



Temperature modulates *Fischerella thermalis* ecotypes in Porcelana Hot Spring

Jaime Alcorta^a, Sebastián Espinoza^a, Tomeu Viver^b, María E. Alcamán-Arias^{a,c,e}, Nicole Trefault^d, Ramon Rosselló-Móra^b, Beatriz Díez^{a,e,*}

^a Department of Molecular Genetics and Microbiology, Pontificia Universidad Católica de Chile, Avenida Libertador Bernardo O'Higgins 340, Casilla 144-D, C.P. 651 3677, Santiago, Chile

^b Marine Microbiology Group, Mediterranean Institute for Advanced Studies (IMEDEA; CSIC-UIB), Esporles, E-07190, Spain

^c Department of Oceanography, University of Concepcion, Concepcion, Chile

^d GEMA Center for Genomics, Ecology and Environment, Universidad Mayor, Camino La Pirámide 5750, Santiago, Chile

^e Center for Climate and Resilience Research (CR)2, Chile

ARTICLE INFO

Article history:

Received 14 December 2017

Received in revised form 18 May 2018

Accepted 23 May 2018

Keywords:

Hot spring

Fischerella

Ecotypes

MALDI-TOF MS

Metagenomic assembled genomes

ANI

ABSTRACT

In the Porcelana Hot Spring (Northern Patagonia), true-branching cyanobacteria are the dominant primary producers in microbial mats, and they are mainly responsible for carbon and nitrogen fixation. However, little is known about their metabolic and genomic adaptations at high temperatures. Therefore, in this study, a total of 81 *Fischerella thermalis* strains (also known as *Mastigocladus laminosus*) were isolated from mat samples in a thermal gradient between 61–46 °C. The complementary use of proteomic comparisons from these strains, and comparative genomics of *F. thermalis* pangenomes, suggested that at least two different ecotypes were present within these populations. MALDI-TOF MS analysis separated the strains into three clusters; two with strains obtained from mats within the upper temperature range (61 and 54 °C), and a third obtained from mats within the lower temperature range (51 and 46 °C). Both groups possessed different but synonymous *nifH* alleles. The main proteomic differences were associated with the abundance of photosynthesis-related proteins. Three *F. thermalis* metagenome assembled genomes (MAGs) were described from 66, 58 and 48 °C metagenomes. These pangenomes indicated a divergence of orthologous genes and a high abundance of exclusive genes at 66 °C. These results improved the current understanding of thermal adaptation of *F. thermalis* and the evolution of these thermophilic cyanobacterial species.

© 2018 Elsevier GmbH. All rights reserved.

Introduction

Cyanobacteria are key members of hot spring microbial mat communities. They are part of the photoautotrophic layer, located in the first few millimeters under the surface, and several members are able to fix atmospheric nitrogen (Castenholz [6]; Wickström [64]; Miller et al. [32]; Alcamán et al. [2]). However, despite their fundamental importance to biogeochemical dynamics in extreme systems, their genomic and metabolic adaptations to different temperatures are poorly understood. *Fischerella thermalis* species (also known in the literature as *Mastigocladus laminosus*) are true-branching heterocystous members that dominate the thermal

gradient in many hot springs with temperatures <60 °C (Miller et al. [32,33]; Finsinger et al. [12]; Kaštovský and Johansen [21]; Alcamán et al. [2]).

F. thermalis and *M. laminosus* strains, all of which have been isolated from hot springs, form a monophyletic clade by 16S rRNA phylogeny (Kaštovský and Johansen [21]). The nomenclature of these strains has been problematic due to the historical morphological descriptions of these species (for a detailed revision see Kaštovský and Johansen [21]); however, considering genomic data, almost all genomes from thermal strains deposited in the NCBI database have been named *F. thermalis* or *Fischerella* sp. (Dagan et al. [8]; Sano et al. [50]). Recently, the species name of *M. laminosus* for strains and genomes (Miller et al. [31]; Hutchins and Miller [18]) has been reclassified to *F. thermalis* by the comparative genomics study of Sano et al. [50]. This group comprises globally isolated strains with a seemingly low evolutionary divergence, suggesting a recent speciation event (Miller et al. [31]). In accordance with this fine-

* Corresponding author at: Department of Molecular Genetics and Microbiology, Pontificia Universidad Católica de Chile, Avenida Libertador Bernardo O'Higgins 340, Casilla 144-D, C.P. 651 3677, Santiago, Chile.

E-mail address: bdiez@bio.puc.cl (B. Díez).

tuned evolution, *F. thermalis* strains isolated from thermal mats in Yellowstone National Park (YNP) in the USA were found to exhibit genetic and phenotypic differences according to whether they were isolated from higher or lower temperatures (Miller et al. [33]; Wall et al. [58]; Hutchins and Miller [18]). In populations and strains of *F. thermalis* from YNP, this divergence was observed in short genomic regions. This may be explained by the gene flow within the thermal gradient, as revealed by metagenomic data (Wall et al. [58]) and a study of the genomes from 18 *F. thermalis* strains (Hutchins and Miller [18]). In general, previous genetic studies performed on *F. thermalis* strains have mainly focused on their performance in carbon and nitrogen metabolism (Miller et al. [32,33]; Alcámán et al. [2,1]).

In this current study, a matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) proteomic approach was initially applied to evaluate some of the ecological adaptations that distinguish the thermophilic cyanobacteria residing in a thermal gradient. This technique has been used in *Cyanobacteria* mainly to characterize secondary metabolites (Erhard et al. [11]) and chemotype toxic strains (Welker and Christiansen [61]; Welker and Erhard [62]; Kurmayer et al. [25]). However, it has also proved to be useful in typing and differentiating between environmental strains, as it compares the entire pattern of cell protein expression (e.g. Siegrist et al. [52]; Munoz et al. [35]; Viver et al. [57]; Welker and Moore [63]). Recently, it was used to distinguish *Microcystis* strains using their ribosomal proteomic patterns (Sun et al. [53]).

Furthermore, in the microbial mat from Porcelana Hot Spring (North Patagonia, Chile), it was determined that *F. thermalis* cyanobacteria, and members of the *Oscillatoriales*, dominated the thermal gradient (Mackenzie et al. [28]). In this thermal system, cyanobacterial communities were found to be responsible for all N_2 fixation and oxygenic photosynthesis, including carbon assimilation during the daytime (Alcámán et al. [2]). Metagenomic and metatranscriptomic analyses showed that these processes were also predominantly performed by *F. thermalis* at 58 and 48 °C (Alcámán-Arias et al. submitted [3]). Previous results also showed that only one allele of the 16S rRNA gene was detected for *F. thermalis* cyanobacteria in the Porcelana microbial mat, whereas there were three alleles of the *nifH* gene (the alpha subunit of the nitrogenase enzyme), indicating a potential divergence within *F. thermalis* populations (Alcámán et al. [2]).

Therefore, the primary goal of this study was to characterize 81 strains of *F. thermalis* obtained from Porcelana Hot Spring between 2011 and 2013 along a 61–46 °C thermal gradient. Strains were compared according to morphology, specific marker genes (16S rRNA, ITS and *nifH*) and MALDI-TOF MS profiles. These data, complemented with analyses of *F. thermalis* pangenomes, retrieved after binning from metagenomes obtained in 2013 from temperatures of 66, 58 and 48 °C, improved the understanding of the consequences of ecological speciation in these cyanobacteria along the thermal gradient, following the ecotype definition from Cohan and Perry [7]: “a group of bacteria that are ecologically similar to one another, so similar that genetic diversity within the ecotype is limited by a cohesive force, either periodic selection or genetic drift, or both”. Hence, this study focused on proteomic and genomic features that could explain the adaptations of these bacteria in these extreme systems.

Materials and methods

Study site and cyanobacterial strains

Microbial mat samples were collected from Porcelana Hot Spring located at 42°27'29.1"S–72°27'39.32"W in Northern Patag-

onia, Chile. Samples were collected in 2011, 2012 and 2013 along a 38–69 °C temperature gradient with a pH from 5.1 to 7.1. Other physico-chemical parameters, such as O_2 saturation (42–108%), NO_3^- (0.8–6.5 $\mu\text{mol L}^{-1}$), NH_4^+ (0.01–0.1 $\mu\text{mol L}^{-1}$), PO_4^{3-} (29.7–115.0 $\mu\text{mol L}^{-1}$) and Fe (0.02–0.14 $\mu\text{mol L}^{-1}$) have been described previously by Alcámán et al. [2]. Mat samples were dried at 60 °C for 24 h, then transported to the laboratory for post-analysis. All dried samples were resuspended in 100 mL of sterile medium BG11 (Rippka et al. [47]) and maintained at 37 °C with a photoperiod of 12 h/12 h day-night at a light intensity of 50 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$.

The serial dilution method was implemented to isolate filaments from each timed sample by mixing 2 μL drops from each sample with 2 μL of BG11 medium in 200 μL multi-well dishes. The presence of single filaments in each well was observed using an inverted microscope and then scaled up to 10 mL multiplates. Each isolated strain was named according to its isolation temperature point, the year of collection and the number of the isolated filament. The reference strain *F. thermalis* CHP1 (ex *Mastigocladus* sp. CHP1) was obtained from 46 °C thermal water at Porcelana Hot Spring in 2009 (Alcámán et al. [1]). Three strains obtained from the Pasteur Institute collection of cyanobacteria (PCC): *Fischerella* sp. PCC 9339, *Fischerella* sp. PCC 7520 and *Fischerella* sp. PCC 7522, were used as additional reference strains for comparative analyses.

Identification of cyanobacterial isolates

Strains isolated in culture were stored for determining their morphological characterization after three weeks growth (during the mid-exponential growth phase). Aliquots of 1 mL isolated cultures were fixed with glutaraldehyde at a 2.5% (v/v) final concentration and observed through a Labophot-2 microscope (Nikon, Japan). The morphological identification of the strains was based on Anagnostidis and Komárek [5], while the dimensions of vegetative cells in the branching and main filaments were taxonomically assigned based on Kaššovský and Johansen [21]. Cell size was characterized by measuring the length (as the size of the main axis of the principal or secondary filament) and width (as the size of the perpendicular axis of each filament) of the cells. This analysis was performed for each strain using measures for 20 cells of main and secondary filament, respectively.

