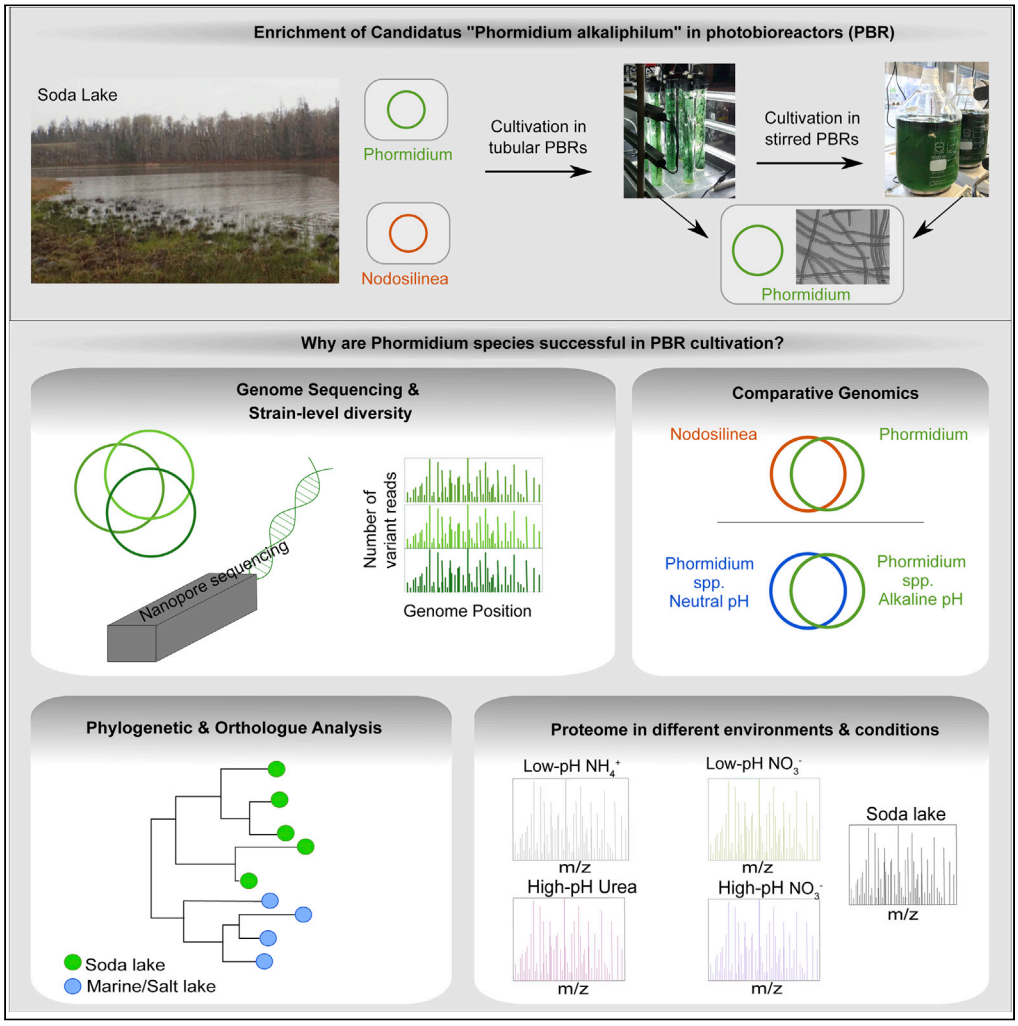
iScience

CellPress
OPEN ACCESS

Article

Proteome and strain analysis of cyanobacterium *Candidatus* "Phormidium alkaliphilum" reveals traits for success in biotechnology



Maryam Ataeian, Agasteswar Vadlamani, Marianne Haines, ..., Manuel Kleiner, Marc Strous, Alyse K. Hawley

alyse.hawley@ubc.ca

Highlights

Closed genome of the cyanobacteria *Ca.* P. alkaliphilum from high-pH photobioreactor

Genetic factors lead this *Phormidium* to outcompete other cyanobacteria in photobioreactor

Adaptation to high pH and alkalinity is not linked to specific genes

Strain-level diversity contributes *Ca.* P. alkaliphilum success in changing conditions

Ataeian et al., iScience *24*, 103405
December 17, 2021 © 2021
https://doi.org/10.1016/j.isci.2021.103405

iScience

CellPress
OPEN ACCESS

Article

Proteome and strain analysis of cyanobacterium *Candidatus* "Phormidium alkaliphilum" reveals traits for success in biotechnology

Maryam Ataeian,[1] Agasteswar Vadlamani,[1] Marianne Haines,[1] Damon Mosier,[1] Xiaoli Dong,[1] Manuel Kleiner,[2] Marc Strous,[1] and Alyse K. Hawley[1,3,4,*]

SUMMARY

Cyanobacteria encompass a diverse group of photoautotrophic bacteria with important roles in nature and biotechnology. Here we characterized *Candidatus* "Phormidium alkaliphilum," an abundant member in alkaline soda lake microbial communities globally. The complete, circular whole-genome sequence of *Ca.* "P. alkaliphilum" was obtained using combined Nanopore and Illumina sequencing of a *Ca.* "P. alkaliphilum" consortium. Strain-level diversity of *Ca.* "P. alkaliphilum" was shown to contribute to photobioreactor robustness under different operational conditions. Comparative genomics of closely related species showed that adaptation to high pH was not attributed to specific genes. Proteomics at high and low pH showed only minimal changes in gene expression, but higher productivity in high pH. Diverse photosystem antennae proteins, and high-affinity terminal oxidase, compared with other soda lake cyanobacteria, appear to contribute to the success of *Ca.* "P. alkaliphilum" in photobioreactors and biotechnology applications.

INTRODUCTION

Cyanobacteria are responsible for a large portion of carbon and nitrogen fixation on the Earth (García-Pichel et al., 2003; Partensky et al. 1999; Zehr et al., 2008). Substantial genetic diversity within cyanobacteria enables them to occupy a wide range of ecosystems, including extreme environments, such as deserts, saline lakes, glaciers, hot lakes, and alkaline soda lakes (Jungblut et al., 2005, Lopez-Cortes et al., 2001, Paerl et al., 2000; Sanchez-Baracaldo et al., 2005, Taton et al., 2006, Ward et al., 1998). The ability of cyanobacteria to fix carbon dioxide (CO_2) through oxygenic photosynthesis, with minimal nutritional requirements, has led to their adoption for bioproduction of pharmaceuticals and food ingredients (Beal et al., 2018; Gronenberg et al., 2013). Although cyanobacteria-based biotechnologies hold much potential for the production of bioproducts, there remain several challenges for the efficient and cost-effective implementation of this technology at scale. The high cost of supplying cyanobacteria with CO_2 for growth is a major drawback to the economic feasibility of these technologies. At scale, CO_2 is typically supplied to production systems, such as open raceway ponds, in the form of sodium bicarbonate, up to 150 μmol/L (Vonshak and Richmond 1988). During growth, CO_2 assimilation increases pond pH and leads to depletion of bicarbonate. The resulting spent medium is discarded and replaced. Another drawback of current technologies is the need to seed the ponds with cyanobacterial monocultures. This increases costs because seed cultures need to be maintained under aseptic conditions. The monoculture additionally leaves the ponds susceptible to invasion by other microbes, such as protists or cyanophages, causing collapse of the culture, and can result in loss of the crop (Cauchie et al., 1995; Forehead & O'Kelly 2013; Oswald 1980).

Alkaline soda lakes represent natural analogs of raceway ponds, with comparatively higher pH and carbonate concentration. Indeed, these conditions can be so extreme that CO_2 from the air spontaneously dissolves into the lake water. This process has driven planetary-scale carbon capture in Earth's geological past (Tutolo and Tosca 2018). Apparently, nature does not need a monoculture to realize robust and rapid growth. High productivity is enabled by the high concentration of dissolved bicarbonate. Carbonate (CO_3^{2-}) is not used directly by microbes, whereas bicarbonate (HCO_3^-) is their main source of CO_2. Bicarbonate concentrations in soda lakes are 10–100 mM, compared with <2 mM typically found in the oceans and 150 mM at inoculation of a raceway pond (Fabry et al., 2008; Raven et al., 2008). Bicarbonate is used by

[1]Department of Geoscience, University of Calgary, Calgary, AB, Canada

[2]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695, USA

[3]School of Engineering, University of British Columbia Okanagan, Kelowna, BC, Canada

[4]Lead contact

*Correspondence: alyse.hawley@ubc.ca

https://doi.org/10.1016/j.isci.2021.103405

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

cyanobacteria through the carbon concentrating mechanism (CCM), which provides a high local concentration of CO_2 for optimal carbon fixation by the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). This mechanism includes the biogenesis of carboxysome compartments within the cell where HCO_3^- is converted to CO_2 by the enzyme carbonic anhydrase. The CCM also includes transporters for active import of HCO_3^- across the cell membrane. Five different inorganic carbon uptake systems have been shown in cyanobacteria. Three are active HCO_3^- transport systems: two Na^+-dependent transport facilitators (BitA and SbtA) and one ATP-dependent transporter (BCT1 encoded by the cmpABCD operon). Two additional CO_2 uptake systems are based on modified NADPH dehydrogenase complexes (NDH-I_3 and NDH-I_4) (Omata et al., 1999; Price 2011; Price et al., 2007; Shibata et al. 2001, 2002). The uptake systems generate a cytosolic pool of HCO_3^-, which allows for higher CO_2 concentration in the carboxysomes where (RuBisCO) carries out the initial steps of the Calvin-Benson-Bassham cycle for carbon fixation (Kerfeld and Melnicki 2016; Rae et al., 2013b). In soda lakes found on the Cariboo plateau in British Colombia, Canada, the CCM supports the growth of high-cell-density cyanobacterial mats (Brady et al., 2013). A core microbiome of <100 shared bacterial lineages was recently described for geographically distant alkaline soda lakes (Vavourakis et al., 2018; Zorz et al., 2019). This shared alkaline soda lake microbiome included nine cyanobacterial populations, each represented by multiple metagenome assembled genomes (MAGs). Two filamentous cyanobacteria from different lineages were highly abundant in soda lakes in British Colombia, one affiliated with *Nodosilinea* (NCBI accession GCA_007692715.1) and one with *Phormidium* (GCA_007693465.1) (Zorz et al., 2019).

Previously, we enriched a microbial consortium from alkaline soda lakes on the Cariboo Plateau (BC, Canada) in photobioreactors at high pH and alkalinity (Sharp et al., 2017). This consortium consisted of a cyanobacterium, which makes up >80% of the community. Heterotrophic bacteria and archaea made up the remainder. The consortium displayed rapid and robust growth. We also demonstrated that the spent medium (pH > 11) could be recharged by spontaneous capture of CO_2 from air while maintaining high and robust biomass productivity (15.2 $\pm$ 1.0 g/m^2/day) (Ataeian et al., 2019). These innovations could potentially reduce operational costs of cyanobacterial biotechnology.

Within cyanobacterial biotechnology applications, there is need for high-quality whole-genome sequences. They enable the precise identification and characterization of strains acquired for regulatory and intellectual property applications. They support approval for use in food and pharmaceuticals and the assessment of biosafety. Many cyanobacteria produce toxins, and the potential for toxin production can be assessed if a complete genome is available. However, the high frequency of repetitive elements present within cyanobacterial genomes make them notoriously difficult to assemble even from pure culture (Mazel et al., 1990; Stucken et al., 2010), resulting in a dearth of complete, closed-circular cyanobacterial genomes available.

In the present study we combined Nanopore and Illumina sequencing to obtain the first complete, circular genome for the genus *Phormidium* so far. Using inStrain analysis we show evidence for evolution of the most abundant population as well as selection of different closely related strains in different conditions. We use comparative genomics to understand the ecological success of this species, both in alkaline soda lakes and in photobioreactors for biotechnology. We also provide experimental proteomes under different growth conditions, as well as morphological characterization.

RESULTS AND DISCUSSION

Cultivation and morphology

In June 2017, microbial mats from four Cariboo Plateau (Canada, BC) soda lakes (Brady et al., 2013; Zorz et al., 2019) were sampled, mixed, and inoculated into planar photobioreactors with red cellophane filters (Sharp et al., 2017) (Figure S1). After 3 months, the enrichment culture was transferred to and maintained in tubular photobioreactors (Ataeian et al., 2019) (Figure S1). Between 5 and 7 months after transfer, the culture was used to study the effect of pH and nitrogen source on growth, as previously described (Ataeian et al., 2019). During that time, four metagenomes and four proteomes, with technical quadruplicates, were obtained. These were associated with four different conditions. The first two cultures were grown at an initial pH of 8.3 and received either ammonium or nitrate as the nitrogen source (low pH). The second two cultures were grown at an initial pH of 10.4 with either nitrate or urea as the nitrogen source (high pH). Unfortunately, despite very high sequencing depth of the cyanobacterial genome in the metagenomes, assembly of the cyanobacterial genome was challenging, resulting in a fragmented whole-genome sequence with hundreds of contigs. Downsampling of reads did not solve these assembly problems. Assembly can be

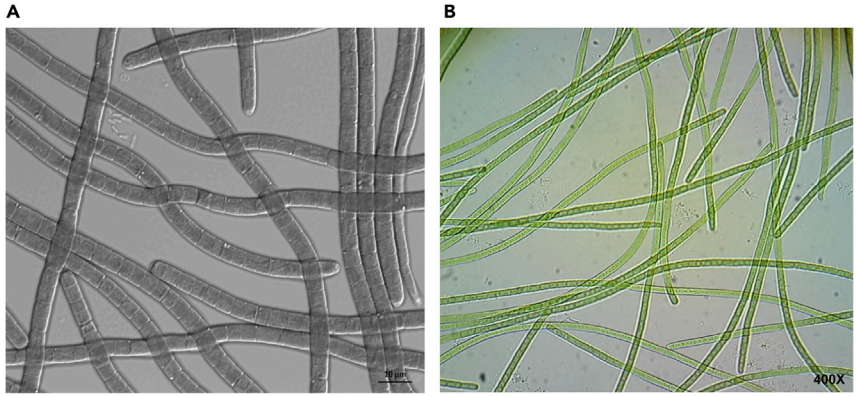


Figure 1. Light microscopic images of *Ca.* P. alkaliphilum

(A) Differential interference contrast microscopy at 1,000× magnification.

