp*I (11623)	M. BlpI (391 aa, 56%) (3756)	GCTNAGC	C5 (u.b.)	Type II methyltransferase M.*Dde*I [Desulfomicrobium norvegicum] P05302.1; 415 aa; 174/386(45%)	AP9108_20450 QQW32123.2 (419 aa)	NIES-39, YZ, Paraca
	*Esp*I (1036)	AplYZORF18635P (289 aa, 100%) (142255)	GCTNAGC		Bpu10I family restriction endonuclease [Arthrospira sp. PLM2. Bin9] TVU52558.1; 289 aa; 286/289(99%)	AP9108_20455 QQW27574.1 (289 aa)	NIES-39, YZ, Paraca
VII	M.*Sac*II (3490)	M.*Ngo*MIII (377 aa, 49%) (3615)	C**C**^{m5}GCGG	C5	DNA cytosine methyltransferase [Arthrospira sp. SH-MAG29] MBS0016931.1; 384 aa; 349/382 (91%)	AP9108_02700 QQW29782.1 (382 aa)	NIES-39, YZ, Paraca
	*Sac*II (1579)	*Ngo*MIII (213 aa, 51%) (3155)	CCGC^GG		Eco29kI family restriction endonuclease [Arthrospira sp. SH-MAG29] MBS0016930.1; 214 aa; 193/214 (90%)	AP9108_02695 QQW29781.1 (214aa)	NIES-39, YZ, Paraca
VIII	M.SplI/ M. *Bsi*WI (3309)	M.Bgl346II (380 aa, 59%) (630499)	CGTA**C**^{m4}G	N4	N-4 cytosine-specific methyltransferase *Sma*I [Serratia marcescens] P14230.1; 292 aa; 74/202(37%)	AP9108_04100 WAK74270.1 (375 aa)	NIES-39, C1
	SplI (1724)	SplZORFDP (270 aa, 100%) (11620)	C^GTACG		Hypothetical protein HFV01_21810 [Limnospira fusiformis SAG 85.79] QJB27931.1; 284 aa; 257/269(96%)	AP9108_04105 WAK74271.1 (270 aa)	All strains
IX	M. HgiCI (3415)	M. HgiCI (420 aa, 53%)	GGYR**C**^{m5}C	C5	Type II methyltransferase M.*Ban*I [Aneurinibacillus aneurinilyticus] P19888.1; 428 aa; 235/413(57%)	AP9108_16900 QQW27034.1 (429 aa)	All strains
	HgiCI (1099)	HgiCI (345 aa, 31%)	G^GYRCC		Type II restriction enzyme *Ban*I [Aneurinibacillus aneurinilyticus] P19887.2; 354 aa; 165/352(47%)	AP9108_16905 QQW27035.1 (348 aa)	All strains
X	M. *Tth*111I (4192)	M.PspABC1III (357 aa, 60%) (406983)	GA^{m6}CNNNGTC	N6A	Type II methyltransferase M.*Hin*fI [Haemophilus influenzae] P20590.1; 359 aa; 29/86(34%)	AP9108_24525 QQW28201.1 (374 aa)	All strains
	*Tth*111I	AplYZMrrP (484 aa, 100%) (142243)	Undetermined, (u.c.)		Restriction endonuclease [Arthrospira sp. PLM2. Bin9] TVU52444.1; 484 aa; 483/484(99%)	AP9108_24520 QQW28200.1 (484 aa)	NIES-39, C1

(*continued on next page*)

Table 3A (*continued*)

System	Prototype in REBASE (Reference)	Homology in REBASE	Hypothetical recognition sequence	Modification base	Homology NCBI Blastp Database non-redundant (excluding L. platensis) or Swiss-Prot	Locus tag Protein accession number	Presence in other L. platensis strains[1]
Complete system: endonuclease and methylase							
XI	M. HgiDI (3417)	M.HindV (304 aa, 55%) (3574)	GRCGYC	C5 (u.b.)	Type II methyltransferase M.HindV [Haemophilus influenzae Rd KW20] P45000.1; 304 aa; 174/304(57%)	AP9108_01515 QQW29611.1 (314 aa)	NIES-39, YZ, Paraca
	HgiDI (1102)	HgiDI (359 aa, 44%)	GR^CGYC		Probable type II restriction enzyme HindVP [Haemophilus influenzae Rd KW20] P44999.1; 334 aa; 169/353(48%)	AP9108_01510 QQW31854.2 (362 aa)	All strains
XII	M.*Pst*I (3483)	M.AplI (345 aa, 100%) (25961)	CTGC^{m5}AG	C5	Type II methyltransferase M.AplI [Arthrospira platensis NIES-39] D4ZX35.1, 345 aa; 345/345(100%)	AP9108_20760 QQW27622.1 (345 aa)	NIES-39, YZ, Paraca
	*Pst*I (1536)	AplI (324 aa, 100%) (25962)	CTGCA^G		Type II restriction enzyme AplI [Arthrospira platensis NIES-39] D4ZX34.1; 324 aa; 324/324(100%)	AP9108_20755 QQW27621.1 (324 aa)	All strains
XIII	M. *Nsp*I (3473)	M.NspHI (397 aa, 68%) (3472)	RC^{m5}ATGY	C5	Type II methyltransferase M.*Hph*IA [Haemophilus parahaemolyticus] P50192.1; 372 aa; 124/327(38%)	AP9108_21810 QQW27785.1 (418 aa)	All strains
	*Nsp*I (1391)	*Nsp*I (244 aa, 50%)	RCATG^Y		Restriction endonuclease [Limnospira fusiformis SAG 85.79] QJB24666.1; 240 aa; 190/194(98%)	AP9108_21805 QQW27784.1 (194 aa)	All strains
	C.*Nsp*I	C.*Mun*I (74 aa, 29%) (3677)			*Bam*HI control element [Bacillus amyloliquefaciens] P23939.2; 81 aa; 29/62(47%)	AP9108_21800 QQW27783 (88 aa)	All strains

1 Compared to the *L. platensis* Paraca, NIES-39, C1 and YZ strains with the largest number of R-M systems (Xu et al., 2016). u.c.: unknown cleavage site; N6A: N6-methyladenosine; C5: 5-methylcytosine; N4: N4-methylcytosine; u.b.: unknown base; n.a-RS: not annotated but identified in RAST-SERVER.

expressed gene of the R-M-I systems (41.2-fold), even more than the Mtase gene of system IV (1.9-fold).

Many of the R-M systems in Anabaena sp. PCC 7120, L. platensis sp. or Helicobacter pylori J99 were found to be partially or completely inactive (Kong et al., 2000; Matveyev et al., 2001; Zhao et al., 2006). However, all Type II R-M systems of PCC 9108 are clearly expressed although at different levels. In those systems in which the MTase encoding gene is transcribed in the opposite direction to the restriction and regulatory genes, like in the *Nsp*I system (Fig. 2), the transcription profile shows similar levels for the latter (7.6 and 11.2-fold) and a higher expression for the former (47.5-fold). The single exception is *Sna*BI system, with 17.6-fold versus 23.5-fold, respectively. The systems in which the MTase and restriction genes are cotranscribed, provided similar expression levels between them (Fig. 4B): SplI, *Ava*I, AplI, *Esp*I, *Nsp*V, or *Sac*II systems. In other cases, in which both genes are cotranscribed, the fold-values for the expression of the MTase and the restriction genes were different: *Ava*III system 19.1-fold versus 11.7-fold, HgiDI system with 5.0-fold versus 17.0-fold and HgiCI system 3.9-fold versus 9.0-fold for the MTase and the restriction genes, respectively. These differences in the expression levels could be attributed to different promoters for each gene.

There is no apparent correlation between the number of putative methylation sites found in the PCC 9108 genome (Table 4) and the expression levels found in the transcriptome, but this can be explained by the fact that we are evaluating mRNA levels and no enzymatic activities. The restriction enzymes genes more expressed corresponds to SplI (32.7-fold)>*Sna*BI (23.5-fold)>HigiCI (17.0-fold) or *Ava*I (17.6-fold)>*Ava*III (11.7-fold)>*Nsp*I (7.5-fold).

Both regulators of R-M-systems appear cotranscribed to the restriction gene in the PCC 9108 genome. The expression of the *Sna*BI regulator is the highest (60.7-fold), quite different to the restriction gene nearby (23.5-fold). These proteins of 79 aa (QQW30042.1) or 88 aa (QQW27783.1) respectively, belong to the xenobiotic response element family transcriptional regulators. Therefore, these R-M regulatory proteins could be involved in delaying the expression of the endonuclease until the host DNA is appropriately methylated, as it has been suggested (McGeehan et al., 2005).

