Reproducible spatial structure formation and stable community composition in the cyanosphere predicts metabolic interactions.

Sarah J.N. Duxbury[1,*,†], Sebastien Raguideau[2,†], Jerko Rosko[1], Kelsey Cremin[1], Mary Coates[1], Christopher Quince[2,3,4], and Orkun S. Soyer[1,*]

Affiliations: [1]School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK. [2]Organisms and Ecosystems, Earlham Institute, Norwich, NR4 7UZ, UK, Gut Microbes and Health, [3]Quadram Institute, Norwich, NR4 7UQ, UK, [4]Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK

† These authors contributed equally to this work.

***Correspondings Authors:** Sarah Duxbury (Sarah.Duxbury@warwick.ac.uk) and Orkun S Soyer (o.soyer@warwick.ac.uk), University of Warwick, Coventry, CV4 7AL, UK, + 44 (0)24 7657 4251.

Keywords: Cyanobacteria, granules, photogranules, cyanobacterial aggregates, microbial communities, freshwater ecosystems, spatial organisation, gliding motility, microbial mats.

Authors contributions: SD and OSS designed the study. SD, KC, JR, and MC conducted experiments and contributed analytical methods. SD, SR, CQ, and OSS analysed data. All authors contributed to manuscript writing.

Funding key: This project is funded by the Gordon and Betty Moore Foundation (grant https://doi.org/10.37807/GBMF9200). We acknowledge additional funding by the UK Natural Environment Research Council (NERC) Environmental Omics Facility (NEOF) Pilot Genomics Competition grant (grant number 1394). CQ was funded through the MRC Methodology Grant 'Strain resolved metagenomics for medical microbiology' MR/S037195/1.

Acknowledgements: We acknowledge the help of past group members Xue Jiang, Jing Chen, and Kieran Randall with early culture maintenance and protocol development, and cooperation of Seven Trent and Draycote Water Sailing Club with sample collection. We acknowledge the help of NEOF facility members including Lucy Knowles, Rachel Tucker, Deborah Dawson and Christopher Owen, for high molecular weight DNA extractions and long-read sequencing project design.

ABSTRACT

Cyanobacterial granules and aggregates can readily form in aquatic environments. The microbial communities found within and around these structures can be referred to as the cyanosphere, and can enable collective metabolic activities relevant to biogeochemical cycles. Cyanosphere communities are suggested to have different composition to that in the surrounding environment, but studies to date are mostly based on single time point samples. Here, we retrieved samples containing cyanobacterial granules from a freshwater reservoir and maintained a culture through sub-culture passages under laboratory conditions for over a year. We show that cyanobacteria-dominated granules form readily and repeatedly in this system over passages, and that this structure formation process seems to be associated with

cyanobacterial motility. Performing longitudinal short-read sequencing over several culture passages, we identified a cyanosphere community comprising of 17 species with maintained population structure. Using long-read sequencing from two different time point samples, we have re-constructed full, circular genomes for 15 of these species and annotated metabolic functions within. This predicts several metabolic interactions among community members, including sulfur cycling and carbon and vitamin exchange. Using three individual species isolated from this cyanosphere, we provide experimental support for growth on carbon sources predicted to be secreted by the cyanobacterium in the system. These findings reinforce the view that the cyanosphere can recruit and maintain a specific microbial community with specific functionalities embedded in a spatially-organised microenvironment. The presented community will act as a key model system for further understanding the formation of the structured cyanosphere, its function and stability, and its metabolic contribution to biogeochemical cycles.

INTRODUCTION

Microbial communities in Nature are often spatially organised. Microbes are commonly found in 2- or 3-dimensional structures such as biofilms, mats, aggregates, and granules, that form in diverse environments including soil, freshwater and oceanic environments, sediments, and human and animal guts (1-4). Within these environments, microbial communities play important roles in biogeochemical cycling (5) and in human health and disease (3). Specific microbial interactions are predicted to underpin long-term community function (2), however the dynamics of these processes are under-studied in spatial communities. In aqueous environments, an environment wherein diverse microbial structures can form, two major routes leading to spatially- structured microbial communities are colonisation of nutritious organic particles (for example marine snow) (6), and formation of microbial aggregates and granules (1, 2). The latter process remains relatively difficult to study due to lack of model systems and has been explored primarily through analysis of natural and bioreactor samples – usually in single-time points studies. These analyses have shown aggregates and granules to usually contain photosynthetic cyanobacteria and to comprise of a resident community of diverse microbes (7-14). Where analysed, these communities are found to be distinct from the microbial composition of the surrounding environment (10, 14). These observations align with the suggestion that there exists a cyanosphere, where specific microbes associate closely with cyanobacteria as they form granule and aggregate structures (14-18).

One of the better studied natural cyanobacterial aggregates are those formed by the filamentous *Trichodesmium* species, found commonly in the ocean and coastal waters. For this species, it has been suggested that the process of aggregate formation might contribute to nitrogen fixation (19, 20), and to iron acquisition (21). It is shown that *Trichodesmium* aggregates harbour a diverse microbial community (12), however, the contribution of this resident community to functional aspects of aggregates is not well understood. More broadly, and considering other cyanobacterial aggregates and blooms, only a few studies have investigated metabolic activity and interactions within these systems (10, 14), and the emergence and stability of metabolic interactions between community members are not studied over longer temporal scales.

In the absence of detailed, longitudinal functional studies on cyanobacterial granules, it might be possible to draw parallels to cyanobacterial mats, which are also usually dominated by filamentous cyanobacteria and tend to form in tidal sediments and freshwater springs (22, 23). Several studies have shown that cyanobacterial mats harbour internal metabolic gradients, particularly in oxygen and sulfate, and contribute significantly to carbon burial and hydrogen production (5, 24-27). The metabolic gradients and hydrogen production in mats are

hypothesised to contribute to specific microbial interactions, which are partially elucidated for specific systems (24, 28). However, the broad set of metabolic interactions in cyanobacterial mats and the overall community dynamics and stability across time remains unclear. It is also unclear how findings from cyanobacterial mats correspond, or not, to aquatic cyanobacterial granules.

Here, we study formation and compositional dynamics in a nature-derived cyanobacterial granule to better understand community dynamics in such systems. Obtained originally from a freshwater reservoir, we were able to maintain this cyanobacterial granule system in the laboratory for over 1-year of successive culturing passages and preserve and revive it from cryostocks. Using longitudinal short-read shotgun sequencing, we show that the system is dominated by a novel, filamentous cyanobacterium and harbours a stable, tractable microbial community of 16 additional species, some with stable strain diversity. Both the composition of this community and the formation of spatially-organised granules, are stable over culture passages under defined laboratory conditions. Using long-read sequencing, we obtained full circular genomes for 15 of the species in the system, and identified several key metabolic functions encoded in them, including anoxygenic photosynthesis and sulfur cycling. We further predict metabolic interactions among the resident species, including carbon exchange between a cyanobacterium and heterotrophs through slime production and consumption, and vitamin exchanges among several species. We were able to provide experimental support for some of these interactions using isolates from the community. Taken together, our findings show that cyanobacterial granules can provide a specific microcosm that favour stable interactions among key microbial functional groups. Furthermore, the presented, nature-derived system provides a significant model system with medium compositional complexity and reproducible spatial organisation dynamics. As such, it will enable further studies on the relations between spatial organisation and community function, stability, and evolution.

RESULTS AND DISCUSSION

Nature-derived cyanobacterial community displays reproducible granule formation. We have collected natural cyanobacterial granules from a local freshwater reservoir and maintained these in a minimal medium with initially irregular, and subsequently regular, culture passaging (see *Methods*) (Fig. 1A). This process revealed stable and reproducible granule formation under laboratory conditions, where we obtain both spherical and more irregularly-shaped granules, as well as aggregates and connected biofilms, dominated by a filamentous cyanobacterium. In the remainder of the text, we refer to these structures collectively as 'cyanobacterial granules'. (Fig. 1B). The cyanobacterium in this system (see next section) displays a characteristic gliding motility (results not shown) involving slime excretion (41), and it is possible that these traits associate with the formation of the granules (9).

Cyanobacterial granules harbour a diverse, medium complexity microbial community. In the regular maintenance period presented here, we have performed over 10 culture passages – equivalent to over 1-year of culture maintenance. We focus here on samples from the five passages collected at similar culture age: (P)assage P0 and P3-6. At each passage, we extracted DNA and performed short-read Illumina sequencing, and then co-assembled the resulting raw sequencing data using bioinformatics pipelines (see *Methods*). Total amount of raw sequence data (Gb) per sample (from paired-end 150bp reads) was as follows: P0 = 10.8, P3 = 15.3, P4 = 14.4, P5 = 14.7, P6 = 15.2.

In all these samples, we found the same set of 16 metagenome assembled genomes (MAGs), which we have taxonomically assigned using GTDB-Tk (29) (see *Methods*, Fig. 1C and Table 1). Besides the mentioned cyanobacterium, the remaining 15 species were

distributed in the phyla Actinobacteriota (5 species) and Proteobacteria (10 species) (Fig. 1C and Table 1). The latter group was spread between the alpha- (8 species) and gamma-Proteobacteria (2 species, *Pseudomonas E. composti* and a species of the *Hydrogenophaga* genus). Within the alpha-Proteobacteria, one, four, and three species were found within the orders Rhodobacterales, Rhizobiales, and Acetobacterales respectively. For the Acetobacterales order, one species was from the *Roseococcus* genus and two from the *Roseomonas* genus.

Of the identified MAGs, most were taxonomically close to either a cultured species or to an uncultured MAG found in databases (see Table 1). The cyanobacterium had only one close homolog in the databases, a single MAG obtained from sampling of extant stromatolites (30). This has prompted us to further analyse the phylogeny of these two cyanobacteria (see *Methods*). We found that the presented cyanobacterium and its close homolog fall under a new monophyletic clade within the Cyanobacteriales order, possibly forming a novel family (Fig. S1). Based on its geographical origin and characteristic motility, we propose the name *Fluctiforma draycotensis* gen. nov., sp. nov. for the species found in our system and suggest a family name of *Fluctiformaceae*.