(B) Bright-field microscopy at 400× magnification. The images were taken from stirred cultures grown at pH (10.4–11.2) and alkalinity (0.5 mol/L) with 1 mM NH_4Cl, 3 mM $NaNO_3$ as nitrogen source.

a challenge for cyanobacteria. For example, among the 15 whole-genome sequences of isolated filamentous cyanobacteria shown in Table S1, only a single whole-genome sequence has been closed. This has been explained by a high frequency of repetitive sequences in these cyanobacterial genomes (Mazel et al., 1990; Stucken et al., 2010). Alternatively, because the target cyanobacterium was part of a microbial consortium and not available as a pure culture, presence of multiple, closely related strains may have compromised the assembly.

After 1 year in tubular bioreactors, the culture was transferred to stirred photobioreactors for ease of maintenance and sampling (Figure S1). The stirred cultures consisted of illuminated 10-L glass bottles stirred with a magnetic stirrer. In these cultures, the consortium grew as loose flocs or aggregates. When the stirrer stopped, biomass spontaneously coagulated and amassed both at the top and the bottom of the bottle. Inspection of the aggregates with microscopy showed abundance of a filamentous cyanobacterium, with cells of 5 μm average width, appearing as cylindrical, unbranched filaments in green shades under bright-field illumination (Figure 1). Observed morphology was consistent with the *Phormidium* previously enriched from the same soda lakes in planar photobioreactors (Sharp et al., 2017). In addition, filaments displayed motility by gliding along each other and the surface of the slide (Video S1).

Although culture methods were changed as described, the culture medium always remained the same and contained 0.5 mol/L combined bicarbonate and carbonate at a pH of 8.3–11.2. No culture "crashes" were ever observed, and a high growth rate was reported in both previous studies (Ataeian et al., 2019; Sharp et al., 2017).

Genomic sequencing and phylogenetic analysis

During summer 2020, 3 years after sampling the biomass from the soda lakes, DNA was extracted for parallel Nanopore and Illumina sequencing. Assembly and base-calling of the Nanopore reads yielded a circular whole-genome sequence associated with a *Phormidium* species. The whole-genome sequence was further polished and perfected with corresponding short reads generated by Illumina sequencing of DNA from the same biomass. Remaining assembly problems, manifesting themselves as regions with anomalous sequencing depth, were manually resolved. The complete, circular genome was 5,000,054 bp in size. Sequencing depth was ~600× for Nanopore reads and 80× for Illumina reads. For the whole-genome sequence, 4,261 protein coding genes were predicted. The coding density was 88%, fairly low, which was attributed to the prolific presence of repeat sequences. We found two rRNA operons, both containing 16S rRNA, 23S rRNA, 5S rRNA genes. The two operons were not identical, showing 99.12%, 99.65%, and 100% sequence identities respectively. No genes or gene clusters encoding production of any known toxin were detected in the genome.

A typical origin of replication (*oriC*) for the complete *Phormidium* genome was not found with Ori-Finder (Gao and Zhang 2008). A plot of the GC skew did not show the conventional symmetrical pattern with a peak at the midpoint of the genome, but rather a much lower skew without a smooth peak (Figure 2A).

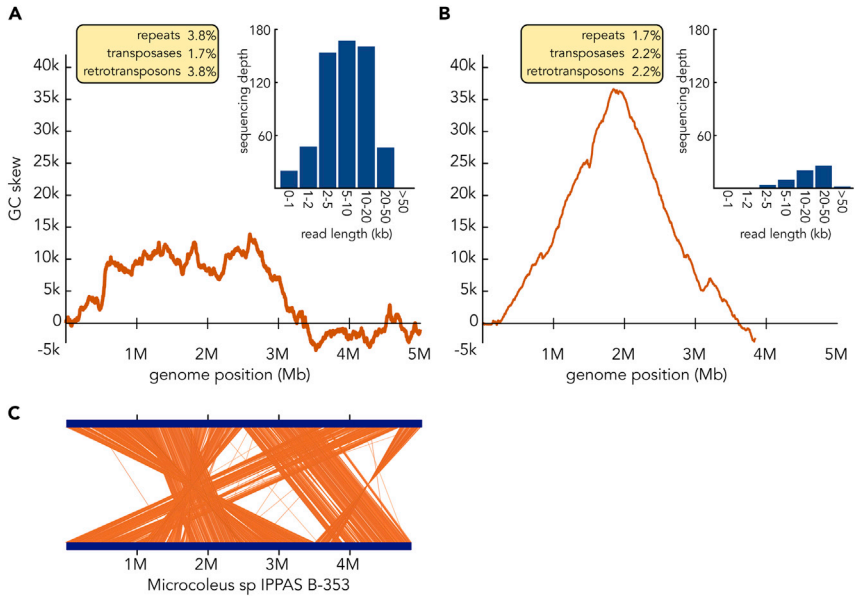


Figure 2. Genome assembly and structure

(A) GC skew of the *Ca.* P. alkaliphilum genome.

(B) GC skew of a *Wenzhouxiangella* genome, sequenced and assembled in the same laboratory with the same methods. The insets in (A and B) show the nanopore read length distribution and the percentage of each genome occupied by repetitive elements.

(C) Genome synteny between *Ca.* P. alkaliphilum and *Microcoleus* sp. IPPAS B-353.

As such the start of the genome was set at the *dnaA* gene. To determine if the lack of GC skew was the result of assembly artifacts caused by repeat sequences, we compared the *Phormidium* assembly to the assembly of a Gammaproteobacterium also sequenced in our laboratory using combined Nanopore and Illumina sequencing (Figures 2B; Sorokin et al., 2020). Even though the Nanopore sequencing depth for the Gammaproteobacterium was lower and repetitive elements were also abundant (Figure 2B, inset), the GC skew showed clear trends with a clear peak and a clear origin of replication. To further investigate the fidelity of the assembly process, we compared the synteny (gene order) of the *Phormidium* with the closely related *Microcoleus* sp IPPAS B-353 (Kupriyanova et al., 2016). The two genomes shared 92.6% average nucleotide identity (ANI), and the quality of the *Microcoleus* genome was excellent, as it had only a single scaffold with four contigs. It appeared that gene order was preserved extremely well, with only four large-scale rearrangements setting the two species apart (Figure 2C). Thus, the lack of canonical GC skew trends and *oriC* did not appear to be caused by faulty assembly or genome rearrangements by hyperactive mobile elements. Instead, it might result from a different mechanism of replication recently proposed for filamentous cyanobacteria (Ohbayashi et al., 2020; Watanabe 2020). Indeed, irregular GC skew profiles have been shown in several other cyanobacteria lacking oriC that likely initiate replication at multiple points within their genome (Nikolaou and Almirantis 2005; Ohbayashi et al., 2016; Worning et al., 2006).

To determine whether the genome underwent evolutionary changes during the 2.5 years between the previous tubular photobioreactor experiments (Ataeian et al., 2019) and the stirred photobioreactor Nanopore sequencing, we carried out analysis with inStrain (Olm et al., 2021). Using inStrain with the Nanopore-sequenced *Phormidium* whole-genome sequence as a reference and reads from the four tubular photobioreactors and stirred photobioreactor metagenomes, we assessed the proportion of variant sites under each condition. Mapping results indicated that the same sequenced *Phormidium* was abundant in the four tubular and the stirred PBRs, recruiting between 42% (low-pH nitrate) and 90% (high-pH urea) of all metagenomic reads passing quality control (Table S3). The analysis revealed a low level of strain-level diversity present in the *Phormidium* population throughout all the conditions. A low number of variant sites was found ranging from 0.2 sites per kb (817 sites in total, high-pH urea) to 11 sites per kb (53,399 sites in total, low-pH nitrate) (Figure S2, Table S3). Interestingly, more variant sites were shared between the two low-pH conditions than were found in the higher pH conditions (Figure S2), indicating a higher level of variation in the *Phormidium* populations grown at low pH.

Over 85% of all variant sites across conditions occurred with <50% of the reads containing a variant base (single nucleotide variants in Table S3). These same low levels of variation were also found within the stirred bioreactor (0.4 variant sites per kb), showing that this natural variation persisted during multiple years of laboratory cultivation. Analysis of the abundance of unique k-mers present in the metagenome for each condition (Figure S2 inset) provides insight into *Phormidium* populations under different conditions. Both high-pH experiments display a symmetrical peak at 600–1,000× sequencing depth corresponding to a ~4.7-Mb genome. This corresponds well to the size of the *Phormidium* whole-genome sequence, excluding repeat sequences (4.8 Mb). The low-pH ammonium condition displays the same peak with a shoulder and slightly larger estimated genome size, indicating the presence of a mixture of two strains, consistent with an intermediate number of reads mapped (71%) to the *Phormidium* whole-genome sequence. In the low-pH nitrate condition, two peaks are seen and the *Phormidium* genome could not be clearly recognized. This is consistent with a much lower percentage of reads mapped (only ~50%) and considerable strain-level diversity detected by inStrain analysis in the low-pH nitrate condition.

To assess the possibility of strain evolution we looked at three areas in the genome where single-nucleotide substitutions (SNS, with ≥95% reads with variants) and consensus variants (with 50%–94% reads with variants) were more frequent (Figure S2). The first area, from 754,780 bp to 765,901 bp in the genome, encompassed 60 SNSs across 12 predicted open reading frames (pha|00884–pha|00899, Table S2). The encoded gene products appeared to have regulatory functions. One of the genes appeared to be truncated and might be a pseudogene, which was no longer functional. Another encoded the iron uptake protein OprB. The second area, from 1,348,096 bp to 1,353,803 bp, encompassed 39 SNSs across four predicted open reading frames (pha|01550–pha|01558, Table S2). One of the genes appeared truncated and may be a pseudogene that previously encoded a heavy metal efflux protein. The third, from 4,619,625 bp to 4,634,481 bp, encompassed 50 consensus variants across seven predicted open reading frames. An additional SNS site found only in the low-pH nitrate condition consisted of two SNS within the *urtC* gene (pha|00357; 303,382 and 303,484 bp, Table S2) encoding a subunit of a urea ABC importer. For all SNS and consensus variant sites occurring in multiple conditions, the variant base was the same in all conditions. This would be consistent with the presence of different strains, not just random mutations. Furthermore, the higher frequency of variants within these regions, some of which have a probable pseudogene, suggests evolution by way of gene loss for genes no longer beneficial within the photobioreactors. These combined results support two processes at work underlying the observed diversity of the *Phormidium* population within the photobioreactors. One, a diversity of strains present emerges under different conditions and contributes to robustness with regard to changes in operating and environmental conditions. This is especially clear in the low-pH nitrate bioreactor, where the strain represented in the sequenced *Phormidium* whole genome appears to perform less well. Two, the *Phormidium* population is evolving to eliminate gene content that is no longer useful for life within the photobioreactors. In demonstrating how strain-level diversity within the cyanobacterial population offers robust responses to changing conditions and evolution toward the loss of non-beneficial genes, we demonstrate the value of maintaining strain-level diversity from natural environments within cyanobacteria-based biotechnologies.

Given that the enriched *Phormidium* originated from microbial mats in Cariboo soda lakes, we compared the whole-genome sequence to the MAGs previously obtained from these mats. The enriched *Phormidium* genome displayed 96.7% ANI to GEM2.Bin31, **C5** [GCA_007693465.1] (Zorz et al., 2019). GEM2.Bin31 was associated with the most abundant bacterium in Lake Goodenough. Among Central Asian (Kulunda Steppe) soda lake cyanobacteria, MAG *T3Sed10_304* [GCA_003566575.1] (Vavourakis et al., 2018) was most closely related, at 88.4% ANI (Table S1). These close similarities indicated that the enriched, genome-sequenced *Phormidium* represents a cyanobacterial genus that is widespread and abundant in soda lakes worldwide.

