A naturally occurring lytic bioprocess improves sustainability of phycocyanin production from cyanobacteria

Authors:

Jackie Zorz[1]*, Alexandre J. Paquette[1], Timber Gillis[1], Angela Kouris[1,2], Cigdem Demirkaya[3], Hector De La Hoz Siegler[3], Marc Strous[1], Agasteswar Vadlamani[1,2]

Affiliations:

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

[2]Synergia Biotech Inc., Calgary, Canada

[3]Department of Chemical and Petroleum Engineering, University of Calgary, Calgary, Canada

*Corresponding author. Email: jacqueline.zorz@ucalgary.ca

Abstract

Phycocyanin is a natural blue pigment produced by cyanobacteria and is a valuable compound for food and cosmetic industries. At present, phycocyanin is manufactured with expensive and resource-heavy biotechnology, impeding its widespread use as a blue dye substitute. Here we show that cells of an alkaliphilic cyanobacterium lyse spontaneously in dark incubations mimicking natural soda lake environments, releasing concentrated phycocyanin. Proteogenomics showed that lysis likely resulted from a programmed response triggered by a failure to maintain osmotic pressure in the wake of severe energy limitation. This response explains the high turnover rates of cyanobacterial cells observed in soda lakes. Cells of *Arthrospira platensis* (Spirulina), currently used worldwide for phycocyanin production, lyse and release their pigments in the same manner. We propose this natural form of programmed cell death could reduce the costs and resources needed to produce phycocyanin, enabling displacement of current artificial blue colourants associated with adverse health effects.

One-Sentence Summary:

Failure to maintain osmotic balance in the dark forces blue-green algae to share their bounty with the world.

Introduction

Phycocyanin is a valuable, naturally produced blue dye substitute for cosmetic and food industries, found within light-harvesting phycobiliproteins of cyanobacteria. It is currently commercially produced from the cyanobacterial genus *Arthrospira* (Spirulina), which has GRAS (generally recognized as safe) status and is grown in outdoor raceway ponds. These ponds contain bicarbonate and other nutrients and are sparged with concentrated carbon dioxide. After harvesting, the cells are dried, washed and mechanically lysed. The lysate is filtered and the phycocyanin is recovered in the filtrate which is generally spray-dried to yield the final product, a blue powder. The two drying steps, the need for mechanical disruption, occasional "crashing" of Spirulina populations in the ponds as well as water losses because of evaporation, compromise the sustainability of the overall process and contribute to the high cost. For example, per kg of phycocyanin, as much as 5,000 L of water can be consumed (1), five times the amount needed for production of coffee beans.

To improve sustainability and cost-effectiveness of phycocyanin production, growing cyanobacteria at much higher pH and alkalinity enabling direct capture of CO_2 from air (2), and using a cyanobacterial consortium to improve process robustness by avoiding population crashes (3), were previously proposed. Alkaliphilic cyanobacterial consortia could be sourced from alkaline soda lakes, environments with naturally high pH (>10) and alkalinity (>0.5 M), that in some cases feature productive and dense microbial mats dominated by cyanobacteria (4). Interestingly, the prolific growth of cyanobacteria in these mats does not appear to translate into a buildup of biomass or sediment. Instead, the presence of steep sulfide gradients might indicate that in these mats cyanobacteria are rapidly turned over (5, 6). The combination of robust, prolific growth with, potentially, rapid turnover could be the ideal natural starting point for a sustainable phycocyanin biotechnology.

Recently, we isolated a cyanobacterial consortium from alkaline soda lakes (7, 8), dominated by "*Candidatus* Phormidium alkaliphilum" (9). The consortium grows optimally at a pH above 11 permitting direct capture of CO_2 from air (10) and displayed robust, crash-free growth during a 130-day outdoor pilot plant trial (11). Here, we first demonstrate the turnover of cyanobacterial biomass in natural alkaline soda lake microbial mats and sediments using 16S rDNA amplicon sequencing of sliced push-cores. Next, we show that the rapid lysis of cells of the cyanobacterium *Ca.* P. alkaliphilum observed in mat and sediment cores can be replicated in the laboratory, by incubating the cells in the dark. During this lysis, high quality phycocyanin is released into the medium. Monitoring lysis and release using metagenomics and metaproteomics yields no evidence for ecological interactions such as predation by other bacteria or viruses as the cause of cell lysis. Instead, it appears that lysis results from programmed cell death provoked by energy starvation. Finally, we show that a one-week, static, dark incubation of Spirulina also

results in phycocyanin release without any need for mechanical disruption of cells. We propose that this newly discovered bioprocess could reduce costs and improve sustainability of phycocyanin production.

Results

To demonstrate rapid turnover of cyanobacterial biomass, we collected cores from the alkaline Lake Goodenough (Canada), 30 cm into the mats and underlying sediment. 16S rRNA gene amplicon sequencing of sectioned cores showed high abundance of cyanobacteria at the top of the mat (Fig. 1). The abundance of cyanobacteria like *Phormidium* and *Nodosilinea* decreased rapidly and became essentially negligible two cm below the sediment surface. Other bacterial taxa, better adapted to dark and anoxic conditions, became abundant (Fig. 1). Rapid turnover of cyanobacterial biomass explained the previously observed steep sulfide gradients (6). Sulfide likely builds up below the mats after the depletion of oxygen because sulfur reducing bacteria oxidize fatty acids, hydrogen, and other cyanobacterial degradation products and reduce sulfate and other sulfur-compounds to sulfide. Amplicon sequencing showed the ecological success at depth of thiosulfate and elemental sulfur reducing *Dethiobacter* (12, 13), and the sulfate, thiosulfate, and sulfite reducing *Desulfonatronovibrio* (14) (Fig. 1).