The cyanobacterial granule community is stable over culture passages and displays species- and strain-level co-existence. To see how, and if, the community composition changed over the period of the regular culture passages, we used the co-assembly and coverage analysis pipeline STRONG (31) (see *Methods*). The coverage across passage samples P0 to P6, normalised per Giga-base pair (Gbp) of sequencing to avoid biases from differences in sample sizes, qualitatively showed a stable composition when analysed at the order level (Fig. 2A). At the level of MAGs, the same analysis showed some variability over passages (Fig. 2B), but there was no significant correlation between passage number and coverage for any of the bins (Table S2). *F. draycotensis* was found in much higher abundance than all other bins (coverage 225(-fold) per Gbp) and decreased only 1.5-fold between P0 and P6. The STRONG pipeline is designed to identify strain diversity from longitudinal datasets such as presented here. This allowed us to identify two strains in the *Allorhizobium* species, and also two strains each in the species assigned to the *Pseudoxanthobacteraceae* family and *Chryseoglobus* genus (Table 1). All of these strains were detected in all passages presented in Fig. 2, but for each species and passage, one of the two strains was most dominant in each case (Fig. S2).

While the presence of the same 16 MAGs and strains over passages suggests that a high level of compositional stability is reached in this community, we noticed that three species showed a larger magnitude of decrease in abundance over passages, suggesting that these species might still be lost from the system with subsequent culturing. Two of these, belonging to the *Microbacterium* genus (Bin 3) and the *Beijerinckiaceae* family (Bin 16), decreased in coverage by more than 10^2- and 10^5-fold between P0 and P6, respectively. The species *Rhodococcus B sp000954115* (Bin 14) had low but stable coverage over passages but was not detected in P6. This observation might relate to a younger sample age of P6 (155 days old compared with 184 to 254 days for the other samples), as coverage is expected to change over culture development time.

Overall, we found that almost all reads were mapped to MAGs (above 99%) for all samples covered in this study, showing that we identified all species extracted from the community. We have also noted that the system was generally robust to cryo-preservation, as revived community samples developed granules and showed the same community composition in terms of species composition (Table S3).

Long-read sequencing provides circular genomes of community members. To obtain complete genome assemblies, we used long-read PacBio Hifi sequencing of two samples (P0

and P7) (see *Methods*). Co-assembly of the resulting sequences resulted in 16 genomes (Table S1). These contained two *Allorhizobium* species, compared to one in the short-read assembly, and one of these species, *Allorhizobium rhizophilum,* had two strains resolved (as was the case for the single *Allorhizobium* bin identified in the short-read based analysis). The latter findings corroborate the short-read based analysis by STRONG, whilst the detection of an additional species of the *Allorhizobium* genus shows that the higher resolution of the long-read sequencing allowed us to disentangle this additional diversity. Two species from the short-read sequencing (Bins 3 and 14) were missing in the long-read sequencing. The coverages for these species decreased over passages in our longitudinal analysis (see above and Fig. 2B), therefore were likely missed from the co-assembly of two samples. The overall agreement between short-read and long-read assemblies further indicates that we have sufficiently catalogued the species diversity in this structured cyanobacterial community.

Of the resulting 16 long-read genome assemblies (including two *Allorhizobium rhizophilum* genomes for different strains), 5 were contained in a single, circular contig indicating high quality assemblies, whereas 11 were split over multiple contigs (see Table S1). For two species (*F. draycotensis* - referred to as JAAUUE01 in Table S1- and the species from the *Chryseoglobus* genus), the assemblies resulted in high variance, which was resolved through use of a consensus path algorithm resulting in the presented final genome sequences (see *Methods*). Overall, these long-read genomes showed high completeness with eleven genomes showing 100% completeness based on presence of a set of single copy genes (SCGs) (see *Methods*). Bin genome sizes ranged between 2.9Mb for the species from the *Chryseoglobus* genus to 6.8Mb for the species from the *Roseomonas A* genus, while GC content ranged between 44.3% for *F. draycotensis* to 71.3% for the species from the *Roseomonas B* genus.

Full genome annotations predict key metabolic functions and interactions in the cyanosphere. The stable species coverage in this structured cyanobacterial community, together with the fact that our culture media lacked any carbon sources, implies the presence of metabolic interactions among community members. To explore this hypothesis, we used the binned genomes from both short- and long-read sequencing and analysed metabolic capabilities of each species (see *Methods*). This analysis has confirmed the photosynthetic capability of *F. draycotensis*, as expected, but also revealed anoxygenic photosynthesis capability in nine species; the species from the *Hydrogenophaga* genus, four out of five species from the Rhizobiales order, and the three species from the Acetobacterales order and the species from the Rhodobacterales order (Fig. 3A). Anoxygenic photosynthesis is an alternative type of photosynthesis whereby sulphide or other electron donors are used, instead of water, resulting in no oxygen production (32-34). Anoxygenic photosynthesis was first discovered in anaerobic bacteria (32) however it is possible that functions are maintained in the presence of oxygen (in aerobic anoxygenic phototrophs) (35). In this latter case, the role of anoxygenic photosynthesis is unclear, particularly in heterotrophic organisms (32, 35), but it is known that this type of photosynthesis specialises on longer (above 700 nm) wavelengths (32, 36, 37). Thus, it is possible that the shading effects and oxygen gradients arising from structure formation in this granule system create a suitable microenvironment for anoxygenic photosynthesising species to scavenge higher wavelengths of light than those used by *F. draycotensis,* for energy generation or carbon fixation.

We have also found presence of several sulfur assimilation and cycling genes (Fig. 4). Two species, those from the *Hydrogenophaga* genus and the *Beijerinckiaceae* family (genus BN140002), showed capability of thiosulfate oxidation (complete SOX complex), while several species, including *Pseudomonas E. composti* showed capability for sulfate and sulfite reduction. This indicates presence of a sulfur cycle in the system, through sulfate reduction to

sulfite and hydrogen sulfide, and subsequent oxidation of these compounds – and related thiosulfate and sulfur (38, 39) – back to sulfate via the SOX complex. Additionally, we found that several species, including *F. Draycotensis* contain alkanesulfonate and thiosulphate transport genes, suggesting interspecies sulfur exchanges via alkanesulfonate secretion and uptake.

In addition to these light- and sulfur-related key metabolic functions, we assessed the completeness of carbon degradation pathways with predicted function in the cyanosphere. We found that glucuronate, galactonate, and galacturonate degradation pathways are largely present in most of the species belonging to the alpha-Proteobacteria phylum (*Rhizobiales* and *Rhodobacterales* orders and *Roseomonas* genus) and galactose degradation is present in two species of the Actinobacteria phylum (Fig. 4). This finding is interesting because galacturonic acid and galactose have been previously identified as major and minor components, respectively, of cyanobacterial slime (40), which is essential for gliding motility (41). Thus, it is possible that a major interaction in the granule system is the degradation of cyanobacterial slime and its use as a carbon source by other bacteria in the community (see also next section). We find that several of the species with genetic capacity for these degradation pathways also have the genes for the acetate and propionate fermentation pathways (Fig. 3A), suggesting that they might be fermenting some of the slime-derived carbon into these organic acids, which can then be used by other species.

Previous studies have shown vitamin-based interactions between algae or cyanobacteria and other bacteria (42-44). Analysing the completeness of vitamin biosynthesis pathways in the presented genomes (Fig. S3), we identified two key vitamins – vitamin B7 and B12 – for which pathway completion varied greatly. Vitamin B7 (Biotin) biosynthesis consists of a two-part pathway (illustrated in Fig. S4B) in which the biotin precursor pimeloyl-ACP is firstly synthesised and then converted into a biotin ring (45). We found a complete biotin pathway only in *Pseudomonas E. composti* and low completeness across the other species (Fig. S4A). The vitamin B12 (cobalamin) biosynthesis pathway consists of more than 30 genes, involving corrin ring synthesis and the nucleotide loop assembly (46, 47) (Fig. S4B). Corrin ring synthesis forms the upper pathway and can proceed via an aerobic (KEGG module M00925) or anaerobic (M00924) pathway (Fig. S4B). Additionally, synthesis of the lower ligand of vitamin B12, DMB (5,6-dimethylbenzimidazole), is required and feeds into the nucleotide loop assembly (KEGG module M00122) (48, 49). Cells can alternatively uptake corrinoid compounds (such as the intermediate cobinamide) and convert these via the scavenge pathway for entry into the nucleotide loop assembly (47). Cells may also transport exogeneous cobalamin directly into the cell, via transporters that have been differentially characterised for gram-negative and positive bacteria (47, 50). Vitamin B12 biosynthesis is so far known to be restricted to certain bacterial species (51, 52) however species variation in genes encoding biosynthesis and transport pathways is under-characterised. This could explain apparent absence of genes encoding cobalamin transporters particularly for species that lack biosynthesis. In our system, we found that *P. composti* and *A. rhizophilum* have complete or near-complete modules for all sections of the vitamin B12 pathway suggesting full biosynthesis capability in these two species (Fig. S4A). All other species have partially complete B12 pathways, some also missing DMB synthesis but with cobinamide scavenging complete, including *F. draycotensis.* Vitamin B12 biosynthetic genes are almost completely absent in the two species from the *Microbacteriaceae* family and in the three species from the *Acetobacteraceae* family (Fig. S4A).

Presence of vitamin B12 biosynthetic genes, except for DMB synthesis and conversion, in *F. draycotensis* supports previous reports that cyanobacteria commonly retain the ability to synthesise pseudocobalamin (44), an alternative form of cobalamin. Overall, the patchiness of biosynthetic genes across the different species in our system suggests possible

metabolic interactions mediated by vitamin B7 and B12 exchanges not only between cyanobacteria and the other bacteria, but among the latter as well.