On the other hand, expression values of the main genes belonging to the five CRISPR-Cas clusters are shown in Fig. 4C. Cluster 1 showed the lowest expression. Cluster 2 and 3 genes also showed low expression level, from 0.1 to 11.7-fold and 1.7–7.6-fold respectively. Cluster 4 (Type IIIA) and cluster 5 (Type IIIB) displayed higher levels of expression (from 7.0 to 23.4-fold and from 4.4 to 57.3-fold respectively). The differences in the expression levels within the genes of the cluster suggest different promoters and/or regulation mechanisms. The highest induction was found in the cmr3 gene (57.3-fold induction, cluster 5) coding a non-catalytic subunit of a type III CRISPR-Cas system that facilitates immunity and avoids autoimmunity by coupling with the interaction between crRNA and its tag (Guo et al., 2019). In contrast, most of the genes within cluster 4, belonging to the CRISPR-Cas type IIIA system were induced (Fig. 4C). The higher expression of RNA/DNA targeting capabilities of Type III CRISPR-Cas systems may provide a versatile immune response against many different viruses, plasmids, and other mobile genetic elements.

Since L. platensis is naturally competent for transformation (Jester et al., 2022), we have also studied the expression of the core natural competence genes (Wendt and Pakrasi, 2019; Nies et al., 2022) (Fig. 4D). The two most conserved genes in subsection V cyanobacteria, PilT1, an ATPase responsible for pilus retraction, and the ubiquitous RecA (Nies et al., 2022), displayed values of 11.4-fold and 15.1-fold respectively. The highest induction was observed for the gene coding for PilQ (41.61-fold), a monomer that in combination with PilC, PilM, PilN, and PilO conforms the DNA uptake and processing machinery (Schirmacher et al., 2020). On the other hand, the genes that facilitate the transfer of DNA through the inner membrane into the cytoplasm (comEA, comEC, and comF), and that are not highly conserved within cyanobacteria (Nies et al., 2022), are poorly expressed in PCC 9108.

4. Discussion

The genome sequence of L. platensis PCC 9108 revealed the presence of a large number of CRISPR-Cas and R-M systems as it has been reported for other L. platensis strains (Cai et al., 2013; Hou et al., 2019; Jungblut et al., 2021). However, although the distribution of CRISPR loci in the genome of PCC 9108 is comparable to that of L. platensis NIES-39 and YZ strains (Fujisawa et al., 2010; Xu et al., 2016; Silas et al., 2017), the number of spacers and direct repeats differ in CRISPR3 from L. platensis YZ, and in CRISPR5 and CRISPR6 from L. platensis NIES-39.

Table 3B
Incomplete type II R–M systems found in L. platensis PCC 9108.

Prototype in REBASE (Reference) Only methylase	Homology in REBASE	Hypothetical recognition sequence	Modification base	Homology NCBI Blastp Database non-redundant (excluding L. platensis) or Swiss-Prot	Locus tag Protein accession number	Presence in other L. platensis strains[1]
M.*Mbo*I (1,2) (3667, 3668)	M.AflVII (278 aa, 59%) (163840)	GATC	N6A	Type II methyltransferase M.LlaDCHIA [Lactococcus cremoris] P50179.2; 284 aa; 127/272(47%)	AP9108_20865 QQW27640.1 (281 aa)	All strains
***Pvu*I** (1541)	M.Ssp6803I (424 aa, 58%) (3791)	^{m5}CGATCG	C5	Modification methylase XorII [Xanthomonas oryzae pv. oryzae KACC 10331] P52311.2 (424 aa); 188/418(45%)	AP9108_15705 QQW31762.1 (411 aa)	All strains
***Ava*II** (166)	M.HgiCII (437 aa, 40%) (3416)	GGWCC	C5 (u.b.)	Modification methylase M.HgiCII [Herpetosiphon aurantiacus] P25264.1; 437 aa; 194/428(45%)	AP9108_25595 QQW28355.1 (443 aa)	All strains
M.CsaIII (12253)	M.CsaIII (253 aa, 17%)	^{m6}AB S^{m6}AAM	N6A	DNA adenine methylase [Microcystis sp. 49638_E5] MCE2671800.1; 281 aa; 129/266(48%)	AP9108_17565 QQW27137.1 (290 aa)	NIES-39, YZ, Paraca
	M.CsaIII (253 aa, 19%)			DNA adenine methylase [Microcystis sp. 49638_E5] MCE2671800.1; 281 aa; 136/266(51%)	AP9108_27410 QQW28621.1 (289 aa)	
	M.CsaIII (253 aa, 22%)			DNA adenine methylase [Microcystis sp. 49638_E5] MCE2671800.1; 281 aa; 134/262(51%)	AP9108_31020 WAK74115.1 (261 aa)	
M.Tam77409IV (141166)	M.Tam77409IV (942 aa, 49%)	CCR^{m4}CTC	N4	DUF1156 domain-containing protein [Arthrospira sp. PLM2. Bin9] TVU52848.1; 963 aa; 877/972(90%)	AP9108_24165 QQW28150.1 (972 aa)	NIES-39, YZ, C1
M.*Nsp*I (3473)	M.*Nsp*I (397 aa, 27%)	R^{m5}CATGY	C5	Modification methylase XorII [Xanthomonas oryzae pv. oryzae KACC 10331] P52311.2; 424 aa; 137/370(37%)	AP9108_28425 QQW28747.1 (390 aa)	All strains
TauI (2579)	M.CglI (363 aa, 54%) (3577)	GCSGC	C5 (u.b.)	Cytosine-specific methyltransferase NgoFVII [Neisseria gonorrhoeae] Q59606.1; 374 aa; 209/355(59%)	AP9108_15125 QQW31665.1 (406 aa)	All strains
M.PfrJS9IV (177192)	M.PfrJS9aIV (968 aa, 24%) (218604)	ARCC^{m4}CC	N4	DUF1156 domain-containing protein [Desertifilum sp. FACHB-1129] MBD2311126.1; 941 aa; 690/968(71%)	AP9108_15215 QQW31681.1 (972 aa)	NIES-39, YZ, Paraca
Endonuclease and methylase merged forming only one enzyme with two functions						
RM.AplYZ	RM.AplYZ (1053 aa, 100%) (284171)	Undetermined	N6A	N-6 DNA methylase [Limnospira maxima CS-328] EDZ92055.1; 1054 aa; 956/1054(91%)	AP9108_20650 QQW27605.1 (1053 aa)	All strains
RM.AplP	RM.AplP (1109 aa, 100%) (92440)	Undetermined	N6A	Type II restriction endonuclease [Arthrospira sp. PLM2. Bin9] TVU53426.1; 1098 aa; 991/1116(89%)	AP9108_03245 QQW29865.1 (1109 aa)	NIES-39, YZ, Paraca
RM.Eli8509	RM.Eli8509 (918 aa, 28%) (156312)	CCGG^{m6}AG (u.c.)	N6A	DNA methyltransferase RM. BsuMORF6760P [Bacillus subtilis subsp. subtilis str. 168] O31504.1; 879 aa; 145/383(38%)	AP9108_29180 QQW28860.1 (381 aa)	NIES-39, YZ, Paraca
RM.Pru8113	RM.Pru8113 (1644 a, 36%) (289328)	CAG^{m6}ANGC (u.c.)	N6A	Hypothetical protein HFV01_30305 [Limnospira fusiformis SAG 85.79] QJB29335.1; 1612 aa; 1237/1287(96%)	AP9108_02715 WAK74222.1 (1285 aa)	C1
RM.AmaCSI	RM.AmaCSI (1026 aa; 39%) (26714)	GCTCCA (11/9)	N6A (u.b.)	DNA helicase [Arthrospira sp. PLM2. Bin9] TVU55120.1; 1024aa; 472/512(92%)	AP9108_11835 WAK74513.1 (519 aa)	Paraca, NIES-39, C1
RM.Nbr128	RM.Nbr128 (932 aa, 17%) (124782)	ACCG^{m6}AC ACCGAC (u.c.)	N6A	DNA methyltransferase RM. BsuMORF6760P [Bacillus subtilis subsp. subtilis str. 168] O31504.1; 879 aa; 76/217(35%)	AP9108_29055 QQW28843.1 (233 aa)	NIES-39, YZ, Paraca

1 Compared to the L. platensis Paraca, NIES-39, C1 and YZ strains with the largest number of R-M systems (Xu et al., 2016). u.c.: unknown cleavage site; N6A: N6-methyladenosine; C5: 5-methylcytosine; N4: N4-methylcytosine; u.b.: unknown base; n.a-RS: not annotated but identified in RAST-SERVER.