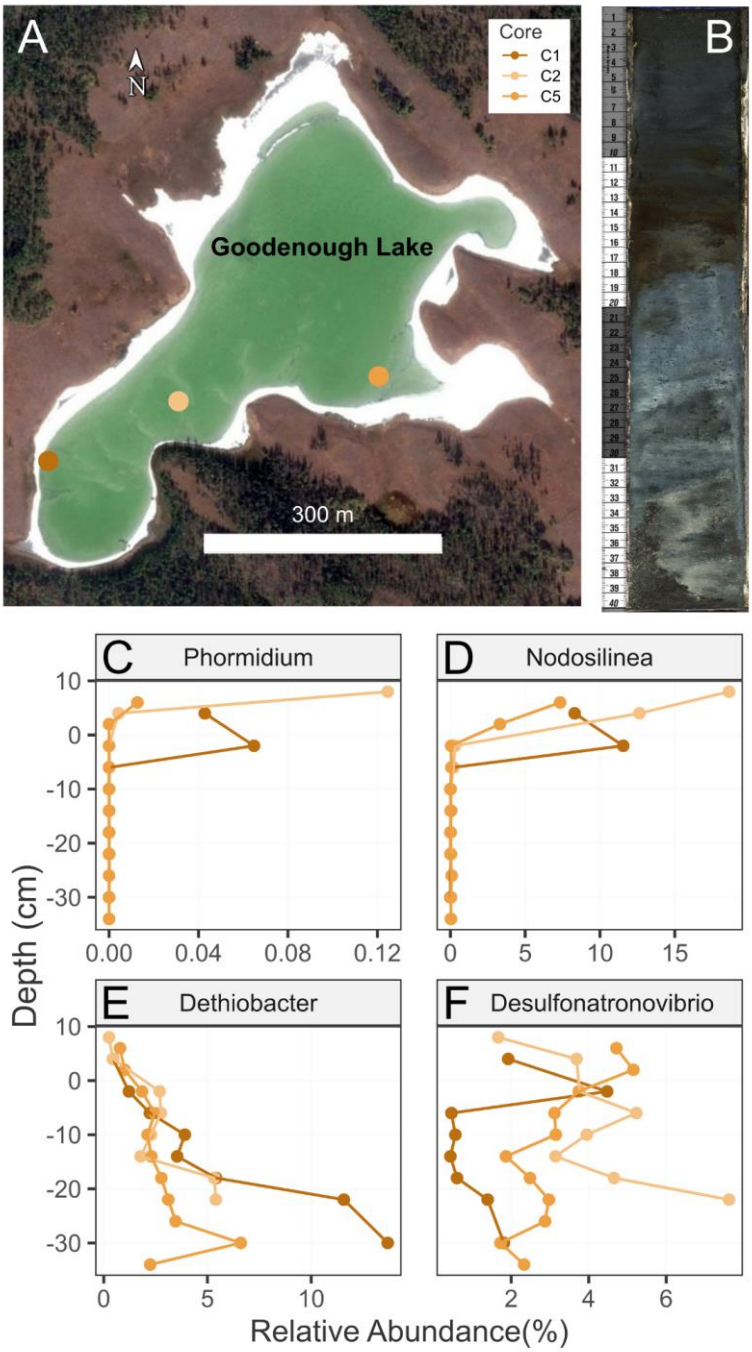

Fig. 1. Cores obtained from Lake Goodenough sediment show rapid displacement of cyanobacteria with sulfate and sulfur reducing bacteria. (**A**) Map of sampling locations within Goodenough Lake, British Columbia, Canada. (B) Cross section image of a representative core. Distribution of genera from 16S rRNA gene abundance: (C) Cyanobacteria *Phormidium*, (D) Cyanobacteria *Nodosilinea*, (E) Sulfidogenic thiosulfate and elemental sulfate reducing bacteria *Dethiobacter*, (F) Sulfidogenic, sulfate, sulfite and thiosulfate reducing bacteria *Desulfonatronovibrio*. Positive depth values in C-F represent the cyanobacterial mat and negative centimetres represent distance below the sediment surface.

To investigate if the rapid degradation observed *in situ* could be replicated in the laboratory, we used a cyanobacterial consortium previously isolated from Cariboo Plateau soda lake mats (7,8). This consortium mainly consisted of *Ca.* P. alkaliphilum (9), also shown in Fig. 1C. The consortium was subjected to a 12 day dark and anoxic incubation in soda lake media (0.5 M inorganic carbonates, initial pH > 10). Samples were taken every two days and were centrifuged, yielding solid and supernatant fractions. For solid fractions, we determined ash free dry weight and performed microscopy, metagenomics and metaproteomics. For supernatant fractions, we determined the pH and concentrations of organic acids and phycocyanin, and also performed microscopy, metagenomics and metaproteomics.

Initially, 72 % of the DNA extracted from the solid fraction originated from "*Ca.* P alkaliphilum*"*, but only 3.6 % of this DNA remained after 6-8 days in the dark. After 12 days, DNA from "*Ca.* P alkaliphilum" was barely detectable (0.15%) (Fig. 2B). A similar pattern was observed in the supernatant fraction, as initially 20% of DNA could be attributed to *Ca.* P. alkaliphilum, which decreased to 1.6% by day 12 (Fig. 2C). In contrast, cyanobacterial proteins persisted, always making up at least 65% of the protein composition in the solids fraction (Fig. 2E) and increasing to >80% of the supernatant fraction (Fig. 2F), suggesting a discrepancy in the way that the two biomolecules were degraded. Coinciding with the decrease of cyanobacterial DNA there was an increase in concentrations of fermentation products such as acetate and propionate (fig. S1). Because acetate mainly accumulated before cyanobacterial lysis, *Ca.* P. alkaliphilum itself was likely responsible for its production. *Ca.* P. alkaliphilum was previously shown to have the genetic capability for dark fermentation to acetate (9). Propionate increased later in the incubation, and was likely produced by other bacteria fermenting compounds within the cyanobacterial lysate. By day 6, the supernatant was coloured intensely blue (Fig. 3A) and contained a large amount of phycocyanin based on UV/Vis spectrometry (fig. S2). Proteomics showed that phycocyanin made up 22%-32% of the protein in supernatant samples (fig. S3). Microscopy showed that the cyanobacterial cells were lysing and breaking apart by the sixth day of incubation, explaining the presence of phycocyanin in the external medium (Fig. 3BC). Cell lysis of densely populated cyanobacterial blooms and the subsequent blue colour change caused by the release of phycocyanin is a phenomenon that has been observed previously in freshwater lakes, but the mechanism of those lysis events remains unknown (15, 16).

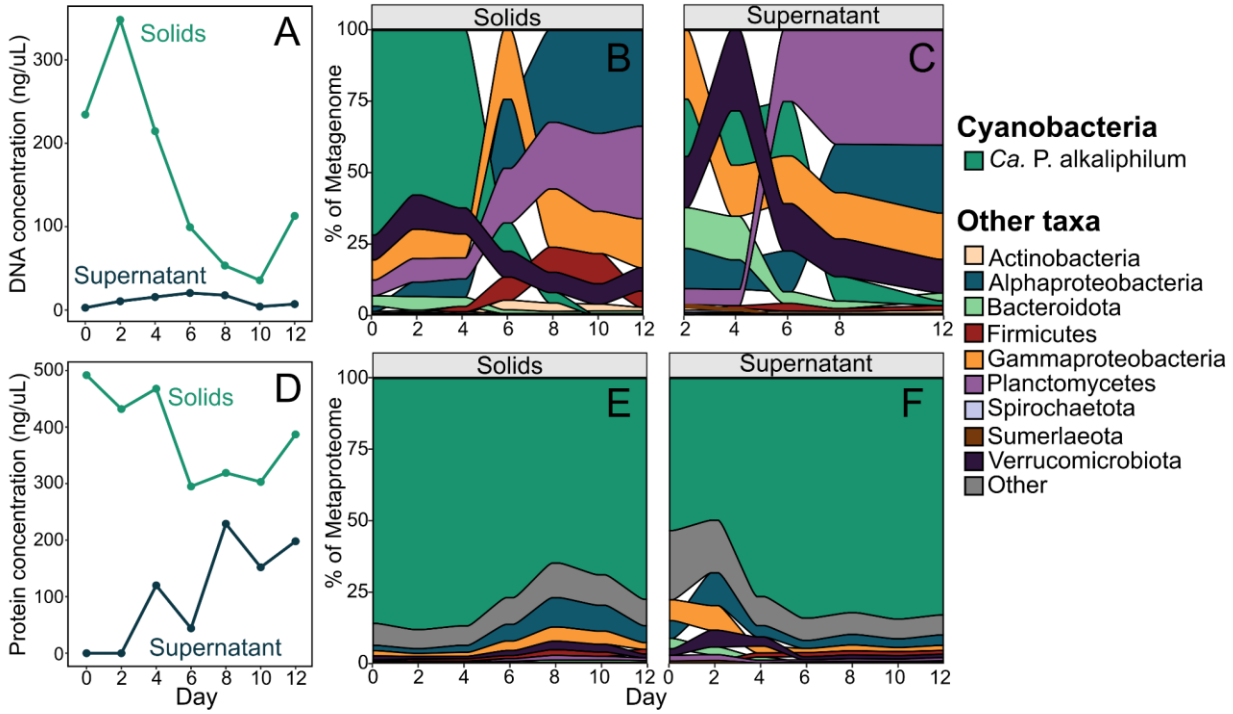

Fig. 2. DNA and protein dynamics during the 12-day dark and anoxic incubation. Concentration of DNA (**A**) and protein (D) in the solid and supernatant fractions. Microbial composition of consortium determined from the DNA in the solids (B), and supernatant (C) fractions, and from the protein content in the solids (E), and supernatant (F) fractions.