DNA extraction and DGGE analysis

From two-week-old multiplate cultures, 200 μL were used for DNA extractions using the protocol described by Alcámán et al. [2]. Briefly, samples were placed in Lysing Matrix E tubes (Qbiogene, Carlsbad, CA, USA) containing XS buffer (Tillet and Neilan [54]) and solid glass beads (1 mm) to obtain cell lysates by bead beating twice at 4.0 ms^{-1} for 1 min. Then, 10% SDS was added and samples were incubated at 65 °C for 2 h. Samples were vortexed briefly, incubated on ice for 30 min and centrifuged at maximum speed at 4 °C for 10 min. Subsequently, two steps of phenol:chloroform:IAA (25:24:1) extraction and one with chloroform:IAA (25:1) were performed. The nucleic acids were precipitated at –20 °C for 2 h with absolute isopropanol together with a 1/10 volume of ammonium acetate (4 M). The pellet was washed in 70% ethanol and finally resuspended in 50–100 μL nuclease-free H_2O .

A 674-bp fragment of the ribosomal 16S rRNA gene was then amplified using a specific cyanobacterial set of primers (CYA106F, CYA781aR and CYA781bR) (Nübel et al. [36]; Uku et al. [56]). A fragment of approximately 400 bp was amplified from the internal transcribed spacer region of the rRNA operon (ITS) using the CSIF and 340R (ULS) primers (Janse et al. [19]), and a 358-bp fragment was amplified from the α -subunit of the nitrogenase iron-protein (*nifH* gene) using a specific cyanobacterial set of primers (CNF and

CNR) (Olson et al. [39]; Díez et al. [9]). All forward primers were attached to a leading 40-bp 5'-GC-tail for denaturing gradient gel electrophoresis (DGGE) analysis.

The resulting amplicons were resolved in a D-code system (BioRad), as previously described by Díez et al. [9] in a 40–70% denaturing gradient for the 16S ribosomal rRNA gene, a 30–50% gradient for ITS, and a 50–60% gradient for the *nifH* genes. The DGGE bands obtained using each set of primers were viewed on electrophoresis gels. When they corresponded to the same position, they were considered the same allele. A band was cut out for each different position on the DGGE gels. The product was then eluted in 20 μ L DNase/RNase-free water, incubated at 4 °C overnight, reamplified with the respective primers lacking GC-tails, and sequenced by Sanger sequencing (Macrogen Inc., Korea).

MALDI TOF MS and statistical analysis

Whole cell biomass was obtained from 81 Porcelana isolates and reference strains. These were grown for 2 weeks at 37 °C, then transferred to a 384-well steel plate (Bruker Daltonics) and analyzed in an Autoflex III MALDI-TOF/TOF MS (Bruker Daltonics, Leipzig, Germany) equipped with a 200-Hz Smartbeam laser, as previously described by Viver et al. [57]. The proteomic patterns were clustered by Biotyper 3.0 software (Bruker Daltonics).

To identify the most relevant proteins, the relative intensity (RI) was determined for the 100 most important mass peaks obtained for each MALDI-TOF MS spectrum. The mass peaks with an RI exceeding 1% and within a 4-Da range in at least 15 strains were selected for further analysis. Average mass was calculated and compared to predicted molecular weight using the predicted proteins from *Fischerella* sp. JSC-11 and *Fischerella* sp. NIES-3754 proteomes (UNIPROT UP000004344 and UP000068400, respectively), and from the three *F. thermalis* metagenome assembled genomes from this study. The following post-translational modifications (PTM) were considered for the search to be the most frequent; N-end rule (N-terminal methionine presence), the presence of N-formyl methionine, one phosphorylation, one acetylation, one glycosylation, and one attached heme group.

The RI's of the detected mass peaks were compared together by the isolation temperature and MALDI-TOF MS clusters, and all statistical analyses were performed with Graphpad Prism 7.02 software (CA, USA). The recovered data were not equilibrated between the groups, but normality was tested using the D'Agostino and Pearson test for each temperature (61 °C n_{\max} = 20; 54 °C n_{\max} = 12; 51 °C n_{\max} = 25 and 46 °C n_{\max} = 24), and a MALDI-TOF MS cluster analysis including the four reference strains (HT n_{\max} = 13, MT n_{\max} = 22 and LT n_{\max} = 50). Groups with peaks in less than 50% of the strains were not considered for further analysis. A one-way ANOVA with Tukey's multiple comparison was used to compare the three groups with a normal distribution, however, when a normal distribution was not present in at least one group, a non-parametric Kruskal–Wallis test with Dunn's multiple comparison was used. When only two groups were compared, a Student's t-test with Welch's correction for different SD was used, whereas a Kolmogorov–Smirnov test was performed when no normal data was being compared.

Metagenomic *Fischerella* recovery

Three metagenomic sequenced samples (Alcamán-Arias et al. submitted [3]) that were obtained in 2013 from Porcelana Hot Spring microbial mats with temperatures of 66, 58 and 48 °C (NCBI BioProject accession no. NA382437) were used for the *Fischerella* genomic estimations. DNA was extracted as previously described (Alcamán-Arias et al. submitted [3]) and sequenced using Illumina HiSeq technology (Research and Testing Laboratory, Texas, USA). The length of the reads obtained was approximately

120 bp. FastQC (www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) software was used to assess the quality of the metagenomic reads, and the reads were trimmed using the Cutadapt tool (Martin [29]), with sequences shorter than 30 bp being discarded. A 3' end trimming was performed for bases with a Phred quality score (Q score) below 28. A hard clipping was performed for the first four left-most bases, and a perfect match of at least 10 bp against standard Illumina adaptors. This procedure reduced the total number of sequences from 452.7 to 394.7 million. *De novo* assemblies of trimmed reads were generated using IDBA assembler (Peng et al. [41]) with the option “pre-correction”. Contigs longer than 1000 bp were grouped in metagenome assembled genomes (MAGs) using MaxBin v2.2 (Wu et al. [65]) with default parameters. The CheckM tool v1.0.5 (Parks et al. [40]) was used to calculate the completeness and contamination (quality) of the recovered MAGs. The abundance of the recovered draft genomes was assessed as read recruitment from the trimmed reads of each sample using a 98% cut-off identity. The sequencing depth of each MAG was calculated using the script BlastTab.seqdepth.pl provided by the enveomics collection (Rodríguez-R and Konstantinidis [48]). To calculate the closest available genome, a tetra correlation search (TCS) analysis was performed, along with average nucleotide identity (ANI) calculated using the JSpecies tool (Richter et al. [46]). Average amino acid identity (AAI) between all genomes were determined according to Konstantinidis and Tiedje [67], using the webserver available at <http://enve-omics.gatech.edu/> (Rodríguez-R and Konstantinidis [68]). The *Fischerella* MAG contigs were annotated with the RAST server (<http://rast.nmpdr.org/>; Meyer et al. [30]). The molecular weight of the predicted proteins was calculated using the Compute pI/Mw on the ExPASy web server (Gasteiger et al. [15]). The orthologous coding DNA sequences (CDS) for annotated MAGs and reference genomes in the NCBI database were defined by an all-versus-all BLAST in all pairwise genomes, using a 35% sequence similarity cut-off over a region of 50% of the query size.

16S rRNA reads, obtained from the metagenomes, were extracted using the Parallel-META v 2.4 tool and clustered at 98.7% similarity using the QIIME script pick.closed.reference.otus.py to obtain operational taxonomic units (OTUs). The representative sequences for each OTU were aligned using SINA (Pruesse et al. [42]) and then added to the SILVA REF123 reference database using the parsimony method, as implemented in the ARB software package (Ludwig et al. [27]). OTUs were clustered in operational phylogenetic units (OPUs) that represented the smallest monophyletic group of sequences containing OTU representatives together with the closest reference sequence, including the sequence of a type strain when possible (Mora-Ruiz et al. [34]). In general, due to the divergence in the internal 16S rRNA sequence, those separated by one OPU are considered to represent the same bacterial species (Mora-Ruiz et al. [34]).

Almost complete 16S rRNA genes were recovered from the *Fischerella* MAGs using the RNAmmer 1.2 Server (Lagesen et al. [26]). Phylogenetic reconstructions were performed using ARB software version 5.5 (Ludwig et al. [27]). Sequences were aligned using the SINA tool, and the 16S rRNA genes were added by parsimony to the SILVA_NR99_128 and LSURef_123 reference datasets. The closest relative sequences and nearly complete 16S rRNA genes obtained from the MAGs were used to reconstruct *de novo* trees using the neighbor-joining algorithm with Jukes–Cantor correction.

Results and discussion

Isolates and morphological characterization

Over three consecutive years, four different microbial mats were sampled from different locations along the thermal gradi-

ent between temperatures of 61 to 46 °C in Porcelana Hot Spring. During this time period, a total of 81 cyanobacterial isolates were recovered (Supplementary data; Table S1). Twenty strains of filamentous true-branching cyanobacteria were obtained from the 61 °C sample location in 2011, whereas 12 were obtained at 54 °C in 2013, 25 were obtained at 51 °C in 2012, and another 24 strains were obtained at 46 °C in 2012. Only 64 strains were analyzed using the three methods (see below), with this number varying since the culture conditions meant that some strains were no longer available (morphological characterization $n=65$; DGGE genetic marker profiling $n=75-77$ and MALDI-TOF MS $n=81$).

Morphological characterization of 65 isolates showed that strains exhibited uniseriate main trichomes and were true-branching, which indicated affiliation to the Order *Nostocales* and the *Hapalosiphonaceae* family (Komárek et al. [22]). The cell size parameters defined by Kaštovský and Johansen [21] classified all the Porcelana strains into the *M. laminosus* Cohn ex Kirchner 1898 morphology, which included a width of 8–10 μm for the main trichome cells and a width of 5–7 μm for branching cells. However, Kaštovský and Johansen [21] discussed that *M. laminosus* and *F. thermalis* were “almost certainly the same species typified by the same population, with only 39 years of temporal variation in sampling” and that the drawings of *F. thermalis* (type species for *Fischerella*) were subsequently used to characterize soil (mesophilic) strains. With regard to the secondary filaments, the T-branching type was the most common, and fewer Y-branching types were observed in all strains from Porcelana, including *F. thermalis* strain CHP1 (Alcamán et al. [1]). Y-di and T-branching was a distinctive feature of *Mastigocladus* genera for a long time, as described by Anagnostidis and Komárek [5], compared to just T-branching in *Fischerella* spp. However, this feature has been challenged by a more recent ultrastructural and developmental study performed by Nurnberg et al. [37], since the authors demonstrated that Y- and T-branching types were topologically similar and that this feature was the result of a randomized direction of cell elongation.