Additional related cyanobacteria were identified through a BLASTp search of the RpoB protein (DNA-directed RNA polymerase subunit B) against the NCBI non-redundant protein sequence (nr) database. The three most closely related species included *Microcoleus* sp IPPAS B-353 [GCF_009846485], from Kulunda soda lakes (Kupriyanova et al., 2016); *Phormidium* OSCR [GCA_001314905.1], from hypersaline, Hot Lake, Washington (Nelson et al., 2016); and *Phormidium lacuna* HE10JO [GCA_900,149,785.2], from North Sea rock pools (Nies et al., 2017). The whole-genome sequences of these organisms showed 93%, 86%, and 84% ANI to our newly sequenced genome, respectively. As these ANI numbers are less than the often used species threshold of 95% (Olm et al. 2021), the enriched cyanobacterium represented a

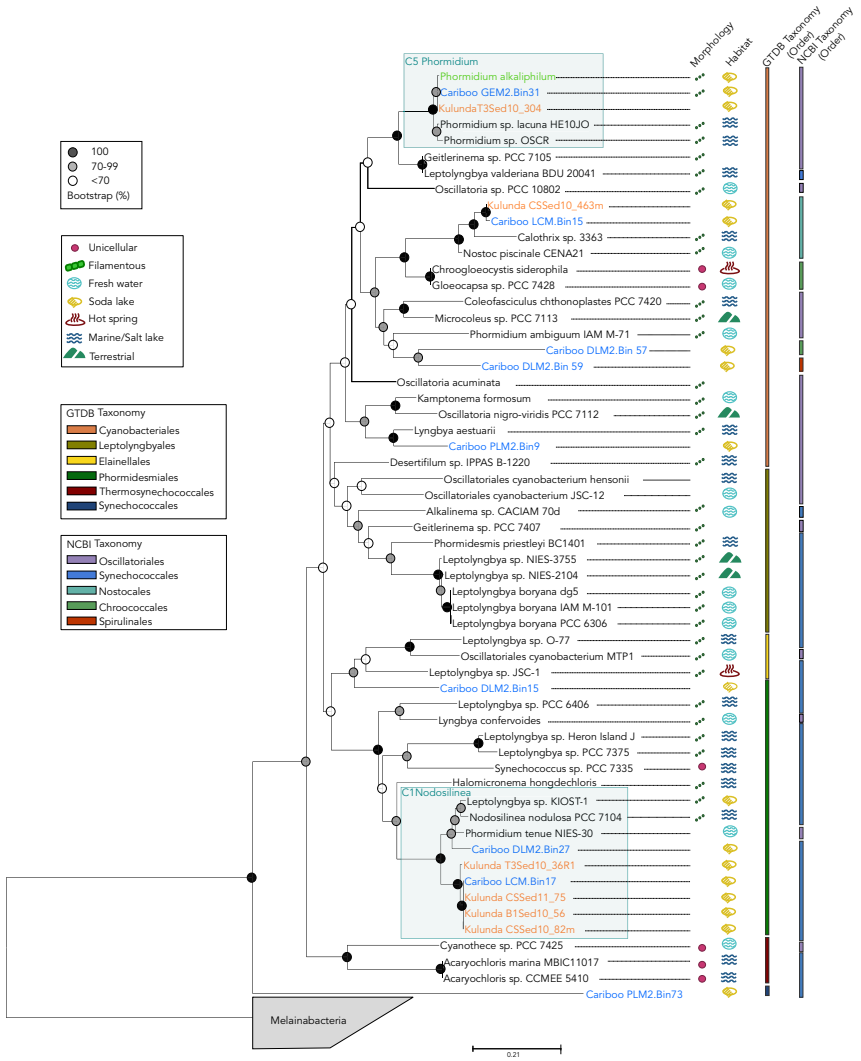


Figure 3. Maximum likelihood phylogenetic tree based on *rpo*ABC amino acid sequences

Organism names are according to the NCBI taxonomy database. Taxa names in blue and orange are associated with Metagenome Assembled Genomes (MAGs) obtained from Cariboo Plateau and Kulunda Steppe soda lakes, respectively. Two subclades of *Phormidium* and *Nodosilinea* are shown in teal boxes labeled as **C5** and **C1**, consistent with Zorz et al. (2019). Lifestyle and habitat information is encoded to the right of the taxa. Nodes are color coded based on the % of bootstrap support. *Melainabacteria* species were used as outgroup. Table S4 contains all genomes with full NCBI and GTDB names.

new species within the genus *Phormidium*. Owing to its ability to grow at high pH and alkalinity and increased relative abundance and productivity in high-pH experiments (Ataeian et al., 2019), we propose to provisionally name this species *Candidatus* "Phormidium alkaliphilum." Isolation into pure culture will be required for this name to become officially recognized.

Figure 3 shows the phylogenetic relationship of *Ca.* "P. alkaliphilum" with other cyanobacteria based on a concatenated alignment of RpoABC2 (Table S4). The tree also includes the cyanobacterial MAGs obtained from Cariboo mats and Kulunda Steppe sediments. Three species from *Melainabacteria* were used as the outgroup. These share a common ancestor with cyanobacteria but are not photosynthetic (Soo et al., 2017). The two most abundant cyanobacteria found in the Cariboo soda lakes, *Phormidium* (clade **C5**) and *Nodosilinea* (clade **C1**) (Zorz et al., 2019), are from quite distant phylogenetic lineages within the cyanobacteria (Shih et al., 2013). We compared the topology of our rpoABC tree with the GTDBtk reference tree as well as with a previously published tree (Shih et al., 2013) and found the same topology in all trees. According to

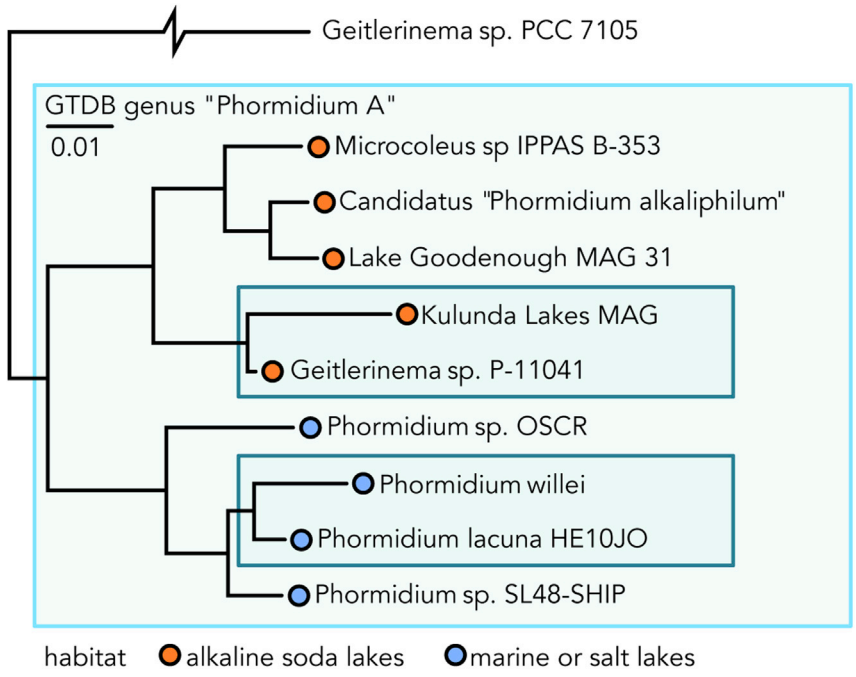


Figure 4. Maximum likelihood phylogenetic tree

Based on a concatenated alignment of 125 conserved marker genes used by the Genome Taxonomy Database (GTDB). The tree shows members with >84% average nucleotide identity (ANI) to *Ca.* P. alkaliphilum. Organism names are according to the NCBI taxonomy database. Bootstrap support is 100% for every branch.

the Genome Taxonomy Database (GTDB), clade **C5** including *Ca.* P. alkaliphilum is affiliated with the order Cyanobacteriales, family Geitlerinemaceae, and genus *Phormidium*. **C1** is, counter-intuitively, affiliated with the order Phormidesmiales, family Phormidesmiaceae, and genus *Nodosilinea*. The historical naming of cyanobacteria, based on morphology, has unfortunately resulted in confusing taxonomic names.

To explore the evolutionary relationships of **C5** *Phormidium* in more detail, a phylogenetic tree with higher resolution was constructed using the 125 GTDB conserved marker genes (Figure 4). This phylogenetic tree showed that within **C5**, *Phormidium* from soda lakes form a separate evolutionary branch from cyanobacteria originating from marine environments and salt lakes. *Ca.* P. alkaliphilum clustered with other *Phormidia* from alkaline soda lake environments, including GEM2.Bin31, *Microcoleus* sp. IPPAS B-353, *Cyanobacterium T3Sed10*, and *Geitlerinema* sp. P-1104, at 96.7%, 92.6%, 88.6%, and 88.4% ANI, respectively. For comparison, the four *Phormidium* genomes from marine and inland salt lake environments displayed lower ANI to *Ca.* P. alkaliphilum at 84%–86%. Bacteria living in soda lakes have to adapt to sodium (bi)carbonate brines with high pH, whereas marine/salt lake bacteria live in pH-neutral habitats containing mainly sodium chloride (Jones et al., 1977). Comparing **C5** genome content may show which adaptations contribute to the ecological success at high pH and high alkalinity.

Comparative genomics and orthologue analysis

To investigate the relationship between the soda lake and marine/salt lake *Phormidium* species in clade **C5** we identified and analyzed sets of orthologous genes found in the five soda lake and four marine/salt lake *Phormidium* genomes (Tables S1 and S2).

Orthologue analysis revealed approximately 1,376 core genes shared between both groups. Shared genes included those required for DNA, RNA, and protein biosynthesis; homeostasis of the gram-negative cell envelope; and lipid biosynthesis. Genes encoding photosystem I (psaABCDEFIJGLMXK), photosystem II (psbADBCEFOHNPQUVZ), and cytochrome *b6f* (petADBCEJ) were also part of the core genome, as well as genes needed for production of chlorophyll a (chlELNB). Both groups have genes encoding superoxide dismutase, peroxiredoxin, glutaredoxin, peroxidase, and tocopherol for scavenging reactive oxygen

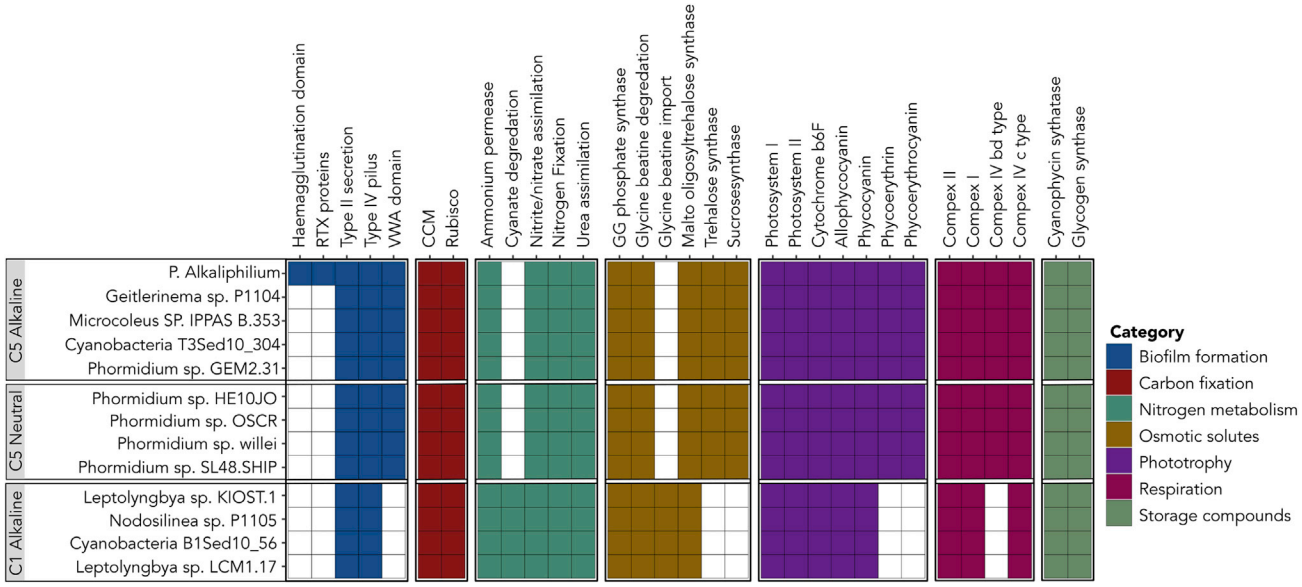


Figure 5. Presence of orthologous genes by functional category among *Phormidium* (C5) and *Nodosilinea* (C1) from different habitats

Organism names are according to the NCBI taxonomy database. "C5 Alkaline" refers to *Phormidium* species collected from alkaline soda lakes. "C5 Neutral" refers to *Phormidium* species from marine and salt lake habitats. "C1 Alkaline" refers to *Nodosilinea* species collected from alkaline soda lakes. Table S2 contains full list of *Ca.* P. alkaliphilum genes with annotations and proteomics values.

species (ROS) (Latifi et al., 2009). Genes for orange carotenoid protein, which is involved in photoprotection by dissipating excess energy, were also detected in both groups (Wilson et al., 2006). All genomes contained a complete set of genes for carbon fixation via the Calvin-Benson-Bassham cycle. Genes of the CCM, including carbonic anhydrases, bicarbonate transport system, CO_2 uptake systems, and genes required for carboxysome assembly (ccaA, bicA, ictB, NDH-I_3, NDH-I_4, and ccmKMNOL) were present in both groups (Figure S3). The CCM is responsible for transport of bicarbonate into cells and providing a steady supply of carbon dioxide (CO_2) to the RuBisCO enzyme, to maximize the rate of CO_2 fixation into biomass (Kerfeld and Melnicki 2016; Price et al., 2007; Raven et al., 2008). Genes for dark carbon catabolism, including glycolysis, the citric acid cycle, the pentose phosphate pathway, as well as a respiratory electron transport chain (ETC) were observed in all genomes. All the genes required for pyruvate fermentation to lactate, acetate, ethanol, formate, and acetoin (Figure S4) were shared, indicating the capability for energy generation under anoxic or low-oxygen conditions. Fermentation pathways may also support managing the overflow of metabolic intermediates as a result of high metabolic rates at high CO_2 availability in intense sunlight (Cano et al., 2018). Shared genes for production of osmotic solutes indicated the capability to produce glycerol, sucrose, trehalose, and glucosyl-glycerol (Figure 5). These genes are used for maintaining cellular ionic and osmotic balance in response to environmental fluctuations in salt concentration. These osmotic solutes can also serve as substrates for energy storage and fermentation.