Nevertheless, although CRISPR5 from L. platensis NIES-39 shows short sequences without DR resembling three different CRISPR arrays, the DRs are the same for these three parts (Xu et al., 2016).

Considering that the spacer repertoire at each CRISPR locus represents a history of previous invasions, we compared the 137 unique spacers identified in all CRISPR loci of the PCC 9108 genome (Table S5) with available databases of bacteria, plasmids, viruses or metagenome sequence data to search for possible proto-spacers. After excluding the hits from L. platensis, only four spacers showed similarity to chromosomal sequences of other bacteria (Table S6). Three out of the four spacers had similarity to other Limnospira, or other cyanobacteria (e.g., Nostoc). Only the spacer CTTGTTCTTGTTTTGGTTCTTGTTCT TTAAAAATATTATAA from CRISPR-4 shares similarity with a non-cyanobacterium, i.e., a facultative anaerobic bacterium (Hathewaya massiliensis strain Marseille-P3545). The other spacers did not show similarity to those of other organisms described in databases. This result reflects the bias in the available data on databases, as previously reported (Pleckaityte et al., 2012; Yang et al., 2015).

The cas clusters described in PCC 9108 mainly belong to type III CRISPR–Cas systems, the most diverse and complex CRISPR, able to cleave RNA and DNA, and the only known that use three nuclease activities (Burmistrz et al., 2020; Molina et al., 2020) (Table 1). This type possesses the signature gene cas10 (csm1 or cmr2) which encodes a multidomain protein that acts as the large subunit of effector complexes of type III systems (Tamulaitis et al., 2017). Csm (Type III-A) and Cmr (Type III-B) complexes function as RNA-activated single-stranded (ss)

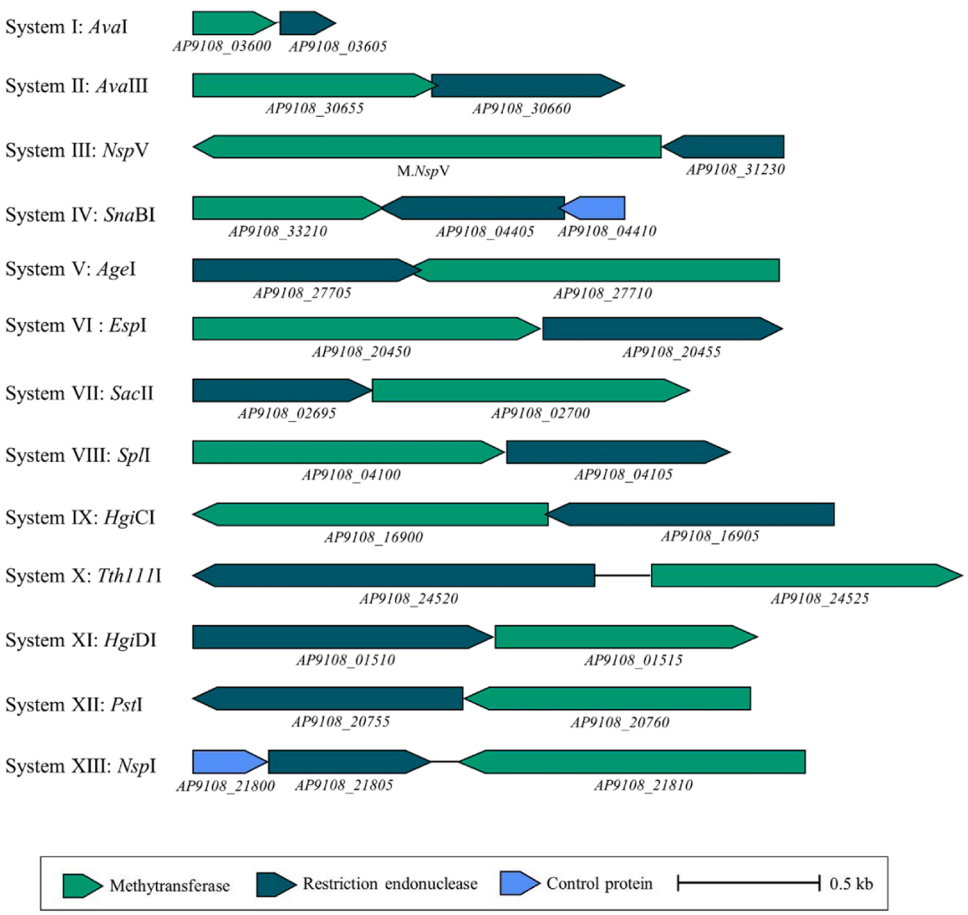

Fig. 2. Scheme showing the organization of the genes encoding the hypothetical type II R-M systems of L. platensis PCC 9108. Arrows represent the direction of translation and the sizes of the open reading frames deduced from the nucleotide sequence analysis.

DNases that couple the target RNA binding/cleavage with ssDNA degradation (Tamulaitis et al., 2017).

We have identified subtypes III-A (Csm module) and III-B (Cmr module) clusters in the PCC 9108 genome. The transcriptomic analysis showed that all genes within the cas clusters were expressed, along with IIIA genes (cluster 3 and 4) exhibiting higher expression levels than IIIB (cluster 2 and 5), except for gene cmr3 of cluster 5, which showed a 57.3-fold higher and is involved in facilitating immunity and avoiding autoimmunity in the Type III-B system. The presence of both subtypes A and B in the genome could contribute to the defense systems in PCC 9108.

R-M systems are widespread in sequenced bacterial genomes, and they are reported to be a significant impediment to transformation processes (Loenen and Raleigh, 2014; Roberts et al., 2015). The search in the PCC 9108 genome of R-M systems revealed that it contains 4 type I R-M systems and 13 type II R-M systems. Type I R-M consists of three subunits: HsdM, HsdS and HsdR (encoded by different genes, referred to as hsd for host specificity determinant), which form an enzyme complex that performs DNA methylation, DNA restriction and DNA sequence-recognition (Bickle and Krüger, 1993; Zhao et al., 2006). Among the 4 complete sets of type I R-M systems found in the PCC 9108 genome, some of them are also present in other L. platensis strains (Fujisawa et al., 2010) (Table 2). The hsdS gene of System II is reported by Zhao et al. (2006) as a single gene with a non-sense codon, but in PCC 9108, it corresponds to two different hsdS genes. Zhao et al. (2006) proposed a hsdR gene in System III that codes for a 146 aa protein, while in PCC 9108, this gene belongs to a DEAD/DEAH box helicase family and codes for a 774 aa protein. System IV in PCC 9108 contains two uma2 genes between the mtase gene and the hsdR gene that is not mentioned by Zhao et al. (2006). The uma2 gene belongs to a family of proteins greatly expanded in cyanobacteria that are likely to be acting as an endonuclease (https://www.ncbi.nlm.nih.gov/Structure/cdd/PF05685).

R-M type II are a drawback in cyanobacterial conjugation, as type II endonucleases in recipient bacteria decrease the efficiency of DNA transfer (Wolk et al., 1984; Elhai et al., 1997; Taton et al., 2012). The MTase activity of 13 R-M systems found in PCC 9108 was verified in vitro and demonstrated that its genome is completely protected by these methyltransferases as 11 out of 12 methyltransferases tested were found to be functional in heterologous systems. The single exception was MT. *Ava*I, which provided a partial protection. In this sense, the putative *Ava*I found in PCC 9108 is a smaller version compared to the commercially used derived from Anabaena variabilis (ATCC 27892) (100 aa versus 486 aa). This result suggests that the M.*Ava*I-like system of PCC 9108 that is present in several Limnospira strains could display a different role in the genome.

We have also identified 8 orphan methylases and 6 single-enzyme R-M systems (Table 3). The presence of orphan MTases is common in prokaryotic genomes, and although most of them are not well characterized, their participation in regulation of gene expression, DNA replication, repair, and others, has been proposed (Sánchez-Romero and Casadesús, 2020). On the other hand, the single-enzymes R-M is a typical feature of Type IIG, Type IIB, or Type IIC systems, some of which have likely evolved from Type I systems (Anton and Roberts, 2021). Another explanation could be that some of these R-M systems could be derived from mobile genetic elements that had lost the endonuclease gene allowing the methylase to remain as an orphan activity.

