CrossMark

REVIEW

Cyanobacterial ultrastructure in light of genomic sequence data

C. R. Gonzalez-Esquer[1] · J. Smarda[2] · R. Rippka[3] · S. D. Axen[4] · G. Guglielmi[5] ·
M. Gugger[3] · C. A. Kerfeld[1,6,7,8,9]

Abstract Cyanobacteria are physiologically and morphologically diverse photosynthetic microbes that play major roles in the carbon and nitrogen cycles of the biosphere. Recently, they have gained attention as potential platforms for the production of biofuels and other renewable chemicals. Many cyanobacteria were characterized morphologically prior to the advent of genome sequencing. Here, we catalog cyanobacterial ultrastructure within the context of genomic sequence information, including high-magnification transmission electron micrographs that represent the diversity in cyanobacterial morphology. We place the image data in the context of tabulated protein domains—which are the structural, functional, and evolutionary units of proteins—from the 126 cyanobacterial genomes comprising the CyanoGEBA dataset. In particular, we identify the correspondence between ultrastructure and the occurrence of genes encoding protein domains related to the formation of cyanobacterial inclusions. This compilation of images and genome-level domain occurrence will prove useful for a variety of analyses of cyanobacterial sequence data and provides a guidebook to morphological features.

Introduction

Cyanobacteria are photosynthetic prokaryotes that have played and continue to play a prominent role in shaping the Earth's biosphere. Cyanobacteria were the main drivers of oxygenation in the early atmosphere (Canfield 2005) and gave rise to chloroplasts via endosymbiosis (Raven and Allen 2003). Today, they draw attention for their roles in CO_2 fixation and biogeochemical cycles, as well as key members of the global food web. Some cyanobacteria have

R. Rippka and G. Guglielmi have both retired

✉ C. A. Kerfeld
ckerfeld@lbl.gov

1 MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI, USA

2 Department of Biology, Faculty of Medicine, Masaryk University, University Campus, Building A6, Kamenice 5, 625 00 Brno, Czech Republic

3 Unité des Cyanobactéries, Institut Pasteur, Centre National de la Recherche Scientifique (CNRS) Unité de Recherche Associée (URA) 2172, 75724 Paris Cedex 15, France

4 Bioinformatics Graduate Group, University of California, San Francisco, CA 94158, USA

5 Institut de Biologie de l'ENS, IBENS, Inserm, U1024, CNRS, UMR 8197, Ecole Normale Supérieure, 75005 Paris, France

6 Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

7 Department of Plant and Microbial Biology, UC Berkeley, Berkeley, CA, USA

8 Berkeley Synthetic Biology Institute, UC Berkeley, Berkeley, CA, USA

9 Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, USA

shown promise as sources of renewable commodities (Machado and Atsumi 2012) or as a source of proteins and pathways that could be used to improve photosynthesis in plants (Zarzycki et al. 2013; Price et al. 2013). Moreover, many cyanobacteria are also known to produce compounds with pharmaceutical potential (Shenoy et al. 2001; Hess 2011; Dittmann et al. 2015).

Microscopy techniques have long been used to visualize the subcellular organization and morphology of cyanobacteria. Ultrastructure analysis of cyanobacteria by transmission electron microscopy (TEM) of thin sections of cells dates back to 1960s (Ris and Singh 1961; Gantt and Conti 1969). By 1977, cyanobacteria were well studied by light microscopy (Stanier and Bazine 1977), and by the 1980s, their ultrastructural characteristics were well described (Allen 1984; Jensen 1984; Stanier 1988). In this paper, we used the classical bacteriological fixation by osmium tetroxide according to Kellenberger et al. (1958) for visualizing cyanobacterial cell ultrastructure at the first time. The criteria for the morphological classification of cyanobacteria (subsections I–V) were established in 1979 (Rippka et al. 1979). More recently, genetics and ultrastructure were used to reconsider and to establish diacritical guidelines for the modern classification of cyanobacteria (Hoffmann et al. 2005).

The availability of 16S rRNA gene sequences and, subsequently, genomic sequence data has provided a new foundation for the classification of the cyanobacteria. The genome of Synechocystis sp. PCC 6803 was the first cyanobacterial genome to be fully sequenced (Kaneko et al. 1996), quickly accompanied by the first multicellular representative Nostoc sp. PCC 7120 (Kaneko et al. 2001). As cyanobacterial genome sequences continued to become available, it soon became evident that cyanobacterial morphology does not always correlate with phylogeny (Komárek and Kaštovský 2003). Researchers have been able to correlate morphology with phylogenetics in the case of the origin of multicellularity in cyanobacteria (Schirrmeister et al. 2011); however, until 2013, the available sequence data did not fully represent the five subsections. Coverage was greatly improved by the cyanobacterial genomic encyclopedia of bacteria and archaea (CyanoGEBA) dataset (Shih et al. 2013). This sequencing initiative more than doubled the amount of available cyanobacterial sequence data and included representatives of all subsections to more fully represent the phylum at the genomic level. Modern taxonomical classification of cyanobacteria involves both morphological features and evolutionary relationships uncovered by genomics; it has been proposed recently that such taxonomy should be based on monophyletic analysis (Komárek et al. 2014).