Genes required for a variety of nitrogen acquisition pathways were found in all genomes including nitrogen fixation, nitrate/nitrite assimilation, urea uptake and hydrolysis, ammonium transport and assimilation, and genes for transport of neutral, branched, and polar amino acids (Figure 5). In the Cariboo Plateau soda lakes, ammonium and nitrate concentrations were very low (<23 μM ammonium and <0.16 μM nitrate) (Zorz et al., 2019). Genes for nitrogen fixation via nitrogenase and urea uptake were highly expressed in Cariboo soda lakes (Zorz et al., 2019). Cyanobacteria use different strategies to protect the oxygen-sensitive nitrogenase. One strategy is the physical separation of nitrogen fixation enzymes from oxygenic photosynthesis via the formation of heterocysts. Heterocysts are specialized cells dedicated to nitrogen fixation, protecting the nitrogenase from oxygen (Berman-Frank et al., 2003). Two genes required for initiation of heterocyst formation (*hetR* and *ntcA*) were found in all genomes, but other essential genes, including *hetP, hetC, hetF, hetL,* and *patA* were missing. Additionally, no visual evidence of heterocyst development was ever observed for *Ca.* P. alkaliphilum (Figure 1). It was suggested that *hetR* may be involved in other forms of regulation in non-heterocyst-forming filamentous bacteria (Zhang et al., 2009) and *ntcA* is known to be involved in the regulation of nitrogen uptake as well as heterocyst formation (Herrero and Flores 2019). Therefore, these

cyanobacteria may protect nitrogenase through a different mechanism, such as temporal separation of photosynthesis (day) and nitrogen fixation (night). Nitrogen fixation produces hydrogen (H_2) as a byproduct (Almon and Böger 1988; Bothe et al., 2010). To salvage this hydrogen, nitrogen-fixing cyanobacteria often possess hydrogenases. Genes for uptake hydrogenase (*hypABCD*) as well as bidirectional hydrogenase (*hoxEUS, hndC*) were found in all genomes. H_2 recycling is an important process for non-heterocystous cyanobacteria, as it can provide an anoxic environment for nitrogenase via the respiratory "oxyhydrogen reaction" (uptake and aerobic respiration of H_2 in the dark) (Bothe et al., 2010; Dixon 1972).

Genes required for formation and utilization of the storage molecules glycogen and cyanophycin were detected in both groups (Figure 5). In cyanobacteria, glycogen is the major carbon reserve molecule for excess photosynthetic carbon. It is then used as a carbon and energy source in the dark, for survival and homeostasis (Cano et al., 2018; Lehmann and Wober 1976; Preiss 1984). Cyanophycin (multi-L-arginyl-poly-L-aspartate) serves as both nitrogen and carbon storage. In non-heterocystous cyanobacteria, cyanophycin is synthesized at night when nitrogen fixation occurs and there is excess nitrogen available. During the day when photosynthesis resumes and nitrogen fixation is inhibited, cyanophycin is used as a nitrogen source (Finzi-Harta et al., 2009, Sherman et al., 1998). Genes for polyhydroxybutyrate biosynthesis were not present in any of the genomes.

Surprisingly, **C5** soda lake bacteria shared only eight orthologous genes not present in **C5** marine/salt lake bacteria. Only 30 additional genes were present in all but one soda lake genome and missing in all but one **C5** marine/salt lake genome. Apart from a single gene, a phytoene/squalene synthetase, none of these genes had a clear association with a known phenotype relevant to survival in alkaline conditions. Two toxin-antitoxin gene pairs were found specific to the **C5** alkaliphiles. These may be involved in regulation of cell growth versus persistence, biofilm formation, and tolerance to antibiotics and radical oxygen species (Khomutovska et al., 2020; Marsan et al., 2017; Wen et al., 2014). Previous studies identified marker genes for adaptation to alkaline conditions. For example, multiple studies have implicated production of hopanoids, such as squalene, in survival under alkaline conditions by inhibiting cation leakage through the membrane bilayers (Haines 2001; Hauβ et al., 2002). Genomes of alkaliphilic but not neutrophilic gammaproteobacteria *Wenzhouxiangella* encode specific outer membrane porins, squalene biosynthesis, as well as the sodium-dependent NADH:ubiquinone oxidoreductase complex (Sorokin et al., 2020). Alkaliphilic *Bacillus lentus* synthesize a secondary acidic cell membrane consisting of peptidoglycan and teichuronopeptide to support the generation of the proton motive force in the hydroxide-rich alkaline environment (Aono et al., 1999). Genomes of *Bacillus halodurans* encode a multi-subunit Mrp Na^+/H^+ antiporter for alkaline pH homeostasis (Hamamoto et al., 1994). Apart from the single phytoene/squalene synthetase gene, no gene involved in any conceivable form of adaptation was found to set the alkaliphilic *Phormidia* apart from their neutrophilic relatives. One explanation for lack of differences in gene content could be that the neutrophilic *Phormidia* are already innately adapted to high pH. These filamentous cyanobacteria are often found in close physical association, such as microbial mats in Cariboo soda lakes. In such arrangements, photosynthesis will quickly increase local pH, due to consumption of CO_2. Fast-growing filamentous cyanobacteria would likely need mechanisms to manage this local pH increase even in neutral environments.

Lack of distinguishing gene content does not preclude evolutionary adaptation. The observation that alkaliphilic **C5** *Phormidia* form a separate monophyletic clade (Figure 4) hints at the presence of some kind of adaptation. For example, variations of amino acid sequences and structural diversification in cell envelope proteins could result from differences between habitats (Panja et al., 2020). After all, without such adaptations, we would have expected the alkaliphilic species to be scattered phylogenetically among their neutrophilic relatives.

Photobioreactor selection of *Candidatus* "Phormidium alkaliphilum"

Cultivation of the Cariboo soda lake mats in the photobioreactors at high pH and high alkalinity repeatedly enriched for a cyanobacterial consortium with high abundance of the same *Phormidium* species. Previous studies have reported the selection of related (>84% ANI) neutrophilic *Phormidium* species in photobioreactor environments (Cole et al., 2014; Nies et al., 2017). This suggests the capabilities of this genus to adapt to new conditions and outcompete other, naturally coexisting cyanobacteria in the laboratory or photobioreactor environments. In other words, the genus has potential in biotechnology. To investigate the genetic basis of this repeated selection we explored gene content specific to **C5** *Phormidium* when

compared with another, distantly related, abundant cyanobacteria from the same soda lakes, **C1** *Nodosilinea* (Figure 5 and Table S1).

A shared core genome of 892 genes was detected between **C1** and **C5** species and included genes involved in DNA, RNA, and protein biosynthesis as well as carbon and nitrogen metabolism. Genes for the production of the phycobilisome antenna complex differed between the *Phormidium* and *Nodosilinea* lineages. The phycobilisome complex contains a suite of pigment proteins to absorb light between 550 and 680 nm. The efficient fluorescence of this light energy from the phycobilisome to chlorophyll a of photosystem II, provides access to wavelengths of light otherwise inaccessible to chlorophyll (Ting et al., 2002). Although genes for biosynthesis of the phycobilisome protein allophycocyanin (maximum absorbance 650 nm) and phycocyanin (maximum absorbance 620 nm) were found in both lineages, **C1** had at least three times more copies of allophycocyanin genes. **C5**, however, contained genes for two additional pigments: phycoerythrin (maximum absorbance at 495 and 560 nm) and phycoerythrocyanin (maximum absorbance at 575 nm) (Figure 4). Having additional pigments to harvest a wider spectrum of light could likely result in higher overall efficiency of photosynthesis and growth, which could provide selective advantage in a photobioreactor.

Additionally, differences in the genes involved in oxygen respiration were observed (Figure 4). Genes for aa3-type cytochrome *c* oxidoreductase (*cox* genes) existed in both groups. **C5** species contained additional genes for the bd-type quinol oxidase complex. The high-affinity bd-type quinol oxidase is generally found only in bacteria that experience low-O_2 conditions (D'mello et al., 1996; Gong et al., 2018). In the photobioreactors, this feature would benefit *Ca.* P. alkaliphilum in competing for oxygen during dark cycles by outcompeting other species, while simultaneously enabling nitrogen fixation.

The ability to grow as a biofilm may also contribute to the enrichment of *Ca.* P. alkaliphilum in photobioreactors. Cyanobacterial biofilm formation has been attributed to type II secretion/type IV pilus assembly systems (Schatz et al., 2013), and these genes are present in all **C5** *Phormidium* and **C1** *Nodosilinea* species (Figure 5 and Table S1). Interestingly, *Ca.* P. alkaliphilum contained three additional groups of proteins with potential roles in cell adhesion and biofilm formation. The first group consisted of nine von Willebrand factor type A (VWA)-domain-containing proteins (Konto-Ghiorghi et al., 2009; Kuchma et al., 2015; Ponting et al., 1999), present in all **C5** species. The VWA domain in the pilus-associated adhesin (PilA) in *Streptococcus agalactiae* is required for adherence to epithelial cells and biofilm formation (Konto-Ghiorghi et al., 2009). The second group consisted of 17 proteins with RTX calcium binding domains. The RTX proteins are transported out of the cell in an unfolded state using ABC transporters. Later, by binding to calcium (Ca^{2+} ions they form a folded, stable structure that could function in biofilm formation (Pérez et al., 2010; Sánchez-Magraner et al., 2007). Last, seven proteins with hemagglutination-like domains, also with a potential role in cell adhesion (Syed et al., 2009), were detected. Having a higher number of mechanisms to assist in cell adhesion and biofilm formation in **C5** species could enable them to form microbial mats in their natural habitat and biofilms or flocs in photobioreactors. Aggregation provides, for example, effective self-shading and protection from direct sunlight, protection from predation by protists, and may enable nitrogen fixation by depletion of oxygen during dark cycles.

Comparison of genes involved in signal response mechanisms indicated that **C5** bacteria were likely more responsive to their environment than **C1** bacteria. Within the *Ca.* P. alkaliphilum genome, approximately 94 genes encoded for two-component signal transduction systems (Table S2). These genes consisted of sensor histidine kinases and response regulators. Genomes of **C1** species encoded less than 40 of those genes. In this mechanism, the response regulator is phosphorylated by a histidine kinase in response to specific environmental stimuli, functioning as a molecular switch, activating downstream processes such as transcription, translation, or enzyme activity (Dikiy et al., 2019; Galperin 2010). Proteomic analysis showed that these genes were relevant to growth of **C5** in bioreactors, as expression was demonstrated for 40 genes at high-pH compared with only 11 genes at low pH (Tables S5 and S6).

Coping with change in salinity is important in many natural habitats. Differences between the type of osmolytes used by **C1** and **C5** species were observed (Figure 4). Genes for synthesis of glucosylglycerol (Figure S5) was observed in both groups. Genes of sucrose biosynthesis (Figure S5) were only detected in **C5** species. Genes for two trehalose biosynthesis (Figure S5) pathways (maltooligosyl-trehalose-synthase and trehalose synthase) were also only present in **C5** species. In contrast, only **C1** species had the genes for

maltooligosyl-trehalose-synthesis. Even though both groups had genes required for glycine betaine (GB) degradation (Figure S5), **C1** but not **C5** species had genes for import of GB into the cell. As soda lakes are often shallow, their microbial inhabitants are impacted by desiccation and rehydration cycles. Survival in these ecosystems likely depends on effective responses to changes in salt concentration, for example, when brines get diluted during rainfall. It remains unclear whether the observed differences in osmolyte production also contribute to success in photobioreactors.

Ca. P. alkaliphilum proteome under different pH and N source and within the soda lakes

To further explore the capacity of *Ca.* P. alkaliphilum to tolerate changing environments we carried out experiments within the tubular photobioreactors with different pH and N sources, namely, low pH (initial pH of 8.3) with either ammonia or nitrate and high pH (initial pH of 10.4) with either nitrate or urea. Analysis of the four tubular photobioreactors showed expression of 2,658 *Ca.* P. alkaliphilum genes under the four different pH and nitrogen conditions. The most abundant proteins were phycobilisome pigments, such as phycocyanin, making up ~20% of the cyanobacterial proteome. Surprisingly, proteomes were very similar across the four different conditions. To identify proteins that differed significantly in expression between different conditions, Student's t test was applied and corrected for multiple testing with permutation-based false discovery rate (FDR) to all replicates. The pH significantly affected the expression of 121 genes (shown by log Student's t test p value >2, FDR of 0.05, Tables S5 and S6).

In low-pH bioreactors, where the bicarbonate concentration was higher, *Ca.* P. alkaliphilum allocated more resources toward increased expression of the Calvin cycle proteins and of the large chain and small chain of RuBisCO, to assimilate the abundant CO_2 (Figure 6, Tables S5 and S6). Correspondingly, increased expression of CcmK and CcmM, the structural components of the carboxysome (Badger and Price 2003), was seen, allowing for accommodation of higher amounts of RuBisCO (Rae et al., 2013a). Conversely, in high-pH bioreactors, where bicarbonate concentrations were lower, higher expression of the bicarbonate transporter (BicA) and the photosystem I reaction center subunit IV (PsaE), which powers bicarbonate transport (Sültemeyer et al., 1997), was observed. Additionally, expression of the CO_2 hydration protein (ChpX), also involved in CO_2 uptake and hydration (Price 2011), was increased. The increased expression of BicA, PsaE, and ChpX likely works to increase the concentration of cytosolic bicarbonate, relative to a lower external bicarbonate concentration (Figure 6, Table S6). Additionally, the high pH led to a significant increase in the expression of photosynthetic ferredoxin-NADP reductase (Figure 6, Tables S5 and S6). At the end of the photosynthetic ETC, ferredoxin-NADP reductase transfers the energized electrons from ferredoxin to NADPH (Chitnis 2001). The conversion of CO_2 to HCO_3^- by ChpXY is driven by electron transport and proton translocation by dedicated NADPH dehydrogenase complexes (NDH-I_3 and NDH-I_4) (Badger and Price 2003; Maeda et al., 2002). This way, excess NADPH is used to retain CO_2 inside the cell by hydration to bicarbonate. A significant increase was also observed in expression of the ETC complexes including NdhFKH (Figure 6, Tables S5 and S6). Higher nighttime energy demands for cytoplasmic pH homeostasis of *Ca.* P. alkaliphilum under high pH could be a possible explanation.

Expression of genes for nitrogen acquisition was dependent on the source of nitrogen. When provided with nitrate, expression of nitrate uptake and assimilation genes increased, regardless of the pH of the medium. A significant increase was observed for NrtA, the nitrate/nitrite binding protein involved in nitrate/nitrite uptake, and ferredoxin-nitrite reductase (NirA) converting nitrite to ammonium. When provided with urea, urea import proteins (Urt) as well as urease subunits were highly expressed (Figure 6, Table S6). Nitrogenase genes were expressed in all experiments, despite the presence of fixed nitrogen in the form of ammonia, nitrate, or urea. The form of nitrogen source did not show any significant change in the expression of central carbon metabolism genes, as was previously discussed (Anderton et al., 2020). This observation could be due to different timing of sampling between these two studies, particularly as sample for this study were taken at the point of depletion of nitrogen in culture media.