To date, only a few studies have been conducted on the global DNA

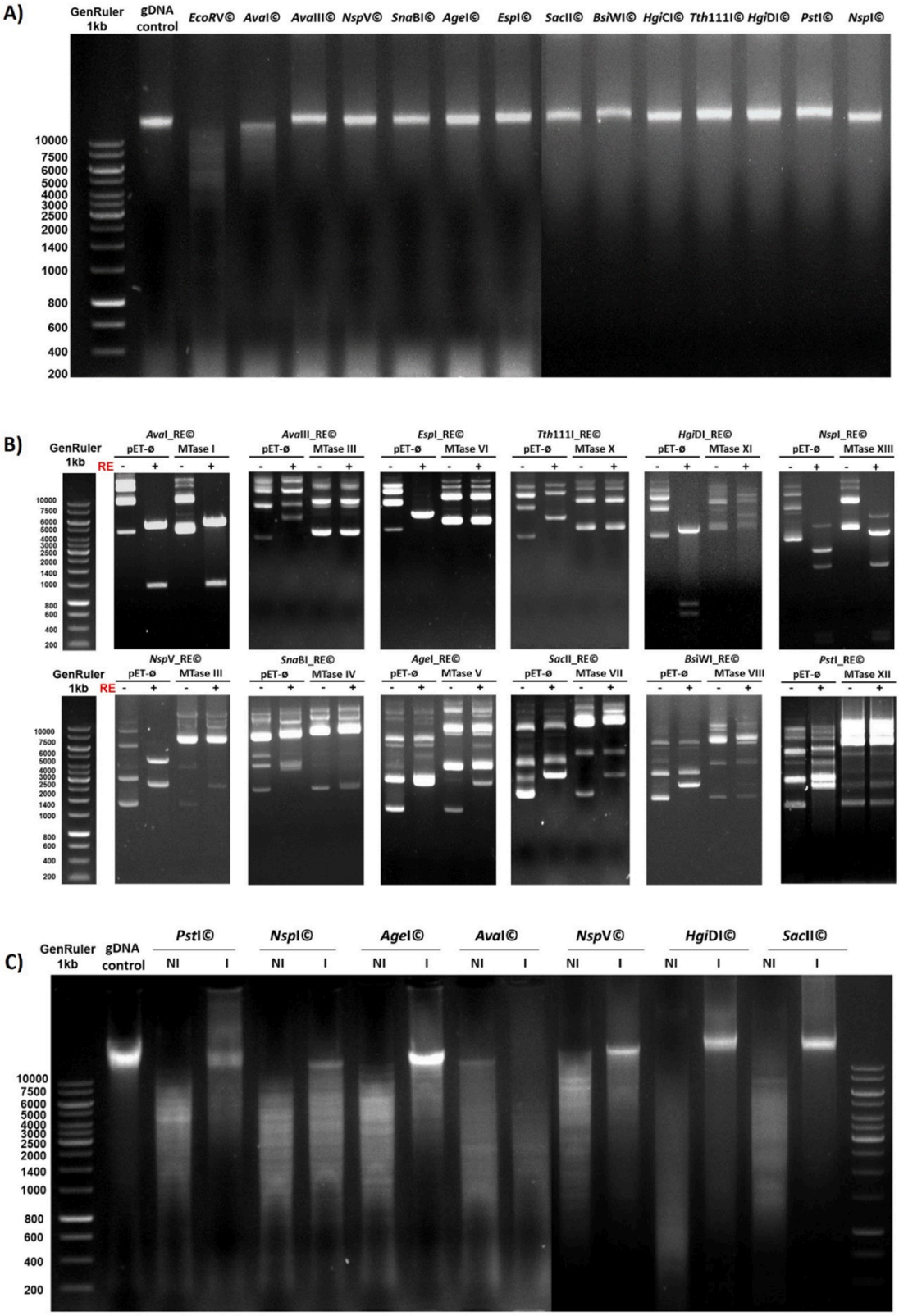

Fig. 3. Activity of the 13 MTases found in L. platensis PCC 9108. A) Methylation status of L. platensis genomic DNA. Cyanobacterial DNA was isolated, purified, treated with commercial restriction endonucleases identified in L. platensis PCC 9108 genome, and analyzed by gel electrophoresis. *Eco*RV was used as digestion control. **B) MTase activity of methyltransferase genes cloned in pET29a+ by in vivo methylation protection assays.** Plasmids containing genes encoding the MTases and a control plasmid without MTases genes were isolated from E. coli. Methylated plasmid DNAs + lines were challenged with equivalent restriction endonucleases (RE). - lanes are the non-digested plasmid control in each case. **C) MTase activity of cloned methyltransferase genes in pTIP-QC1 using in vivo methylation protection assays.** Genomic DNA from Rhodococcus after in vivo methyltransferase assay was isolated and digested with corresponding commercial restriction enzymes. NI: non-induced gDNA digestion, I: Induced gDNA digestion.

Table 4
Most common methylated recognition sites in L. platensis PCC 9108 identified by MinION sequencing.

Homology in REBASE	Modification base	Number of putative recognition sites	Number of methylated recognition sequences	% Methylated sites	Most common recognition sequence
M. *Age*I	5-methylcytosine (m5C) (base undetermined)	361	340	94.2	ACCGGT
M. AplI	C5-methylcytosine (m5C)	248	168	67.7	CTGCm5AG
M. *Ava*III	N6- methyladenine (m6A)	84	65	77.4	ATGCAm6T
M. HgiCI /*Ban*I	C5-methylcytosine (m5C)	775	749	96.6	GGTGCm5C
M. HgiDI /*Acy*I	5-methylcytosine (m5C)	136	127	93.4	GRCm5GYC
M. *Nsp*I	C5-methylcytosine (m5C)	93	22	23.7	RCm5ATGY
M. *Sac*II	C5-methylcytosine (m5C)	37	34	91.9	CCm5GCGG
M. *Tth*111I	N6- methyladenine (m6A)	90	52	57.8	GAm6CNNNGTC