Species isolation and experimental assays support metabolic predictions. To support some of the metabolic functions and interactions obtained from the genome analysis, we attempted to isolate individual species from the cyanosphere community. Via addition of different carbon sources to our original culture media and dilution plating (see *Methods)*, different colony morphotypes were identified and three species isolated. Characterisation of the 16S rRNA V4-V5 hypervariable gene region identified these as *P. composti* (Gammaproteobacteria), *A.rhizophilum* (Alphaproteobacteria) and *Microbacterium maritypicum* (Actinobacteriota). This was based on complete alignment to the 16S rRNA sequences in the assembled genomes from the long-read data.

Using these isolates, we performed growth assays on a wide range of carbon sources (see *Methods*), including those sugars for which degradation pathways were identified, as discussed above. We observed growth on galacturonate and glucuronate for *A. rhizophilum* and *P. composti* (Fig. 3B). Growth on galactose was seen for *A. rhizophilum* and *M. maritypicum*. In addition, growth, particularly for *A. rhizophilum* and *M. maritypicum*, was seen on several other sugars (glucose, galactose, xylose, mannose, fucose, arabinose) that are also described as components of cyanobacterial slime (40, 53, 54). These findings support the genome-based analyses suggesting a possible carbon exchange through production and degradation of slime, by cyanobacteria and heterotrophs respectively.

CONCLUSIONS

We have presented here laboratory maintenance and analysis of a cyanobacterial granule-forming system. Culture passaging and observation of this system has shown repeatable and highly robust structure formation, while longitudinal sampling and sequencing has shown that it comprises of a medium-complexity community of 17 species. We found that this community has a stable composition with both species and strain-level diversity across culture passages. The compositional stability as well as structure formation, is also observed through cryostocking and revival. We have retrieved fully complete genomes for 15 species in this system and found these to encode key metabolic functions that are suggested to lead to metabolic interactions through sulfur cycling, and carbon and vitamin exchange. In particular, three isolated species from the system showed growth on sugar monomers predicted to be found in cyanobacterial slime.

Taken together, these results support the notion of a cyanosphere, which is a microbial system dominated by cyanobacteria and co-inhabited by select bacteria with key metabolic functionalities (14-17, 55). In our system, the cyanosphere is structurally-organised, and this organisation seems to be associated with the gliding motility of a filamentous cyanobacterium. A key role for filamentous cyanobacteria is also shown for the formation of photogranules in the open ocean (12, 56) and in bioreactors (9), and these bacteria are always found as dominant species in microbial mats (22, 23). These findings, together, suggest filamentous cyanobacteria either as sole determinants or key components of macro-scale spatial organisation. This spatial organisation, both during its initial formation stages and its later ‘mature’ stage may provide specific environmental niches, including anoxic regions, as well as regions with specific light, for other bacteria. The presence of such micro-niches could also explain the unexpected finding of strain and species co-existence in our system.

In line with the cyanosphere notion, we find a functionally and taxonomically-distinct resident community composition in the presented, spatially-organised granule system. The composition

is similar – at the higher taxonomic levels – to previous single-time point studies of cyanobacterial aggregates and blooms (10-14). In particular, our system is also enriched in alpha- and gamma-Proteobacteria and Actinobacteria, however, some phyla such as the Verrucomicrobiota and Planctomycetia that are implied to be enriched in previous studies are missing in our system. It must be noted, however, that all previous studies to date were of single or a few time-points studies on natural samples, while the presented analysis focussed on a nature-derived system kept in the laboratory over many passages. Thus, it is possible that there is a core cyanosphere community that can be stably recovered under laboratory conditions, while there are also peripheral community members that are more prone to change in a condition-specific manner and that can be lost over time under laboratory conditions.

Interestingly, we find that the stable community arising here encodes many of the key functionalities seen in cyanobacterial mats, including anoxygenic photosynthesis, fermentation, and sulfur cycling (22, 23). Relating to these functions, we predict key interactions regarding sulphur cycling, carbon and vitamin exchanges between the cyanobacterium and bacterial heterotrophs as well as amongst heterotrophs. In line with our findings, a recent study used metagenomics to characterise several *Microcystis* bloom communities and has identified similar heterotrophic composition as found in our system. Functional metabolic complementarities in sulphur cycling and vitamin B12 biosynthesis, in addition to nitrogen and fatty acid metabolism were also identified (10). This suggests possible common functionalities and interactions in the cyanosphere, similar to those found in mats, and in particular involving sulphur cycling and vitamin exchanges.

Nitrogen fixation, as implied in oceanic cyanobacterial aggregates and intertidal mats (15, 25, 26) and in interactions between cyanobacteria and heterotrophic diazotrophs (55), is missing in our system. This could be due to presence of nitrate in the eutrophic lake environment that our samples originate from, or in the culture media used. It would be interesting to repeat the presented laboratory culturing approach in the absence of nitrate, to see if the resulting community would be significantly different. A similarly interesting question is whether metabolic functionalities retained in cyanobacterial granules would change under different environmental conditions and perturbations, including changes in light intensity or wavelength. The presented system will act as a model system to study this maturation process and its drivers in future studies. In particular, the medium complexity and culturability of the presented system offers the possibility of detailed metagenomic and metatranscriptomic studies, exploring both population dynamics and evolutionary genomics (57) of replicated community cultures through time.

METHODS

Sample collection and culture maintenance. Freshwater samples were collected from Draycote Water Reservoir, Warwickshire, UK on 5th October 2013. Samples were initially stored in lake water at room temperature in the laboratory (approximately 21°C) under static conditions and under diel light cycle provided by a fluorescent lamp (PowerPlant Sun Mate Grow CFL Reflector with 250w Warm Lamp).

Irregular culture maintenance period. Culture vessels were sealed with gas permeable film. Samples were sub-cultured in a set of media previously described for culturing of cyanobacteria: initially, samples were cultured in a minimal medium (MM) based on description in (58), with addition of trace metal and vitamin mixes (see Tables S4-6). Subsequently they were transferred to BG11 (59), and finally to BG11+ media (DSMZ media reference number 1593), which differs from BG11 only in vitamin B12 addition. Irregular sub-

culturing was performed in liquid media or on agar plates, however, full records of culture cycles were not kept over this initial period (six years).

Regular sub-culturing period. Cultures were maintained in BG11+, without any carbon source addition. A vitamin mix was added to BG11+ (Table S6) and the final media used is referred to as "BG11+ vitamin mix". Cultures were grown under continuous 12h/12h light/dark cycles with white light illumination provided by a fluorescent lamp (see above). Light intensity was measured using a PAR sensor (LI-COR Quantum Sensor (LI-190R-BNC-5) and Light Meter (LI-250A)) and cultures were grown under 14 – 20 µmol photons m^{-2} s^{-1}. Cultures were kept at room temperature (approximately 21°C) under static conditions and in 150 ml medical flat glass bottles. Long-term sub-culture passages have been maintained over more than two years and are ongoing. In this study, we focus on passages between 'P0' and 'P6' representing a period of one year. For each sub-culture, we performed a 1 in 200 dilution by transferring 150 µl of re-suspended filamentous culture into a final volume of 30 ml of BG11+ vitamin mix (Fig. 1A). Due to the in-homogeneity of the biofilm material, sampled biomass quantity at each transfer could not be fully standardised. Passages were performed every 34-38 days, although a time period of 187 days occurred between Passage 0 and Passage 1. Culturing conditions for the Passage 0 culture (the first sequenced passage) differed slightly from the other cultures in that the medium contained an alternative metal mixture to the trace metal mix of the BG11+ medium (Table S5) and was cultured in a 500 ml medical flat bottle with a culture volume of 100 ml. Similar filament bundles and granules were observed in this culture vessel as in later passage cultures.

Cryopreservation and revival. We tested cryopreservation and revival of a sample taken from Passage 6 of the structured community. After 28 days of growth, filaments were re-suspended in the culture via pipette mixing and 1 ml of culture was sampled. This aliquot was preserved in 10% v/v glycerol, based on previously described protocols (60-62). The 1 ml culture aliquot was centrifuged at 10,000*g* for five minutes until the culture was sufficiently pelleted, followed by removal of supernatant. 1 ml of BG11+ vitamin mix containing 10% v/v glycerol was added to the pellet and the pellet was re-suspended by pipetting. The sample was left for a 15-minute incubation period at low light intensity (below 5 µmol m^{-2} s^{-1}). This served as an equilibration period to protect the cells from cryoprotectant damage. The cryotube was then stored at -80°C. After 39 days of storage at -80°C, the cryostock was revived by thawing the top of the stock so that a 300 µl aliquot could be pipetted. To wash the cells, this aliquot was added to a microcentrifuge tube then centrifuged at 6,500*g* for 5 minutes. The supernatant was discarded and fresh culture medium (BG11+ vitamin mix) without cryoprotectant was added. This centrifugation and washing step were performed twice, and then cells were re-suspended in 300 µl of fresh BG11+ vitamin mix medium. Light intensity was reduced upon thawing to protect against cell damage. Culture aliquot is then stored at room temperature in the dark for 24 hours. To re-grow the community culture, the 300 µl aliquot was added into 14.7 ml of BG11+ vitamin mix medium in a 50 ml Erlenmeyer flask and re-grown under the same light conditions as the original culture. A sub-culture of the revived culture was prepared after 30 days of growth, confirming longer-term culture health.

Samples used in taxonomy and coverage analyses. *For short-read sequencing:* Cell pellets were collected from a set of eight samples. These eight samples consisted of *(i)* five samples taken from passages 0 and 3-6, selected for sequencing based on availability of older age cultures between 155 and 254 days following sub-culture; *(ii)* a sample from a culture maintained in MM (see media section above), aged at 441 days since sub-culture; and *(iii)* a sample from the cryo-revived culture of passage 6 (described above) after 44 days. This last

sample was compared with a similarly aged sample taken from passage 11 (collected at 49 days). *For long-read sequencing:* Cell pellets were collected from community samples (using same methods as above) for passage 0 (aged 129 days since sub-culture) (sample 1) and passage 7 (aged 84 days since sub-culture) (sample 2), with a minimum pellet wet weight of 100 mg. For passage 0, pellets were stored at -80°C prior to DNA extraction, whereas for passage 7, pellets were snap frozen in liquid nitrogen before being stored at -80°C.