Proteomes acquired for *Ca.* P. alkaliphilum in photobioreactors were compared with proteomes obtained of their close relatives living in soda lakes *Phormidium sp. GEM2.Bin31* (**C5**). The relative gene expressions were very similar, with a few exceptions (Figure 6, Tables S4 and S5). Phycocyanin and phycoerythrocyanin subunits were expressed 2 times more in the lakes, potentially enabling harvesting of more light on both the longer and shorter end of the spectrum. Orange carotenoid protein was expressed 10 times more in the lakes, potentially indicating a higher need for protection against photodamage. These differences might result from adaptation to growth on light-emitting diode lights instead of natural sunlight. Conversely,

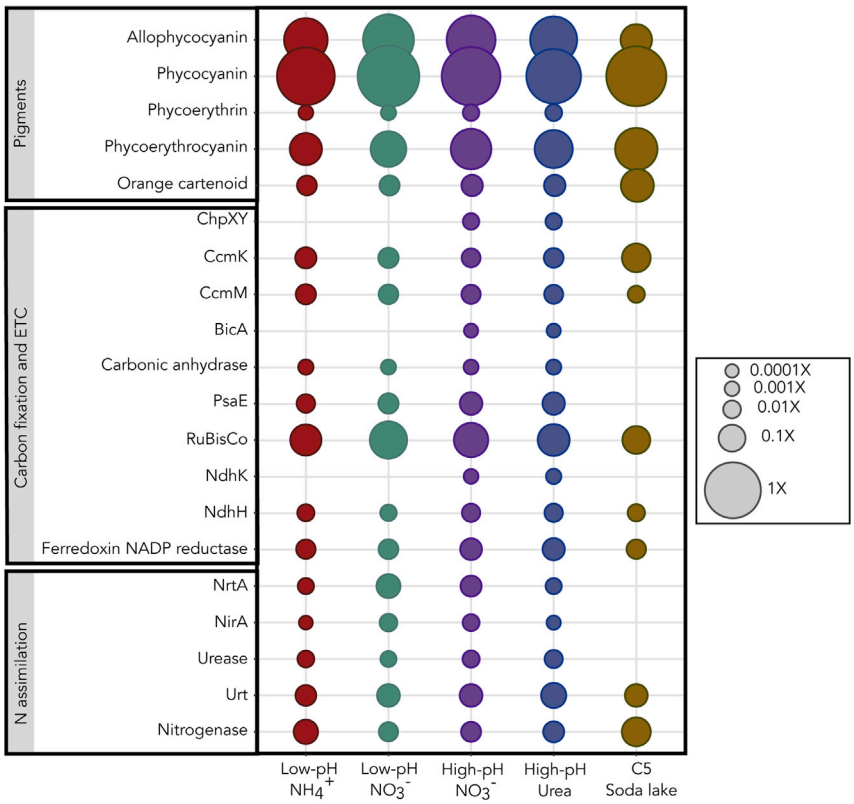


Figure 6. Bubble plot comparing protein expression of *Ca.* P. alkaliphilum and its close relative *Phormidium sp. GEM2.Bin31* (C5) from Cariboo soda lakes

Ca. P. alkaliphilum protein expression is shown for four different conditions (low pH with NH_4^+ or NO_3^-, high pH with NO_3^- or urea). Size of the bubble is the average of normalized spectral abundance factor (NSAF) from quadruplet samples across each condition normalized against the summed NSAF of ribosomal proteins, translation factors, and protein chaperones in the MAG's proteome.
See also Tables S5 and S6.

allophycocyanin was expressed 5 times lower in the soda lakes *GEM2.Bin31* and expression of phycoerythrin subunits, with maximum absorbance at 495 and 560 nm, was not detected in the lakes, potentially due to the light attenuation profile of lakes containing of high dissolved organic matter (Butturini et al., 2020; Croce and Van Amerongen 2014; Markager and Vincent 2000). CcmK was highly expressed in the lakes, contributing to the acquisition of bicarbonate. However, RuBisCO subunits showed higher expression in the photobioreactors compared with the lake (Figure 6, Table S6), possibly due to increased energy demands for homeostasis in the more dynamic lake environment, leaving less energy for carbon fixation.

Expression of nitrate uptake and assimilation genes could not be detected in the lakes, consistent with the absence of detectable nitrate in the lakes (Zorz et al., 2019). Urea transporters were detected in both lake and bioreactors; however, no expression for the urease subunits was detected in the lakes. Expression of nitrogenase was 10 times higher in the lakes, indicating the importance of dinitrogen as a nitrogen source for *GEM2.Bin31* (Figure 6, Tables S5 and S6). The apparent capability of this group of cyanobacteria to allocate resources to the acquisition of a wide range of nitrogen sources could contribute to their survival and ecological success both in natural and engineered environments.

Conclusion

A whole-genome sequence was obtained for an alkaliphilic cyanobacterium affiliated with the genus *Phormidium*. This newly identified species was enriched from alkaline soda lake microbial mats in high-pH photobioreactors. Genome assembly challenges caused by prolific repeat sequences were overcome with Nanopore sequencing. This led to a closed, circular, near-perfect whole-genome sequence, the first of its kind for this genus. The species was provisionally named *Candidatus* "Phormidium alkaliphilum." It

occurs in soda lakes worldwide and is the most abundant microbial species overall in Lake Goodenough, a soda lake in Canada. It also belongs to a clade of closely related *Phormidia* with potential for biotechnology, because of their rapid and robust growth. Evolutionary changes in the genome during 3 years of laboratory cultivation consisted of the elimination of genes that might not be useful in the photobioreactor environment. Selection of different strain variants contributed to photobioreactor performance under different operating conditions. Comparative genomics showed that these favorable assets could be explained by genes for harvesting sunlight across a wide portion of the solar spectrum, respiratory systems tuned to low-oxygen conditions, as well as an expanded set of genes involved in biofilm formation and response mechanisms.

We show that within the genus *Phormidium*, soda lake cyanobacteria form a separate clade from their relatives adapted to neutral pH environments. However, a lack of clear differences in gene content between alkaliphilic and neutral pH species suggested that both *Phormidium* lineages inherently have coping strategies for high-pH conditions. Consistently, comparison of proteomes of *Ca.* P. alkaliphilum grown at pH 9 and 11 showed only modest differences in gene expression in response to pH.

Limitations of study

Portions of this study rely on asynchronous datasets, such that the proteome and strain variation data were obtained from the tubular photobioreactors, and were not contemporaneous to the sample used for the long-read Nanopore assembled genome. As such, it is possible that the protein expression may be different under the stirred photo bioreactor conditions. However, given the consistency in expressed proteins under different tubular photobioreactor conditions (i.e., pH and nitrogen) this is unlikely. There is a difference in sequencing depth with short-read Illumina platform between the tubular photobioreactor, very deeply sequenced, and the stirred photobioreactor, not deeply sequenced. This difference could mean that the low level of strain heterogeneity observed in the stirred conditions may not completely capture all the variant bases within the population. However, the lower level of sequence variation observed for the deeply sequenced high pH-urea conditions indicates that the amount of strain variation is not simply a function of sequencing depth.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHODS DETAILS
  - Cyanobacterial consortium cultivation and sample preparation
  - High molecular weight DNA extraction and long read DNA sequencing
  - High molecular weight DNA extraction and short read DNA sequencing
  - Genome assembly and validation
  - Genome annotation
  - Protein extraction, peptide preparation, and 1D-LC-MS/MS
  - Phylogenetic analysis
  - Microscopy imaging

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.103405.

ACKNOWLEDGMENTS

The authors thank William Richardson and Dan Liu for laboratory support and Angela Kouris for help with metaproteomics sample preparation and analysis. We also would like to thank Jackie Zorz and Srijak Bhatnagar for help with metagenome analysis and image preparation and Maria A. Bautista for phylogenetic tree preparation. This study was supported by the Natural Sciences and Engineering Research Council of Canada, Canada Foundation for Innovation (CFI), Canada First Research Excellence Fund, Alberta

Innovates, the Government of Alberta, the University of Calgary, Israel Binational Science Foundation (BSF, Grant #2019055 to M.K.), and the U.S. National Science Foundation (grant OIA #1934844 to M.K.).

AUTHOR CONTRIBUTIONS

Conceptualization, M.A., A.K.H., and M.S.; methodology, M.A., A.V., D.M., and M.H.; software, M.S., X.D., A.K.H., and M.K.; formal analysis, M.A., M.S., and A.K.H.; Investigation, M.A., M.S., and A.K.H.; data curation, M.A., M.S., and X.D.; writing – original draft, M.A. and A.K.H.; writing – review & editing, M.S., A.K.H., M.K., and M.A.; visualization, M.A. and M.S.; funding acquisition, M.S.; supervision, M.S. and A.K.H.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 28, 2021
Revised: August 27, 2021
Accepted: November 3, 2021
Published: December 17, 2021

REFERENCES

Almon, H., and Böger, P. (1988). Hydrogen metabolism of the unicellular cyanobacterium Chroococcidiopsis thermalis ATCC29380. FEMS Microbiol. Lett. *49*, 445–449.

Anderton, C.R., Mobberley, J.M., Cole, J.K., Nunez, J.R., Starke, R., Boaro, A.A., Yesiltepe, Y., Morton, B.R., Cory, A.B., Cardamone, H.C., et al. (2020). Nitrogen source governs community carbon metabolism in a model hypersaline benthic Phototrophic biofilm. mSystems *5*, e00260-00220.

Aono, R., Ito, M., and Machida, T. (1999). Contribution of the cell wall component teichuronopeptide to pH homeostasis and alkaliphily in the alkaliphile Bacillus lentus C-125. J. Bacteriol. *181*, 6600–6606.

Ataeian, M., Liu, Y., Canon-Rubio, K.A., Nightingale, M., Strous, M., and Vadlamani, A. (2019). Direct capture and conversion of CO_2 from air by growing a cyanobacterial consortium at pH up to 11.2. Biotechnol. Bioeng. *116*, 1604–1611.

Badger, M.R., and Price, G.D. (2003). CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. J. Exp. Bot. *54*, 609–622.

Beal, C.M., Archibald, I., Huntley, M.E., Greene, C.H., and Johnson, Z.I. (2018). Integrating algae with bioenergy carbon capture and storage (ABECCS) increases sustainability. Earths Future *6*, 524–542.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.

Berman-Frank, I., Lundgren, P., and Falkowski, P. (2003). Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. Res. Microbiol. *154*, 157–164.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., and Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. *47*, W81–W87.

Bothe, H., Schmitz, O., Yates, M.G., and Newton, W.E. (2010). Nitrogen fixation and hydrogen metabolism in cyanobacteria. Microbiol. Mol. Biol. Rev. *74*, 529–551.

Brady, A.L., Druschel, G., Leoni, L., Lim, D.S., and Slater, G.F. (2013). Isotopic biosignatures in carbonate-rich, cyanobacteria-dominated microbial mats of the Cariboo Plateau, B.C. Geobiology *11*, 437–456.

Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner. United States. https://www.osti.gov/servlets/purl/1241166.

Butturini, A., Herzsprung, P., Lechtenfeld, O.J., Venturi, S., Amalfitano, S., Vazquez, E., Pacini, N., Harper, D.M., Tassi, F., and Fazi, S. (2020). Dissolved organic matter in a tropical saline-alkaline lake of the East African Rift Valley. Water Res. *173*, 115532.

Cano, M., Holland, S.C., Artier, J., Burnap, R.L., Ghirardi, M., Morgan, J.A., and Yu, J. (2018). Glycogen synthesis and metabolite overflow contribute to energy balancing in cyanobacteria. Cell Rep. *23*, 667–672.

Cauchie, H.-M., Hoffmann, L., Jaspar-Versali, M.-F., Salvia, M., and Thome, J.-P. (1995). Daphnia magna Straus living in an aerated sewage lagoon as a source of chitin: ecological aspects. Belg. J. Zool. *125*, 67–78.

Chitnis, P.R. (2001). Photosystem I: function and physiology. Annu. Rev. Plant Biol. *52*, 593–626.

Cole, J.K., Hutchison, J.R., Renslow, R.S., Kim, Y.M., Chrisler, W.B., Engelmann, H.E., Dohnalkova, A.C., Hu, D., Metz, T.O., Fredrickson, J.K., et al. (2014). Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions. Front Microbiol *5*, 109.

Costa, K.C., Navarro, J.B., Shock, E.L., Zhang, C.L., Soukup, D., and Hedlund, B.P. (2009). Microbiology and geochemistry of great boiling and mud hot springs in the United States Great Basin. Extremophiles *13*, 447–459.

Croce, R., and Van Amerongen, H. (2014). Natural strategies for photosynthetic light harvesting. Nat. Chem. Biol. *10*, 492–501.

D'mello, R., Hill, S., and Poole, R.K. (1996). The cytochrome bd quinol oxidase in Escherichia coli has an extremely high oxygen affinity and two oxygen-binding haems: implications for regulation of activity in vivo by oxygen inhibition. Microbiology *142*, 755–763.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. Gigascience *10*, giab008.

Dikiy, I., Edupuganti, U.R., Abzalimov, R.R., Borbat, P.P., Srivastava, M., Freed, J.H., and Gardner, K.H. (2019). Insights into histidine kinase activation mechanisms from the monomeric blue light sensor EL346. Proc. Natl. Acad. Sci. U S A *116*, 4963–4972.

Dixon, R.O.D. (1972). Hydrogenase in legume root nodule bacteroids: occurrence and properties. Archiv für Mikrobiol. *85*, 193–201.

Dong, X., and Strous, M. (2019). An integrated pipeline for annotation and visualization of metagenomic contigs. Front. Genet. *10*, 999.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinform. *9*, 18.

Fabry, V.J., Seibel, B.A., Feely, R.A., and Orr, J.C. (2008). Impacts of ocean acidification on marine fauna and ecosystem processes. ICES J. Mar. Sci. *65*, 414–432.