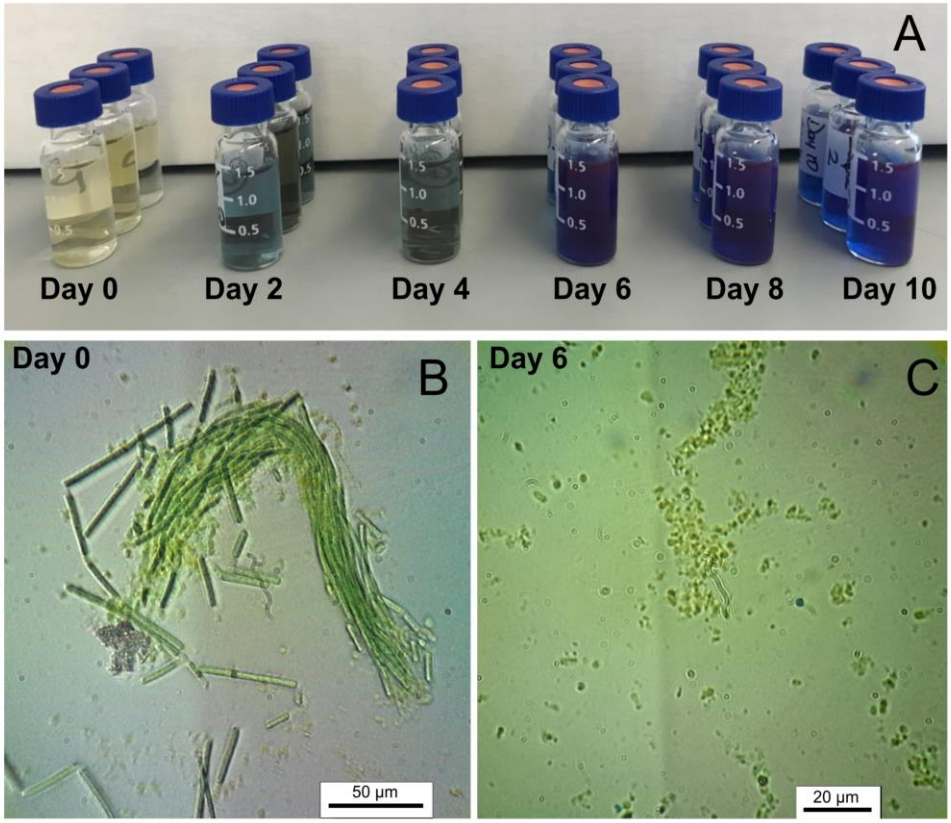

Fig. 3. Release of phycocyanin during a dark and anoxic incubation. Image showing the colour of the supernatant fraction of each sample taken during the incubation **(A)**. Microscopic images of the cyanobacterial consortium taken on day 0 (B), and day 6 (C) of the incubation.

Cell lysis could be attributed to a number of causes including predation, viral attack, or genetically programmed signals (i.e. programmed cell death) (17). Each of these possibilities was evaluated using the proteogenomics data obtained during the incubations. The possibility of predation by a larger eukaryotic cell or an antagonistic bacterial species was assessed using the metagenomics data. Phagocytosis and grazing are predatory strategies performed by eukaryotes. Eukaryotic ciliate grazers affiliated with *Schmidingerothrix* were present in the consortium, however their abundance, calculated through copies of the rDNA gene in the metagenomes, was low (<2%) and did not increase over the course of the dark, anoxic incubation (fig. S4), suggesting that eukaryotic predation was not directly responsible for the collapse of the cyanobacterial population.

Provisional genomes, or metagenome assembled genomes (MAGs), were acquired for the main cyanobacterial species, *Ca.* P. alkaliphilum, in addition to 59 other species from eight bacterial phyla including Proteobacteria, Bacteroidota, Firmicutes, Planctomycetota and Verrucomicrobiota (Fig. 2, table S1). These bacterial dynamics, and the diverse and enhanced expression of carbohydrate active transporters, like the TonB-dependent transporters (expressed by at least 24 different MAGs), are akin to

the succession of bacterial populations observed after phytoplankton blooms in other aquatic systems (18). However, although the relative abundance of heterotrophs increased during the incubation, the abundances of heterotrophs relative to each other changed only modestly (Fig. 2) and there was no single species that benefited in proportion to the magnitude of the lysis event. Total DNA concentrations decreased (Fig. 2A), released cyanobacterial proteins persisted and fermentation (producing acetate) declined after the lysis of the cyanobacteria (fig. S1). This might mean that the mainly aerobic consortium members did not have time to consume the released proteinaceous cyanobacterial lysate anaerobically. Therefore, antagonistic bacteria were likely not the cause of the cell lysis event.

Cyanophages, viruses that specifically target cyanobacteria, play an important role in the fate of cyanobacteria in natural environments, and consequently in global carbon and nutrient cycles (19,20). The cyanobacterial MAG contained seven CRISPR arrays, suggesting previous viral infections. However, no viral proteins were identified in the metaproteome, and no viral contigs matching cyanobacterial contigs were identified in the assembly (fig. S5). The most abundant contig of viral origin, identified by *Virsorter* (21), was inferred to be a prophage of a Planctomycetota MAG, and not associated with the cyanobacteria. It only reached 50% of the average cyanobacterial pre-collapse sequencing depth (fig. S5). In the event of a viral-mediated lysis, the depth of the associated viral contig(s) would be expected to increase at least 10-fold (burst size) in comparison to the depth of the cyanobacterial host contigs (22). The comparatively low abundance of viral associated contigs indicated that a mass viral-induced lysis of the cyanobacterial cells did not occur.

Finally, we explored if a genetically programmed signal, like programmed cell death, was the most likely cause of the cyanobacterial lysis. Programmed cell death might be initiated in response to external stressors, such as the depletion of an internal storage pool and the consequent failure to maintain osmotic pressure, caused by sustained darkness (23, 24).

Total cyanobacterial protein abundance remained relatively unchanged during the 12-day incubation, and included detection of just over 2,000 proteins, accounting for 52% of the predicted proteome of *Ca.* P. alkaliphilum (table S2). Of the expressed proteins, 459 increased by at least two-fold between the beginning and the end of the dark incubation, while 1,039 proteins decreased expression over 50%. In general, cyanobacteria do not drastically change their proteome composition in response to diel cycling (25, 26). Thus, a greater than twofold change in approximately 75% of the expressed proteins suggests that *Ca.* P. alkaliphilum had mounted a stress response that was outside the normal range of proteomic circadian cycles.

We observed the well-known signs of a shift from linear electron flow to cyclic electron flow in the cyanobacteria's proteome (27, 28, table S2). Firstly, the relative expression of photosystem I (PSI) increased nearly four-fold to 12.7% of the metaproteome by day 12, which was three times higher than the expression of photosystem II (PSII), a ratio that began at 1:1 on day 0. In parallel, large increases in

other proteins and complexes required for cyclic electron flow including ATP synthase, cytochrome b6f, and ferredoxin were observed (Fig. 4). Proteins not required for cyclic electron flow, like ferredoxin-NADP+ reductase and the oxygen evolving complex of PSII decreased ten and two-fold, respectively, during the same period (Fig. 4B). Previous studies have identified an increase in cyclic electron flow in response to the stress of dark and anoxic conditions and have hypothesized that it could be used as a defensive strategy protecting PSII in the moments before the CBB cycle is activated (29), or alternatively as a mechanism to jumpstart metabolism through the rapid generation of ATP once light energy returns (30).

There was a rapid disappearance of cyanobacterial DNA observed from day 4 onwards in the metagenomes. Cleavage of DNA to fragments between 150-300 base pairs in length is a common occurrence in programmed cell death in prokaryotes, including cyanobacteria (23, 31). A size selection step in the metagenome library preparation protocol excluded DNA fragments below 300 base pairs, and thus these DNA fragments would not get sequenced in the resulting metagenomes. The potential for programmed cell death was further supported by the presence of multiple (>10) toxin-antitoxin systems in the *Ca.* P. alkaliphilum whole genome sequence. A toxin-antitoxin system consists of a small stable toxin that is responsible for cell death or growth arrest, and an easily degraded antitoxin which blocks its activity (32-36). Peptides assigned to toxin (HicA) and antitoxin (ParD, AbiEii) proteins were identified in the metaproteome (Fig. 4, table S2). Antitoxin abundance declined markedly during the first six days of the incubation while a peak in the abundance of toxins accompanied the cell lysis event. There was also a specific decrease (4-45-fold reduction) in proteins related to translation and transcription, like ribosomes, RNA polymerases, elongation factors, and translation initiation factors (Fig. 4). Corresponding with these observations in the proteome, several toxin-antitoxin systems promote cell death through the global reduction of replication, transcription, and translation in cells (32). Mechanosensitive channels, known to be upregulated in response to changes in osmotic pressure (37), increased in expression five-fold by day 12.