In addition, the delta (Δ) between the length and width of 20 main filament cells (MFCs) and 20 secondary filament cells (SFCs) was analyzed in 65 strains (Supplementary data; Table S1). Sixteen strains were not available as cultures at the time of this analysis, thus it was not possible to carry out morphological characterization. Measurements of MFCs and SFCs were used to separate

different strain morphologies on a four-axis plot based on filament type (Supplementary Fig. S1). Almost all strains fell into morphotypes I and II (>92%, Supplementary Table S1), in which the main filament cells were longer and the secondary filament cells were longer or wider, respectively. Examples of morphotypes I and II included strains P3.11.11 (Fig. 1a) and *F. thermalis* CHP1 (Fig. 1b), respectively. Furthermore, the P2.12.23 isolate corresponded to a third morphotype (III), which was characterized by wider main and secondary filament cells (Fig. 1c), while the P2.12.22 isolate corresponded to a fourth morphotype (IV) that was characterized by wider main cells but longer secondary filament cells (Fig. 1d). None of the retrieved morphotypes (I to IV) had a distribution related to the isolation temperature, as shown in Supplementary Fig. S1, but this could be related to the high amount of pleomorphism observed in true-branching cyanobacteria (Anagnostidis and Komárek [5]).

DGGE fingerprinting of marker genes

Fragments of the 16S ribosomal rRNA gene and the internal transcribed spacer (ITS) region of the ribosomal operon were amplified for 75 isolated strains. Six strains could not be analyzed using these techniques as the cultures were lost for DNA extraction. The resulting amplicons were resolved by DGGE and yielded a single band that represented a single allele for the 16S rRNA gene, and ITS region (Supplementary data; Figs. S2 and S3, respectively). The sequences obtained from the bands had a high identity compared to the sequences of thermal *Fischerella* present in the NCBI database (100% with KJ696694 in Alcamán et al. [2] and 96% with DQ786173 in Finsinger et al. [12], respectively). Furthermore, the *nifH* haplotyping strategy was used as it has proved useful in characterizing the phylogeography of *F. thermalis* strains (Miller et al. [31]). In Alcamán et al. [2], three different *nifH* gene bands were detected in environmental samples from Porcelana Hot Spring, and the sequences were affiliated to the thermal strains of *Mastigocladus/Fischerella* that corresponded to three synonymous alleles. During this study, only two of these alleles were detected from 77 isolates that were tested for the gene (Supplementary data; Fig. S4). This enabled differentiation between isolates from the upper (76% with only the CN1 allele) and lower (85% with only the CN3 allele) temperature sources throughout Porcelana's thermal gradient (Fig. 2a). The synonymous mutations found for the *nifH* gene

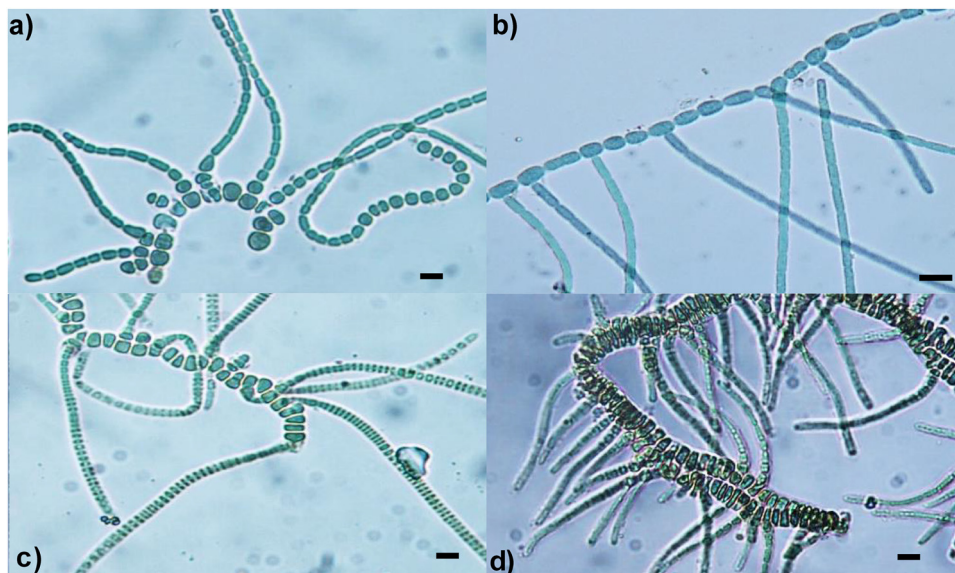


Fig. 1. Morphological characterization of Porcelana *Fischerella thermalis* strains. Images were obtained using light microscopy for the following morphotypes I–IV, with true branching filament representatives: P3.11.11 (a), *Fischerella thermalis* CHP1 (b), P2.12.23 (c) and P2.12.22 (d), respectively. Scale bar represents 10 μm .

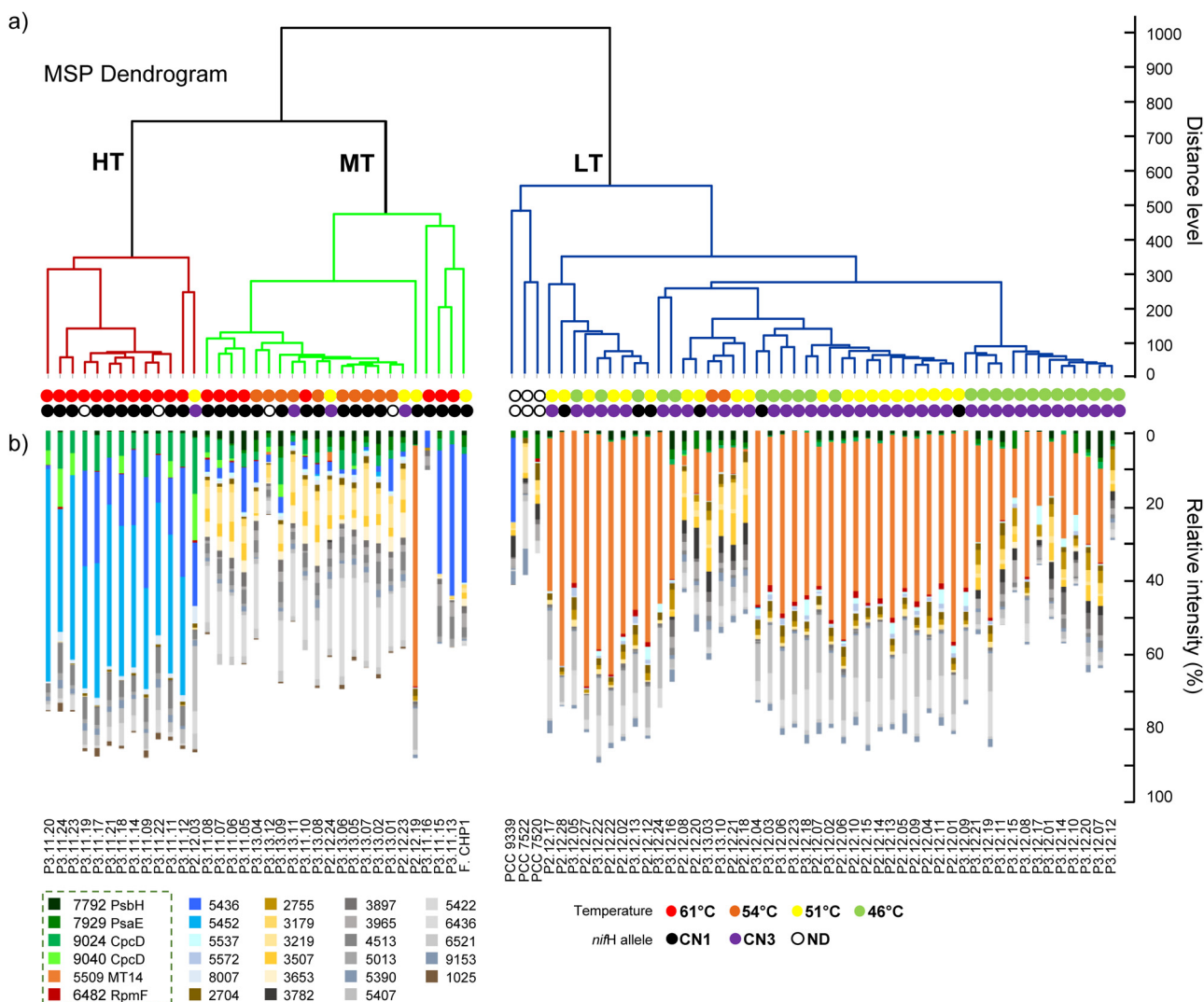


Fig. 2. *Fischerella thermalis* strain analysis by MALDI-TOF MS. (a) MSP dendrogram constructed by Biotyper 3.0 for the 81 *Fischerella* isolates and four reference strains. The respective isolation temperature for each strain is shown in red (61 °C), orange (54 °C), yellow (51 °C) or green (46 °C), and each strain's respective *nifH* gene allele (recovered by DGGE) is also shown, with the CN1 allele marked in black and the CN3 allele marked in purple. (b) Relative abundance of main mass peaks with >1% relative intensity within the 100 most important peaks for each strain. The legend shows the average mass (Da) of each peak and is framed for those that were associated with a predicted mass in the five analyzed proteomes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

could be fixed in these populations and this is a potential consequence of the divergence between strains at different temperatures in Porcelana (see below).

The DGGE CN2 band found in environmental samples was also found in some strains, but it did not represent a unique band in any of the isolates (Supplementary data; Fig. S4). This could be explained by the growth conditions in the laboratory, such as the 37 °C temperature, that could have selected for certain specific populations of *F. thermalis*, thus limiting isolation only to strains able to grow under these controlled laboratory conditions.