Finzi-Harta, J.A., Pett-Ridgeb, J., Weberb, P.K., Popac, R., Fallond, S.J., Gundersona, T., Hutcheonb, I.D., Nealsona, K.H., and Caponea, D.G. (2009). Correction for Finzi-Hart et al., Fixation and fate of C and N in the cyanobacterium Trichodesmium using nanometer-scale secondary ion mass spectrometry. Proc. Natl. Acad. Sci. U S A *106*, 9931.

Florens, L., Carozza, M.J., Swanson, S.K., Fournier, M., Coleman, M.K., Workman, J.L., and Washburn, M.P. (2006). Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. Methods *40*, 303–311.

Forehead, H.I., and O'Kelly, C.J. (2013). Small doses, big troubles: modeling growth dynamics of organisms affecting microalgal production cultures in closed photobioreactors. Bioresour. Technol. *129*, 329–334.

Galperin, M.Y. (2010). Diversity of structure and function of response regulator output domains. Curr. Opin. Microbiol. *13*, 150–159.

Gao, F., and Zhang, C.-T. (2008). Ori-Finder: a web-based system for finding oriC s in unannotated bacterial genomes. BMC Bioinform. *9*, 79.

García-Pichel, F., Belnap, J., Neuer, S., and Schanz, F. (2003). Estimates of global cyanobacterial biomass and its distribution. Algolog. Stud. *109*, 213–227.

Gong, X., Garcia-Robledo, E., Lund, M.B., Lehner, P., Borisov, S.M., Klimant, I., Revsbech, N.P., and Schramm, A. (2018). Gene expression of terminal oxidases in two marine bacterial strains exposed to nanomolar oxygen concentrations. FEMS Microbiol. Ecol. *94*, fiy072.

Gronenberg, L.S., Marcheschi, R.J., and Liao, J.C. (2013). Next generation biofuel engineering in prokaryotes. Curr. Opin. Chem. Biol. *17*, 462–471.

Haines, T.H. (2001). Do sterols reduce proton and sodium leaks through lipid bilayers? Prog. lipid Res. *40*, 299–324.

Hamamoto, T., Hashimoto, M., Hino, M., Kitada, M., Seto, Y., Kudo, T., and Horikoshi, K. (1994). Characterization of a gene responsible for the Na+/H+ antiporter system of alkalophilic Bacillus species strain C-125. Mol. Microbiol. *14*, 939–946.

Hamann, E., Gruber-Vodicka, H., Kleiner, M., Tegetmeyer, H.E., Riedel, D., Littmann, S., Chen, J., Milucka, J., Viehweger, B., Becker, K.W., et al. (2016). Environmental Breviatea harbour mutualistic Arcobacter epibionts. Nature *534*, 254–258.

Hauß, T., Dante, S., Dencher, N.A., and Haines, T.H. (2002). Squalane is in the midplane of the lipid bilayer: implications for its function as a proton permeability barrier. Biochim. Biophys. Acta (BBA)-Bioenerg. *1556*, 149–154.

Herrero, A., and Flores, E. (2019). Genetic responses to carbon and nitrogen availability in Anabaena. Environ. Microbiol. *21*, 1–17.

Jain, C., Rodriguez, -R.L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. *9*, 5114.

Jones, B.F., Eugster, H.P., and Rettig, S.L. (1977). Hydrochemistry of the lake Magadi basin, Kenya. Geochim. Cosmochim. Acta *41*, 53–72.

Jungblut, A.D., Hawes, I., Mountfort, D., Hitzfeld, B., Dietrich, D.R., Burns, B.P., and Neilan, B.A. (2005). Diversity within cyanobacterial mat communities in variable salinity meltwater ponds of McMurdo Ice Shelf, Antarctica. Environ. Microbiol. *7*, 519–529.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kerfeld, C.A., and Melnicki, M.R. (2016). Assembly, function and evolution of cyanobacterial carboxysomes. Curr. Opin. Plant Biol. *31*, 66–75.

Khomutovska, N., Sandzewicz, M., Łach, Ł., Suska-Malawska, M., Chmielewska, M., Mazur-Marzec, H., Cegłowska, M., Niyatbekov, T., Wood, S.A., Puddick, J., et al. (2020). Limited microcystin, anatoxin and cylindrospermopsin production by cyanobacteria from microbial mats in cold deserts. Toxins *12*, 244.

Kim, J.H., Choi, W., Jeon, S.M., Kim, T., Park, A., Kim, J., Heo, S.J., Oh, C., Shim, W.B., and Kang, D.H. (2015). Isolation and characterization of Leptolyngbya sp. KIOST-1, a basophilic and euryhaline filamentous cyanobacterium from an open paddle-wheel raceway Arthrospira culture pond in Korea. J. Appl Microbiol. *119*, 1597–1612.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. *37*, 540–546.

Konto-Ghiorghi, Y., Mairey, E., Mallet, A., Duménil, G., Caliot, E., Trieu-Cuot, P., and Dramsi, S. (2009). Dual role for pilus in adherence to epithelial cells and biofilm formation in Streptococcus agalactiae. PLoS Pathog. *5*, e1000422.

Kuchma, S.L., Delalez, N.J., Filkins, L.M., Snavely, E.A., Armitage, J.P., and O'Toole, G.A. (2015). Cyclic di-GMP-mediated repression of swarming motility by Pseudomonas aeruginosa PA14 requires the MotAB stator. J. Bacteriol. *197*, 420–430.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. *33*, 1870–1874.

Kupriyanova, E.V., Cho, S.M., Park, Y.-I., Pronina, N.A., and Los, D.A. (2016). The complete genome of a cyanobacterium from a soda lake reveals the presence of the components of CO2-concentrating mechanism. Photosynth. Res. *130*, 151–165.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Latifi, A., Ruiz, M., and Zhang, C.-C. (2009). Oxidative stress in cyanobacteria. FEMS Microbiol. Rev. *33*, 258–278.

Lehmann, M., and Wober, G. (1976). Accumulation, mobilization and turn-over of glycogen in the blue-green bacterium Anacystis nidulans. Arch. Microbiol. *111*, 93–97.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

Lopez-Cortes, A., Garcia-Pichel, F., Nubel, U., and Vazquez-Juarez, R. (2001). Cyanobacterial diversity in extreme environments in Baja California, Mexico: a polyphasic study. Int. Microbiol. *4*, 227–236.

Maeda, S.i., Badger, M.R., and Price, G.D. (2002). Novel gene products associated with NdhD3/D4-containing NDH-1 complexes are involved in photosynthetic CO2 hydration in the cyanobacterium, Synechococcus sp. PCC7942. Mol. Microbiol *43*, 425–435.

Markager, S., and Vincent, W.F. (2000). Spectral light attenuation and the absorption of UV and blue light in natural waters. Limnol. Oceanograph. *45*, 642–650.

Marsan, D., Place, A., Fucich, D., and Chen, F. (2017). Toxin-antitoxin systems in estuarine synechococcus strain CB0101 and their transcriptomic responses to environmental stressors. Front. Microbiol. *8*, 1213.

Mazel, D., Houmard, J., Castets, A.M., and Tandeau de Marsac, N. (1990). Highly repetitive DNA sequences in cyanobacterial genomes. J. Bacteriol. *172*, 2755–2761.

Nelson, W.C., Maezato, Y., Wu, Y.W., Romine, M.F., and Lindemann, S.R. (2016). Identification and resolution of microdiversity through metagenomic sequencing of parallel consortia. Appl. Environ. Microbiol. *82*, 255–267.

Nies, F., Wörner, S., Wunsch, N., Armant, O., Sharma, V., Hesselschwerdt, A., Falk, F., Weber, N., Weiß, J., Trautmann, A., et al. (2017). Characterization of Phormidium lacuna strains from the North Sea and the Mediterranean Sea for biotechnological applications. Process Biochem. *59*, 194–206.

Nikolaou, C., and Almirantis, Y. (2005). A study on the correlation of nucleotide skews and the positioning of the origin of replication: different modes of replication in bacterial species. Nucleic Acids Res. *33*, 6816–6822.

Oberg, A.L., and Vitek, O. (2009). Statistical design of quantitative mass spectrometry-based proteomic experiments. J. Proteome Res. *8*, 2144–2156.

Ohbayashi, R., Hirooka, S., Onuma, R., Kanesaki, Y., Hirose, Y., Kobayashi, Y., Fujiwara, T., Furusawa, C., and Miyagishima, S.-y. (2020). Evolutionary Changes in DnaA-dependent Chromosomal Replication in cyanobacteria. Front. Microbiol. *11*, 786.

Ohbayashi, R., Watanabe, S., Ehira, S., Kanesaki, Y., Chibazakura, T., and Yoshikawa, H. (2016). Diversification of DnaA dependency for DNA replication in cyanobacterial evolution. ISME J. *10*, 1113–1121.

Olm, M.R., Crits-Christoph, A., Bouma-Gregson, K., Firek, B.A., Morowitz, M.J., and Banfield, J.F. (2021). inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat. Biotechnol. *39*, 727–736.

Omata, T., Price, G.D., Badger, M.R., Okamura, M., Gohta, S., and Ogawa, T. (1999). Identification of an ATP-binding cassette transporter involved in bicarbonate uptake in the cyanobacterium Synechococcus sp. strain PCC 7942. Proc. Natl. Acad. Sci. *96*, 13571–13576.

Oswald, W.J. (1980). Algal production–problems,achievements and potential. In Algae Biomass: Production and use. Sponsored by the National Council for Research and Development, Israel and the Gesellschaft Fur Strahlen-Und Umweltforschung (GSF), Munich, Germany, G. Shelef and C.J. Soeder, eds. (Elsevier/North-Holland: Biomedical Press), pp. 1–8.

Paerl, H.W., Pinckney, J.L., and Steppe, T.F. (2000). Cyanobacterial-bacterial mat consortia: examining the functional unit of microbial survival and growth in extreme environments. Environ. Microbiol. *2*, 11–26.

Panja, A.S., Maiti, S., and Bandyopadhyay, B. (2020). Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges. Scientific Rep. *10*, 1–9.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. *25*, 1043–1055.

Partensky, F., Hess, W.R., and Vaulot, D. (1999). Prochlorococcus, a marine photosynthetic prokaryote of global significance. Microbiol. Mol. Biol. *63*, 106–127.

Pérez, A.-C.S., Karst, J.C., Davi, M., Guijarro, J.I., Ladant, D., and Chenal, A. (2010). Characterization of the regions involved in the calcium-induced folding of the intrinsically disordered RTX motifs from the Bordetella pertussis adenylate cyclase toxin. J. Mol. Biol. *397*, 534–549.

Petersen, J.M., Kemper, A., Gruber-Vodicka, H., Cardini, U., van der Geest, M., Kleiner, M., Bulgheresi, S., Mussmann, M., Herbold, C., Seah, B.K., et al. (2016). Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. Nat. Microbiol. *2*, 16195.

Ponting, C.P., Aravind, L., Schultz, J., Bork, P., and Koonin, E.V. (1999). Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. J. Mol. Biol. *289*, 729–745.

Preiss, J. (1984). Bacterial glycogen synthesis and its regulation. Annu. Rev. Microbiol. *38*, 419–458.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics (Oxford, England) *21* (*Suppl. 1*), i351–i358.

Price, G.D. (2011). Inorganic carbon transporters of the cyanobacterial CO2 concentrating mechanism. Photosynth. Res. *109*, 47–57.

Price, G.D., Badger, M.R., Woodger, F.J., and Long, B.M. (2007). Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. J. Exp. Bot. *59*, 1441–1461.

Rae, B.D., Long, B.M., Badger, M.R., and Price, G.D. (2013a). Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria. Microbiol. Mol. Biol. Rev. *77*, 357–379.

Rae, B.D., Long, B.M., Whitehead, L.F., Förster, B., Badger, M.R., and Price, G.D. (2013b). Cyanobacterial carboxysomes: microcompartments that facilitate CO_2 fixation. J. Mol. Microbiol. Biotechnol. *23*, 300–307.

Raven, J.A., Cockell, C.S., and De La Rocha, C.L. (2008). The evolution of inorganic carbon concentrating mechanisms in photosynthesis. Philos. Trans. R. Soc. Lond. B Biol. Sci. *363*, 2641–2650.

Rozanov, A.S., Shipova, A.A., Bryanskaya, A.V., and Peltek, S.E. (2019). Metagenome-assembled genome sequence of Phormidium sp. strain SL48-SHIP, isolated from the microbial mat of Salt Lake Number 48 (Novosibirsk Region, Russia). Microbiol Resour Announc *8*, e00651–00619.

Samylina, O., Sapozhnikov, F., Gainanova, O.Y., Ryabova, A., Nikitin, M., and Sorokin, D.Y. (2014). Algo-bacterial communities of the Kulunda steppe (Altai region, Russia) soda lakes. Microbiology *83*, 849–860.

Sanchez-Baracaldo, P., Hayes, P.K., and Blank, C.E. (2005). Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach. Geobiology *3*, 145–165.

Sánchez-Magraner, L., Viguera, A.R., García-Pacios, M., Garcillán, M.P., Arrondo, J.-L.R., de la Cruz, F., Goñi, F.M., and Ostolaza, H. (2007). The calcium-binding C-terminal domain of Escherichia coli α-hemolysin is a major determinant in the surface-active properties of the protein. J. Biol. Chem. *282*, 11827–11835.

Schatz, D., Nagar, E., Sendersky, E., Parnasa, R., Zilberman, S., Carmeli, S., Mastai, Y., Shimoni, E., Klein, E., and Yeger, O. (2013). Self-suppression of biofilm formation in the cyanobacterium S ynechococcus elongatus. Environ. Microbiol. *15*, 1786–1794.

Serang, O., MacCoss, M.J., and Noble, W.S. (2010). Efficient marginalization to compute protein posterior probabilities from shotgun mass spectrometry data. J. Proteome Res. *9*, 5346–5357.