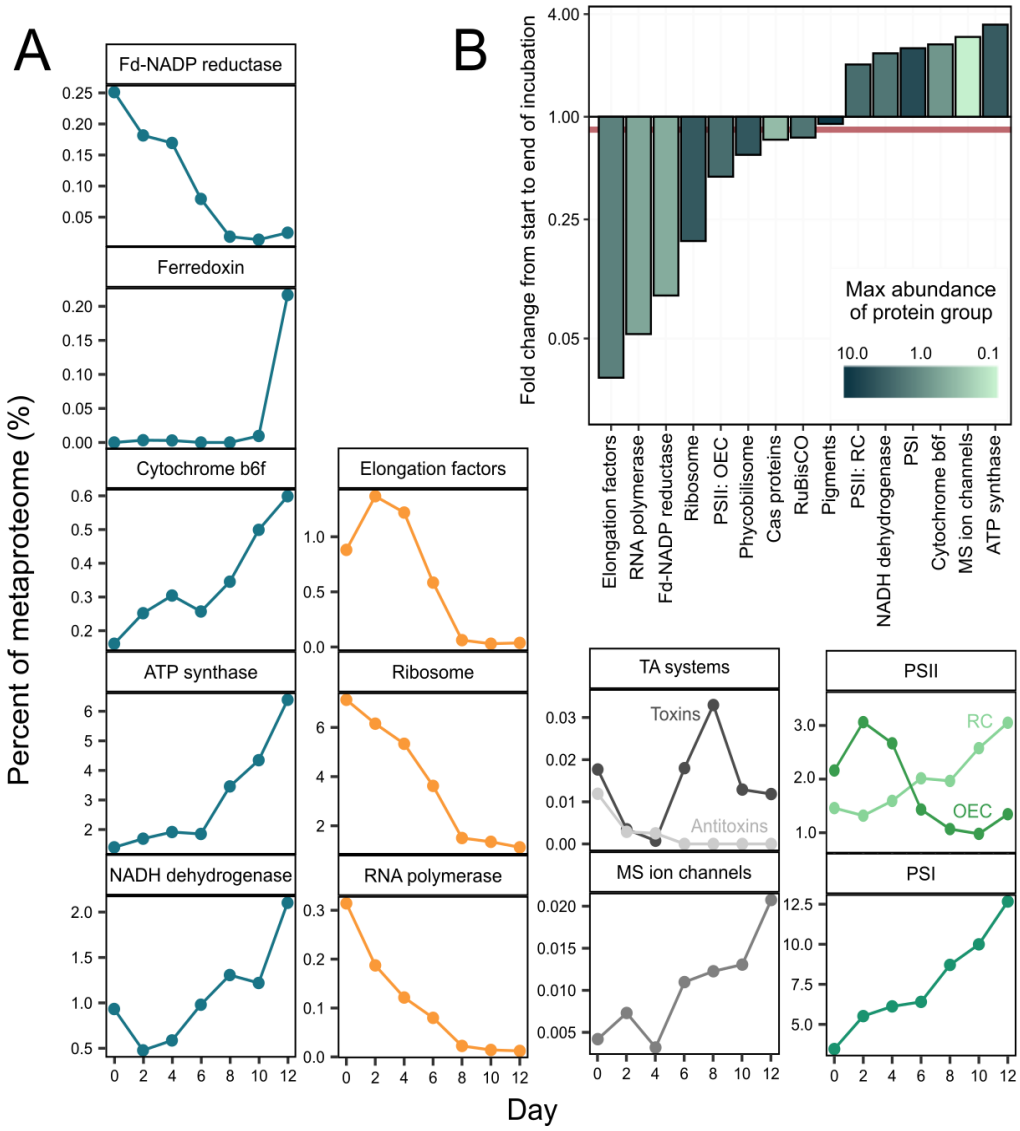

Fig. 4. Response of the *Ca.* P. alkaliphilum proteome to the dark and anoxic incubation. (**A**) Abundance of select functional protein categories in the solids fraction of the *Ca.* P. alkaliphilum proteome. Note the difference in y-axis scales. Colours of points broadly define the protein's functional category. Blue: electron transport chain proteins, orange: translation and transcription proteins, grey: stress proteins, green: photosystem proteins. (B) Fold change of the *Ca.* P. alkaliphilum functional protein groups from the average of days 10 and 12 to the average of days 0 and 2. The red line shows the fold change in all *Ca.* P. alkaliphilum proteins over the same period (0.84). The colour of the bar represents the maximum protein abundance in the incubation for that protein category. PSII: photosystem II, RC: reaction center, OEC: oxygen evolving complex, PSI: photosystem I, TA systems: toxin-antitoxin systems, Fd: ferredoxin, MS ion channels: mechanosensitive ion channels, Cas: CRISPR associated.

If energy depletion was the root cause of cyanobacterial cell lysis, increasing salinity prior to a dark and anoxic incubation should result in an earlier cellular lysis. The cells would need to spend more maintenance energy to cope with the higher osmotic stress and would consequently deplete their reserves sooner. This hypothesis was tested by performing separate dark and anoxic incubations of the cyanobacterial consortium at higher (1M Na^+) and lower (0.25M Na^+) salinity.

The dark and anoxic incubation at 0.25M Na^+, resulted in a cyanobacterial lysis event that occurred later and resulted in a much lower concentration of released phycocyanin (1.5 mg/mL) compared to the original incubation at 0.5M Na^+ (6.6 mg/mL) (Fig. 5). In the incubation with 1M Na^+, cell lysis occurred sooner, by day 5 (Fig. 5), and the final concentration of 7.2 mg/mL phycocyanin was higher than the original incubation. These results support the hypothesis that cyanobacterial cell lysis in these dark and anoxic incubations is due to the depleted energy reserves used to maintain osmotic equilibrium. Cells in an environment of higher salinity require more energy to maintain osmotic equilibrium, and thus deplete energy reserves faster. These results also supported the earlier conclusion that heterotrophs were slow to consume released cyanobacterial proteins; the faster the lysis, the more phycocyanin remained.

Lastly, we demonstrated that phycocyanin extraction by programmed cell death can be directly applied to commercial operations. A culture of Spirulina was incubated in dark and anoxic conditions with 1M Na^+. After 10 days of incubation, a similar phycocyanin release was observed (fig. S6).

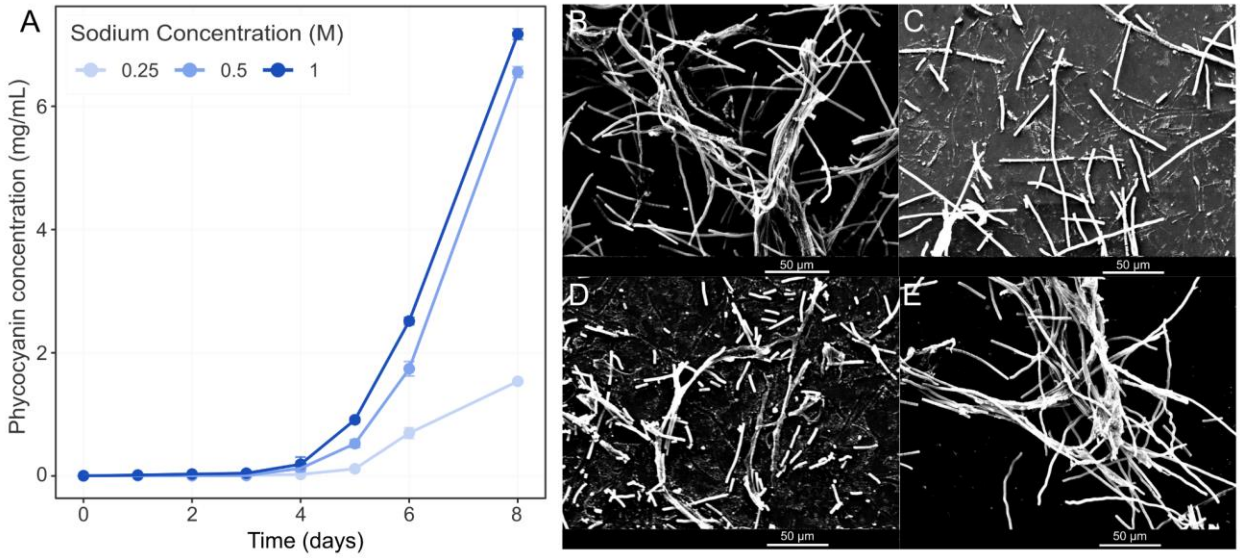

Fig. 5. Dark and anoxic incubations with varying sodium concentrations. (**A**) Phycocyanin concentration in the supernatant fraction of dark and anoxic incubations with varying sodium concentrations in the media. Electron micrographs of cyanobacterial cells on day 0 (B), and day 5 of dark

and anoxic incubations, with the original 0.5 M Na^+ media (C), and 1 M (D), and 0.25 M (E) Na^+ media from the additional experiments.