DNA extraction and sequencing of community samples.

Shotgun sequencing. Cell pellets were collected from samples using the Qiagen PowerSoil Pro kit (Hilden, Germany, Cat. No. 47014). Biomass (suspension or biofilm) was sampled in 1 or 1.5 ml volume for liquid culture suspension or by suction onto the end of a pipette tip for large biofilm aggregates. Tubes were centrifuged at 10,000*g* for five minutes in a microcentrifuge (Stuart Microfuge SCF2: Bibby Scientific, Staffordshire, UK). The liquid phase was discarded. Wet weights of the pellets were recorded and adjusted to the range of 0.04 – 0.25gr. by sampling additional culture aliquots if necessary. When pellets were not used immediately for DNA extractions, samples were stored in the -80°C freezer until extraction. For DNA extractions, the kit protocol was followed with the following modifications. Beads from each PowerBead tube provided in the kit were carefully transferred into clean microcentrifuge tubes. Cell pellets were re-suspended in 0.5 ml of sterile water then added to each PowerBead tube and centrifuged at 10,000*g* for five minutes. The liquid phase was removed before adding beads back into the PowerBead tubes. For the bead-beating step, a Vortex Genie-2 vortexer was used (Merck, Darmstadt, Germany, Cat No. Z258423) with a 24-tube adaptor (Qiagen, Cat No. 13000-V1-24) and all samples were vortexed for 15 minutes. For all centrifugation steps, the maximum speed of the microcentrifuge (12,300*g*) was set for 3 minutes. A negative control sample was included from step 1 of the protocol, by adding Solution CD1 to an empty PowerBead tube. DNA was stored in 75 µl volume of Solution C6 at -80°C before sending for sequencing. DNA concentration was quantified using a Nanodrop spectrophotometer (NanoPhotometer N60, Implen, München, Germany) and a Qubit fluorimeter (ThermoFisher Scientific, Waltham, USA, Cat No. Q33226). DNA was diluted to 15 – 30 ng/µl and a total of 450 – 900 ng of DNA was sent to Novogene (Cambridge, UK) for shotgun Illumina metagenomics sequencing involving DNA sample QC, library preparation and Illumina NovaSeq paired-end 150 base pair read sequencing, producing approximately 10 GB raw data per sample.

PacBio HiFi sequencing. DNA extractions were performed by the Natural Environment Research Council (NERC) Environmental Omics Facility (NEOF). In brief, high molecular weight (HMW) DNA was extracted using the Macherey-Nagel NuceloBond HMW DNA kit with liquid nitrogen grinding using mortar and pestle for more than 15 minutes. The frozen culture sample was added directly to the kit buffer in a 50ml tube to reduce any HMW DNA degradation. Proteinase K volume was doubled, compared to kit instructions. The samples were then cleaned using AMpure PB beads with four repeated cycles. DNA was extracted in small volumes to reduce contaminants and DNA degradation. DNA quality scores were 1.2 – 1.5, measured by absorbance at 260/230nm and 1.75 – 2.0 measured by absorbance at 260/280nm. PacBio DNA libraries and sequencing were completed at the Centre for Genomics Research (CGR), from extracted genomic DNA samples. Following sample QC, low input library preparation was used for sample 1 (sample P0) and the ultra-low input protocol was used for sample 2 (sample P7). Sequencing was performed on the Sequel II SMRT Cell in CCS run mode and raw sequence data was delivered for downstream bioinformatics analyses.

Short-read sequence assembly, binning, and coverage analysis. The STRONG pipeline was used to co-assemble all eight short-read samples with metaSPAdes, calculate coverage depths normalised by sample sequencing depth (per Gbp), bin the sample into MAGs, and finally use the time-series to decompose MAGs into strains (31). The STRONG pipeline is available online at: https://github.com/chrisquince/STRONG.

Long-read sequence assembly and binning. The two PacBio HiFi samples were assembled using hifiasm-meta (63) and the resulting unitig assembly graphs were used in the downstream analyses. ORFs were called on unitigs with Prodigal (V2.6.3) (64) with option meta, and single-copy core genes (SCGs) were annotated through RPS-BLAST (v2.9.0) (65) using the pssm formatted COG database (66), which is made available by the CDD (67) as in the STRONG pipeline. However, to take into account strain diversity and non bluntification of the assembly graph, SCGs were clustered with MMseqs2 (68) (v13.45111) with options --min-seq-id 0.99 -c 0.80 --cov-mode 2 --max-seqs 10000. The unitigs were then binned based on coverage in the two samples and based on their composition, using both binning software CONCOCT (69) and metabat2 (70). MAGs were dereplicated using custom scripts resulting in 18 high quality bins (greater than 75% completeness of SCGs in single-copy). Two high abundance species were fragmented by noise and strain diversity, *F. draycotensis* (denoted JAAUUE01 in Table S1) and the species from the *Chryseoglobus* genus. In these cases, we found circular consensus paths using maximum aggregate coverage depths. We taxonomically classified MAGs with GTDB-Tk (29), revealing two long-read MAGs that were strains of *Allorhizobium rhizophilum*. For the MAG identified as deriving from the *Phreatobacter* genus, genome size was small and completeness based on Check M using the DFAST annotation platform (71) (see below) was only 48.36 %. We improved on this by using the connected graph component for this MAG instead (Comp 243), which had higher completeness (see Table S1).

To assess the fraction of the total diversity represented by this collection of MAGs, 16S genes where annotated using Barrnap (72) and clustered into OTUs using VSEARCH (73) with 97% identity. Each of the resulting 14 distinct OTUs (species and strains of the *Allorhizobium* genus were clustered together) could be mapped to one or more MAGs, which allows to ascertain that the full genomic diversity in the dataset was converted into MAGs.

Genome annotations. Genome annotations were performed using the DFAST annotation platform (releases 1.2.15/1.2.18; (71)). DFAST was run per genome with settings additional to the defaults as follows: perform taxonomy and completeness checks, use Prodigal for annotation of the coding sequence, set E-value to 1e-10 and enable both HMM scan and RPSBLAST. For the cyanobacterial genome, the Cyanobase organism-specific database was selected. Resulting protein sequences from genome annotation were further annotated for KEGG orthologs (KOs) using KofamKOALA (release 102.0/103.0; (74)). Resulting KO lists from short and long-read data were concatenated to create a unique KO list (where possible) for each of the MAGs/bins identified from the PacBio data. Metabolic module completeness based on KOs was completed using select modules present in the MetQy database (75) with a threshold for completeness at 0.25 across genomes, for a given module. KO lists were combined for the two bins predicted to represent two different strains of *Allorhizobium rhizophilum*. Individual KOs were manually searched for the vitamin B12 and sulphur pathways analyses. Data was plotted using R version 4.1.2 (76) and package MetQy (version 1.1.0).

Phylogenetic analyses. Taxonomic assignment of metagenomes and circularised genomes is done using GTDB-Tk v2.1.1 (77) and data version r207, using standard settings on GTDB-Tk. This approach involves placing user MAGs on a pre-compiled phylogenetic tree of existing

546 species and MAGs found on GTDB-Tk (29). The resulting tree contains over 3600 tips, which
547 represent the GTDB database species (or MAGs) and the MAGs from our sample. To facilitate
548 its visualisation, in Fig. 1C, we have randomly sampled 5 percent or 1 representative
549 (whichever was larger) of GTDB database tips at the order-level, except for orders that had 10
550 or fewer representatives, which were included with all their representatives. This process
551 resulted in 2590 species (or MAGs), including the 16 MAGs discovered in this study. Out of
552 these 16 MAGs, 13 shows high similarity to cultured or sequenced genomes and were assigned
553 by GTDB-Tk at genus level, while the other three are assigned at higher levels (see Table 1).
554 In the case of the cyanobacterium found in the presented community, there was high sequence
555 similarity to only one other uncultured metagenome in the databases (GTDB id; JAAUUE01
556 sp012031635). This prompted us to further explore the taxonomic placement and run an
557 additional phylogenetic analysis. To do so, we used the multiple sequence alignment of the
558 single core copy genes – as created by the GTDB-Tk platform – to create an alignment of the
559 *Pseudomonas_E composti* bin (to be used as outgroup), the Cyanobacterium bin, and all the
560 other GTDB species from the Cyanobacteriales order. We then used this alignment to build a
561 maximum likelihood tree with FastTree (78) using the default options. The resulting tree, re-
562 rooted at the *P. composti* bin, is shown in Fig. S1.
563
564 **Species isolation.** Isolation of species from the community was achieved using three types of
565 agar with carbon source supplementation: BG11+ vitamin mix with 0.1% w/v glucose, yeast
566 mannitol and BG11+ vitamin mix with 0.05% w/v riboflavin. Yeast mannitol was selected to
567 enrich for species from the *Allorhizobium* genus and was prepared following the methods of
568 (79). Riboflavin was chosen to select for *Microbacterium maritypicum* based on previous
569 studies describing riboflavin degradation by this species (80, 81). Addition of glucose was
570 expected to generally enrich for heterotrophs.
571 Later passage cultures were selected to attempt isolation, using pipette mixing to
572 resuspend filaments. For BG11+ vitamin mix with 0.1% glucose, a 10 µl aliquot from passage
573 8, aged 42 days old, was sampled and successively streaked across an agar plate. For yeast
574 mannitol, a dilution of 1000-fold was firstly prepared from passage 7, aged 60 days old. A
575 spread plate was prepared with a 50 µl aliquot from this dilution. For BG11+ vitamin mix with
576 0.05% w/v riboflavin, a 100 µl aliquot was sampled from passage 12 at 57 days old and a
577 spread plate was prepared. Agar plates were either incubated under the same lighting conditions
578 as the community culture (see above) for BG11+ vitamin mix + 0.1% glucose, or in the dark
579 at 30°C for 2 – 3 days, or up to 9 days for BG11+ vitamin mix with riboflavin due to slower
580 colony growth. Single colonies observed on each agar type were re-streaked onto new agar
581 plates of the same type either once more (riboflavin agar), three times (glucose agar) or four
582 times (yeast mannitol agar) until maintenance of the colony morphotype was clear. Plates were
583 incubated at 30°C in the dark. On the glucose plate, a cream-coloured smooth margin colony
584 type was observed. On the yeast mannitol plate, round cream colonies were observed with a
585 gelatinous texture. On the riboflavin plate, small, cream, round colonies were visible and
586 bleaching of the plate occurred. The bleaching response occurs due to breakdown of riboflavin
587 to form lumichrome and ribose (80).
588 A liquid culture of each colony type was prepared to harvest material for DNA
589 sequencing and to cryopreserve each isolate. A single colony was inoculated into 30 ml of each
590 respective medium and grown in a 100 ml Erlenmeyer flask, incubated at 30°C and 180 rpm
591 for 48 hours. For the species isolated on BG11+ vitamin mix with 0.05% w/v riboflavin, the
592 liquid culture was prepared in LB medium as growth was unsuccessful in liquid riboflavin
593 medium. Cryostocks were prepared with a final concentration of 15% v/v glycerol for the
594 cultures grown in BG11+ media and with 30% v/v glycerol (as of (82)) for the culture grown
595 in yeast mannitol medium. Cryostocks were stored at -80°C. The remaining cultures were