Sharp, C.E., Urschel, S., Dong, X., Brady, A.L., Slater, G.F., and Strous, M. (2017). Robust, high-productivity phototrophic carbon capture at high pH and alkalinity using natural microbial communities. Biotechnol. Biofuels *10*, 84.

Sherman, L.A., Meunier, P., and Colon-Lopez, M.S. (1998). Diurnal rhythms in metabolism- A day in the life of a unicellular. Photosynth. Res. *58*, 25–42.

Shibata, M., Katoh, H., Sonoda, M., Ohkawa, H., Shimoyama, M., Fukuzawa, H., Kaplan, A., and Ogawa, T. (2002). Genes essential to sodium-dependent bicarbonate transport in cyanobacteria: function and phylogenetic analysis. J. Biol. Chem. *277*, 18658–18664.

Shibata, M., Ohkawa, H., Kaneko, T., Fukuzawa, H., Tabata, S., Kaplan, A., and Ogawa, T. (2001). Distinct constitutive and low-CO2-induced CO2 uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms. Proc. Natl. Acad. Sci. *98*, 11789–11794.

Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., et al. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc. Natl. Acad. Sci. U S A. *110*, 1053–1058.

Soo, R.M., Hemp, J., Parks, D.H., Fischer, W.W., and Hugenholtz, P. (2017). On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. Science *355*, 1436–1440.

Sorokin, D.Y., Mosier, D., Zorz, J.K., Dong, X., and Strous, M. (2020). Wenzhouxiangella strain AB-CW3, a proteolytic bacterium from hypersaline soda lakes that preys on cells of gram-positive bacteria. Front. Microbiol. *11*, 597686.

Stucken, K., John, U., Cembella, A., Murillo, A.A., Soto-Liebe, K., Fuentes-Valdés, J.J., Friedel, M., Plominsky, A.M., Vásquez, M., and Glöckner, G. (2010). The Smallest Known Genomes of Multicellular and Toxic cyanobacteria: Comparison, Minimal gene Sets for Linked Traits and the Evolutionary Implications. PLoS One *5*, e9235.

Sültemeyer, D., Dean Price, G., Bryant, D.A., and Badger, M.R. (1997). PsaE- and NdhF-mediated electron transport affect bicarbonate transport rather than carbon dioxide uptake in the cyanobacteriumSynechococcus sp. PCC7002. Planta *201*, 36–42.

Syed, K.A., Beyhan, S., Correa, N., Queen, J., Liu, J., Peng, F., Satchell, K.J., Yildiz, F., and Klose, K.E. (2009). The Vibrio cholerae flagellar regulatory hierarchy controls expression of virulence factors. J. Bacteriol. *191*, 6555–6570.

Taton, A., Grubisic, S., Balthasart, P., Hodgson, D.A., Laybourn-Parry, J., and Wilmotte, A. (2006). Biogeographical distribution and ecological ranges of benthic cyanobacteria in East Antarctic lakes. FEMS Microbiol. Ecol. *57*, 272–289.

Ting, C.S., Rocap, G., King, J., and Chisholm, S.W. (2002). Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. Trends Microbiol. *10*, 134–142.

Tutolo, B.M., and Tosca, N.J. (2018). Experimental examination of the Mg-silicate-carbonate system at ambient temperature: implications for alkaline chemical sedimentation and lacustrine carbonate formation. Geochim. Cosmochim. Acta *225*, 80–101.

Van Wychen, S., and Laurens, L.M. (2016). Determination of Total Carbohydrates in Algal Biomass: Laboratory Analytical Procedure (LAP) (National Renewable Energy Lab.(NREL)).

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. *27*, 737–746.

Vavourakis, C.D., Andrei, A.S., Mehrshad, M., Ghai, R., Sorokin, D.Y., and Muyzer, G. (2018).

A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. Microbiome *6*, 168.

Vonshak, A., and Richmond, A. (1988). Mass production of the blue-green alga Spirulina: an overview. Biomass *15*, 233–247.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., and Young, S.K. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLos one *9*, e112963.

Ward, D.M., Ferris, M., Nold, S., and Bateson, M. (1998). A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microb. Mol. Biol. *62*, 1353–1370.

Watanabe, S. (2020). Cyanobacterial multi-copy chromosomes and their replication. Biosci. Biotechnol. Biochem. *84*, 1309–1321.

Wen, Y., Behiels, E., and Devreese, B. (2014). Toxin-Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity. Pathog. Dis. *70*, 240–249.

Wick, R.R., Judd, L.M., and Holt, K.E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. *20*, 1–10.

Wilson, A., Ajlani, G., Verbavatz, J.-M., Vass, I., Kerfeld, C.A., and Kirilovsky, D. (2006). A soluble carotenoid protein involved in phycobilisome-related energy dissipation in cyanobacteria. Plant Cell *18*, 992–1007.

Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359–362.

Worning, P., Jensen, L.J., Hallin, P.F., Stærfeldt, H.H., and Ussery, D.W. (2006). Origin of replication in circular prokaryotic chromosomes. Environ. Microbiol. *8*, 353–361.

Zehr, J.P., Bench, S.R., Carter, B.J., Hewson, I., Niazi, F., Shi, T., Tripp, H.J., and Affourtit, J.P. (2008). Globally distributed uncultivated oceanic N2-fixing cyanobacteria lack oxygenic photosystem II. Science *322*, 1110–1112.

Zhang, J.-Y., Chen, W.-L., and Zhang, C.-C. (2009). hetR and patS, two genes necessary for heterocyst pattern formation, are widespread in filamentous nonheterocyst-forming cyanobacteria. Microbiology *155*, 1418–1426.

Zorz, J.K., Sharp, C., Kleiner, M., Gordon, P.M.K., Pon, R.T., Dong, X., and Strous, M. (2019). A shared core microbiome in soda lakes separated by large distances. Nat. Commun. *10*, 4230.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
microbial mats from soda lakes in British Colombia, Canada	This study	N/A
Chemicals, peptides, and recombinant proteins		
Mag-Bind® TotalPure NGS	Omega biotech	SKU: M1378-01
Critical commercial assays		
Qubit dsDNA HS Assay Kit	Invitrogen, Thermo Fisher Scientific	Q32851
Ligation Sequencing Kit	Oxford Nanopore Technologies	SQK-LSK109
FastDNA SPIN Kit for Soil protocol	MP Biomedicals	SKU:116560200-CF
Nextera DNA Flex Library prep kit	Illumina	20018705 20018708
NEB Ultra II library preparation kit	New England Biolabs	E7103L
filter-aided sample preparation (FASP)	Wisniewski et al., (2009)	https://proteomicsresource.washington.edu/protocols03/FASPprotocols.php
Pierce Micro BCA assay	Thermo Scientific Pierce	Catalog number: 23,235
Deposited data		
metagenome raw reads of the Illumina and Nanopore sequencing	This study	BioProject PRJNA377096 SAMN18025958, SAMN18025959 SAMN18025960
metagenome raw reads of the tubular bioreactors	This study	BioProject PRJNA377096 SAMN17969417, SAMN17969418, SAMN17969419, SAMN17969420
mass spectrometry proteomics data	This study	dataset identifier PXD024393
Gem2 Bin 31 (C5)	Zorz et al., (2019)	GCA_007693465
Microcoleus sp IPPAS B-353	Kupriyanova et al., (2016)	GCF_009846485
Cyanobacterium T3Sed10	Vavourakis et al., (2018)	GCA_003566575
Geitlerinema sp. P-1104	Samylina et al., 2014	GCF_012911965
Phormidium sp. OSCR	Cole et al., 2014	GCA_001314905
Phormidium sp. SL48-SHIP	Rozanov et al., 2019	GCA_004299065
Phormidium lacuna HE10JO	Nies et al., (2017)	GCF_900149785
Phormidium willei BDU 130791	N/A	GCF_001637315
LCM Bin17 (C1)	Zorz et al., (2019)	GCA_007692715
Cyanobacterium B1Sed10_56	Vavourakis et al., (2018)	GCA_003550805
Nodosilinea sp. P-1105	Samylina et al., 2014	GCF_012911975
Leptolyngbya sp. KIOST-1	Kim et al., 2015	GCF_000763385
Ca. P. alkaliphilum	This study	accession number CP075902
Software and algorithms		
Guppy v3.2.6	Wick et al., 2019	https://github.com/nanoporetech/medaka/issues/153
Porechop 0.2.4	https://bioweb.pasteur.fr/packages/pack@Porechop@0.2.4	https://github.com/rrwick/Porechop
flye v2.7	Kolmogorov et al., (2019)	https://github.com/fenderglass/Flye

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BWA-MEME v0.7.17	Li and Durbin (2009)	https://www.biostars.org/p/449598/
Racon v1.4.3	Vaser et al., (2017)	https://github.com/isovic/racon/releases
Medaka v0.11.5	N/A	https://github.com/nanoporetech/medaka/blob/master/CHANGELOG.md
Pilon v1.23	Walker et al., (2014)	https://github.com/broadinstitute/pilon/releases/
Ori-fFinder	Gao and Zhang (2008)	https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-79
CheckM	Parks et al., 2015	https://kbase.us/applist/apps/kb_Msuite/run_checkM_lineagetp_wf/release?gclid=CjwKCAjw95yJBhAgEiwAmRrutD9ncHIbhcZNcfddU-XBJEv-W864912sLT0c6VJnHcvgea5W3VeW-RoCmtIQAvD_BwE
BBMap	https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/	https://software.broadinstitute.org/software/igv/home
inStrain	Olm et al., (2021)	https://github.com/MrOlm/inStrain
bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
samtools	Danecek et al., (2021)	http://www.htslib.org/
metaErg	Dong and Strous (2019)	https://github.com/xiaoli-dong/metaerg
Antismash	Blin et al., (2019)	https://antismash.secondarymetabolites.org/#!/start
tandem repeat finder	Benson (1999)	https://github.com/Benson-Genomics-Lab/TRF
RepeatScout	Price et al., (2005)	https://github.com/mmcco/RepeatScout
LTR Harvest	Ellinghaus et al., (2008)	http://genometools.org/tools/gt_ltrharvest.html
FastANI	Jain et al., (2018)	https://kbase.us/applist/apps/FastANI/fast_ani/release?gclid=CjwKCAjw95yJBhAgEiwAmRrutMzVTtGzuIPOhWgbxLiMx-fARtjhQbqdRC5UgkLL6U-5uQVlaGf1KRoC87UQAvD_BwE
Mafft	Katoh and Standley (2013)	https://mafft.cbrc.jp/alignment/software/
MEGA7	Kumar et al., (2016)	https://www.megasoftware.net/
BBDuk v38.06	Bushnell (2014)	https://sourceforge.net/projects/bbmap/

RESOURCE AVAILABILITY

Lead contact

Further requests for information should be directed to the lead contact Alyse Hawley (alyse.hawley@ubc.ca).

Materials availability

This study did not generate unique reagents.

Data and code availability

- Sequence data generated in the study is available at GenBank under BioProject: PRJNA377096 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA377096).
- Biosamples for the metagenome raw reads of the tubular bioreactors are Biosample: SAMN17969417, SAMN17969418, SAMN17969419, SAMN17969420.
- Biosamples for the metagenome raw reads of the Illumina and Nanopore sequencing are Biosample: SAMN18025958, SAMN18025959, and SAMN18025960.

- Biosample for the *Ca.* P. alkaliphilum is Biosample: SAMN17928605 with accession number GenBank: CP075902.
- Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PRIDE: PXD024393.
- This study does not report original code.

METHODS DETAILS

Cyanobacterial consortium cultivation and sample preparation

Selective enrichment of the cyanobacterial consortium from soda lake microbial mats was initially performed in planar photobioreactors with different light filters in 2017. Biomass from the photobioreactor with the red filter was subsequently enriched for two years in tubular photobioreactors (Figure S1 (Ataeian et al., 2019)). The growth medium contained: K_2HPO_4 (1.44 mM), $MgSO_4{\cdot}7H_2O$ (1 mM), $CaCl_2{\cdot}2H_2O$ (0.17 mM), KCl (6 mM), NaCl (0.43 mM), ferric ammonium citrate (10mg$\cdot l^{-1}$). 1 ml L^{-1} of trace metal solution (Titriplex III (EDTA) (500 mg), $FeSO_4.7H_2O$ (200 mg), $ZnSO_4.7H_2O$ (10 mg), $MnCl_2.4H_2O$ (3mg), H_3BO_3 (30 mg), $CoCl_2.6H_2O$ (20 mg), $CuCl_2.2H_2O$ (1 mg), $NiCl_2.6H_2O$ (2 mg), $Na_2MoO_4.2H2O$ (3 mg) per 1000ml) was added after autoclaving of above medium. To investigate impacts of pH and nitrogen source, cyanobacterial consortium was grown under four different photobioreactor conditions in duplicate; low-pH ammonium: 4 mM NH_4Cl, initial pH 8.3; low-pH nitrate: 1 mM NH_4Cl, 3 mM $NaNO_3$, initial pH 8.3; high-pH nitrate: 1 mM NH_4Cl, 3 mM $NaNO_3$, initial pH 10.4; high-pH urea: 1 mM NH_4Cl, 3 mM Urea, initial pH 10.4. The initial pH of the media was set by mixing concentrated $NaHCO_3$ and Na_2CO_3 (Low-pH: 460 mmol/L sodium bicarbonate and 20 mmol/L sodium carbonate, high-pH: 35 mmol/L sodium bicarbonate and 230 mmol/L sodium carbonate) Soluble nitrogen species, bicarbonate, and carbonate concentrations were measured daily (Ataeian et al., 2019). When soluble nitrogen in the medium was depleted (usually after 4 to 5 days) the final pH reached approximately 9.3 for low-pH conditions and 11.2 for high-pH conditions. All experiments where carried out at ~25°c and were illuminated with two full spectrum LED lights (Model T5H0; 6400K, Sunblaster Holdings ULC, Langley, BC, Canada) with a 16:8 light to dark cycle. The LED lights mimic natural sunlight, with a spectrum between 380-780nm with a sharp peak at 453nm and a broader peak at 560nm, as measured using a LI-180 spectrometer (LI-COR Biosciences, Lincoln, NE, USA. 4L of growth medium was circulated through each bioreactor at a flow rate of 10 mL$\cdot min^{-1}$. At nitrogen depletion the biomass was harvested, and 10% of the harvested biomass was used for re-inoculation and start of a new cycle. This way, each culture was grown for at least two sequential cycles. All the harvested biomass samples were immediately frozen at −80°C. Ash free dry weight was measured for total productivity of the systems as previously described (Van Wychen and Laurens 2016).