Discussion

The incubation exposed the cyanobacterial consortium to a myriad of stresses including energy starvation through prolonged darkness and anoxia. Proteogenomics showed that initially the cells altered their proteome to combat these stresses by the rearrangement of protein complexes in the thylakoid membrane to favour cyclic electron flow. Fermentation of available endogenous carbohydrates, like glycogen, cyanophycin, or the osmolytes sucrose, glucosyl glycerol, and trehalose, initially provided energy for sustaining cellular integrity and resulted in the observed increase in acetate (38, 39, fig. S1). After four days, proteins involved in transcription and translation were severely diminished, a first signal of decreased metabolism and arrested growth. After six days without relief from darkness, the supply of endogenous carbohydrates and osmolytes was likely depleted (5), and the ensuing starvation and osmotic stress may have triggered a programmed cell death response, possibly through toxin-antitoxin systems, that resulted in lysis of the cyanobacterial cells. The anoxic conditions of the incubation stunted the degradation of released phycocyanin pigment proteins by the predominantly aerobic heterotrophs making up the consortium.

A similar lysis phenomenon was previously observed in dark and anoxic incubations of a thermophilic cyanobacterium, *Oscillatoria terebriformis*, isolated from hot spring microbial mats (40). In that experiment, cell survival could be prolonged by the addition of an exogenous carbohydrate source (fructose), and/or a reductant (e.g. sodium thioglycolate). This might be another example where dwindling energy stores cause the lysis of a cyanobacterium under dark and challenging conditions. Addition of reducing agents could quench reactive oxygen species (ROS), and a link between ROS and programmed cell death in both eukaryotic and prokaryotic cells has long been known (41-43). Furthermore, the production of ROS has previously been associated with the activation of toxin-antitoxin systems upon stress (44, 45).

Cyanobacterial cellular lysis due to dark and anoxic incubation provided a way to access the internal pigment phycocyanin without costly and energetically intensive mechanical disruption. Evidence of this lytic bioprocess was found in the industrially cultivated species, Spirulina, as well as in cyanobacterial species from hot spring microbial mats (40), and the sediments of the original haloalkaline environment of *Ca.* P. alkaliphilum, suggesting that this phenomenon could be widespread among cyanobacteria in both engineered and natural systems. Thus, this intrinsically occurring bioprocess could

be harnessed as a novel, more cost-effective and sustainable way to produce the natural blue pigment phycocyanin.

Materials and Methods

Sediment sample collection and preparation

Duplicate sediment cores were collected in April 2019 from Goodenough Lake (51.330°N, 121.64°W). The sediment cores were taken from 3 different locations within each lake (Fig. 1) using a 1.5-m single-drive Griffith corer from LacCore: National Lacustrine Core Facility (University of Minnesota). The sediment cores ranged in length from 25-50 cm. To reduce the mixing of water and upper sediment layers in the cores, Zorbitrol was used as a gelling agent to stabilize the sediment-water interface during transport. Cores were then stored upright at -20 °C. For the analysis, positive centimetres represent the benthic cyanobacterial mat and negative centimetres represent distance below the sediment surface.

Cores were removed from the -20 °C freezer and defrosted at room temperature (22 °C) for 2 hours. Cores were then horizontally sliced into 2 cm disks using a Dremel Multi-Max MM50 oscillating saw (Dremel, USA) at the lowest speed, used to reduce blade contact with the sediment. The blade was sterilized with 70% ethanol before each core section was sampled. To avoid the potential risk of contamination from the core liner or during sectioning, sediment in contact with the core liner was removed and the inner core was transferred to a 50 mL tube, sealed, and stored at -20 °C. The sediment from each disk was subsampled for DNA extraction and stored at -80°C.

Experimental setup and sampling

An alkaliphilic cyanobacterial consortium enrichment culture containing a single, abundant, cyanobacteria species, *Candidatus* Phormidium alkaliphilum (7-9), was used as inoculum for the dark incubation. This consortium was originally sourced from alkaline soda lakes in the Cariboo Plateau region of Canada (4). The cyanobacterial consortium was grown in continuous light (200 µmol photons/(m^2·s)) in 10 L stirred glass vessels. The growth medium was previously described (10) and contained 0.5 M sodium (bi)carbonate alkalinity, at an initial pH of 10.3. After six days of photoautotrophic growth the culture was centrifuged for 30 minutes at 4,500 rpm to concentrate the biomass (Allegra X-22R, Beckman Coulter, USA). The wet biomass was then divided into 20 mL serum bottles sealed with butyl-rubber septa. Two grams of wet biomass were added to each serum bottle. The bottles were purged with N_2 gas to create an anoxic headspace, and then placed at room temperature (21°C) in the dark. At 0, 2, 4, 6, 8, 10, and 12 days after the start of the incubation, two sacrificial samples were taken. To each sample, 5 mL of

pH 7, phosphate-buffered saline solution was added and then the sample was centrifuged for 10 min at 4500 rpm, to separate biomass and supernatant.

For biomass pellets, the ash free dry weight of each sample was measured using NREL laboratory analytical procedures protocol (46). For the supernatant, the concentration of the organic acids succinate, formate, propionate, butyrate, and lactate were measured using an UltiMate 3000 HPLC system (ThermoFisher Scientific, USA) equipped with an Aminex HPX-87H column and a UV detector, as previously described (47). The phycocyanin and total protein concentration in the supernatants were measured as absorption at 620 nm and 280 nm respectively (48) using an Evolution 260 Bio UV-Visible Spectrophotometer (ThermoFisher Scientific, USA), with a standard curve prepared from laboratory-grade phycocyanin (Sigma-Aldrich, USA). Bright-field microscope images were taken using a Zeiss Axio Imager A2 Microscope (Carl Zeiss AG, Germany).

DNA extraction

DNA was extracted directly from soda lake sediment and incubation solid samples using the Fast DNA Extraction Kit for Soil (MP Biomedicals, USA). For incubation supernatant samples, 250 µL of sample was used for each extraction. The extraction protocol of the manufacturer was followed, but additional purification steps were performed with 5.5 M guanidine thiocyanate (7). For supernatant samples, the elution buffer was heated to 50°C prior to the elution step to increase yield. Still, supernatant samples from Day 0 and Day 10 did not yield enough DNA for metagenome analysis.

16S rRNA gene PCR and sequencing

Amplicon sequencing and library preparation of the DNA sediment samples was performed as previously described (7) using primer sets 926wF (5′-AAACTYAAAKGAATTGRCGG3′) and 1392R (5′-ACGGGCGGTGTGTRC3′) targeting bacteria (49, 50). Prepared libraries were sequenced on the MiSeq Personal Sequencer (Illumina, USA) using the 2 × 300 bp MiSeq Reagent Kit v3. Amplicon sequencing results were processed using MetaAmp Version 3.0 (51), and the Silva database version 132 (52). Paired-end reads were merged if they had less than eight mismatches in the overlap region and an overlap of >100 base pairs (bp) (7, 53). The merged reads were further filtered by removing reads that were missing the forward or reverse primer and had more than one mismatch in the primer region. All reads were trimmed to a final of 350 bp and clustered into operational taxonomic units (OTUs) of >97% sequence identity (7, 53).