pelleted at 4000 rpm for 10 minutes. DNA was extracted from the cell pellet using the Qiagen PowerBiofilm kit (Cat. No. 24000-50) and stored at -80°C. DNA concentration was quantified using the Nanodrop spectrophotometer.

Species characterisation via 16S rRNA gene region amplification. Isolates were characterised via Sanger sequencing of the bacterial 16S rRNA V3-V4 gene region using primers 341F and 806R as described in (8, 83). To distinguish between the two species of the *Allorhizobium* genus, a longer region was amplified using primer 341F with 1391R (84) as sequence differences occurred between the two species in this region. DNA was diluted to a concentration of approximately 10 ng/µl and added in 5 µl volume to the PCR reaction mix of 25 µl final volume. The reaction mix consisted of 12.5 µl of GoTaq G2 Green Master Mix (2X) (Promega Product Code: M7822), 1 µl of each of the forward and reverse primers (10 µM concentration) and 20 µl of sterile MilliQ water. For some PCR runs, a total volume of 50 µl was used and in these cases, the reagent volumes were doubled.

The PCR reaction was run using an Applied Biosystems Veriti Thermal Cycler (California, USA) with the following cycling conditions. Firstly, an initial denaturation of three minutes at 95°C was run. This was followed by 30 cycles consisting of denaturation for thirty seconds at 95°C, annealing at 49°C (50°C for 341F/1391R) for thirty seconds and elongation for ninety seconds (sixty seconds for 341F/1391R) at 72°C. A final elongation step of ten minutes (seven minutes for 341F/1391R) at 72°C was run followed by infinite hold at 4°C and refrigeration.

Products were run on a 1% w/v agarose gel stained with GelRed dye (Biotium, product code: 41003) using 1 x TAE buffer, including a 100 bp ladder (NEB, product code: N3231S) mixed with purple loading dye. The gel was run at 100 V for at least 45 minutes. Bands were visualised with the gel imaging system U:Genius3 (Syngene, Cambridge, UK) via a blue LED transilluminator. A band of the correct size could be visualised between 400 and 500 base pairs for primer pair 341F/806R and around 1 kilo base pair for primer pair 341F/1391R. PCR products were purified using the GeneJet gel extraction kit (Thermo Scientific, K0691) with the following modifications. Centrifugations were performed at 12,300*g* for 60 seconds. Sodium acetate was not added. Purified products were eluted in 20 – 50 µl of elution buffer then DNA concentration was quantified on the Nanodrop spectrophotometer. DNA was Sanger-sequenced (GATC) in the forward direction using the same forward primer as in the PCR reaction. Taxonomic identity was characterised using sequence alignment to the annotated 16S rRNA gene sequences from the long-read shotgun metagenomics community data.

Characterisation of the 16S rRNA gene regions in this way, revealed the three isolates to be the following three species. *P. composti*, *A. rhizophilum* and *M. maritypicum*. These species were grown successfully on BG11+ vitamin mix with 0.1% w/v glucose, yeast mannitol and BG11+ vitamin mix with 0.05% w/v riboflavin, respectively.

Physiological assays. To assess growth of the three isolated species on different carbon sources, the Biolog phenotypic microassay 'PM1 96 Carbon Utilisation Assay' (Biolog, Hayward, CA, USA) was used. This assay includes a range of possible carbon sources, including a set of amino acids, nucleotides, organic acids, polymers, sugars, sugar alcohols and sugar phosphates (85). Growth was measured via optical density, rather than profiling respiration with a redox dye. Each species was revived from its cryostock on LB agar. Plates were incubated at 30°C for 24 hours (for *P. composti*) and for 48 hours for the other two species (until sufficient colonies had grown). Each species was re-streaked onto a new LB agar plate before use in the Biolog assay. The inoculating fluid 1.0 x IF-0a to be used to suspend colonies swabbed from the agar plate, was prepared from 1.2 x IF-0a (Technopath, UK, product code: 72268) by dilution. For the Gram-positive protocol, an additive solution was added to the

inoculating fluid consisting of the following components in their final concentrations: 2mM magnesium chloride hexahydrate, 1mM calcium chloride dihydrate, 25µM L-arginine HCl, 50µM L-glutamate Na, 12.5µM L-cystine pH8.5, 25µM 5'-UMP 2Na, 0.005% yeast extract, 0.005% tween-85.

Colonies were removed from the LB agar plate using a sterile swab and added to IF-0a to form a lightly turbid suspension. The transmittance (%T) at 600 nm wavelength (turbidity) was checked using a benchtop spectrophotometer (Spectronic 200E, Thermo Scientific), and culture density was adjusted until %T was approximately 42% or 81% for the gram-negative and gram-positive species respectively (86). For gram-negative species, cell suspension was then diluted in 1.0 x IF-0 to the starting inoculum of 85% T (+/- 3%), or to 0.07 OD (87, 88). For gram-positive species, cell suspension was further diluted by 13.64-fold with inoculating fluid and additive solution. Plate PM1 was inoculated with 100 µl per well of cell suspension, with a separate plate prepared for each species.

Growth was measured via absorbance (optical density at 600 nm wavelength) with continuous incubation in a plate reader (CLARIOstar, BMG LABTECH GmbH, Ortenberg, Germany) at 30°C for 48 hours. Measurements were taken every fifteen minutes with double orbital shaking at 200 rpm for five minutes before each reading. Endpoint readings (48 hours) were used for data analysis, blank-corrected by the first or second optical density reading per well. Data was plotted using R version 4.1.2 (76).

FIGURES

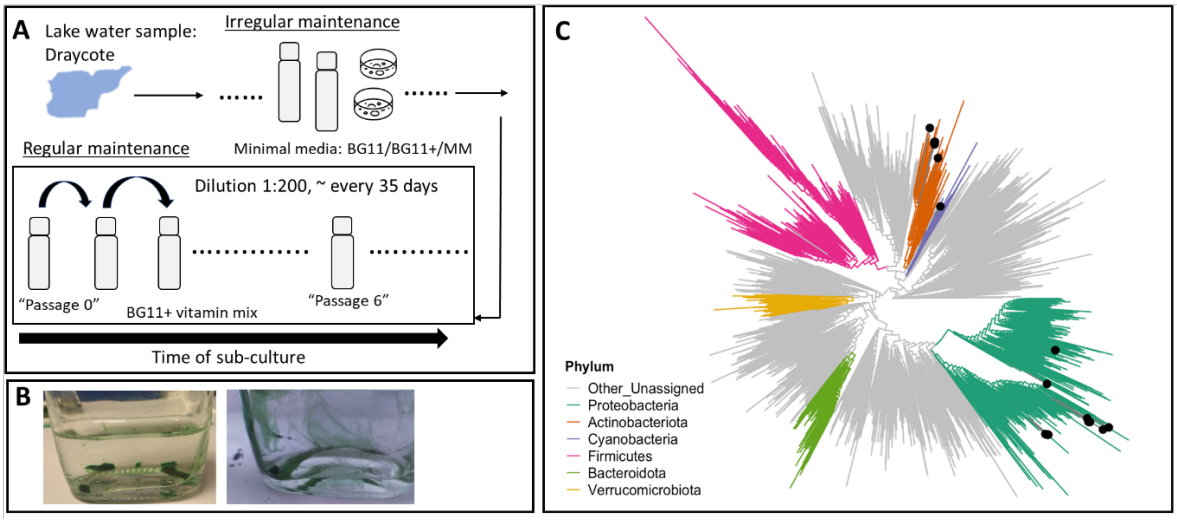

Figure 1. Culturing schematic, structure formation and taxonomic placement of granule-forming freshwater-derived phototrophic communities. (A) Schematic showing culturing regime from natural samples to regular laboratory sub-cultures. See *Methods* for full details. **(B)** Representative images of the structured community. The image on the left show granules of various sizes and on the right, filamentous bundles attaching to the base and walls of the medical flat glass bottle. **(C)** Phylogenetic tree showing taxonomic placement of the identified bacterial species in the cyanosphere community along with select species from the GTDB database (see *Methods*). The species from the community are highlighted with a black dot, while colours indicate different phyla as listed in the legend. 16 species are presented on the tree, described in Table 1, and discussed in the main text. For a maximum likelihood based phylogenetic tree of the cyanobacteria clade, including the cyanobacterium found in the system, see Fig S1.