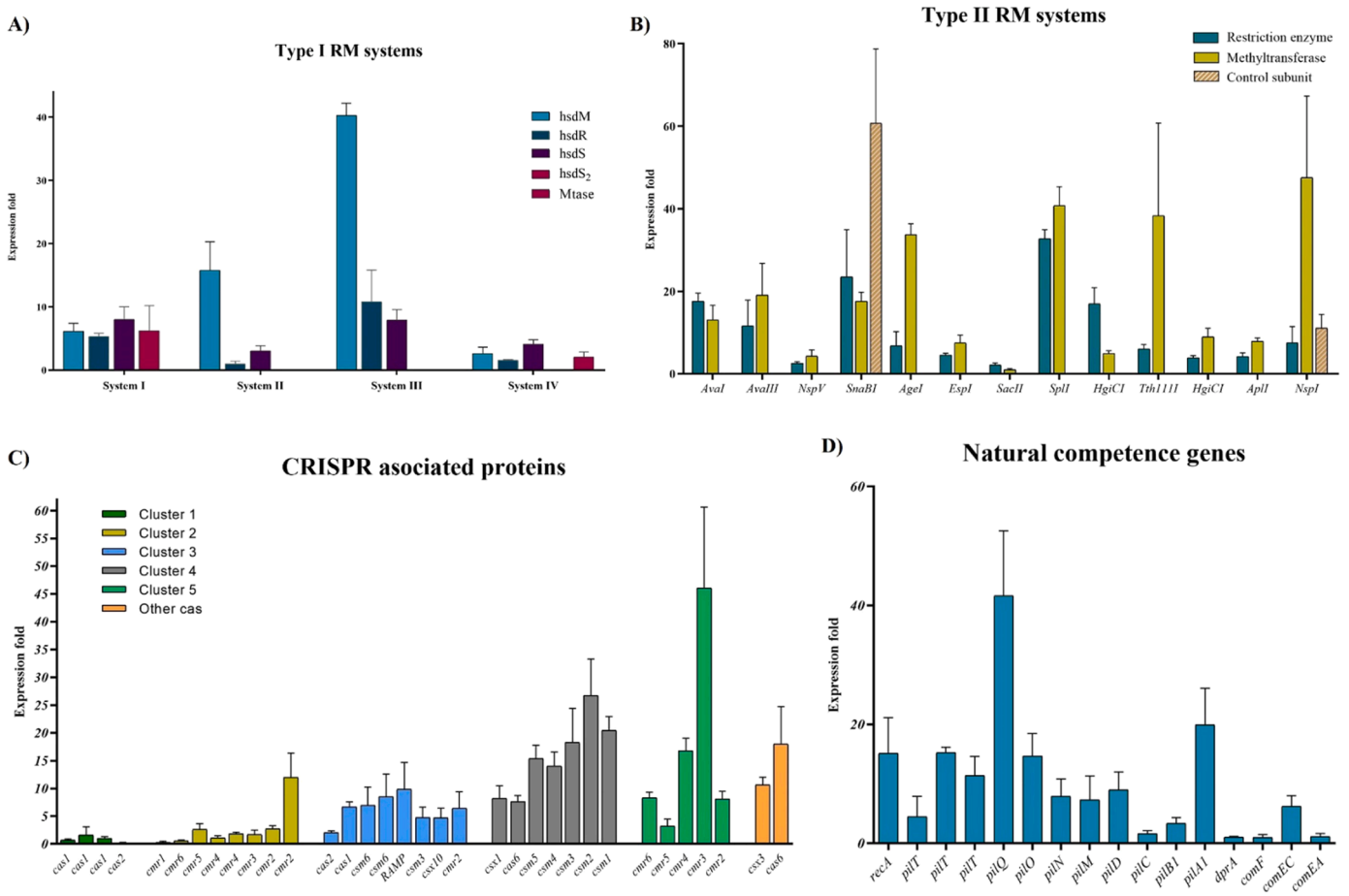

Fig. 4. RNA transcription levels. Gene expression under control conditions (30 °C, 24 h light provided by cool, white lamps at a light intensity of 500 Lux and 120 rpm) in UTEX media. **(A) Type I R-M genes.** Fold changes were normalized against the lowest expression observed (AP9108_35185, hsdR of system II) with a value of 1.2. Data represent means ± SDs from three biological replicates. **B) Type II R-M genes.** Fold changes were normalized against the lowest expression observed (AP9108_02700, M.*Sac*II) with a value of 1.33. Data represent means ± SDs from three biological replicates. **C) CRISPR associated proteins clusters.** Fold changes were normalized against the lowest expression observed (AP9108_34475, Cas1) with a value of 1.03. Data represent means ± SDs from three biological replicates. **D) Competence genes.** Fold changes were normalized against the lowest expression observed (AP9108_05690, ComF) with a value of 3.5. Data represent means ± SDs from three biological replicates.

methylation patterns in cyanobacteria. The methylome and orphan MTases in the model cyanobacterium Synechocystis sp. PCC 6803 were determined by single molecule real-time (SMRT) sequencing revealing four new DNA methylation recognition sequences in addition to the previously known motif m5CGATCG (Hagemann et al., 2018). Furthermore, differences in the global DNA methylation pattern of cyanobacteria have been reported between normal and nutrient stress conditions, such as nitrogen starvation (Hu et al., 2018). The methylation pattern could be also a system to regulate gene expression or even inherited under developmental stress as previously described (Hu et al., 2018; Walworth et al., 2021). Here, we present the first genome-wide methylation map for Limnospira, highlighting the need for further exploration of methylation patterns in different strains and conditions to gain insights into the epigenetic characteristics in this genus. These epigenetic traces could be considered when selecting a strain for transformation.

The presence of multiple R-M systems in some bacteria has been extensively discussed. For example, it was proposed they might confer the ability to colonize many different niches (Vasu et al., 2013; Dimitriu et al., 2020). Also, bacteria with larger genomes as spirulina might have

more diverse lifestyles, select for more diverse types of genes, inhabit more environments, or accommodate more mobile genetic elements, and consequently, they must accumulate more R-M systems to survive. In addition, it is known that naturally competent bacteria, as occurs in spirulina, contain more R-M systems. R-M systems are more abundant in promiscuous species, wherein they establish preferential paths of genetic exchange within and between lineages with cognate R-M systems. On the other hand, contrary to the prevalent view that R-M systems limit the flux of genetic material, it has been proposed that restriction favors strain evolution by producing DNA double-stranded ends that can recombine. R-M systems favor transfer of genetic material between cells by generating restriction breaks that stimulate recombination between homologous sequences (Kobayashi, 2001; Oliveira et al., 2014; Oliveira et al., 2016).

Besides, although putative methylation sites have been identified in the PCC 9108 genome, there is not a clear correlation between the number of methylation sites with the expression levels observed in the transcriptome. This discrepancy can be attributed to the evaluation of mRNA levels rather than enzymatic activities. Notably, the most highly expressed restriction enzyme genes in PCC 9108 are SplI (32.7-fold), *Sna*BI (23.5-fold) and HgiCI (17.0-fold). Considering the abundance and expression of R-M systems in PCC 9108, it is evident that the genetic transformation of this L. platensis strain requires further improvement and that these R-M systems should be taken into account for transformation protocols of this strain.

Several studies have shown that most cyanobacteria, including Limnospira sp., possess essential natural competence genes, indicating their inherent capability for natural transformation (Nies et al., 2022; Wendt and Pakrasi, 2019). The presence and expression of these competence genes in PCC 9108 further implies that this strain might also be amenable to modification through natural transformation, potentially with the assistance of other bacteria, as described in the case of NIES-39 and UTEX LB1926 (Jester et al., 2022). Although the presence of robust genetic barriers in PCC 9108, including restriction-modification systems and CRISPR, as highlighted in this study, can represent a drawback to achieving a successful modification of this cyanobacteria, the deep knowledge of the system could allow the development of technology to overcome this problem.

Finally, it initially seems surprising that an extremophilic bacterium that can grow at very high alkaline pH and high salt concentrations has developed such powerful restriction and CRISPR systems when it would not be expected to have many invasive viruses in its environment.

5. Conclusions

In this work we have identified the restriction-modification and CRISPR-Cas systems from the filamentous cyanobacterium Limnospira platensis. We have found four complete sets of type I R-M and thirteen type II R-M systems, several incomplete type I or II R-M systems (e.g. only methylase) and 6 type II R-M systems formed by one single enzyme with both capabilities. We have cloned and characterized thirteen of the type II methylases found to prove its activity.

We present here the first genome-wide methylation map for L. platensis. We have found that there is no apparent correlation between the putative methylation sites found in the PCC 9108 genome and the expression levels found in the transcriptome, suggesting a complex regulation of the whole system. The restriction enzymes genes more expressed corresponded to SplI (32.7-fold)>*Sna*BI (23.5-fold)>HigiCI (17.0-fold) or *Ava*I (17.6-fold)>*Ava*III (11.7-fold)>*Nsp*I (7.5-fold). These findings will be a valuable resource to understand how this strain avoids being genetically manipulated and for further genomics studies.

Funding

This work was supported by projects S2013/ABI-2783 (INSPIRA1-CM), S2018/BAA-4532 (ALGATEC-CM) from “Comunidad de Madrid /ESF-ERDF”; RTI2018–094399-A-I00 (SETH) from the Spanish Ministry of Economy and Competitivity; RobExplode PID2019-108458RB-I00 (AEI/10.13039/501100011033) and by Sycosys TED2021–130689B-C33 from Spanish Ministry of Science and Innovation (MICINN) grants. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CRediT authorship contribution statement

Díaz Carrasco Asunción: Methodology, Formal analysis, Data curation. **Nogales Juan:** Validation, Investigation, Funding acquisition, Conceptualization. **Agudo Lucía:** Methodology, Formal analysis, Data curation. **Guevara Govinda:** Methodology, Investigation, Formal analysis. **Castillo María:** Methodology, Investigation, Formal analysis. **Navarro Llorens Juana María:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Funding acquisition, Conceptualization. **Suárez Rodríguez Patricia:** Methodology. **Baldanta Sara:** Methodology, Investigation, Formal analysis. **P é rez-P é rez Julián:** Methodology, Formal analysis, Data curation. **Arribas-Aguilar Fernando:** Methodology, Formal analysis, Data curation. **Galan Beatriz:** Writing – review & editing, Supervision, Investigation, Conceptualization. **García José Luís:** Supervision, Investigation, Funding acquisition, Conceptualization.