Library preparation and metagenome sequencing

All biomass samples, and supernatant samples from days 2, 4, 6, 8 and 12 were prepared for metagenomic sequencing as previously described (4). Briefly, DNA was sheared to fragments of ~300 bp and libraries were created using the NEBNext Ultra DNA Library Prep Kit (New England Biolabs, Ipswich, MA). An Illumina NextSeq 500 sequencer (Illumina, San Diego, CA) was used for sequencing using a 300 cycle (2x150bp) high-output sequencing kit at the Center for Health Genomics and Informatics in the Cumming School of Medicine, University of Calgary, Canada.

Metagenome Assembly and Binning

Raw, paired-end Illumina reads were filtered for quality using BBDuk (https://jgi.doe.gov/data-and-tools/bbtools/). Quality control consisted of trimming reads to 150 bp, trimming off adapter sequences, filtering out contaminants, such as the PhiX adapter, and clipping off low quality ends, all as previously described (4). Paired-end reads from each sample were then merged with BBmerge (54). Separate assemblies of the reads from each sample were performed using metaSPAdes version 3.12.0 with default parameters (55). To increase binning success, one large co-assembly using the unmerged reads from all samples was conducted using MegaHit v1.2.2 (56). Only contigs greater than 500 bp in length were processed further. The MetaErg pipeline (57) was used for prediction and annotation of genetic elements on each assembled contig.

Binning of assembled reads into metagenome-assembled-genomes (MAGs) was completed using MetaBat2 version 2.12.1 (58). The binning step was performed on each sample's assembly separately as well as the co-assembly. To generate sequencing depth data for binning, quality-controlled reads of each sample were mapped to the assembly of each sample using BBMap v38.84 (https://sourceforge.net/projects/bbmap/). Mapping results were summarized using the script, "*jgi_summarize_bam_contig_depths*", part of the MetaBat package (59). After binning, the program dRep (60) in conjunction with CheckM v1.0.11 (61) was used to determine the best (highest estimated completeness, and lowest estimated contamination) MAGs associated with each population. In total, 60 MAGs (>80% completeness, and <5% contamination) were identified for further processing and analysis. The program, gtdbtk v0.3.2 was used for the taxonomic assignment of each MAG (62).

The relative abundance of individual MAGs in each metagenome was calculated by mapping quality controlled raw reads from each sample onto the contigs of each MAG as well as the dereplicated contigs that remained unbinned. Again, BBMap (minid = 0.98) was used for this. Unbinned contigs were dereplicated using cd-hit-2d (63). In this step, all contigs sharing > 90% sequence identity with a binned contig were eliminated. The number of reads that mapped to each contig was counted, and then the total counts for each contig of each MAG were summarized. To determine relative abundance, counts were

normalized to MAG genome size and the number of mapped reads per sample. For the unbinned contigs, reads were normalized to the number of base pairs in all dereplicated unbinned contigs.

The program phyloFlash v3.3 (*Emirge* assembly) was used to obtain full length 16S and 18S ribosomal RNA (rRNA) gene sequences and their sequencing depth from the metagenomes (64). The sequencing depths of rRNA sequences were used primarily to determine the population dynamics of species that did not assemble or form MAGs well, mainly eukaryotic protists.

Analysis of viral contigs

Contigs potentially associated with viruses were identified from the metagenome co-assembly using VirSorter v1.0.6 (21). BLASTn was then used to match the DNA sequence of viral contigs to the 60 MAGs. CRISPR (clustered regularly interspaced short palindromic repeats) arrays in the cyanobacterial MAG were identified from the MetaErg output (57), with the program MinCED (github.com/ctSkennerton/minced).

Protein Extraction and LC-MS/MS mass spectrometry

Protein was extracted from biomass and supernatant samples as previously described (65), using the filter aided sample preparation (FASP) protocol (66). To lyse cells, samples were added to lysing matrix E bead tubes (MP Biomedicals, USA) with SDT-lysis buffer (0.1 M DTT) in a 1:10 sample to buffer ratio. The tubes were then subjected to bead-beating in an OMNI Bead Ruptor (Omni International, USA) 24 for 45 s at 6 m s^{-1}. For supernatant samples, 500 µL of supernatant was used for lysis. Supernatant samples from days 0 and 2 had low yields and so lysate was concentrated prior to protein extraction.

Peptides were separated by an UltiMateTM 3000 RSLC nano Liquid Chromatograph (Thermo Fisher Scientific, USA), using a 75cm x 75µm analytical column and analyzed in a QExactive Plus hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, USA) as previously described (67). A total of 2,000 ng of peptide was loaded, and each sample was run for 4 hours.

Metaproteomics data analysis

The database used for protein identification was manually created using the predicted and annotated proteins from the binned and unbinned metagenomic sequences. Cd-hit was used to remove redundant sequences from the database using an identity threshold of 95% (63), giving preference to sequences that came from metagenome assembled genomes (MAGs). Sequences of common contaminating proteins were added to the final database (http://www.thegpm.org/crap/). The final database contained 454,164 proteins. For protein identification MS/MS spectra were searched against the database using the Sequest HT node in Proteome Discoverer version 2.2.0.388 (Thermo Fisher Scientific, USA) as described

previously (68). Only proteins with one unique peptide, and with a protein false discovery rate (FDR) confidence of at least a level of "medium", were kept for further analysis.

Relative protein abundances were estimated using the normalized spectral abundance factor (NSAF) (69). MAG abundance in the metaproteome was estimated by adding the NSAF abundance of all proteins belonging to that MAG. In total, 3,286,730 MS/MS spectra were obtained, yielding 632,137 peptide spectral matches (PSMs), which corresponded to 10,408 expressed proteins after quality control.

Sodium and Spirulina experiments

The dark and anoxic incubation of the cyanobacterial consortium was repeated using dewatered biomass with different concentrations of sodium. Initially, biomass obtained after growth was first dewatered and then gently washed with deionized water to remove the salts. This step was repeated five times to ensure that all the salts were removed. The washed biomass was then separated into three aliquots. Each aliquot was washed with sodium carbonate solution with varied concentrations (0.25M, 0.5M and 1M). Then, approximately 2 grams of wet paste from each aliquot was placed in sterile serum bottles. The headspace in the serum bottles was vacuumed and filled with nitrogen gas up to atmospheric pressure to create anoxic conditions. These serum bottles were then incubated in dark at room temperature for 8 days. Every day two bottles were removed from the incubation and analysed for phycocyanin as described above. Electron microscopy was conducted as described previously (70), without performing ethanol washes and without using gold-sputtered filters.

The dark and anoxic incubation was repeated using a culture of Spirulina (*Arthrospira platensis*). A bulk culture containing 250 grams of wet paste (solid concentration 20% w/w) was incubated in dark and anoxic conditions with 1 M sodium carbonate for 12 days. The culture was visually monitored for signs of lysis and phycocyanin release (fig. S6).

Acknowledgments

We thank the University of Calgary's Center for Health Genomics and Informatics for sequencing and informatics services. We thank Carmen Li for help with MiSeq sequencing. We would like to thank Jayne Rattray and Martin Pabst for help with measurements and identification of organic acids. We thank Varada Khot for help with the analysis of viruses in the metagenomes.

Funding

This study was supported by the Natural Sciences and Engineering Research Council (NSERC), Canada Foundation for Innovation (CFI), Canada First Research Excellence Fund (CFREF), Genome Canada,

Western Economic Diversification, the International Microbiome Center (Calgary), Alberta Innovates, the Government of Alberta, and the University of Calgary.

Author Contributions

J.Z. extracted DNA and protein, and analyzed metagenomics and metaproteomics data. A.J.P. collected soda lake sediment samples, prepared them for 16S rRNA gene amplicon sequencing, and analyzed data. A.V., T.G., C.D., and A.J.P. performed the incubation experiments. A.K. maintained and ran the mass spectrometer required for metaproteomics. J.Z., A.V., and M.S. drafted the manuscript with input from all authors. M.S. and H.D. secured funding for this research.