MALDI-TOF MS clustering

MALDI-TOF MS methodology has proved useful for clustering environmental strains (Viver et al. [57]), thus it was implemented to assess the phenotypic diversity of the 81 isolates in the current study. The main spectra profile (MSP) dendrogram obtained for all 81 isolates, and four reference strains, indicated a division of the isolates into three main clusters (Fig. 2a). These clusters were named according to the temperature at which the strains were

isolated. The high temperature (HT) cluster was composed almost exclusively of strains isolated at 61 °C ($n = 12$), together with a 51 °C strain; the mid temperature (MT) cluster was composed of strains obtained from a wider isolation temperature range (8 at 61 °C, 10 at 54 °C and 3 at 51 °C; including *F. thermalis* strain CHP1); and the low temperature (LT) cluster was mostly composed of strains isolated at lower temperatures (2 at 54 °C, 21 at 51 °C and 24 at 46 °C; including the three PCC reference strains), as shown in Fig. 2a. The patterns observed for the *nifH* allele corresponded to different MALDI cluster proteomes (Fig. 2), with HT and MT clusters representing strains with mainly the CN1 *nifH* DGGE band, and LT clusters predominantly correlating with strains of the CN3 *nifH* allele (Fig. 2a). A synonymous polymorphism at the third base of Leu80 of the *nifH* gene was identified. This differentiated the HT-MT from the LT temperature strains and was similar to a spatial divergence that has been previously observed in another hot spring study (Miller et al. [33]). The CN3 (A) polymorphism in Leu80 was shared by 69% of the Yellowstone lower temperature strains (43 and 39 °C, $n = 13$), whereas the CN1 (G) polymorphism was the same in 80% of the upper temperature strains (54, 51 and 46.5 °C, $n = 20$) obtained from

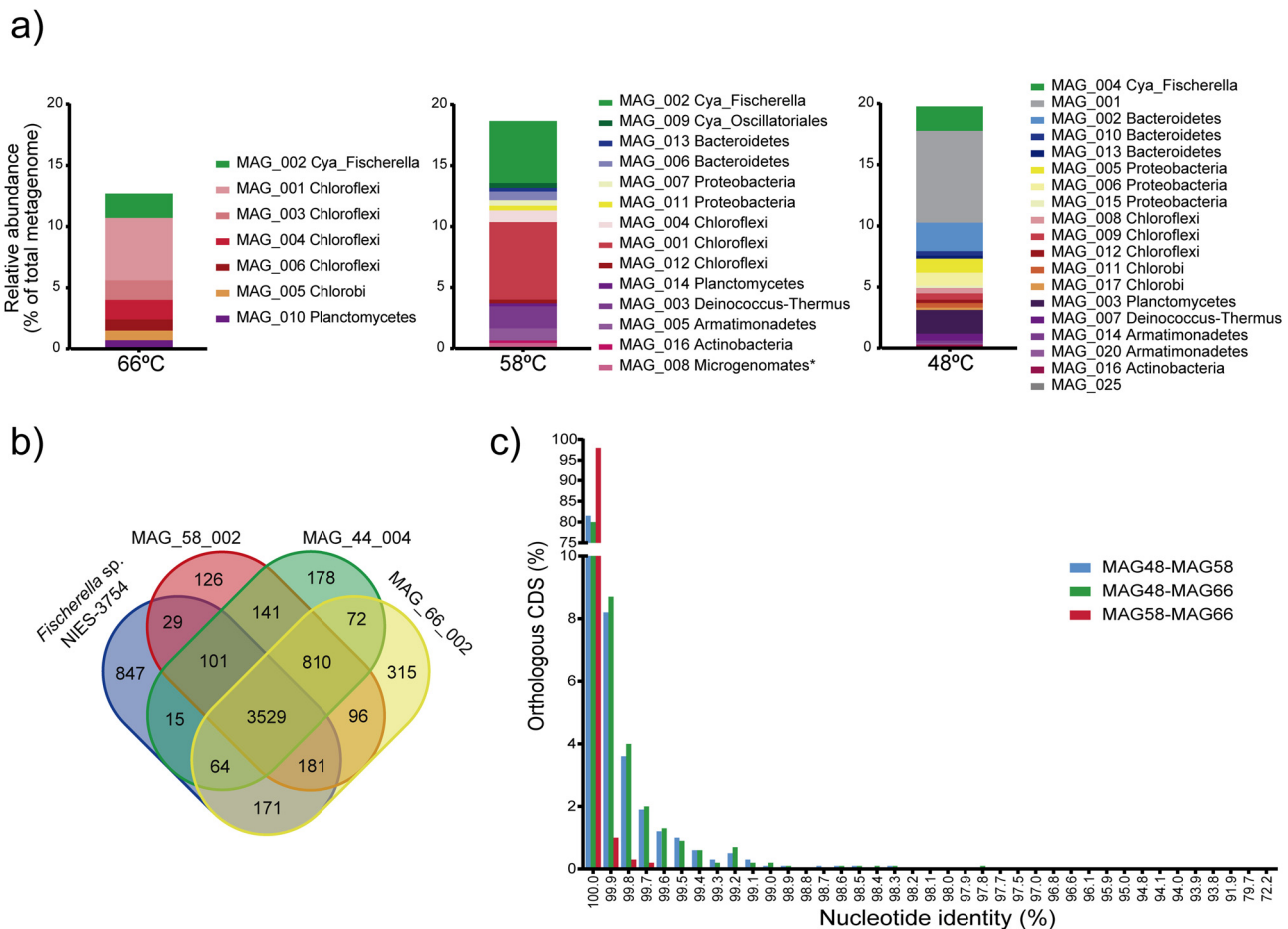


Fig. 3. *Fischerella thermalis* MAGs recovery from Porcelana Hot Spring metagenomes. (a) Relative abundance of MAGs in Porcelana Hot Spring. These were recovered in the *de novo* assembly by MaxBin, assigned by JSpecies, and the abundance was calculated using the script BlastTab.seqdepth.pl (see Methods). (b) Venn diagram of orthologous (all-versus-all BLAST in all pairwise genomes using a 35% sequence similarity cut-off over a region of 50% of the query size) and unique predicted CDS between the closely related genomes of *Fischerella* sp. NIES-3754 and Porcelana MAGs. (c) Orthologous CDS nucleotide identity between *Fischerella thermalis* MAGs. Based on BLAST all-versus-all comparisons, orthologous CDS between each pair of MAGs were grouped across a range of 0.1 nucleotide identity. Only the nucleotide identity is shown for which there was a pair of orthologous CDS with that value. The total orthologous CDS compared for each pair of MAGs was 4313 for 66–58 °C, 4157 for 66–48 °C and 4392 for 58–48 °C.

White Creek hot spring in YNP, USA (Miller et al. [33]). In the same YNP hot spring, divergence was also observed in other genomic regions, including the *rfbC* gene involved in LPS biosynthesis (Wall et al. [58]).

F. thermalis populations along the Porcelana thermal gradient based on metagenomics

The above results suggested that there was potential divergence of *F. thermalis* along the thermal gradient and that this ecological speciation could be related to specific proteomic adaptations. Three mat metagenomes collected in 2013 at 66, 58, and 48 °C (NCBI BioProject accession no. NA382437; Alcamán-Arias et al. submitted [3]) were analyzed to provide a genomic perspective of the similarities and differences of cyanobacterial populations in microbial mats from Porcelana. First, the global bacterial and cyanobacterial diversities were determined in OPUs, according to 16S rRNA gene counts. For cyanobacteria, the relative abundances of the total 16S rRNA fragments were 7.6, 13.3 and 11.4%, respectively, for the 66, 58, and 48 °C metagenomes (Supplementary data; Fig. S5a). Other bacterial phyla, including *Proteobacteria* (17.5–40.2%), *Chloroflexi* (35.1–19.8%) and *Acidobacteria* (4.4–6.5%), also co-occurred in the metagenomes. Within cyanobacterial OPUs, *Stigonematales* (true-branching cyanobacteria or classic subsection V from Rippka et al. [47]) represented 44.6, 46.7 and 26.8% of the

sequences, respectively (Supplementary data; Fig. S5b). *Oscillatoriales* and *Thermosynechococcales* sequences were also observed, which was similar to previous results from Porcelana Hot Spring (Mackenzie et al. [28]; Alcamán-Arias et al. submitted [3]).

The three metagenomes from 66, 58, and 48 °C (Alcamán-Arias et al. submitted [3]) were *de novo* assembled, and 7, 14 and 19 metagenome assembled genomes (MAGs) were recovered, respectively. One MAG in each metagenome could be binned and identified as *F. thermalis*. The total reads mapped to each *F. thermalis* MAG showed that the relative abundances were 2.0%, 5.1% and 2.0% of the total metagenomes, respectively (Fig. 3a). The MAGs contained approximately 699 ± 23 contigs that had a genome completeness above 96%, a size between 5.1 to 5.6 and a GC mol% of $41.03 \pm 0.09\%$ (Supplementary data; Table S2). These characteristics were similar to the only complete genome of thermal true-branching cyanobacteria sequenced to date corresponding to *Fischerella* sp. NIES-3754 (one closed chromosome, 3 plasmids with a total size of 5.8 Mb, and a GC mol% content of 40.97%; Hirose et al. [17]). In this current study, the complete 16S rRNA gene sequences obtained from MAGs were associated with *Fischerella*/*Mastigocladus* 16S rRNA gene sequences obtained from the SILVA database, which included environmental sequences collected from diverse hot springs from around the world (Supplementary data; Fig. S6). The 16S rRNA sequences from MAGs also showed 100% identity with the 16S rRNA DGGE band sequence

obtained from a Porcelana mat and from the isolated strains (NCBI KJ696694). These data indicated that all thermophilic *F. thermalis* sequences were closely affiliated (Kaštovský and Johansen [21]; Alcámán et al. [1]).

It is likely that the binning approach, rather than retrieving one single clonal variety, retrieved genomes representing a “pangenome” of the major *F. thermalis* populations thriving within each temperature in the Porcelana mat (Konstantinidis et al. [24]). The ANI calculated between these MAGs exceeded 99.8%, indicating high identity between the *F. thermalis* pangenomes obtained from each temperature over the thermal gradient (Table 1). Additionally, *F. thermalis* CCMEE 5194, a strain from Puyehue Hot Spring, Chile (Miller et al. [31]), was the closest available draft genome in the NCBI database with >99.6% ANI (Table 1; Supplementary Fig. S7). Since most of the available *Fischerella* spp. genomes are in a draft state, the ANI and other genomic features (see below) were compared with the only complete genome from this genus, *Fischerella* sp. NIES-3754 with an ANI >96.8% (Table 1). From a genomic perspective, by the ANI (Supplementary Fig. S7 and Table 1) and AAI (Table 1) values, the 35 *F. thermalis* genomes, the *Fischerella* sp. NIES-3754 and the three Porcelana MAGs might represent a single species (95–96% identity threshold; Richter and Rosselló-Móra [45]; Rosselló-Móra and Amann [49]), and could be named according to *F. thermalis* PCC 7521 denominated as the representative genome in the NCBI database. When clustering was performed with available genomes using ANI values, the results indicated the presence of two clades. One represented by thermophilic *Fischerella* genomes, and another included the mesophilic *Fischerella* species with ANI values of ~91% (Supplementary Fig. S7; Table 1). These results supported the thermophilic-mesophilic separation observed in the phylogeny presented by Kaštovský and Johansen [21] and in Supplementary Fig. S6 by 16S rRNA phylogeny. The ANI identity between the thermal and mesophilic clades of *Fischerella* (that include *Fischerella* sp. and *F. muscicola* genomes) was lower than 95–96%, and the thermophilic strains had >96% ANI; thus, according to proposed genomic thresholds (Rosselló-Móra and Amann [49]), these results suggested that both groups were different species from the same genus *Fischerella*, with the thermophilic strains being identified as *F. thermalis* (Table 1).