JLG, JN, BG and JMN contributed to conception and design of the study. GG, PS, MC, LA and SB performed the main experiments and provided with analytical tools. ADC and JPP developed the ON sequencing, and FAA and JPP did the methylome issues. GG, MC, SB, JMN and BG analyzed data and wrote the first draft of the manuscript. JLG, JN contributed to manuscript revision. All authors read and approved the submitted version.

Declaration of Competing Interest

The authors are unaware of any conflicting interests and thereby declare no conflict of interest.

Data Availability

All data (genome, RNAseq and methylome reads)has been deposited in public data bases (NCBI).

Acknowledgements

We are grateful to Amaya Blanco Rivero for initiating the first steps of this study. We would like to thank to members of the INSPIRA-1 Consortium (CIB, CSIC, UAM, IMDEA, URJC) and to the CYTED network RENUWAL 320RT0005 for their technical support. We also thank the collaboration of Genomics and Massive Sequencing Service of the Severo Ochoa Molecular Biology Center (CBMSO, CSIC-UAM, Madrid, Spain) for the genome assembly.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.micres.2023.127572.

References

AlFadhly, N.K.Z., Alhelfi, N., Altemimi, A.B., Verma, D.K., Cacciola, F., Narayanankutty, A., 2022. Trends and technological advancements in the possible food applications of Spirulina and their health benefits: a review. Molecules 27, 5584. https://doi.org/10.3390/molecules27175584.

Anton, B.P., Roberts, R.J., 2021. Beyond restriction modification: epigenomic roles of DNA methylation in prokaryotes. Annu. Rev. Microbiol. 75, 129–149. https://doi.org/10.1146/annurev-micro-040521-035040.

Anton, B.P., Heiter, D.F., Benner, J.S., Hess, E.J., Greenough, L., Moran, L.S., Slatko, B.E., Brooks, J.E., 1997. Cloning and characterization of the *Bgl*II restriction–modification

system reveals a possible evolutionary footprint. Gene 187, 19–27. https://doi.org/10.1016/s0378–1119(96)00638–5.
Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., et al., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genom. 9, 75 https://doi.org/10.1186/1471-2164-9-75.
Bickle, T.A., Krüger, D.H., 1993. Biology of DNA restriction. Microbiol. Rev. 57, 434–450. https://doi.org/10.1128/mr.57.2.434-450.1993.
Blow, M.J., Clark, T.A., Daum, C.G., Deutschbauer, A.M., Fomenkov, A., Fries, R., et al., 2016. The epigenomic landscape of prokaryotes. PLoS Genet 12, e1005854. https://doi.org/10.1371/journal.pgen.1005854.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma. (Oxf., Engl.) 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
Burmistrz, M., Krakowski, K., Krawczyk-Balska, A., 2020. RNA-Targeting CRISPR-Cas systems and their applications. Int. J. Mol. Sci. 21, 1122. https://doi.org/10.3390/ijms21031122.
Cai, F., Axen, S.D., Kerfeld, C.A., 2013. Evidence for the widespread distribution of CRISPR-Cas system in the Phylum Cyanobacteria. RNA Biol. 10, 687–693. https://doi.org/10.4161/rna.24571.
Cao, J., Xu, Z., Qiu, G., Li, B., 1996. Effects of Mg^{2+} on the growth and DNase activity of Spirulina platensis, a cyanobacterium. Bioresour. Technol. 67, 287–290. https://doi.org/10.1016/S0960-8524(98)00116-3.
Cesnaviciene, E., Mitkaite, G., Stankevicius, K., Janulaitis, A., Lubys, A., 2003. Esp1396I restriction-modification system: structural organization and mode of regulation. Nucleic Acids Res 31, 743–749. https://doi.org/10.1093/nar/gkg135.
Cheevadhanarak, S., Paithoonrangsarid, K., Prommeenate, P., Kaewngam, W., Musigkain, A., Tragoonrung, S., et al., 2012. Draft genome sequence of Arthrospira platensis C1 (PCC9438). Stand. Genom. Sci. 6, 43–53. https://doi.org/10.4056/sigs.2525955.
Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., et al., 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10, 563–569. https://doi.org/10.1038/nmeth.2474.
Dehghani, J., Adibkia, K., Movafeghi, A., Barzegari, A., Pourseif, M.M., Maleki Kakelar, H., et al., 2018. Stable transformation of Spirulina (Arthrospira) platensis: a promising microalga for production of edible vaccines. Appl. Microbiol. Biotechnol. 102, 9267–9278. https://doi.org/10.1007/s00253-018-9296-7.
Dimitriu, T., Szczelkun, M.D., Westra, E.R., 2020. Evolutionary ecology and interplay of prokaryotic innate and adaptive immune systems. Curr. Biol. 30, R1189–R1202. https://doi.org/10.1016/j.cub.2020.08.028.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinforma. (Oxf., Engl.) 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
Elhai, J., Vepritskiy, A.A., Muro-Pasto,r, A.M., Flores, E., Wolk, C.P., 1997. Reduction of conjugal transfer efficiency by three restriction activities of Anabaena sp. strain PCC 7120. J. Bacteriol. 179, 1998–2005. https://doi.org/10.1128/jb.179.6.1998-2005.1997.
Esquivel-Hernández, D.A., Pennacchio, A.E., Torres-Acosta, M.A., Parra-Saldívar, R., de Souza Vandenberghe, L.P., Faraco, V., 2021. Multi-product biorefinery from Arthrospira platensis biomass as feedstock for bioethanol and lactic acid production. Sci. Rep. 11, 19309 https://doi.org/10.1038/s41598-021-97803-5.
Fernández de Las Heras, L., Mascaraque, V., García Fernández, E., Navarro-Llorens, J.M., Perera, J., Drzyzga, O., 2011. ChoG is the main inducible extracellular cholesterol oxidase of Rhodococcus sp. strain CECT3014. Microbiol. Res. 166, 403–418. https://doi.org/10.1016/j.micres.2010.05.003.
Fernández de Las Heras, L., Perera, J., Navarro Llorens, J.M., 2014. Cholesterol to cholestenone oxidation by ChoG, the main extracellular cholesterol oxidase of Rhodococcus ruber strain Chol-4. J. Steroid Biochem. Mol. Biol. 139, 33–44. https://doi.org/10.1016/j.jsbmb.2013.10.001.
Fujisawa, T., Narikawa, R., Okamoto, S., Ehira, S., Yoshimura, H., Suzuki, I., et al., 2010. Genomic structure of an economically important cyanobacterium, Arthrospira (Spirulina) platensis NIES-39. DNA Res. 17, 85–103 ttps://doi.org/10.1093/dnares/dsq004.
Garrett, R.A., Vestergaard, G., Shah, S.A., 2011. Archaeal CRISPR-based immune systems: exchangeable functional modules. Trends Microbiol 19, 549–556. https://doi.org/10.1016/j.tim.2011.08.002.
Green, M.R., Sambrook, J., 2012. Molecular cloning: a laboratory manual N.Y, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
Grissa, I., Vergnaud, G., Pourcel, C., 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35 (Web Server issue), W52–W57. https://doi.org/10.1093/nar/gkm360.
Gunasundari, E., Senthil Kumar, P., 2017. Higher adsorption capacity of Spirulina platensis alga for Cr(VI) ions removal: parameter optimisation, equilibrium, kinetic and thermodynamic predictions. IET Nanobiotechnol 11, 317–328. https://doi.org/10.1049/iet-nbt.2016.0121.
Guo, T., Zheng, F., Zeng, Z., Yang, Y., Li, Q., Sh,e, Q., et al., 2019. Cmr3 regulates the suppression on cyclic oligoadenylate synthesis by tag complementarity in a Type III-B CRISPR-Cas system. RNA Biol. 16, 1513–1520. https://doi.org/10.1080/15476286.2019.1642725.
Hagemann, M., Gartner, K., Scharnagl, M., Bolay, P., Lott, S.C., Fuss, J., et al., 2018. Identification of the DNA methyltransferases establishing the methylome of the cyanobacterium Synechocystis sp. PCC 6803. DNA Res 25, 343–352. https://doi.org/10.1093/dnares/dsy006.
Heidmann S., Seifert, W., Kessler, C., Domdey, H., 1989. Cloning, characterization and heterologous expression of the *Sma*I restriction-modification system. Nucleic Acids Res. 17, 9783–9796. https://doi.org/10.1093/nar/17.23.978.
Horvath, P., Coute-Monvoisin, A.C., Romero, D.A., Boyaval, P., Fremaux, C., Barrangou, R., 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. Int. J. Food Microbiol. 131, 62–70. https://doi.org/10.1016/j.ijfoodmicro.2008.05.030.
Hou, S., Brenes-Alvarez, M., Reimann, V., Alkhnbashi, O.S., Backofen, R., Muro-Pastor, A.M., et al., 2019. CRISPR-Cas systems in multicellular cyanobacteria. RNA Biol. 16, 518–529. https://doi.org/10.1080/15476286.2018.1493330.
Houmard, J., de Marsac, N.T., 1988. Cyanobacterial genetic tools: current status. Methods Enzym. 167, 808–847. https://doi.org/10.1016/0076-6879(88)67092-3.
Hu, L., Xiao, P., Jiang, Y., Dong, M., Chen, Z., Li, H., et al., 2018. Transgenerational epigenetic inheritance under environmental stress by genome-wide DNA methylation profiling in cyanobacterium. Front. Microbiol. 9, 1479 https://doi.org/10.3389/fmicb.2018.01479.
Ives, C.L., Nathan, P.D., Brooks, J.E., 1992. Regulation of the *Bam*HI restriction-modification system by a small intergenic open reading frame, bamHIC, in both Escherichia coli and Bacillus subtilis. J. Bacteriol. 174, 7194–7201. https://doi.org/10.1128/jb.174.22.7194–7201.1992.
Jeamton, W., Dulsawat, S., Tanticharoen, M., Vonshak, A., Cheevadhanarak, S., 2017. Overcoming intrinsic restriction enzyme barriers enhances transformation efficiency in Arthrospira platensis C1. Plant Cell Physiol. 58, 822–830. https://doi.org/10.1093/pcp/pcx016.
Jeltsch, A., 2002. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem 3, 274–293 https://doi.org/10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S.
Jester, B.W., Zhao, H., Gewe, M., Adame, T., Perruzza, L., Bolick, D.T., et al., 2022. Development of spirulina for the manufacture and oral delivery of protein therapeutics. Nat. Biotechnol. 40, 956–964. https://doi.org/10.1038/s41587-022-01249-7.
Jungblut, A.D., Raymond, F., Dion, M.B., Moineau, S., Mohit, V., Nguyen, G.Q., et al., 2021. Genomic diversity and CRISPR-Cas systems in the cyanobacterium Nostoc in the High Arctic. Environ. Microbiol. 23, 2955–2968. https://doi.org/10.1111/1462-2920.15481.
Kawata, Y., Yano, S., Kojima, H., Toyomizu, M., 2004. Transformation of Spirulina platensis strain C1 (Arthrospira sp. PCC9438) with Tn5 transposase-transposon DNA-cation liposome complex. Mar. Biotechnol. (NY) 6, 355–363. https://doi.org/10.1007/s10126-003-0037-1.
Kobayashi, I., 2001. Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution. Nucleic Acids Res 29, 3742–3756. https://doi.org/10.1093/nar/29.18.3742.
Kong, H., Lin, L.F., Porter, N., Stickel, S., Byrd, D., Posfai, J., et al., 2000. Functional analysis of putative restriction-modification system genes in the Helicobacter pylori J99 genome. Nucleic Acids Res. 28, 3216–3223. https://doi.org/10.1093/nar/28.17.3216.
Koonin, E.V., Makarova, K.S., Zhang, F., 2017. Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. 37, 67–78. https://doi.org/10.1016/j.mib.2017.05.008.
Kose, A., Ozen, M.O., Elibol, M., Oncel, S.S., 2017. Investigation of in vitro digestibility of dietary microalga Chlorella vulgaris and cyanobacterium Spirulina platensis as a nutritional supplement. 3 Biotech 7, 170. https://doi.org/10.1007/s13205-017-0832-4.
Kunin, V., Sorek, R., Hugenholtz, P., 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol. 8, R61. https://doi.org/10.1186/gb-2007-8-4-r61.
Kuno, S., Yoshida, T., Kaneko, T., Sako, Y., 2012. Intricate interactions between the bloom-forming cyanobacterium Microcystis aeruginosa and foreign genetic elements, revealed by diversified clustered regularly interspaced short palindromic repeat (CRISPR) signatures. Appl. Environ. Microbiol. 78, 5353–5360. https://doi.org/10.1128/AEM.00626-12.
Lefort, F., Calmin, G., Crovadore, J., Falquet, J., Hurni, J.P., Osteras, M., et al., 2014. Whole-Genome shotgun sequence of Arthrospira platensis strain Paraca, a cultivated and edible cyanobacterium. e00751-14 Genome Announc 2. https://doi.org/10.1128/genomeA.00751-14.
Liao, Y., Smyth, G.K., Shi, W., 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res 47, e47. https://doi.org/10.1093/nar/gkz114.
Lin, Y., Yuan, J., Kolmogorov, M., Shen, M.W., Chaisson, M., Pevzner, P.A., 2016. Assembly of long error-prone reads using de Bruijn graphs. Proc. Natl. Acad. Sci. USA 113, E8396–E8405. https://doi.org/10.1073/pnas.1604560113.
Loenen, W.A., Raleigh, E.A., 2014. The other face of restriction: modification-dependent enzymes. Nucleic Acids Res 42, 56–69. https://doi.org/10.1093/nar/gkt747.
Lupatini, A.L., Colla, L.M., Canan, C., Colla, E., 2017. Potential application of microalga Spirulina platensis as a protein source. J. Sci. Food Agric. 97, 724–732. https://doi.org/10.1002/jsfa.7987.
Makarova, K.S., Koonin, E.V., 2015. Annotation and classification of CRISPR-Cas systems. Methods Mol. Biol. (Clifton, NJ) 1311 47–75. https://doi.org/10.1007/978-1-4939-2687-9_4.
Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., et al., 2015. An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. 13, 722–736. https://doi.org/10.1038/nrmicro3569.
Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., et al., 2005. CDD: a conserved domain database for protein classification. Nucleic Acids Res 33 (Database issue), D192–D196. https://doi.org/10.1093/nar/gki069.
Matveyev, A.V., Young, K.T., Meng, A., Elhai, J., 2001. DNA methyltransferases of the cyanobacterium Anabaena PCC 7120. Nucleic Acids Res 29, 1491–1506. https://doi.org/10.1093/nar/29.7.1491.