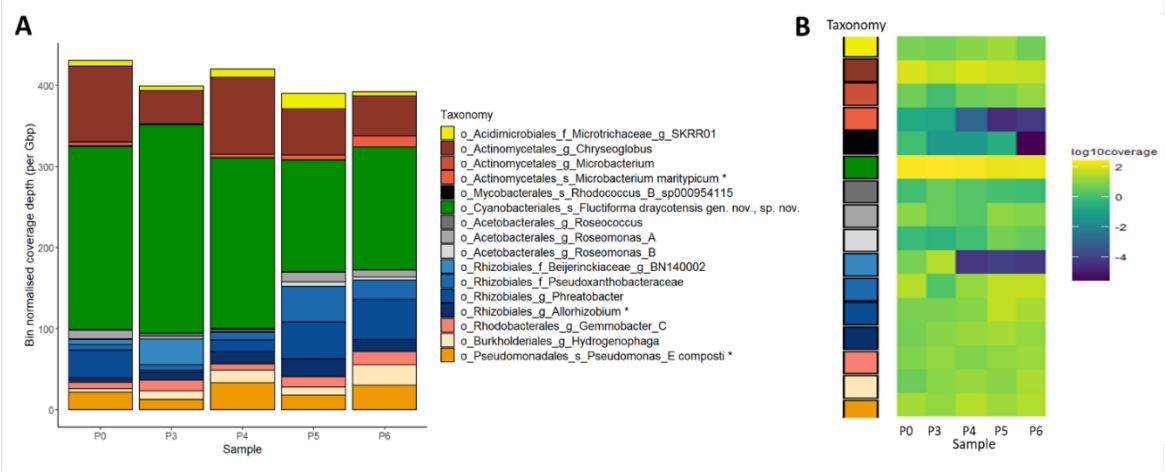

Figure 2. Coverage of taxonomic bins over serial passages. (**A**) Community composition, i.e. species coverage, over sub-culture passages (samples), colour-grouped at the taxonomic order level (listed in Table 1 and shown on the legend). Colour shades indicate MAGs within orders and the highest taxonomic resolution is described in the legend on the figure. Passage number is labelled on the x-axis, denoting different samples. Note that passages 1 and 2 are not shown as late age culture pellets were not collected (see *Methods* for details of samples used). The y-axis shows bin normalised coverage depth per Gbp of sequencing. Isolated species are indicated with an asterisk on the legend. (**B**) Same coverage data as in (A) but per individual bin and shown as log10-transformed. Taxonomy of bins is colour-coded as in (A). For Passage 6, no coverage of the *Rhodococcus_B_sp000954115* bin was detected so a pseudo-coverage was applied by multiplying the smallest bin normalised coverage value found across all bins and passages, by 0.1.

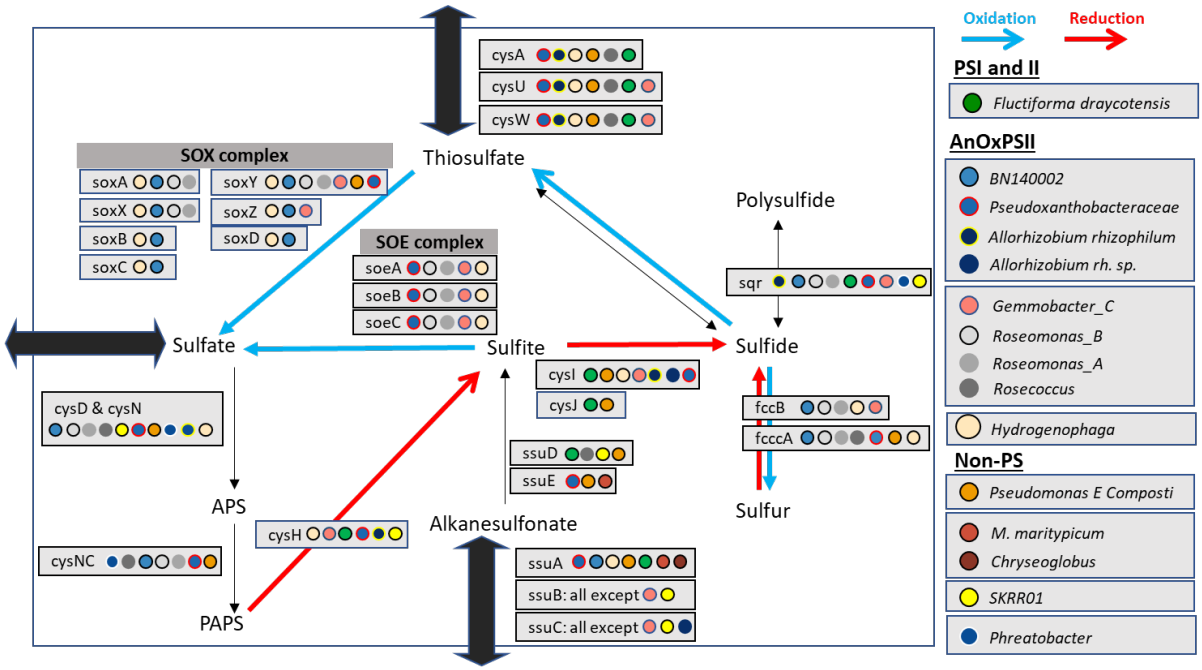

Figure 3. Analysis of sulphur-related genes' presence in assembled genomes. Schematic representation of the sulphur assimilation, scavenging, and oxidation pathways based on KEGG map 00920 on the KEGG database (89). The box outline represents the cell membrane, indicating uptake or secretion of sulphur compounds (indicated with thick, grey arrows), while

blue and red coloured arrows indicate sulphur-compound oxidation and reduction respectively. Gene presence in different genomes obtained in this study are mapped onto this schematic, indicated with coloured circles (see key). Some genes analysed were absent in all genomes. Full species names and taxonomy are described in Table 1. Species are grouped according to presence of photosystems I and II or presence of anoxygenic photosystem II. See *Methods* for full details of gene annotation and analysis.

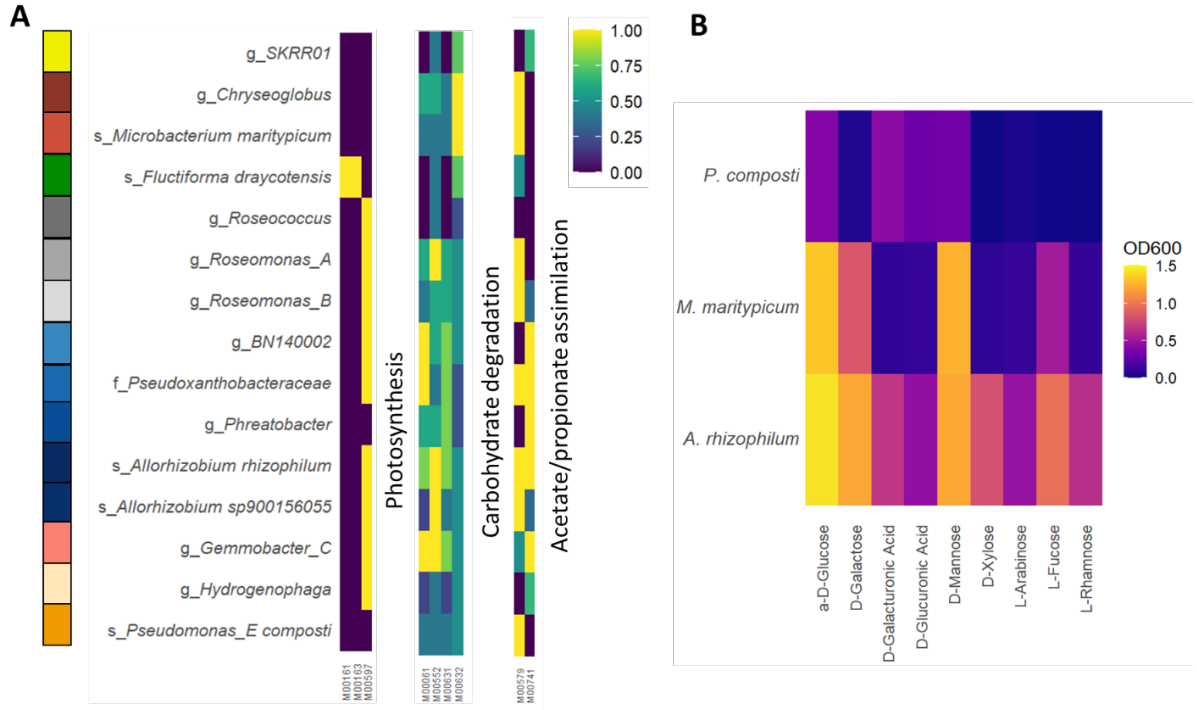

Figure 4. Metabolic modules associated with key carbohydrate and energy metabolic pathways and growth on key carbohydrates. (A) Completeness of select metabolic modules as defined by KEGG orthology (KO) blocks in the KEGG database (89) (see *Methods*). Only modules with completeness above 0.25 in at least one species are shown. Bins are colour-categorised by order (as in Fig. 1) and named by the highest taxonomic resolution resolved per bin (family, genus or species). Full taxonomy is given in Table 1. Module names are as follows. *Photosynthesis:* M00161 = PSII, M00163 = PSI, M00597 = anoxygenic PSII. *Carbohydrate degradation:* M00061 = glucuronate, M00552 = galactonate, M00631 = galacturonate, M00632 = galactose. *Fermentation pathways:* M00579 = conversion of acetyl-CoA to acetate. M00741 = conversion of propanoyl-CoA to succinyl-CoA. See *Methods* for full details of gene annotation and analysis. **(B)** Growth of three isolated species on carbon substrates associated with cyanobacterial 'slime'. Endpoint Optical Density (OD600) following 48 hours of growth is presented across carbon sources assayed in a Biolog PM1 microplate (see *Methods*). Those selected in this figure are based upon prior described components of cyanobacterial slime (40, 54).