Additionally, the major features of *F. thermalis* MAGs and *Fischerella* sp. NIES-3754 (Hirose et al. [17]) were compared. The latter had 4599 predicted CDS, while the MAGs had 4975, 4886 and 4739 predicted CDS at 66, 58 and 48 °C, respectively. The core genome was represented by 3529 orthologous CDS between four cyanobacterial genomes (Fig. 3b). The three MAGs from Porcelana shared 810 predicted CDS that were not present in the NIES-3754 genome, and had 315, 126 and 178 unique CDS at 66 °C, 58 °C and 48 °C, respectively. In addition, the Cluster of Orthologous Genes database (COG; Galperin et al. [14]) was used to compare the number of exclusive predicted CDS in the three MAGs. The number of exclusive genes with predicted functions was observed to be higher in the 66 °C MAG (Supplementary data; Table S3). The 66 °C MAG contained 23% of the exclusive genes related to metabolism, whereas only 13% and 10% were found in the 58 and 48 °C MAGs, respectively, mainly related to inorganic compound transport and CO₂ uptake (e.g. the *hatR* gene). Furthermore, more genes related to cellular processes and signaling exclusive CDS were present in the 66 °C MAG (16%, compared to 7% and 2% in the 58 and 48 °C MAGs, respectively), and they were dominated by genes with regulatory functions, such as *hrcA*, a heat inducible transcriptional repressor. Additionally, more genes related to genetic information processing were found in the 66 °C MAG, compared to MAGs at other temperatures (7%, compared to 3% and 1% in the 58 and 48 °C MAGs, respectively), including a high number of mobile element proteins and DNA repair genes. Therefore, the 66 °C MAG contained an

Table 1
Wide genomic identity comparison. Wide genomic comparisons between eight *Fischerella* genomes available in the NCBI database and *Fischerella* MAGs recovered from Porcelana metagenomes. *Nostoc* sp. PCC 7120 was used for comparison with a heterocystous non-branching cyanobacteria. Average nucleotide identity (ANI) and [% aligned genomes] between genomes are given in the lower triangle; average amino acid identity (AAI) and [number of genes shared] are given in the upper triangle. ^M Strains isolated from non-thermal environments (Rippka et al., 1979 [47]).

	MAG.48	MAG.58	MAG.66	CCMEE 5194	PCC 7521	JSC-11	NIES-3754	PCC 7414	PCC 73103	PCC 9431	PCC 9339	PCC 7120
MAG.48	*											
MAG.58	99.90 [94.89]											
MAG.66	99.86 [90.61]	99.73 [45.82]										
<i>Fischerella thermalis</i> CCMEE 5194	99.67 [83.87]	99.63 [85.26]	99.55 [87.15]									
<i>Fischerella thermalis</i> PCC 7521	98.69 [86.22]	98.59 [89.41]	98.48 [91.51]	98.80 [84.35]								
<i>Fischerella</i> sp. JSC-11	98.69 [86.80]	98.59 [89.97]	98.50 [92.12]	98.80 [83.95]	99.77 [98.72]							
<i>Fischerella</i> sp. NIES-3754	96.97 [79.64]	96.89 [82.77]	96.80 [85.08]	97.14 [75.00]	97.24 [87.20]	97.22 [86.01]						
<i>Fischerella muscicola</i> PCC 7414	93.62 [62.70]	93.62 [65.25]	93.59 [67.32]	93.71 [55.93]	93.50 [68.73]	93.45 [67.87]	93.74 [72.03]					
<i>Fischerella muscicola</i> PCC 73103 ^M	87.92 [45.59]	87.89 [46.74]	87.86 [48.60]	86.25 [36.58]	87.87 [50.29]	87.83 [49.64]	87.82 [50.36]	88.12 [56.54]				
<i>Fischerella</i> sp. PCC 9431 ^M	87.50 [44.66]	87.48 [45.52]	87.44 [47.21]	85.67 [33.83]	87.41 [48.31]	87.43 [48.09]	87.46 [48.56]	87.68 [53.85]	91.76 [71.22]	*		
<i>Fischerella</i> sp. PCC 9339 ^M	87.70 [40.49]	87.66 [41.46]	87.63 [42.86]	85.88 [33.74]	87.61 [43.78]	87.61 [43.31]	87.92 [46.34]	87.95 [50.26]	92.49 [64.55]	91.22 [59.91]	89.57 [43.99]	
<i>Nostoc</i> sp. PCC 7120 ^M	83.26 [2.95]	83.23 [3.14]	83.18 [3.17]	76.53 [4.82]	82.64 [3.22]	82.47 [3.14]	87.81 [5.55]	84.77 [4.37]	83.28 [3.50]	83.32 [3.38]	87.83 [5.71]	*

apparently distinctive and specific set of accessory genes exhibiting genomic divergence along the thermal gradient.

MALDI-TOF MS mass peak comparison against *Fischerella* genomes and Porcelana MAGs

A bioinformatic approach was implemented to determine the most abundant proteins observed in the MALDI-TOF MS profiles. This methodology has recently been used to separate toxic and non-toxic strains of the cyanobacteria *Microcystis aeruginosa* according to ribosomal proteins and their masses predicted from a reference genome (Sun et al. [53]). In this current study, protein profiles of 81 Porcelana strains were analyzed by comparing the more abundant masses with predicted proteomes from the *Fischerella* sp. JSC-11 and *Fischerella* sp. NIES-3754 genomes, together with the three *F. thermalis* MAGs from the study.

The most abundant proteins (29 mass peaks with >1% relative intensity, within a 4-Da range and present in at least 15 MALDI-TOF MS profiles) only comprised an average of $76.78\% \pm 7.68$ of the relative intensity (RI), due to the abundance of lower intensity peaks. When the 29 mass peaks (Table 2) were compared by the MALDI-TOF MS clusters, a total of 9, 10 and 10 peaks had a higher RI for the HT, MT and LT strains, respectively. For the temperature of isolation, the 54 °C strains had the most peaks with a higher RI, with a total of 14 peaks, compared to 4, 3 and 6 peaks in the 46, 51 and 61 °C isolated strains, respectively (Table 2). The number of peaks with masses that matched the molecular weight of proteins predicted from CDS with an associated function (non-hypothetical proteins), was 18 in the JSC-11 or NIES-3754 proteomes, whereas for *F. thermalis* MAGs the total was 17. Table 3 also shows the number of other peaks associated with hypothetical proteins, however, these were not considered for further analysis.

The most distinctive peak was 5509 Da, which predominated in the LT cluster (Table 3). The mass of the peak was associated in the five proteomes analyzed with a predicted metallothionein family 14 protein, reported to be involved in metal resistance and oxidative stress responses (Shimizu et al. [51]; Turner and Robinson [55]). Furthermore, the 6482 Da peak was associated with ribosomal protein L32 from the reference genomes and MAGs (Table 3). In addition, this protein was the most abundant ribosomal protein observed in the freshwater cyanobacteria *Microcystis* sp. from the first classification study using MALDI-TOF MS in cyanobacteria (Sun et al. [53]). This result, considering the presence of the most abundant ribosomal protein, supported the previous cyanobacterial study, even though strains came from different orders and environments.

Furthermore, four recurring mass peaks represented up to 30.51% of the total RI in the strains (e.g. strain P3.11.20), and they were associated with the photosynthetic apparatus in reference and MAG proteomes, as shown in Fig. 2b. The mean RI for each cluster, based on what was assumed to be photosynthetic proteins (7792 Da, PsbH/ApcC; 7929 Da, Psae and 9024–9040 Da, CpcD; Table 3), significantly differed between groups (p value <0.0001, Kruskal–Wallis test), with the RI for HT, MT, and LT being 12.60% ($SD \pm 6.67$), 7.49% ($SD \pm 3.71$), and 3.04% ($SD \pm 2.55$), respectively. An elevated abundance of these proteins could be related to the sensitivity of phycobiliproteins and Chl-*a* to high temperatures (Zhao and Brand [66]; Radway et al. [43]). The need for stabilization functions in cyanobacteria has been reported previously (e.g. Psae (Jeanjean et al. [20]) and PsbH (Komenda et al. [23]) in *Synechocystis*; CpcD in all cyanobacteria, including early studies in *M. laminosus* (Guan et al. [16]; Füglistaller et al. [13]; Reuter et al. [44]); and ApcD in *Synechococcus* (Dong et al. [10]).

Other mass peaks were associated with the photosynthetic apparatus and carbon fixation in the MAGs, but not in the reference genomes: the 5436 Da (NdhC, NAD(P)H-quinone oxidoreductase

subunit (3) and the 8008 Da (CP12, protein CP12 involved in regulation of Calvin cycle) were predominant in the HT cluster; and the 5537 Da (PsbF, cytochrome b559 beta chain) together with the 5572 Da (HatR, high-affinity carbon uptake protein) were more abundant in the LT cluster. Conversely, the 5452 Da mass peak was the most abundant protein in the HT cluster, and was associated with the Psal protein (photosystem I reaction center subunit VIII) in the reference proteomes, but was not related to any masses in the MAG proteomes. For most of the additional 18 mass peaks identified in this study, each representing less than 8% of the total RI, multiple matches were found (see Table 3). The high frequency of CDS in the genomes without a known function (~60% of the <20 kDa predicted CDS of the two reference proteomes analyzed) led to difficulties in associating the peak mass with a predicted functional protein, thus preventing the application of an effective bioinformatic approach to the MALDI-TOF MS results.