McGeehan, J.E., Streeter, S.D., Papapanagiotou, I., Fox, G.C., Kneale, G.G., 2005. High-resolution crystal structure of the restriction-modification controller protein C.AhdI from Aeromonas hydrophila. J. Mol. Biol. 346, 689–701. https://doi.org/10.1016/j.jmb.2004.12.025.

Mohamadzadeh Shirazi, H., Karimi-Sabet, J., Ghotbi, C., 2017. Biodiesel production from Spirulina microalgae feedstock using direct transesterification near supercritical methanol condition. Bioresour. Technol. 239, 378–386. https://doi.org/10.1016/j.biortech.2017.04.073.

Molina, R., Sofos, N., Montoya, G., 2020. Structural basis of CRISPR-Cas Type III prokaryotic defence systems. Curr. Opin. Struct. Biol. 65, 119–129. https://doi.org/10.1016/j.biortech.2017.04.073.

Morin, N., Vallaeys, T., Hendrickx, L., Natalie, L., Wilmotte, A., 2010. An efficient DNA isolation protocol for filamentous cyanobacteria of the genus Arthrospira. J. Microbiol. Methods 80, 148–154. https://doi.org/10.1016/j.mimet.2009.11.012.

Nakashima, N., Tamura, T., 2004. Isolation and characterization of a rolling-circle-type plasmid from Rhodococcus erythropolis and application of the plasmid to multiple-recombinant-protein expression. Appl. Environ. Microbiol. 70, 5557–5568. https://doi.org/10.1128/AEM.70.9.5557-5568.2004.

Nies, F., Springstein, B.L., Hanke, D.M., Dagan, T., 2022. Natural competence in the filamentous, heterocystous Cyanobacterium Chlorogloeopsis fritschii PCC 6912. mSphere 7, e0099721. https://doi.org/10.1128/msphere.00997-21.

Nowicka-Krawczyk, P., Muhlsteinova, R., Hauer, T., 2019. Detailed characterization of the Arthrospira type species separating commercially grown taxa into the new genus Limnospira (Cyanobacteria). Sci. Rep. 9, 694 https://doi.org/10.1038/s41598-018-36831-0.

Oliveira, P.H., Touchon, M., Rocha, E.P., 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res 42, 10618–10631. https://doi.org/10.1093/nar/gku734.