TABLES

MAG	nb_haplo	RED	Phylum	Class	Order	Familly	Genus	Species
Bin_9	1	0.911643	Actinobacteriota	Acidimicrobiia	Acidimicrobiales	*Microtrichaceae*	*SKRR01*	/
Bin_24	2	0.992554	Actinobacteriota	Actinomycetia	Actinomycetales	*Microbacteriaceae*	*Chryseoglobus*	/

Bin_3	1	0.981747	Actinobacteriota	Actinomycetia	Actinomycetales	*Microbacteriaceae*	*Microbacterium*	/
Bin_10	1	1	Actinobacteriota	Actinomycetia	Actinomycetales	*Microbacteriaceae*	*Microbacterium*	*Microbacterium maritypicum*
Bin_14	1	1	Actinobacteriota	Actinomycetia	Mycobacteriales	*Mycobacteriaceae*	*Rhodococcus_B*	*Rhodococcus_B sp000954115*
Bin_6	1	0.985369	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	*JAAUUE01*	*JAAUUE01*	***Fluctiforma draycotensis* gen. nov., sp. nov.**
Bin_22	1	0.980006	Proteobacteria	Alphaproteobacteria	Acetobacterales	*Acetobacteraceae*	*Roseococcus*	/
Bin_4	1	0.958809	Proteobacteria	Alphaproteobacteria	Acetobacterales	*Acetobacteraceae*	*Roseomonas_A*	/
Bin_0	1	0.987196	Proteobacteria	Alphaproteobacteria	Acetobacterales	*Acetobacteraceae*	*Roseomonas_B*	/
Bin_16	1	0.930681	Proteobacteria	Alphaproteobacteria	Rhizobiales	*Beijerinckiaceae*	*BN140002*	/
Bin_20	1	0.981853	Proteobacteria	Alphaproteobacteria	Rhizobiales	*Phreatobacteraceae*	*Phreatobacter*	/
Bin_23	2	0.894109	Proteobacteria	Alphaproteobacteria	Rhizobiales	*Pseudoxanthobacteraceae*	/	/
Bin_19	2	0.99932	Proteobacteria	Alphaproteobacteria	Rhizobiales	*Rhizobiaceae*	*Allorhizobium*	/
Bin_7	1	0.954287	Proteobacteria	Alphaproteobacteria	Rhodobacterales	*Rhodobacteraceae*	*Gemmobacter_C*	/
Bin_17	1	0.990819	Proteobacteria	Gammaproteobacteria	Burkholderiales	*Burkholderiaceae*	*Hydrogenophaga*	/
Bin_21	1	1	Proteobacteria	Gammaproteobacteria	Pseudomonadales	*Pseudomonadaceae*	*Pseudomonas_E*	*Pseudomonas_E composti*

Table 1. Taxonomy of short-read assembled MAGs found in the cyanosphere community. Assigned taxonomy of the MAGs obtained from Illumina short-read shotgun metagenomics and using GTDB-Tk (29). See *Methods* for details of samples and bin assembly. "RED" stands for relative evolutionary distance and indicates phylogenetic similarity of the newly assembled bins to those existing in the GTDB database (29). "nb_haplo" refers to the number of haplotypes/strains detected in each MAG by the STRONG pipeline (31). The highest taxonomic resolution is presented for each bin. The novel species of cyanobacteria in our system is assigned by GTDB-Tk to an automatically generated family called 'JAAUUE01' (see main text for further discussion of the phylogeny of this species). This species name is highlighted in red in the table.

REFERENCES

1. F. Azam, F. Malfatti, Microbial structuring of marine ecosystems. *Nat Rev Microbiol* **5**, 782-791 (2007).
2. O. X. Cordero, M. S. Datta, Microbial interactions and community assembly at microscales. *Curr Opin Microbiol* **31**, 227-234 (2016).
3. C. Tropini, K. A. Earle, K. C. Huang, J. L. Sonnenburg, The Gut Microbiome: Connecting Spatial Organization to Function. *Cell Host Microbe* **21**, 433-442 (2017).
4. J. K. Jansson, K. S. Hofmockel, The soil microbiome-from metagenomics to metaphenomics. *Curr Opin Microbiol* **43**, 162-168 (2018).
5. T. M. Hoehler, B. M. Bebout, D. J. Des Marais, The role of microbial mats in the production of reduced gases on the early Earth. *Nature* **412** (2001).
6. R. E. Szabo *et al.*, Historical contingencies and phage induction diversify bacterioplankton communities at the microscale. *Proc Natl Acad Sci U S A* **119**, e2117748119 (2022).
7. B. Borer, I. Zhang, A. E. Baker, G. A. O'Toole, A. R. Babbin, Porous marine snow differentially benefits chemotactic, motile, and non-motile bacteria. . *bioRxiv* https://doi.org/10.1101/2022.05.13.491378 (2022).
8. L. M. Trebuch *et al.*, Impact of hydraulic retention time on community assembly and function of photogranules for wastewater treatment. *Water Res* **173**, 115506 (2020).
9. K. Milferstedt *et al.*, The importance of filamentous cyanobacteria in the development of oxygenic photogranules. *Sci Rep* **7**, 17944 (2017).

10. Q. Li *et al.*, A Large-Scale Comparative Metagenomic Study Reveals the Functional Interactions in Six Bloom-Forming *Microcystis*-Epibiont Communities. *Front Microbiol* **9**, 746 (2018).
11. B. Parveen *et al.*, Bacterial communities associated with *Microcystis* colonies differ from free-living communities living in the same ecosystem. *Environ Microbiol Rep* **5**, 716-724 (2013).
12. K. R. Frischkorn, M. Rouco, B. A. S. Van Mooy, S. T. Dyhrman, Epibionts dominate metabolic functional potential of *Trichodesmium* colonies from the oligotrophic ocean. *ISME J* **11**, 2090-2101 (2017).
13. M. D. Lee *et al.*, The *Trichodesmium* consortium: conserved heterotrophic co-occurrence and genomic signatures of potential interactions. *The ISME Journal* **11**, 1813-1824 (2017).
14. N. Pascault *et al.*, Insights into the cyanosphere: capturing the respective metabolisms of cyanobacteria and chemotrophic bacteria in natural conditions? *Environ Microbiol Rep* **13**, 364-374 (2021).
15. W. Bell, R. Mitchell, Chemotactic and growth responses of marine bacteria to algal extracellular products. *Biological Bulletin* **143**, 265-277 (1977).
16. J. K. Cole *et al.*, Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions. *Front Microbiol* **5**, 109 (2014).
17. E. Couradeau, A. Giraldo-Silva, F. De Martini, F. Garcia-Pichel, Spatial segregation of the biological soil crust microbiome around its foundational cyanobacterium, *Microcoleus vaginatus*, and the formation of a nitrogen-fixing cyanosphere. *Microbiome* **7**, 55 (2019).
18. C. Nelson, A. Giraldo-Silva, F. W. Thomas, F. Garcia-Pichel, Spatial self-segregation of pioneer cyanobacterial species drives microbiome organization in biocrusts. . *ISME Communications* **2**, 1-9 (2022).
19. H. W. Paerl, B. M. Bebout, Direct measurement of o2-depleted microzones in marine oscillatoria: relation to n2 fixation. *Science* **241**, 442-445 (1988).
20. M. Eichner *et al.*, N2 fixation in free-floating filaments of *Trichodesmium* is higher than in transiently suboxic colony microenvironments. *New Phytol* **222**, 852-863 (2019).
21. M. Eichner, S. Basu, S. Wang, D. de Beer, Y. Shaked, Mineral iron dissolution in *Trichodesmium* colonies: The role of oxygen and pH microenvironments. *Limnology and Oceanography* **65**, 1149-1160 (2019).
22. L. J. Stal, P. Caumette, North Atlantic Treaty Organization. Scientific Affairs Division., *Microbial mats : structure, development, and environmental significance*, NATO ASI series Series G, Ecological sciences (Springer-Verlag, Berlin ; New York, 1994), pp. xviii-463.
23. H. Bolhuis, M. S. Cretoiu, L. J. Stal, Molecular ecology of microbial mats. *FEMS Microbiol Ecol* **90**, 335-350 (2014).
24. L. C. Burow *et al.*, Anoxic carbon flux in photosynthetic microbial mats as revealed by metatranscriptomics. *ISME J* **7**, 817-829 (2013).
25. L. C. Burow *et al.*, Hydrogen production in photosynthetic microbial mats in the Elkhorn Slough estuary, Monterey Bay. *ISME J* **6**, 863-874 (2012).
26. D. Woebken *et al.*, Revisiting N(2) fixation in Guerrero Negro intertidal microbial mats with a functional single-cell approach. *ISME J* **9**, 485-496 (2015).
27. L. Charpy *et al.*, Dinitrogen-fixing cyanobacteria in microbial mats of two shallow coral reef ecosystems. *Microb Ecol* **59**, 174-186 (2010).