After two years cultivation in the tubular photobioreactor, the biomass was transferred to stirred 10 L glass bottle photobioreactors (Figure S1). The high-pH medium (35 mmol/L sodium bicarbonate, 230 mmol/L sodium carbonate) with 1 mM NH_4Cl and 3 mM $NaNO_3$ was used and 1/3 of the volume (6.5-7L) was exchanged with fresh media every 7-10 days. The bioreactors were stirred at 300-330 RPM and were illuminated with a 16:8 light: dark cycle.

High molecular weight DNA extraction and long read DNA sequencing

Genomic DNA was extracted from biomass grown in stirred photobioreactors as previously described (Sorokin et al., 2020). Briefly, a freshly harvested and pelleted sample was used following a modified version of the 'Bacterial genomic DNA isolation using CTAB' protocol (Version 3, Joint Genome Institute). The pellet was completely resuspended in 1x TE buffer at room temperature, treated with lysozyme (100 mg/mL) for 30 minutes at 37°C, then with 10% SDS and proteinase K (20 mg/mL) for 2 hours at 55°C. Following cell lysis, 4.5 M NaCl and 10% CTAB solution were added to assist in DNA purification with a 10 minute incubation at 65°C. Organic extraction was performed using chloroform:isoamyl (24:1 ratio) and phenol:chloroform:isoamyl (25:24:1 ratio). The solution was mixed after each step and centrifuged at high speed for 10 minutes. The final top layer (aqueous phase) was mixed with 0.6 volumes of isopropanol and incubated at −20°C overnight to precipitate genomic DNA followed by 15 minutes centrifugation at 4°C. The pellet was washed with 70% ethanol and air dried at room temperature. The pellet was resuspended in 1x TE buffer and treated with RNase A (10 mg/mL) at 37°C for 1 hour. 3 M Sodium Acetate (1/10 volume) and 100% ethanol (2.5 volumes) were added and the mixture was incubated at −80°C for 30 minutes. The DNA was then pelleted by centrifugation at 4°C for 20 minutes, ethanol washed and air dried. The resuspended

DNA in DNase-free water was quantified using the Qubit dsDNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific) and stored at 4°C. The modified version of the 'Genomic DNA by Ligation (SQK-LSK109)' protocol (Oxford Nanopore Technologies) designed for SpotOn flow cells was used for the genomic DNA library preparation. Genomic DNA was sheared using a 25-gauge needle and syringe. Then NGS beads (NucleoMag) were used for clean-up. Repair and end-prep of the DNA as well as adapter ligation and clean-up were performed following the manufacturers protocol with changes mentioned in (Sorokin et al., 2020)). DNA was quantified using the Qubit dsDNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific) after every major step. Total DNA quantifications after every step was as follows: 4.2 μg post-needle shearing, 2.85 μg post-bead clean-up, 1.66 μg post-DNA repair and end-prep, and 1.07 μg post-adapter ligation and clean-up. The DNA library was loaded onto the SpotOn flow cell according to the manufacturer's instructions. The sequencing run was then started using the MinKNOW program to track and visualize output and a MinIT to store the resulting data. The run progressed for ~72 hours and generated 0.61 GB of data, with an N50 value of ~11,000.

High molecular weight DNA extraction and short read DNA sequencing

Genomic DNA was extracted from the same sample used for Nanopore sequencing, obtained from stirred photobioreactors in 2020. DNA was extracted from pelleted biomass using a modified version of the FastDNA SPIN Kit for Soil protocol (MP Biomedicals) as previously described (Costa et al., 2009), with a minor change being that the samples were processed in a bead beater twice for 30 s at setting 4.5 m/s. To prepare the genomic DNA sequencing library, the Nextera DNA Flex Library Prep protocol (Illumina) was used following the manufacturer's protocol. Shotgun metagenomic sequencing (2 × 300 bp) was performed using an Illumina MiSeq sequencer. BBDuk v38.06 (Bushnell 2014) was used with options "qtrim = rl trimq = 15" to trim reads that contained adapter sequences, remove low quality ends, eliminate contaminants, and filter out reads shorter than 30 bp after cutting the adapters.

Genomic DNA was also extracted from biomass from tubular photobioreactors in 2018: DNA was extracted from 0.4 g of −80 frozen microbial biofilm samples using the FastDNA Extraction Kit for Soil (MP Biomedicals, Santa Ana, CA, USA) according to manufacturer's protocol with minor modifications: centrifugation time was increased to 10 min and 5.5 M guanidine thiocyanate was used for additional purification step before washing with concentrated SEWS-M. DNA concentrations were measured using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Canada). Metagenomic library preparation and DNA sequencing was conducted at the Center for Health Genomics and Informatics in the Cumming School of Medicine, University of Calgary. DNA was sequenced paired end 2 x 150 bp with a 300 cycle mid-output reagent cartridge on the Illumina NextSeq 500 sequencer. The samples were sheared to approximately 350 bp via Covaris sonication, and the libraries were prepared with NEB Ultra II library preparation kit. After adaptor ligation, the average library sizes ranged from 471-483 bp. The adaptors were the NEBNext Multiplex Oligos for Illumina Set 1 and Set 2.

Genome assembly and validation

Assembly of cyanobacterial closed genome from Nanopore and Illumina reads was carried out as described in Sorokin et al. (2020). In brief, Nanopore reads were base-called in real time on MinIT with Guppy v3.2.6, only reads with Qscore of ≥7 were used in further analysis. Adapters were trimmed with Porechop 0.2.4 with default settings. Assembly of adapter-trimmed reads was carried out with flye v2.7 (Kolmogorov et al., 2019). The resulting single circular contig was polished with Nanopore reads four times with BWA-MEME v0.7.17 (Li and Durbin 2009) with '-x ont2d' and Racon v1.4.3 (Vaser et al., 2017) with default parameters and subsequently polished once with Medaka v0.11.5 (ONT) (r941_min_high_g303 model). Final polishing consisted of four rounds of BWA-MEM and Pilon v1.23 (Walker et al., 2014) with '–fix all –changes' using the quality-controlled Illumnia reads from the same biological sample. The circular contig was then rearranged with the start position at the *dnaA* gene. The entire whole-genome sequence was then curated manually. Cumulative GC skew was calculated and absence of a canonical origin of replication was shown using Ori-Finder (Gao and Zhang 2008)https://genskew.csb.univie.ac.at/. Distribution of marker genes was inspected using CheckM. Mapping (BBMap with options ambiguous=random minid=0.99, BWA) of both Nanopore and Illumina reads was inspected using the Integrated Genome Viewer (https://software.broadinstitute.org/software/igv/home). Near-absence of mapped reads in 30 regions could be attributed to assembly artefacts, leading to misplacement of reads to other regions. These problems were fixed manually. The final assembly had near perfect read placement of both Illumina and Nanopore reads with very few remaining problems, all localized to a few regions with long stretches of repetitive sequences.

To confirm that the *Ca.* P. alkaliphilum strain sequenced by Nanopore in the stirred bioreactor was the same in the tubular bioreactor conditions, analysis with inStrain (Olm et al., 2021) was carried out. The metagenomic database was constructed from the complete co-assembly of four tubular bioreactor metagenomes. Contigs matching to the *Ca.* P. alkaliphilum genome with >90% identity over 200bp were removed and the *Ca.* P. alkaliphilum genome appended to the metagenomic database. Quality controlled reads from each condition in the tubular photobioreactors were mapped separately to the complete metagenomic database using bowtie2 with default parameters. From the resulting sam files, reads mapping to the *Ca.* P. alkaliphilum genome were selected into a separate .bam file using samtools (Danecek et al., 2021). The resulting.bam file and the *Ca.* P. alkaliphilum whole genome sequence fasta (the reference genome) were provided as input to "inStrain profile". The option –min_cov 20 for the tubular bioreactors was used to account for the very deep sequencing that was carried out for these samples and –min_cov 5 for the stirred bioreactor. Only sites that displayed the expected sequencing depth across all samples were considered. Sites with anomalously low or high sequencing depth were always associated with low sequence complexity or repeat motifs, leading to misplacement of reads during mapping. Variant bases were classified into three groups based on proportion of reads containing the variant base relative to the reads containing the reference base. Sites with >95% of reads containing a variant base were considered single-nucleotide-substitutions (inStrain 'SNS'); sites with >50 - 95% of reads containing a variant base were considered consensus-variants (inStrain 'con_SNV'); sites with 5 – 49% of reads containing a variant base were considered single-nucleotide-variants (inStrain 'SNV'). Average read depth, proportion of genome covered and total number of quality controlled reads were determined by inStrain. K-mer analysis of quality-controlled reads was performed with kmercountexact, included with bbmap, with parameters "histcolumns=3".

Genome annotation

The genome was annotated using metaErg (Dong and Strous 2019). Absence of known genes involved in production of secondary metabolites and toxins was shown using Antismash (Blin et al., 2019). Orthologous genes shared between *Ca.* P. Alkaliphilum and a set of relevant other cyanobacteria were identified using blast and markov clustering as previously described (Sorokin et al., 2020). Repetitive sequences, transposases and retrotransposons were identified using tandem repeat finder (Benson 1999), RepeatScout (Price et al., 2005) and LTR Harvest (Ellinghaus et al., 2008). Genome synteny and average nucleotide identities between genomes were calculated using FastANI (Jain et al., 2018).

Protein extraction, peptide preparation, and 1D-LC-MS/MS

Protein extraction of all samples with technical quadruplicates was done using the filter-aided sample preparation (FASP) protocol described by Wisniewski et al. (Wisniewski et al., 2009) with modification from Hamann et al. (Hamann et al., 2016). Peptide concentrations were determined using the Pierce Micro BCA assay (Thermo Scientific Pierce, Rockford, IL, USA).

Samples were analyzed by one-dimensional LC-MS/MS using a block-randomized design (Oberg and Vitek 2009). To reduce carry over, two wash runs and one blank run were programed between samples. For each run, 1200 ng of extracted peptides were loaded onto a 5 mm, 300 μm ID C18 Acclaim® PepMap100 pre-column (Thermo Fisher Scientific) using an UltiMate^{TM} 3000 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific) and desalted on the pre-column. Peptides were transferred to a 50 cm x 75 μm analytical EASY-Spray column packed with PepMap RSLC C18, 2μm material (Thermo Fisher Scientific), and heated to 45°C. The analytical column was connected directly to the Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) via an Easy-Spray source. Peptides were separated on the analytical column at a flow rate of 225 nL/min using a 260 min gradient and mass spectra acquired in the Orbitrap as described by Petersen et al. (Petersen et al., 2016). Protein identification, quantification and statistics.

Peptide identification was performed by searching MS/MS spectra against the *Ca.* P. alkaliphilum genome amended with the cRAP protein sequence database (http://www.thegpm.org/crap/) of common laboratory contaminating proteins, using the SEQUEST HT node in Proteome Discoverer version 2.0.0.802 (Thermo Fisher Scientific), as described previously (Petersen et al., 2016). Results were combined into a multiconsensus report using the FidoCT node in Proteome Discoverer and the protein-level false discovery rate (FDR) was restricted to below 5% (FidoCT q-value <0.05) (Serang et al., 2010) for high confidence identifications (FidoCT q-value <0.01) and medium confidence identifications (FidoCT q-value 0.01-0.05). Based

on these filtering criteria a total of 2127 unique cyanobacterial proteins were identified across all samples. Protein quantification was performed using the normalized spectral abundance factors (NSAFs) method (Florens et al., 2006). The relative abundance of proteins was calculated based on number of peptide spectral matches (PSMs) per protein that were normalized for length of the protein and the total number of spectra. NSAFs values were loaded into Perseus software (version 1.5.8.5, https://link.springer.com/protocol/10.1007%2F978-1-4939-7493-1_7) and a cutoff number of proteins was reduced by only including proteins that had at least four NSAF values greater than 0 in all replicates of one condition. All NSAF values were transformed by taking log2. Missing values produced by log2(0) were replaced by sampling from a normal distribution assuming that the missing values were on the lower end of abundance. A student t-test was carried out on transformed NSAF to identify proteins with significant difference in expression level between samples using a permutation-based FDR calculation to account for multiple hypothesis testing, with parameters: groupings were not preserved for randomizations, both sides, 250 randomizations, FDR of 0.05 and s0 of 0.

Phylogenetic analysis

Phylogenetic tree of *Ca.* P. alkaliphilum genome was generated using amino acid sequences of DNA-directed RNA polymerase subunit A, B, and C2 (rpo A, B, C2). Sequences were aligned using Mafft (Katoh and Standley 2013) with –maxiterate 1000 –localpair options. Poorly aligned regions were identified with jalview and removed and remaining sequences were concatenated. Maximum-likelihood phylogenetic tree was constructed in MEGA7 (Kumar et al. 2016) using the LG+G model with 100 bootstraps. Three species from *Melainabacteria* were selected as the outgroup. The inset tree (Figure 3) was made as previously described (Sorokin et al., 2020).

Microscopy imaging

All microscopy was performed using the upright Axio Imager.A2 (Zeiss, Germany). Grayscale images were direct contrast interference microscopy at 100 X magnification and captured using the Axiocam 506 mono camera (Zeiss, Germany) with ZEN 2012 blue edition software (Zeiss, Germany). Coloured images were brightfield microscopy at 400 X magnification and captured using the DFK MKU226-10x22USB 3.0 Color Microscope Camera (The Imaging Source, Germany) with IC Capture 2.5 software (The Imaging Source, Germany).