Data and Materials Availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Metaproteomes were deposited to the PRIDE database under accession PXD023504. Metagenomes can be found in BioSamples SAMN17264972-SAMN17264985 (BioProject: PRJNA377096). MAGs from the study were deposited into the NCBI Whole Genome Shotgun submission database under BioSamples SAMN17266165-SAMN17266224.

Conflict of interest

JZ, AJP, AK, AV, CD, HD, and MS report a relationship with Synergia Biotech Inc. that includes equity or stocks. AV, JZ, CD, HD, and MS have patent #WO2021102563 - Alkaliphilic consortium shifting for production of phycocyanins and biochemicals pending to Synergia Biotech Inc. AJP, AV, and MS have a provisional patent - Method for cost and energy effective production of cyanobacterial consortium pending to Synergia Biotech Inc.

References and Notes

1. R. Davis, B. Klein, Algal biomass production via open pond algae farm cultivation: 2020 state of technology and future research. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5100-79931. (2021) https://www.nrel.gov/docs/fy21osti/79931.pdf

2. Z. Chi, Y. Xie, F. Elloy, Y. Zheng, Y. Hu, S. Chen. Bicarbonate-based integrated carbon capture and algae production system with alkalihalophilic cyanobacterium. *Bioresour. Technol*. **133**, 513-521 (2013).

3. S.R. Subashchandrabose, B. Ramakrishnan, M. Megharaj, K. Venkateswarlu, R. Naidu. Consortia of cyanobacteria/microalgae and bacteria: Biotechnological potential. *Bioresour. Technol.* **29**, 896-907 (2011).

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

5. L.J. Stal. Physiological ecology of cyanobacteria in microbial mats and other communities. *New Phytol.* **131**, 1-32 (1995).

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

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

8. M. Ataeian, Y. Liu, A. Kouris, A.K. Hawley, M. Strous. Ecological interactions of cyanobacteria and heterotrophs enhances the robustness of cyanobacterial consortium for carbon sequestration. *Front. Microbiol.* **13**, 780346 (2022)

9. M. Ataeian, A. Vadlamani, M. Haines, D. Mosier, X. Dong, M. Kleiner, M. Strous, A.K. Hawley. Proteome and strain analysis of cyanobacterium *Candidatus* “Phormidium alkaliphilum” reveals traits for success in biotechnology. *iScience*. **24**, 103405 (2021).

10. M. Ataeian, Y. Liu, K.A. Canon-Rubio, M. Nightingale, M. Strous, A. Vadlamani. Direct capture and conversion of CO2 from air by growing a cyanobacterial consortium at pH up to 11.2. *Biotechnol. Bioeng*. **116**, 1604-1611 (2019).

11. M. Haines, A. Vadlamani, W.D.L. Richardson, M. Strous. Pilot-scale outdoor trial of a cyanobacterial consortium at pH 11 in a photobioreactor at high latitude. *Bioresour. Technol.* **354**, 127173 (2022).

12. D.Y. Sorokin, T.P. Tourova, M. Mussman, G. Muyzer. Dethiobacter alkaliphilus gen. nov. sp. nov., and Desulfurivibrio alkaliphilus gen. nov. sp. nov.: two novel representatives of reductive sulfur cycle from soda lakes. *Extremophiles*. **12**, 431-439 (2008).

13. E.D. Melton, D.Y. Sorokin, L. Overmars, A.L. Lapidus, M. Pillay, N. Ivanova, T.G. Del Rio, N.C. Kyrpides, T. Woyke, G. Muyzer. Draft genome sequence of *Dethiobacter alkaliphilus* strain AHT1^{T}, a gram-positive sufidogenic polyextremophile. *Stand. Genomic Sci.* **12**, 57 (2017).

14. T.N. Zhilina, G.A. Zavarzin, F.A. Rainey, E.N. Pikuta, G.A. Osipov, N.A. Kostrikina. *Desulfonatronovibrio hydrogenovorans* gen. nov., sp. nov., an alkaliphilic, sulfate-reducing bacterium. *Int. J. Syst. Evol. Microbiol.* **47**, 144-149 (1997).

15. R.D. Fallon, T.D. Brock. Lytic organisms and photooxidative effects: influence on blue-green algae (cyanobacteria) in Lake Mendota, Wisconsin. *Appl. Environ. Microbiol.* **38**, 499-505 (1979).

16. S. Arii, K. Tsuji, K. Tomita, M. Hasegawa, B. Bober, K. Harada. Cyanobacterial blue color formation during lysis under natural conditions. *Appl. Environ. Microbiol.* **81**, 2667-2675 (2015).

17. D.C. Sigee, A. Selwyn, P. Gallois, A.P. Dean. Patterns of cell death in freshwater colonial cyanobacteria during the late summer bloom. *Phycologia.* **46**, 284-292 (2007).

18. H. Teeling, B.M. Fuchs, D. Becher, C. Klockow, A. Gardebrecht, C.M. Bennke, M. Kassabgy, S. Huang, A.J. Mann, J. Waldmann, M. Weber, A. Klindworth, A. Otto, J. Lange, J. Bernhardt, C. Reinsch, M. Hecker, J. Peplies, F.D. Bockelmann, U. Callies, G. Gerdts, A. Wichels, K. H. Wiltshire, F.O. Glockner, T. Schweder, R. Amann. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science*. **336**, 608-611 (2012).

19. S. Roux, J.R. Brum, B.E. Dutilh, S. Sunagawa, M.B. Duhaime, A. Loy, B.T. Poulos, N. Solonenko, E. Lara, J. Poulain, S. Peasant, S. Kandels-Lewis, C. Dimier, M. Picheral, S. Searson, C. Cruaud, A. Alberti, C.M. Duarte, J.M. Gasol, D. Vaque, Tara Oceans Coordinators, P. Bork, S.G. Acinas, P. Wincker, M.B. Sullivan. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*. **537**, 689-693 (2016).

20. R.J. Puxty, D.J. Evans, A.D. Millard, D.J. Scanlan. Energy limitation of cyanophage development: implications for marine carbon cycling. *ISME J.* **12**, 1273-1286 (2018).

21. S. Roux, F. Enault, B.L. Hurwitz, M.B. Sullivan. VirSorter: mining viral signal from microbial genomic data. *PeerJ.* **28**, e985 (2015).

22. M.G. Weinbauer, C.A. Suttle. Potential significance of lysogeny to bacteriophage production and bacterial mortality in coastal waters of the gulf of Mexico. *Appl. Environ. Microbiol.* **62**, 4374-4780 (1996).

23. S.-B. Ning, H.-L. Guo, L. Wang, Y.-C. Song. Salt stress induces programmed cell death in prokaryotic organism *Anabaena*. *J. Appl. Microbiol.* **93**, 15-28 (2002).

24. S. Bhattacharjee, A.K. Mishra. The tale of caspase homologues and their evolutionary outlook: deciphering programmed cell death in cyanobacteria. *Journal of Experimental Botany.* **71**, 4639–4657 (2020).

25. Welkie D. X. Zhang, M.L. Markillie, R. Taylor, G. Orr, J. Jacobs, K. Bhide, J. Thimmapuram, M. Gritsenko, H. Mitchell, R.D. Smith, L.A. Sherman. Transcriptomic and proteomic dynamics in the metabolism of a diazotrophic cyanobacterium, *Cyanothece* sp. PCC 7822 during a diurnal light-dark cycle. *BMC Genom.* **15**, 1185 (2014).

26. D.G. Welkie, B.E. Rubin, S. Diamond, R.D. Hood, D.F. Savage, S.S. Golden. A hard day's night: Cyanobacteria in diel cycles. *Trends Microbiol.* **27**, 231-242 (2019).

27. T. Hibino, B.H. Lee, A.K. Rai, H. Ishikawa, H. Kojima, M. Tawada, H. Shimoyama, T. Takabe. Salt enhances photosystem I content and cyclic electron flow via NAD(P)H dehydrogenase in the halotolerant cyanobacterium *Aphanothece halophytica*. *Aust. J. Plant Physiol.* **23**, 321-330 (1996).