Oliveira, P.H., Touchon, M., Rocha, E.P., 2016. Regulation of genetic flux between bacteria by restriction–modification systems. Proc. Natl. Acad. Sci. USA 113, 5658–5663. https://doi.org/10.1073/pnas.1603257113.

Pleckaityte, M., Zilnyte, M., Zvirbliene, A., 2012. Insights into the CRISPR/Cas system of Gardnerella vaginalis. BMC Microbiol 12, 301. https://doi.org/10.1186/1471-2180-12-30.

Rand, A.C., Jain, M., Eizenga, J.M., Musselman-Brown, A., Olsen, H.E., Akeson, M., et al., 2017. Mapping DNA methylation with high-throughput nanopore sequencing. Nat. Methods 14, 411–413. https://doi.org/10.1038/nmeth.4189.

Ren, J., Lee, H.M., Shen, J., Na, D., 2022. Advanced biotechnology using methyltransferase and its applications in bacteria: a mini review. Biotechnol. Lett. 44, 33–44. https://doi.org/10.1007/s10529-021-03208-9.

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., Peplies, J., 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32, 929–931. https://doi.org/10.1093/bioinformatics/btv681.

Rimšeliené, R., Vaišvila, R., Janulaitis, A., 1995. The eco72IC gene specifies a trans-acting factor which influences expression of both DNA methyltransferase and endonuclease from the Eco72I restriction-modification system. Gene 157, 217–219. https://doi.org/10.1016/0378-1119(94)00794-s.

Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., et al., 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res 31, 1805–1812. https://doi.org/10.1093/nar/gkg274.

Roberts, R.J., Vincze, T., Posfai, J., Macelis, D., 2015. REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 43 (Database issue), D298–D299. https://doi.org/10.1093/nar/gku1046.

Rolando, M., Silvestre, C.D., Gomez-Valero, L., Buchrieser, C., 2022. Bacterial methyltransferases: from targeting bacterial genomes to host epigenetics. microLife 3, uqac014. https://doi.org/10.1093/femsml/uqac014.

Sánchez-Romero, M.A., Casadesús, J., 2020. The bacterial epigenome. Nat. Rev. Microbiol. 18, 7–20. https://doi.org/10.1038/s41579-019-0286-2.

Saveria, T., Parthiban, C., Seilie, A.M., Brady, C., Martinez, A., Manocha, R., et al., 2022. Needle-free, spirulina-produced Plasmodium falciparum circumsporozoite vaccination provides sterile protection against pre-erythrocytic malaria in mice. NPJ Vaccin. 7, 113 https://doi.org/10.1038/s41541-022-00534-5.

Schirmacher, A.M., Hanamghar, S.S., Zedler, J.A.Z., 2020. Function and benefits of natural competence in cyanobacteria: from ecology to targeted manipulation. Life 10 (11). https://doi.org/10.3390/life10110249.

Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., et al., 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc. Natl. Acad. Sci. USA 110, 1053–1058. https://doi.org/10.1073/pnas.1217107110.

Shmakov, S.A., Utkina, I., Wolf, Y.I., Makarova, K.S., Severinov, K.V., Koonin, E.V., 2020. CRISPR Arrays Away from cas Genes. CRISPR J. 3, 535–549. https://doi.org/10.1089/crispr.2020.0062.

Silas, S., Makarova, K.S., Shmakov, S., Paez-Espino, D., Mohr, G., Liu, Y., et al., 2017. On the origin of reverse transcriptase-using CRISPR-Cas systems and their hyperdiverse, enigmatic spacer repertoires. mBio 8, e00897-17. https://doi.org/10.1128/mBio.00897-17.

Stucken, K., Koch, R., Dagan, T., 2013. Cyanobacterial defense mechanisms against foreign DNA transfer and their impact on genetic engineering. Biol. Res. 46, 373–382. https://doi.org/10.4067/S0716-97602013000400009.

Suzuki, S., Yamaguchi, H., Kawachi, M., 2019. The draft genome of a hydrogen-producing cyanobacterium, Arthrospira platensis NIES-46. J. Genom. 7, 56–59. https://doi.org/10.7150/jgen.38149.

Tabakh, H., Jester, B.W., Zhao, H., Kuestner, R., Khuong, N., Shanitta, C., et al., 2023. Protocol for the transformation and engineering of edible algae Arthrospira platensis to generate heterologous protein-expressing strains. STAR Protoc. 4, 102087 https://doi.org/10.1016/j.xpro.2023.102087.

Tamulaitis, G., Venclovas, Č., Siksnys, V., 2017. Type III CRISPR-Cas immunity: major differences brushed aside. Trends Microbiol 25, 49–61. https://doi.org/10.1016/j.tim.2016.09.012.

Taton, A., Lis, E., Adin, D.M., Dong, G., Cookson, S., Kay, S.A., et al., 2012. Gene transfer in Leptolyngbya sp. strain BL0902, a cyanobacterium suitable for production of biomass and bioproducts. PLoS One 7, e30901. https://doi.org/10.1371/journal.pone.0030901.

Vasu, K., Nagaraja, V., 2013. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol. Mol. Biol. Rev. 77, 53–72. https://doi.org/10.1128/MMBR.00044-12.

R.M. Vijesurier L. Carlock R.M. Blumenthal J.C. Dunbar Role and mechanism of action of C. *Pvu*II, a regulatory protein conserved among restriction-modification systems J. Bacteriol. 182 2000 477 487 doi: 10.1128/JB.182.2.477-487.2000.

Walworth, N.G., Lee, M.D., Dolzhenko, E., Fu, F.X., Smith, A.D., Webb, E.A., et al., 2021. Long-Term m5C methylome dynamics parallel phenotypic adaptation in the cyanobacterium Trichodesmium. Mol. Biol. Evol. 38, 927–939. https://doi.org/10.1093/molbev/msaa256.

Wendt, K.E., Pakrasi, H.B., 2019. Genomics approaches to deciphering natural transformation in cyanobacteria. Front. Microbiol. 10, 1259 https://doi.org/10.3389/fmicb.2019.01259.

Wilson, G.G., Murray, N.E., 1991. Restriction and modification systems. Annu. Rev. Genet. 25, 585–627. https://doi.org/10.1146/annurev.ge.25.120191.003101.

Wolk, C.P., Vonshak, A., Kehoe, P., Elhai, J., 1984. Construction of shuttle vectors capable of conjugative transfer from Escherichia coli to nitrogen-fixing filamentous cyanobacteria. Proc. Natl. Acad. Sci. USA 81, 1561–1565. https://doi.org/10.1073/pnas.81.5.1561.

Xu, T., Qin, S., Hu, Y., Song, Z., Ying, J., Li, P., et al., 2016. Whole genomic DNA sequencing and comparative genomic analysis of Arthrospira platensis: high genome plasticity and genetic diversity. DNA Res 23, 325–338. https://doi.org/10.1093/dnares/dsw023.

Yadav, M., Kumar, V., Sandal, N., Chauhan, M.K., 2022. Quantitative evaluation of mercury adsorption and removal efficacy of Spirulina (Arthrospira platensis) powder in mice. Arch. Microbiol. 204, 387 https://doi.org/10.1007/s00203-022-03005-6.

Yang, C., Lin, F., Li, Q., Li, T., Zhao, J., 2015. Comparative genomics reveals diversified CRISPR-Cas systems of globally distributed Microcystis aeruginosa, a freshwater bloom-forming cyanobacterium. Front. Microbiol. 6 https://doi.org/10.3389/fmicb.2015.00394.

Zhao, F., Zhang, X., Liang, C., Wu, J., Bao, Q., Qin, S., 2006. Genome-wide analysis of restriction-modification system in unicellular and filamentous cyanobacteria. Physiol. Genom. 24, 181–190. https://doi.org/10.1152/physiolgenomics.00255.2005.

Zheleznaya, L.A., Kainov, D.E., Yunusova, A.K., Matvienko, N.I., 2003. Regulatory C protein of the *Eco*RV modification-restriction system. Biochem. Biokh. 68, 105–110. https://doi.org/10.1023/a:1022105804578.

Zinicovscaia, I., Chiriac, T., Cepoi, L., Rudi, L., Culicov, O., Frontasyeva, M., et al., 2017. Selenium uptake and assessment of the biochemical changes in Arthrospira (Spirulina) platensis biomass during the synthesis of selenium nanoparticles. Can. J. Microbiol. 63, 27–34. https://doi.org/10.1139/cjm-2016-0339.