The expanded coverage of cyanobacterial genome sequence data, and consequently improved phylogenetic trees, prompted the revisiting of questions about the evolution of cyanobacterial morphological features, for example the study of carbonate inclusions and polyphosphate bodies (Benzerara et al. 2014) and of multicellularity in the phylum (Schirrmeister et al. 2015). Similarly, the analysis of an extended pool of genes/genomes, together with the knowledge of the associated physiology, morphology, and ecology, has resulted in the discovery of previously unrecognized metabolic potential in cyanobacteria. For example, by "subtractive genome analysis," Schirmer et al. (2010) were able to identify potential candidate genes for the production of alkanes (a drop-in biofuel).

Here, we present a comparative morphological study of selected cyanobacterial strains in the context of genomic information from the CyanoGEBA dataset. In addition to the representation of five morphotypes [corresponding to the classification by Rippka et al. (1979)] in a phylogenomic context, this dataset captures the ecophysiological diversity within the phylum, as described in the associated metadata (detailed metadata compilation can be found in Shih et al. 2013). We compared proteins by tabulating the encoded Pfam domains (Finn et al. 2014), the structural and functional building blocks of proteins. The generation of novel proteins with different functions as a result of domain divergence or (re)combination can provide evolutionary information (Vogel et al. 2004). Our intent is to provide an online resource of cyanobacterial morphology representative of sequenced strains, and a source of information for the prediction of the capacity to form cyanobacterial inclusions, many of which hold biotechnological potential.

Micrographs of representatives from five morphological types

According to the structural and developmental classification proposed by Rippka et al. (1979), cyanobacteria can be classified into five groups (subsections I–V). Subsections I and II encompass single-celled cyanobacteria (unicellular and baeocystous, respectively), while subsections III–V comprise the cyanobacteria formed by a row of cells (filamentous, heterocystous, and ramified, respectively). We compiled micrographs from 37 cyanobacterial strains (33 of the 37 were sequenced for the CyanoGEBA dataset) illustrating all five subsections and morphologies (hereafter referred to as Morphotypes) from the Pasteur Culture Collection of Cyanobacteria (PCC). Specifically, the micrographs include eight strains from Morphotype I, three strains from Morphotype II, 18 strains from Morphotype III, five strains from Morphotype IV, and three strains from Morphotype V (Figs. 1, 2, 3, 4, and 5 and Online

Resources 1–5). Given that there is a rich history of morphological description of cyanobacteria through imaging by transmission electron microscopy, we include a partial list of references containing electron micrographs to complement the images presented here (Online Resource 6). For the majority of strains included here, cultures were grown at the PCC and processed during the late exponential phase of growth. The micrographs were prepared by standard methods (see Online Resource 7).

Morphotype I (unicellular) encompasses spherical, cylindrical, or oval single cells that reproduce by binary fission (as shown in Fig. 1—PCC 7428) or by budding (Rippka et al. 1979). Several examples are illustrated by *Gloeobacter violaceus* PCC 7421 and *Gloeocapsa* sp. PCC 7428 in Fig. 1 and by *Synechococcus* sp. PCC 6312, *Chamaesiphon minutus* PCC 6605, *Synechococcus* sp. PCC 7336, *Halothece* sp. PCC 7418, *Synechococcus* sp. PCC 7502, and *Gloeocapsa* sp. PCC 73106 in Online Resource 1. The thylakoid-less *G. violaceus* PCC 7421 (Rippka et al. 1974) compared to a thylakoid-
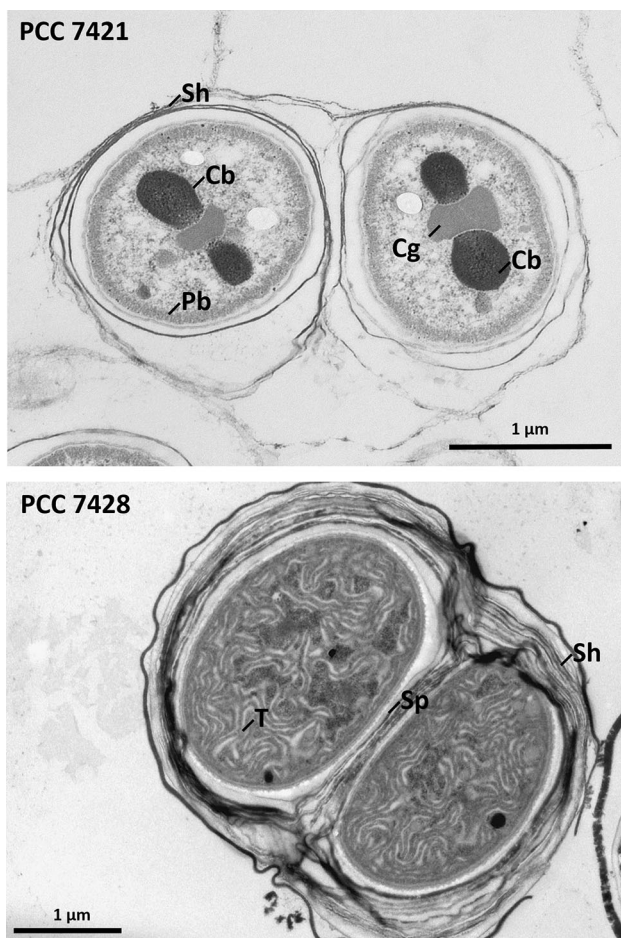