Sequences of the CDS associated with a mass peak from the reference proteomes were identified and compared with the three MAGs. A total of 11 orthologous sequences (*psbT*, *DHS*, *glnB*, *hatR*, *MT-14*, *rpmF*, *csbD*, *apcD*, *psaE*, *cpcD*, *rpoZ* and *acpP*) had 100% identity at the amino acid level (Supplementary Table S4). Considering that the strains were isolated from Porcelana mats along the same thermal gradient as the metagenomic *Fischerella* MAGs, it was expected that these *F. thermalis* isolates would represent populations also observed in the metagenomes. Subsequently, similar masses were observed between proteins detected and identified from the strains using MALDI-TOF MS and the orthologous sequences of the MAGs, reinforcing this method as an effective way to identify proteins.

Distinguishing *F. thermalis* ecotypes along the temperature gradient in Porcelana

According to the ecotype definition presented by Cohan and Perry [7], the *F. thermalis* population could represent at least two different ecotypes. This is based on observations from the present analysis predicting three different proteomic and two *nifH* gene allele clusters observed in the 81 strains. Similarly, Porcelana *F. thermalis* pangenomes (which reflect the *in situ* occurrence of the strains on the mat) differed along the thermal gradient. For example, in addition to being within the ANI species threshold, 97% of the observed orthologous CDS between the 66 and 58 °C MAGs shared 100% identity at the nucleotide level. Moreover, when comparing 58 and 48 °C or 66 and 48 °C, greater divergence was observed between these two pairs, with only ~80% of CDS showing complete matches (Fig. 3c). Hence, at the genomic level, two ecotypes could be distinguished. The group obtained at 66 and 58 °C (the HT pangenome) was distinct from the group obtained at 48 °C (the LT pangenome), which supported the observed difference in the proteomic and *nifH* gene allele results.

These data collectively indicated that filamentous true-branching cyanobacteria inhabiting hot springs were likely to be organisms subjected to ecological speciation along the temperature gradient. Similar observations have been made for the unicellular *Synechococcales* (Ward et al. [60,59]). In *Synechococcales*, different ecotypes have been reported over a temperature gradient (Ward et al. [60,59], Allewalt et al. [4]) and according to light intensity (Ward et al. [60,59], Olsen et al. [38]). In the present study, for *F. thermalis*, temperature and light intensity could also be related to ecotypes, since a greater abundance of photosynthetic proteins were observed in strains isolated at higher temperatures. Herein, at least two potential ecotypes of *F. thermalis* strains are described, which exhibited differences in their proteomic profiles and their exclusive gene content across pangenomes.

Table 2
MALDI TOF-MS main peaks identification. The peaks were selected as the masses with a relative intensity >1%, within a 4 Da window and repeated in at least 15 strains from MALDI-TOF MS profiles for 81 isolates and 4 reference *Fischerella*. A statistical analysis was performed according to the recovered MALDI-TOF MS data. The groups did not have the same sample size, but their distributions were corroborated by a D'Agostino and Pearson normality test. Groups in which fewer than 50% of the samples had a detected peak were not considered for further analysis.

Selected peaks	Relative intensity in all strains (n=85)				Presence by MALDI-TOF MS cluster (%)				Presence by temperature of isolation (%)				
Average mass (Da)	Mass SD (+)	Presence (%)	Average intensity (%)	Intensity SD (%)	HT (n=13)	MT (n=22)	LT (n=50)	MALDI-TOF MS cluster predominance	61 °C (n=20)	54 °C (n=12)	51 °C (n=25)	46 °C (n=24)	Temperature predominance
2704	0.72	84.7	1.68	0.93	77	68	88	LT ^c	70	92	92	88	54-51-46 ^b
2755	0.62	52.9	2.18	1.42	–	8	84	LT ^a	5	17	76	92	46 ^d
3179	0.58	45.9	2.52	1.92	–	68	43	LT ^a	30	83	32	48	54 ^a
3219	0.57	88.2	2.29	2.78	69	84	88	MT ^c	75	100	96	84	54 ^b
3507	0.48	43.5	3.33	2.07	8	72	35	MT ^a	35	92	24	40	54 ^a
3653	0.81	34.1	2.50	2.11	15	80	14	MT ^a	45	83	12	28	54 ^a
3782	0.64	25.9	3.35	2.09	–	8	39	–	–	17	20	48	–
3897	0.56	92.9	1.36	1.34	77	88	92	MT ^c	85	100	100	88	NS ^c
3965	0.54	62.4	1.97	1.55	8	88	59	NS ^d	45	100	44	72	54 ^d
4513	1.26	89.4	2.42	2.32	100	88	80	HT-MT ^c	95	100	84	88	61-54 ^c
5013	0.63	38.8	1.28	0.47	92	80	2	NS ^d	90	83	12	4	NS ^d
5390	0.94	82.4	1.07	0.68	100	76	75	HT ^c	90	92	76	80	61-54 ^b
5407	0.73	91.8	6.68	5.36	100	72	92	LT ^c	85	100	96	92	51-46 ^b
5422	0.91	32.9	1.41	1.47	62	–	39	HT ^a	40	17	28	40	–
5436	1.62	38.8	12.39	12.32	85	84	2	HT ^e	85	83	12	4	61 ^e
5452	0.54	20.0	28.31	19.80	92	12	4	HT ^a	55	25	4	4	61 ^a
5509	1.12	57.6	36.94	19.14	–	12	90	LT ^a	–	25	92	92	NS ^d
5537	1.27	63.5	1.19	1.28	–	68	73	LT ^e	25	92	64	84	NS ^b
5572	1.24	37.6	0.95	0.55	–	4	61	LT ^a	–	–	68	60	NS ^d
6436	0.77	91.8	7.39	6.59	100	76	90	MT ^c	85	100	96	84	54 ^b
6482	0.76	61.2	0.70	0.52	77	32	67	LT ^d	60	17	80	64	51-46 ^b
6521	0.47	68.2	1.06	0.93	38	72	69	MT ^d	45	92	84	68	54 ^c
7792	0.73	91.8	1.74	1.46	100	84	86	MT-LT ^c	90	100	100	80	NS ^c
7929	0.67	80.0	1.35	1.56	77	80	75	MT-LT ^c	70	100	92	60	54 ^c
8008	1.34	70.6	0.82	0.79	100	76	55	HT ^c	85	92	72	52	61-54 ^c
9024	1.10	84.7	2.76	3.38	100	84	75	HT-MT ^c	95	100	92	68	61-54 ^c
9040	0.72	18.8	3.41	3.77	46	28	6	–	45	25	8	8	–
9153	0.81	56.5	1.74	1.39	–	4	92	LT ^a	–	17	88	88	NS ^e
10026	0.89	34.1	0.80	0.64	100	60	2	HT ^d	75	83	12	–	NS ^d

^a Predominance assumed by presence in only one group.

^b One-way ANOVA test with Tukey's multiple comparison test ($p < 0.05$).

^c Non-parametric Kruskal–Wallis test with Dunn's multiple comparison test ($p < 0.05$).

^d Student's unpaired t-test with Welch's correction (different SD) ($p < 0.05$).

^e Kolmogorov–Smirnov test ($p < 0.05$).

Table 3

MALDI-TOF MS masses against predicted molecular weight from CDS of reference and MAGs genomes. The masses were searched in a 6 Da window in the predicted molecular weight of CDS from the two reference genomes (*Fischerella* sp. JSC-11 and *Fischerella* sp. NIES-3754) and the three *F. thermalis* MAGs obtained in this study, including the following PTMs: N-terminal methionine presence, N-formyl methionine presence, one phosphoryl, one acetyl, one hydroxyl, one glycosyl and one attached heme group. ^a Parentheses represents the number of genomes where the function was found. For reference genomes (n = 2) and *Fischerella* MAGs (n = 3). The putative protein names were abbreviated as follow: PstT, Photosystem II reaction center protein T; SDR, Short-chain dehydrogenase reductase; DHS, Deoxyhypusine synthase-like protein; GlnB, Nitrogen regulatory protein P-II; PsaI, Photosystem I reaction center subunit VIII; MT14, Metallothionein family 14; WD40, WD-40 repeat protein; RpmF, 50S ribosomal protein L32; TatA, Sec-independent protein translocase protein; CsbD, CsbD family protein; TIS4, Transposase IS4 family protein (Fragment); PstH, Photosystem II reaction center protein H; ApcC, Phycobilisome 7.8 kDa linker polypeptide, allophycocyanin-associated, core; PsaE, Photosystem I reaction center subunit IV; EmrB, EmrB subfamily drug resistance transporter; TIS701, Transposase IS701 family protein; Cas1, CRISPR-associated protein Cas1; CpcD, CpcD phycobilisome linker domain protein; HypC, Hydrogenase assembly chaperone hypC/hupF; RpoZ, DNA-directed RNA polymerase subunit omega (EC 2.7.7.6); NifU, Nitrogen-fixing NifU domain-containing protein; AcpP, Acyl carrier protein; Gst1, Glutathione S-transferase; OrfB, Transposase IS605; GlsA, Glutaminase (EC 3.5.1.2); OppA, Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein (TC 3.A.1.5.1); MEP, Mobile element protein; GrxC, Glutaredoxin 3 (Grx1); CyaA, denylate cyclase (EC 4.6.1.1); XseA, Exodeoxyribonuclease VII large subunit (EC 3.1.11.6); PstS, Phosphate ABC transporter, periplasmic phosphate-binding protein (TC 3.A.1.7.1); all5100, UPF0192 protein all5100 precursor; MetN, Methionine ABC transporter ATP-binding protein; RCM, Regulator of cell morphogenesis and NO signaling; NdhC, NAD(P)H-quinone oxidoreductase subunit 3 (EC 1.6.5.2); CypX1, cytochrome P450; HemJ, Protoporphyrinogen IX oxidase, novel form, HemJ; PstF, Cytochrome b559 beta chain; HatR, High-affinity carbon uptake protein Hat/HatR; CpcG, Phycobilisome small core linker polypeptide; FtsA, Coenzyme F390 synthetase; RbcX, Possible RuBisCo chaperonin; DppC, ABC-type dipeptide/oligopeptide/nickel transport systems, permease components; DevB, heterocyst specific ABC-transporter, membrane fusion protein DevB homolog; CP12, Protein CP12, regulation of Calvin cycle via association/dissociation of PRK/CP12/GAPDH complex; HypC, [NiFe] hydrogenase metallocenter assembly protein; PatB/Orf2, PatB and ORF2 genes; Thij, Thij/Pfpl family protein; RfbW, Glycosyl transferase, group 1.