28. J. Zorz, J.R. Allanach, C.D. Murphy, M.S. Roodvoets, D.A. Campbell, A.M. Cockshutt. The RUBISCO to photosystem II ratio limits the maximum photosynthetic rate in picocyanobacteria. *Life*. **5**, 403-417 (2015).

29. A.J. Golding, G. Finazzi, G.N. Johnson. Reduction of the thylakoid electron transport chain by stromal reductants - evidence for activation of cyclic electron transport upon dark adaptation or under drought. *Planta*. **220**, 356-363 (2004).

30. P. Joliot, A. Joliot. Cyclic electron transfer in plant leaf. *PNAS*. **99**, 10209-10214 (2002).

31. I. Berman-Frank, K.D. Bidle, L. Haramaty, P.G. Falkowski. The demise of the marine cyanobacterium, *Trichodesmium* spp., via an autocatalyzed cell death pathway. *Limnol. Oceanogr*. **49**, 997-1005 (2004).

32. A. Harms, D.E. Brodersen, N. Mitarai, K. Gerdes. Toxins, targets, and triggers: An overview of toxin-antitoxin biology. *Molecular Cell.* **70**, 768-784 (2018).

33. L. Gelens, L. Hill, A. Vandervelde, J. Danckaert, R. Loris. A general model for toxin-antitoxin module dynamics can explain persister cell formation in *E. coli*. *PLOS Computational Biology.* **9**, e1003190.

34. E.W. Sevin, F. Barloy-Hubler. RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. *Genome Biology*. **8**, R155 (2007).

35. Y. Wen, E. Behiels, B. Devreese. Toxin-antitoxin systems: their role in persistence, biofilm formation and pathogenecity. *Pathogens and disease.* **70**, 240-249 (2014).

36. S.H. Peeters, M.I. de Jonge. For the greater good: Programmed cell death in bacterial communities. *Microbiological Research.* **207**, 161-169 (2018).

37. U. Schumann, M.D. Edwards, T. Rasmussen, W. Bartlett, P. van West, I.R. Booth. YbdG in *Escherichia coli* is a threshold-setting mechanosensitive channel with MscM activity. *PNAS*. **107**, 12664-12669 (2010).

38. H. Heyer, L. Stal, W.E. Krumbein. Simultaneous heterolactic and acetate fermentation in the marine cyanobacterium *Oscillatoria limosa* incubated anaerobically in the dark. *Arch. Microbiol.* **151**, 558-564 (1989).

39. J.Z. Lee, L.C. Burow, D. Woebken, C. Everoad, M.D. Kubo, A.M. Spormann, P.K. Weber, J. Pett-Ridge, B.M. Bebout, T.M. Hoehler. Fermentation couples *Chloroflexi* and sulfate-reducing bacteria to *Cyanobacteria* in hypersaline microbial mats. *Front. Microbiol.* **5**, 61 (2014).

40. L.L. Richardson, R.W. Castenholz. Enhanced survival of the cyanobacterium *Oscillatoria terebriformis* in darkness under anaerobic conditions. *Appl. Environ. Microbiol*. **53**, 2151-2158 (1987).

41. M.D. Jacobson, Reactive oxygen species and programmed cell death. *Trends Biochem. Sci*. **21**, 83-86 (1996).

42. A. Hochman. Programmed cell death in prokaryotes. *Crit. Rev. Microbiol*. **23**, 207-214 (1997).

43. Y. Hong, J. Zeng, X. Wang, X. Zhao. Post-stress bacterial cell death mediated by reactive oxygen species. *PNAS.* **116**, 10064-10071 (2019).

44. X. Zhao, K. Drlica. Reactive oxygen species and the bacterial response to lethal stress. *Curr Opin Microbiol*. **21**, 1-6 (2014).

45. H. Van Acker, T. Coenye. The role of reactive oxygen species in antibiotic-mediated killing of bacteria. *Trends Microbiol*. **25**, 456-466 (2017).

46. S. Van Wychen, L.M. Laurens. Determination of total carbohydrates in algal biomass: Laboratory analytical procedure (LAP). *National Renewable Energy Lab*. (2016).

47. B. Novotnik, J. Zorz, S. Bryant, M. Strous. The effect of dissimilatory manganese reduction on lactate fermentation and microbial community assembly. *Front. Microbiol.* **10**, 1007 (2019).

48. G. Patil, S. Chethana, A.S. Sridevi, K.S.M.S. Raghavarao. Method to obtain C-phycocyanin of high purity. *J. Chromatogr. A.* **1127**, 76-81 (2006).

49. C. Rinke, J. Lee, N. Nath, D. Goudeau, B. Thompson, N. Poulton, E. Dmitrieff, R. Malmstrom, R. Stepanauskas, T. Woyke. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* **9**, 1038-1048 (2014).

50. T. Stoeck, D. Bass, M. Nebel, R. Christen, M.D.M. Jones, H-W. Breiner, T.A. Richards. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol*. **19**, 21-31 (2010).

51. X. Dong, M. Kleiner, C.E. Sharp, E. Thorson, C. Li, D. Liu, M. Strous. Fast and simple analysis of MiSeq amplicon sequencing data with MetaAmp. *Front. Microbiol.* **8**, 1461 (2017).

52. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F.O. Glöckner. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Opens external link in new window. *Nucl. Acids Res*. **41,** D590-D596 (2013).

53. A.J. Paquette, C.E. Sharp, P.J. Schnurr, D.G. Allen, S.M. Short, G.S. Espie. Dynamic changes in community composition of *Scenedesmus*-seeded artificial, engineered microalgal biofilms. *Algal Res.* **46**, 101805 (2020).

54. B. Bushnell, J. Rood, E. Singer. BBMerge - Accurate paired shotgun read merging via overlap. *PloS one.* **12**, e0185056 (2017).

55. S. Nurk, D. Meleshko, A. Korobeynikov, P.A. Pevzner. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824-834 (2017).

56. D. Li, C-M. Liu, R. Luo, K. Sadakane, T-W. Lam. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics*. **31**, 1674-1676. (2015).

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

58. D.D. Kang, F. Li, E. Kirton, A. Thomas, R. Egan, H. An, Z. Wang. MetaBAT2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, **7,** e7359 (2019).

59. D. Kang, J. Froula, R. Egan, Z. Wang. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*. **3**, e1165 (2015).

60. M.R. Olm, C.T. Brown, B. Brooks, J.F. Banfield. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME J.* **11**, 2864-2868 (2017).

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

62. P-A. Chaumeil, A.J. Mussig, P. Hugenholtz, D.H. Parks. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. **36**, 1925-1927 (2020).

63. W. Li, A. Godzik. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. **22**, 1658-1659 (2006).

64. H. R. Gruber-Vodicka, B.K.B. Seah, E. Pruesse. phyloFlash: Rapid small-subunit rRNA profiling and targeted assembly from metagenomes. *mSystems*, **5**, e00920-20 (2020).

65. M. Kleiner, E. Thorson, C.E. Sharp, X. Dong, D. Liu, C. Li, M. Strous. Assessing species biomass contributions in microbial communities via metaproteomics. *Nat. Comm.* **8**, 1558 (2017).

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

67. T. Hinzke, A. Kouris, R-A. Hughes, M. Strous, M. Kleiner. More is not always better: Evaluation of 1D and 2D-LC-MS/MS Methods for metaproteomics. *Front. Microbiol.* **10**, 238 (2019).

68. J.M. Petersen, A. Kemper, H. Gruber-Vodicka, U. Cardini, M. van der Geest, M. Kleiner, S. Bulgheresi, M. Mubmann, C. Herbold, B.K.B. Seah, C.P. Antony, D. Liu, A. Belitz, M. Weber. Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nat. Microbiol.* **2**, 16195 (2016).

69. B. Zybailov, A.L. Mosley, M.E. Sardiu, M.K. Coleman, L. Florens, M.P. Washburn. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **5**, 2339-47 (2006).

70. R.J. Barnes, S.P. Voegtlin, C.R.J. Hubert, S.R. Larter, S.L. Bryant. The critical role of environmental synergies in the creation of bionanohybrid microbes. *Appl. Environ. Microbiol.* **88**, e02321-21 (2022).