28. J. Z. Lee *et al.*, Fermentation couples Chloroflexi and sulfate-reducing bacteria to Cyanobacteria in hypersaline microbial mats. *Front Microbiol* **5**, 61 (2014).
29. 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 (2019).
30. S. C. Waterworth, E. W. Isemonger, E. R. Rees, R. A. Dorrington, J. C. Kwan, Conserved bacterial genomes from two geographically isolated peritidal stromatolite formations shed light on potential functional guilds. *Environ Microbiol Rep* **13**, 126-137 (2021).
31. C. Quince *et al.*, STRONG: metagenomics strain resolution on assembly graphs. *Genome Biol* **22**, 214 (2021).
32. S. Hanada, Anoxygenic Photosynthesis -A Photochemical Reaction That Does Not Contribute to Oxygen Reproduction. *Microbes Environ* **31**, 1-3 (2016).
33. J. F. Imhoff, "Systematics of Anoxygenic Phototrophic Bacteria." in Sulfur Metabolism in Phototrophic Organisms. Advances in Photosynthesis and Respiration, R. Hell, C. Dahl, D. Knaff, T. Leustek, Eds. (Springer, Dordrecht, 2008), vol. 27.
34. D. M. George, A. S. Vincent, H. R. Mackey, An overview of anoxygenic phototrophic bacteria and their applications in environmental biotechnology for sustainable Resource recovery. *Biotechnol Rep (Amst)* **28**, e00563 (2020).
35. D. Hauruseu, M. Koblížek, Influence of light on carbon utilization in aerobic anoxygenic phototrophs. *Appl Environ Microbiol* **78**, 7414-7419 (2012).
36. S. Spring, H. Lünsdorf, B. M. Fuchs, B. J. Tindall, The photosynthetic apparatus and its regulation in the aerobic gammaproteobacterium *Congregibacter litoralis* gen. nov., sp. nov. *PLoS One* **4**, e4866 (2009).
37. V. Selyanin, D. Hauruseu, M. Koblížek, The variability of light-harvesting complexes in aerobic anoxygenic phototrophs. *Photosynth Res* **128**, 35-43 (2016).
38. K. Wasmund, M. Mussmann, A. Loy, The life sulfuric: microbial ecology of sulfur cycling in marine sediments. *Environ Microbiol Rep* **9**, 323-344 (2017).
39. B. B. Jorgensen, A. J. Findlay, A. Pellerin, The Biogeochemical Sulfur Cycle of Marine Sediments. *Front Microbiol* **10**, 849 (2019).
40. J. L. Plude *et al.*, Chemical Characterization of Polysaccharide from the Slime Layer of the Cyanobacterium *Microcystis flos-aquae* C3-40. *Appl Environ Microbiol* **57**, 1696-1700 (1991).
41. E. Hoiczyk, Gliding motility in cyanobacterial: observations and possible explanations. *Arch Microbiol* **174**, 11-17 (2000).
42. M. T. Croft, A. D. Lawrence, E. Raux-Deery, M. J. Warren, A. G. Smith, Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**, 90-93 (2005).
43. S. A. Sañudo-Wilhelmy, L. Gómez-Consarnau, C. Suffridge, E. A. Webb, The role of B vitamins in marine biogeochemistry. *Ann Rev Mar Sci* **6**, 339-367 (2014).
44. K. R. Heal *et al.*, Two distinct pools of B12 analogs reveal community interdependencies in the ocean. *Proc Natl Acad Sci U S A* **114**, 364-369 (2017).
45. S. Lin, J. E. Cronan, Closing in on complete pathways of biotin biosynthesis. *Mol Biosyst* **7**, 1811-1821 (2011).
46. M. J. Warren, E. Raux, H. L. Schubert, J. C. Escalante-Semerena, The biosynthesis of adenosylcobalamin (vitamin B12). *Natural product reports* **19**, 390-412 (2002).
47. F. G. Costa, E. Deery, M. Warren, J. C. Escalante-Semerena, New insights into the biosynthesis of cobamides and their use. *Comprehensive Natural Products III*, 364-394 (2020).

48. M. E. Taga, N. A. Larsen, A. R. Howard-Jones, C. T. Walsh, G. C. Walker, BluB cannibalizes flavin to form the lower ligand of vitamin B12. *Nature* **446**, 449-453 (2007).
49. A. B. Hazra *et al.*, Anaerobic biosynthesis of the lower ligand of vitamin B12. *Proc Natl Acad Sci U S A* **112**, 10792-10797 (2015).
50. J. A. Santos *et al.*, Functional and structural characterization of an ECF-type ABC transporter for vitamin B12. *Elife* **7** (2018).
51. H. Fang, J. Kang, D. Zhang, Microbial production of vitamin B_{12}: a review and future perspectives. *Microb Cell Fact* **16**, 15 (2017).
52. X. Jin *et al.*, Eco-phylogenetic analyses reveal divergent evolution of vitamin B12 metabolism in the marine bacterial family 'P*sychromonadaceae'*. *Environmental Microbiology Reports* **14**, 147-163 (2022).
53. M. Nakagawa, Y. Takamura, O. Yagi, Isolation and characterization of the slime from a cyanobacterium, *Microcystis aeruginosa* K-3A. . *Agricultural and biologicalchemistry* **51**, 329-337 (1987).
54. R. De Philippis, M. Vincenzini, Exocellular polysaccharides from cyanobacteria and their possible applications. . *FEMS Microbiology Reviews* **22**, 151-175 (1998).
55. C. Nelson, A. Giraldo-Silva, F. Garcia-Pichel, A symbiotic nutrient exchange within the cyanosphere microbiome of the biocrust cyanobacterium, *Microcoleus vaginatus*. *ISME J* **15**, 282-292 (2021).
56. J. A. Arguedas-Leiva, J. Słomka, C. C. Lalescu, R. Stocker, M. Wilczek, Elongation enhances encounter rates between phytoplankton in turbulence. *Proc Natl Acad Sci U S A* **119**, e2203191119 (2022).
57. R. J. Whitaker, J. F. Banfield, Population genomics in natural microbial communities. *Trends Ecol Evol* **21**, 508-516 (2006).
58. G. V. Murvanidze, V. L. Gabai, A. N. Glagolev, Taxic responses in *Phormidium uncinatum.Microbiology* **128**, 1623-1630 (1982).
59. R. Y. Stanier, R. Kunisawa, M. Mandel, G. Cohen-Bazire, Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriol Rev* **35**, 171-205 (1971).
60. J. G. Day, Cryopreservation of microalgae and cyanobacteria. *Methods Mol Biol* **368**, 141-151 (2007).
61. A. A. Esteves-Ferreira *et al.*, Comparative evaluation of different preservation methods for cyanobacterial strains. *J Applied Phycol* **25**, 919-929 (2013).
62. M. J. Rastoll *et al.*, The development of a cryopreservation method suitable for a large cyanobacteria collection. *J Appl Phycol* **25**, 1483-1493 (2013).
63. X. Feng, H. Cheng, D. Portik, H. Li, Metagenome assembly of high-fidelity long reads with hifiasm-meta. *Nat Methods* **19**, 671–674 (2022).
64. D. Hyatt *et al.*, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
65. S. F. Altschul *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997).
66. R. L. Tatusov, M. Y. Galperin, D. A. Natale, E. V. Koonin, The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**, 33-36 (2000).
67. S. Lu *et al.*, CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* **48**, D265-D268 (2020).
68. M. Steinegger, J. Soding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026-1028 (2017).

69. J. Alneberg *et al.*, Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**, 1144-1146 (2014).
70. D. D. Kang *et al.*, MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
71. Y. Tanizawa, T. Fujisawa, Y. Nakamura, DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* **34**, 1037-1039 (2018).
72. T. Seemann (2018) barrnap 0.9 : rapid ribosomal RNA prediction. (https://github.com/tseemann/barrnap)
73. T. Rognes, T. Flouri, B. Nichols, C. Quince, F. Mahe, VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
74. T. Aramaki *et al.*, KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251-2252 (2020).
75. A. S. Martinez-Vernon, F. Farrell, O. S. Soyer, MetQy-an R package to query metabolic functions of genes and genomes. *Bioinformatics* **34**, 4134-4137 (2018).
76. R. C. Team (2021) R: A language and environment for statistical computing. . (R Foundation for Statistical Computing, Vienna, Austria).
77. P. A. Chaumeil, A. J. Mussig, P. Hugenholtz, D. H. Parks, GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315-5316 (2022).
78. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641-1650 (2009).
79. S. A. Legesse, Isolation, identification and authentication of root nodule bacteria (Rhizobia) in promoting sustainable agricultural productivity: A review. *Journal of Developing Societies* **6**, 87-93 (2016).
80. K. Yamamoto, Y. Asano, Efficient production of lumichrome by *Microbacterium* sp. strain TPU 3598. *Appl Environ Microbiol* **81**, 7360-7367 (2015).
81. H. Xu *et al.*, Identification of the First Riboflavin Catabolic Gene Cluster Isolated from *Microbacterium maritypicum* G10. *J Biol Chem* **291**, 23506-23515 (2016).
82. S. Y. Lin, A. Hameed, H. I. Huang, C. C. Young, *Allorhizobium terrae* sp. nov., isolated from paddy soil, and reclassification of *Rhizobium oryziradicis* (Zhao *et al.* 2017) as *Allorhizobium oryziradicis* comb. nov. *Int J Syst Evol Microbiol* **70**, 397-405 (2020).
83. A. Klindworth *et al.*, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**, e1 (2013).
84. J. J. Walker, N. R. Pace, Phylogenetic composition of Rocky Mountain endolithic microbial ecosystems. *Appl Environ Microbiol* **73**, 3497-3504 (2007).
85. B. R. Bochner, P. Gadzinski, E. Panomitros, Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res* **11**, 1246-1255 (2001).
86. Q. Guo *et al.*, Using a phenotype microarray and transcriptome analysis to elucidate multi-drug resistance regulated by the PhoR/PhoP two-component system in Bacillus subtilis strain NCD-2. *Microbiol Res* **239**, 126557 (2020).
87. L. Zhou, X. H. Lei, B. R. Bochner, B. L. Wanner, Phenotype microarray analysis of *Escherichia coli* K-12 mutants with deletions of all two-component systems. *J Bacteriol* **185**, 4956-4972 (2003).
88. S. S. Kumar, A. Penesyan, L. D. H. Elbourne, M. R. Gillings, I. T. Paulsen, Catabolism of Nucleic Acids by a Cystic Fibrosis. *Front Microbiol* **10**, 1199 (2019).

961 89. M. Kanehisa, M. Furumichi, Y. Sato, M. Kawashima, M. Ishiguro-Watanabe, KEGG
962 for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*
963 10.1093/nar/gkac963 (2022).
964
965
966