Peak average mass (Da)	Predominance in MALDI-TOF MS cluster (Column is derived from Table 2)	Mass match against reference genome (n)	Mass match reference genome – non hypothetical protein	Protein name in reference genomes ^a	Mass match against Porcelana MAGs (n)	Mass match Porcelana MAGs – non hypothetical protein	Protein name in three <i>Fischerella</i> MAGs ^a
2704	LT	–	–	–	–	–	–
2755	LT	–	–	–	–	–	–
3179	LT	–	–	–	–	–	–
3219	MT	1	1	(1) SDR – phosphoryl	–	–	–
3507	MT	2	0	–	–	–	–
3653	MT	1	0	–	–	–	–
3782	–	1	0	–	–	–	–
3897	MT	5	2	(2) PstT – native	1	0	–
3965	NS	3	0	–	3	0	–
4513	HT-MT	7	2	(2) PstT – hemyl	51	4	(3) GlsA – glycosyl; (1) OppA – Nformyl
5013	NS	3	1	(1) DHS – native	53	6	(3) MEP – acetyl; (1) GrxC – hemyl; (1) CyaA – acetyl; (1) XseA – native
5390	HT	5	0	–	33	6	(3) PstS – N-formyl; (3) all5100 – non N-methionine
5407	LT	5	1	(1) GlnB – native	17	3	(3) MetN – glycosyl
5422	HT	8	1	(1) GlnB – methyl	50	1	(1) RCM – native
5436	HT	8	1	(1) GlnB – N-formyl	26	4	(2) NdhC – hemyl; (1) CypX1 – phosphoryl; (1) RCM – hydroxyl
5452	HT	3	2	(2) PsaI – hemyl	11	0	–
5509	LT	7	1	(1) MT14 – non N-methionine	31	4	(3) MT14 – non N-methionine; (1) HemJ – phosphoryl
5537	LT	4	0	–	25	8	(3) PstF – hemyl; (3) all5100 – hydroxyl; (2) PstS – hemyl
5572	LT	2	1	(1) WD40 – acetyl	14	3	(3) HatR – acetyl
6436	MT	5	0	–	18	0	–
6482	LT	8	3	(2) RpmF – non N-methionine; (1) TatA – native	22	3	(3) RpmF – non N-methionine
6521	MT	13	4	(2) CsbD – non N-methionine; (1) TIS4 – hemyl; (1) TatA – native	12	0	–

7792	MT-LT	10	4	(2) PsbH – glycosyl; (2) ApcC – native	18	13	(3) PsbH – glycosyl; (3) CpcG – native; (3) FtsA – non N-methionine; (2) RbcX – hydroxyl; (1) PstS – non N-methionine; (1) DppC – hydroxyl
7229	MT-LT	13	4	(1) PsaE – non N-methionine; (1) EmrB – acetyl; (1) TIS701 – N-formyl; (1) Cas1 – methyl	19	5	(3) PsaE – non N-methionine; (1) DevB – glycosyl; (1) Permease – non N-methionine
8008	HT	6	0	–	6	2	(2) CP12 – non N-methionine
9024	HT-MT	2	4	(2) CpcD – native; (1) TatA – native; (1) HypC – hydroxyl	17	7	(3) HypC – hydroxyl; (2) CpcD – native; (2) PatB/Orf2 – hydroxyl
9040	–	14	5	(2) RpoZ – non N-methionine; (1) NifU – hemy; (1) CpcD – hydroxyl; (1) TatA – methyl	8	3	(3) CpcD – hydroxyl
9153	LT	10	3	(2) AcpP – native; (1) Gst1 – phosphoryl	13	4	(3) ThiJ – methyl; (1) AcpP – native
10026	HT	3	1	(1) OrfB – glycosyl	2	2	(2) RfbW – phosphoryl

Conclusions

True-branching cyanobacterial strains from Porcelana Hot Spring, isolated along a temperature gradient, were identified both morphologically and phylogenetically to be *F. thermalis*. Subsequently, they were divided into two MALDI-TOF MS cluster profiles with different *nifH* alleles. Each division could be associated with specific temperatures along the thermal gradient. Compared to strains adapted to lower temperatures, those adapted to increased temperatures seemed to have more stable photosynthetic machinery due to a high abundance of phycobilisome linkers, or faster turnover of photosystem proteins. These characteristics may assist in the adaptation of the photosynthetic apparatus to combat high temperatures. Moreover, the pangenomes recovered from *F. thermalis* populations revealed that the abundance of exclusive genes was higher in the high temperature MAG. In addition, differences existed in coding sequence mutations between higher and lower temperature pangenomes. These differences could potentially be related to divergence, facilitating better performance at high temperatures or in a highly dynamic temperature niche. The metagenomic comparison of *Fischerella* MAGs, together with the identification of proteomic differences and synonymous fixed mutations, suggested that ecological speciation occurred in *F. thermalis* strains, and that this process had been modulated across the temperature gradient. Similar situations have been observed in other global regions for thermophilic cyanobacteria, such as *Synechococcus* sp., that have evolved in response to environmental factors.

Acknowledgements

We thank Huinay Scientific Field Station for providing technical support in the field. This work was financially supported by FONDECYT N° 1110696 and 1150171, and FONDAP N° 15110009. Additionally, we thank Ana Cifuentes and Mercedes Urdiain for performing sample preparation and the MALDI-TOF MS analysis. RRM and TVP acknowledge the Spanish Ministry of Economy for research project CLG2015.66686-C3-1-P and a pre-doctoral fellowship (N° BES-2013-064420), respectively.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.syapm.2018.05.006>.

References

- [1] Alcamán, M.E., Alcorta, J., Bergman, B., Vázquez, M., Polz, M., Díez, B. (2017) Physiological and gene expression responses to nitrogen regimes and temperatures in *Mastigocladus* sp. Strain CHP1, a predominant thermotolerant cyanobacterium of hot springs. *Syst. Appl. Microbiol.* 40, 102–113.
- [2] Alcamán, M.E., Fernández, C., Delgado, A., Bergman, B., Díez, B. (2015) The cyanobacterium *Mastigocladus* fulfills the nitrogen demand of a terrestrial hot spring microbial mat. *ISME J.* 9, 2290–2303.
- [3] Alcamán-Arias, M.E., Pedrós-Alió, C., Tamames, J., Fernández C., Pérez-Pantoja D., Vázquez M., Díez, B. Diurnal changes in active carbon and nitrogen pathways along the temperature gradient in Porcelana Hot Spring microbial mat. *Front. Microbiol.* (submitted for publication).
- [4] Allewalt, J., Bateson, M., Revsbech, N., Slack, K., Ward, D. (2006) Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the Octopus Spring microbial mat community of Yellowstone National Park. *Appl. Environ. Microbiol.* 72, 544–550.
- [5] Anagnostidis, K., Komárek, J. (1990) Modern approach to the classification system of Cyanophytes 5 – Stigonematales. *Arch. Hydrobiol.* 86 (Suppl), 1–73.
- [6] Castenholz, R. (1969) Thermophilic blue-green algae and the thermal environment. *Bacteriol. Rev.* 33, 476–504.
- [7] Cohan, F., Perry, E. (2007) A systematics for discovering the fundamental units of bacterial diversity. *Curr. Biol.* 17, R373–R384.
- [8] Dagan, T., Roettger, M., Stucken, K., Landan, G., Koch, R., Major, P., Gould, S.B., Goremykin, V.V., Rippka, R., Tandeau de Marsac, N., Gugger, M., Lockhart,

- P.J., Allen, J.F., Brune, I., Maus, I., Puhler, A., Martin, W.F. (2013) Genomes of Stigonematalean cyanobacteria (Subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biol. Evol.* 5, 31–44.
- [9] Díez, B., Bauer, K., Bergman, B. (2007) Epilithic cyanobacterial communities of a marine tropical beach rock (Heron Island, Great Barrier Reef): diversity and diazotrophy. *Appl. Environ. Microbiol.* 73, 3656–3668.
- [10] Dong, C., Tang, A., Zhao, J., Mullineaux, C., Shen, G., Bryant, D. (2009) ApcD is necessary for efficient energy transfer from phycobilisomes to photosystem I and helps to prevent photoinhibition in the cyanobacterium *Synechococcus* sp. PCC 7002. *Biochim. Biophys. Acta* 1787, 1122–1128.
- [11] Erhard, M., Döhren, H., Jungblut, P. (1997) Rapid typing and elucidation of new secondary metabolites of intact cyanobacteria using MALDI-TOF mass spectrometry. *Nat. Biotechnol.* 15, 906–909.
- [12] Finsinger, K., Scholz, I., Serrano, A., Morales, S., Uribe-Lorio, L., Mora, M., Sittenfeld, A., Weckesser, J., Hess, W. (2008) Characterization of true-branching cyanobacteria from geothermal sites and hot springs of Costa Rica. *Environ. Microbiol.* 10, 460–473.
- [13] Füglistaller, P., Suter, F., Zuber, H. (1985) Linker polypeptides of the phycobilisomes from the cyanobacterium *Mastigocladus laminosus*: amino-acid sequences and relationships. *Biol. Chem. Hoppe Seyler* 366, 993–1001.
- [14] Galperin, M., Makarova, K., Wolf, Y., Koonin, E. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acid Res.* 43, D261–D269.
- [15] Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. In: *The Proteomics Protocols Handbook*, Humana Press, pp. 571–607.
- [16] Guan, X., Qin, S., Zhao, F., Zhang, X., Tang, X. (2007) Phycobilisomes linker family in cyanobacterial genomes: divergence and evolution. *Int. J. Biol. Sci.* 3, 434–445.
- [17] Hirose, Y., Fujisawa, T., Ohtsubo, Y., Katayama, M., Misawa, N., Wakazuki, S., Shimura, Y., Nakamura, Y., Kawachi, M., Yoshikawa, H., Eki, T., Kanesaki, Y. (2016) Complete genome sequence of cyanobacterium *Fischerella* sp. NIES-3754, providing thermoresistant optogenetic tools. *J. Biotechnol.* 20, 45–46.
- [18] Hutchins, P., Miller, S. (2017) Genomics of variation in nitrogen fixation activity in a population of the thermophilic cyanobacterium *Mastigocladus laminosus*. *ISME J.* 11, 78–86.
- [19] Janse, I., Meima, M., Kardinaal, E., Zwart, G. (2003) High-resolution differentiation of cyanobacteria by using rRNA-internal transcribed spacer denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* 69, 6634–6643.
- [20] Jeanjean, R., Latif, A., Matthijs, H.C., Havaux, M. (2008) The Psae subunit of photosystem I prevents light-induced formation of reduced oxygen species in the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochim. Biophys. Acta (BBA) – Bioenergetics* 1777, 308–316.
- [21] Kaštovský, J., Johansen, J. (2008) *Mastigocladus laminosus* (Stigonematales Cyanobacteria): phylogenetic relationship of strains from thermal springs to soil-inhabiting genera of the order and taxonomic implications for the genus. *Phycologia* 43, 307–320.
- [22] Komárek, J., Kaštovský, J., Mareš, J., Johansen, J.R. (2014) Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia* 86, 295–335.
- [23] Komenda, J., Lupinková, L., Kopecký, J. (2002) Absence of the psbH gene product destabilizes photosystem II complex and bicarbonate binding on its acceptor side in *Synechocystis* PCC 6803. *FEBS J.* 269, 610–619.
- [24] Konstantinidis, K., Rosselló-Móra, R., Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. *ISME J.* 11 (11), 2399–2406.
- [25] Kurmayer, R., Christiansen, G., Fastner, J., Börner, T. (2004) Abundance of active and inactive microcystin genotypes in populations of the toxic cyanobacterium *Planktothrix* spp. *Environ. Microbiol.* 6, 831–841.
- [26] Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.H., Rognes, T., Ussery, D.W. (2007) Rfam: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108.
- [27] Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lüssmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., Schleifer, K.H. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371.
- [28] Mackenzie, R., Pedrós-Alí, C., Díez, B. (2013) Bacterial composition of microbial mats in hot springs in Northern Patagonia: variations with seasons and temperature. *Extremophiles* 17, 123–136.
- [29] Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* 17, 10–12.
- [30] Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A. (2008) The metagenomic RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 386.
- [31] Miller, S., Castenholz, R., Pedersen, D. (2007) Phylogeography of the thermophilic cyanobacterium *Mastigocladus laminosus*. *Appl. Environ. Microbiol.* 73, 4751–4759.
- [32] Miller, S., Purugganan, M., Curtis, S. (2006) Molecular population genetics and phenotypic diversification of two populations of the thermophilic cyanobacterium *Mastigocladus laminosus*. *Appl. Environ. Microbiol.* 72, 2793–2800.
- [33] Miller, S., Williams, C., Strong, A., Carvey, D. (2009) Ecological specialization in a spatially structured population of the thermophilic cyanobacterium *Mastigocladus laminosus*. *Appl. Environ. Microbiol.* 75, 729–734.
- [34] Mora-Ruiz, M.D.R., Font-Verdera, F., Díaz-Gil, C., Urdiain, M., Rodríguez-Valdecantos, G., González, B., Orfila, A., Rosselló-Móra, R. (2015) Moderate halophilic bacteria colonizing the phylloplane of halophytes of the subfamily *Salicornioideae* (*Amaranthaceae*). *Syst. Appl. Microbiol.* 38, 406–416.
- [35] Munoz, R., López-López, A., Urdiain, M., Moore, E., Rosselló-Móra, R. (2011) Evaluation of matrix-assisted laser desorption/ionization-time of flight hole cell profiles for assessing the cultivable diversity of aerobic and moderately halophilic prokaryotes thriving in solar saltern sediments. *Syst. Appl. Microbiol.* 34, 60–75.
- [36] Nübel, U., Garcia-Pichel, F., Muyzer, G. (1997) PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl. Environ. Microbiol.* 63, 3327–3332.
- [37] Nurnberg, D., Mariscal, V., Parker, J., Mastroianni, G., Flores, E., Mullineaux, C. (2014) Branching and intercellular communication in the Section V cyanobacterium *Mastigocladus laminosus*, a complex multicellular prokaryote. *Mol. Microbiol.* 91, 935–949.
- [38] Olsen, M., Nowack, S., Wood, J., Becraft, E., LaButti, K., Lipzen, A., Martin, J., Schackwitz, W., Rusch, D., Cohan, F., Bryant, D., Ward, D. (2015) The molecular dimension of microbial species: 3. Comparative genomics of *Synechococcus* strains with different light responses and *in situ* diel transcription patterns of associated putative ecotypes in the Mushroom Spring microbial mat. *Front. Microbiol.* 6, 604.
- [39] Olson, J., Steppe, T., Litaker, R., Paerl, H. (1998) N₂-fixing microbial consortia associated with the ice cover of Lake Bonney, Antarctica. *Microb. Ecol.* 36, 231–238.
- [40] Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2014) Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055.
- [41] Peng, Y., Leung, H., Yiu, S.M., Chin, F. (2010) IDBA – a practical iterative De Bruijn graph *de novo* assembler, in: Berger, B. (Ed.), *Research in Computational Molecular Biology. RECOMB 2010. Lecture Notes in Computer Science*, vol 6044, Springer, Berlin, Heidelberg.
- [42] Pruesse, E., Quast, C., Knittel, K., Fuchs, B.N., Ludwig, W., Peplies, J., Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196.
- [43] Radway, J., Weissman, J., Wilde, E., Benemann, J. (1992) Exposure of *Fischerella [Mastigocladus]* to high and low temperature extremes: strain evaluation for a thermal mitigation process. *J. Appl. Phycol.* 4, 67–77.
- [44] Reuter, W., Wiegand, G., Huber, R., Than, M.E. (1999) Structural analysis at 2.2 Å of orthorhombic crystals presents the asymmetry of the allophycocyanin-linker complex, $AP_{LC}^{7.8}$ from phycobilisomes of *Mastigocladus laminosus*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 1363–1368.
- [45] Richter, M., Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Nat. Acad. Sci. U. S. A.* 106, 19126–19131.
- [46] Richter, M., Rosselló-Móra, R., Glöckner, F.O., Peplies, J. (2015) JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931.
- [47] Rippka, R., Deruelles, J., Waterbury, J., Herdman, M., Stanier, R. (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* 111, 1–61.
- [48] Rodríguez-R, L., Konstantinidis, K. (2014) Bypassing cultivation to identify bacterial species. *Microbe* 9, 111–118.
- [49] Rosselló-Móra, R., Amann, R. (2015) Past and future species definitions for Bacteria and Archaea. *Syst. Appl. Microbiol.* 38, 209–216.
- [50] Sano, E.B., Wall, C.A., Hutchins, P.R., Miller, S.R. (2018) Ancient balancing selection on heterocyst function in a cosmopolitan cyanobacterium. *Nat. Ecol. Evol.* 2, 510–519.
- [51] Shimizu, T., Hiyama, T., Ikeuchi, M., Inoue, Y. (1992) Nucleotide sequence of a metallothionein gene of the thermophilic cyanobacterium *Synechococcus vulcanus*. *Plant. Mol. Biol.* 20, 565–567.
- [52] Siegrist, T., Anderson, P., Huen, W., Kleinheinz, G., McDermott, C., Sandrin, T. (2007) Discrimination and characterization of environmental strains of *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). *J. Microbiol. Methods* 68, 554–562.
- [53] Sun, L., Jiang, W., Sato, H., Kawachi, M., Lu, X. (2016) Rapid classification and identification of *Microcystis aeruginosa* strains using MALDI-TOF MS and polygenic analysis. *PLoS One* 11 (5), e0156275.
- [54] Tillett, D., Neilan, B.A. (2000) Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. *J. Phycol.* 36, 251–258.
- [55] Turner, J., Robinson, N. (1995) Cyanobacterial metallothioneins: biochemistry and molecular genetics. *J. Ind. Microbiol.* 14, 119–125.
- [56] Uku, J., Björkl, M., Bergman, B., Díez, B. (2007) Characterization and comparison of prokaryotic epiphytes associated with three East African seagrasses. *J. Phycol.* 43, 768–779.
- [57] Viver, T., Cifuentes, A., Díaz, S., Rodríguez-Valdecantos, G., González, B., Antón, J., Rosselló-Móra, R. (2015) Diversity of extremely halophilic cultivable prokaryotes in Mediterranean Atlantic and Pacific solar salterns: evidence that unexplored sites constitute sources of cultivable novelty. *Syst. Appl. Microbiol.* 38, 266–275.
- [58] Wall, C., Koniges, G., Miller, S. (2014) Divergence with gene flow in a population of thermophilic bacteria: a potential role for spatially varying selection. *Mol. Ecol.* 23, 3371–3383.
- [59] Ward, D., Bateson, M., Ferris, M., Kühl, M., Wieland, A., Koeppel, A., Cohan, F. (2006) Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park Wyoming) as species-like units linking

- microbial community composition structure and function. *Phil. Trans. R. Soc. B* 361, 1997–2008.
- [60] Ward, D., Ferris, M., Nold, S., Bateson, M. (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol. Mol. Biol. Rev.* 62, 1353–1370.
- [61] Welker, M., Christiansen, G. (2004) Diversity of coexisting *Planktothrix* (Cyanobacteria) chemotypes deduced by mass spectral analysis of microcystins and other oligopeptides. *Arch. Microbiol.* 182, 288–298.
- [62] Welker, M., Erhard, M. (2007) Consistency between chemotyping of single filaments of *Planktothrix rubescens* (cyanobacteria) by MALDI-TOF and the peptide patterns of strains determined by HPLC-MS. *J. Mass Spectrom.* 42, 1062–1068.
- [63] Welker, M., Moore, E. (2011) Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *Syst. Appl. Microbiol.* 34, 2–11.
- [64] Wickström, C. (1980) Distribution and physiological determinants of blue-green algal nitrogen fixation along a thermogradient. *J. Phycol.* 16, 436–443.
- [65] Wu, Y., Tang, Y., Tringe, S., Simmons, B., Singer, S. (2014) MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2, 26.
- [66] Zhao, J., Brand, J. (1989) Specific bleaching of phycobiliproteins from cyanobacteria and red algae at high temperature *in vivo*. *Arch. Microbiol.* 152, 447–452.
- [67] Konstantinidis, K.T., Tiedje, J.M. (2005) Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* 187, 6258–6264.
- [68] Rodríguez-R, L.M., Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *Peer J. Preprints* 4, e1900v1.