Fig. 1 Micrographs of select cyanobacterial strains from Morphotype I. *Top*: *Gloeobacter violaceus* PCC 7421 (Fixation: GLA) and *bottom*: *Gloeocapsa* sp. PCC 7428 (Fixation: KEL). *Cs* carboxysome, *Pb* phycobilisomes, *Sp* septum, *Sh* sheath, *T* thylakoid
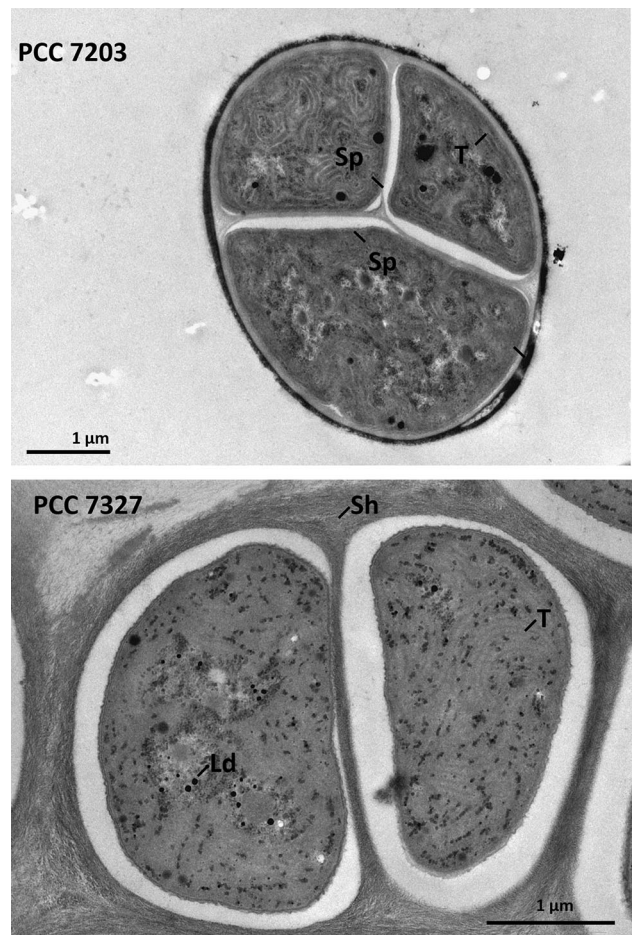

Fig. 2 Micrographs of select cyanobacterial strains from Morphotype II. *Top*: *C. thermalis* PCC 7203 (Fixation: KEL) and *bottom*: *Pleurocapsa* sp. PCC 7327 (Fixation: GLA). *Ld* lipid droplet, *Sh* sheath, *Sp* septum, *T* thylakoid

containing species represented by PCC 7428 (Fig. 1) highlights the ultrastructural organization needed for an efficient energy source to be incorporated into these photosynthetic organisms. In PCC 7421, photosynthesis occurs in the cytoplasmic membrane with a phycobilisome-rich area in its vicinity (Fig. 1—PCC 7421). In contrast, in PCC 7428 and all other cyanobacteria, the photosynthetic reaction centers are embedded in the thylakoid membranes. Morphotype II (baeocystous) reproduce through multiple fission in three planes (as can be observed in Fig. 2—PCC 7203) and are characterized by the formation of an outermost fibrous layer (Fig. 2—PCC 7327). Because of the formation of this layer and the rapid fission of cells, the strains from this morphotype form baeocytes (Rippka et al. 1979). Micrographs from the members of Morphotype II such as *Chroococcidiopsis thermalis* PCC 7203 and *Pleurocapsa* sp. PCC 7327, and *Stanieria cyanosphaera* PCC 7437 are shown in Fig. 2 and in Online Resource 2, respectively.
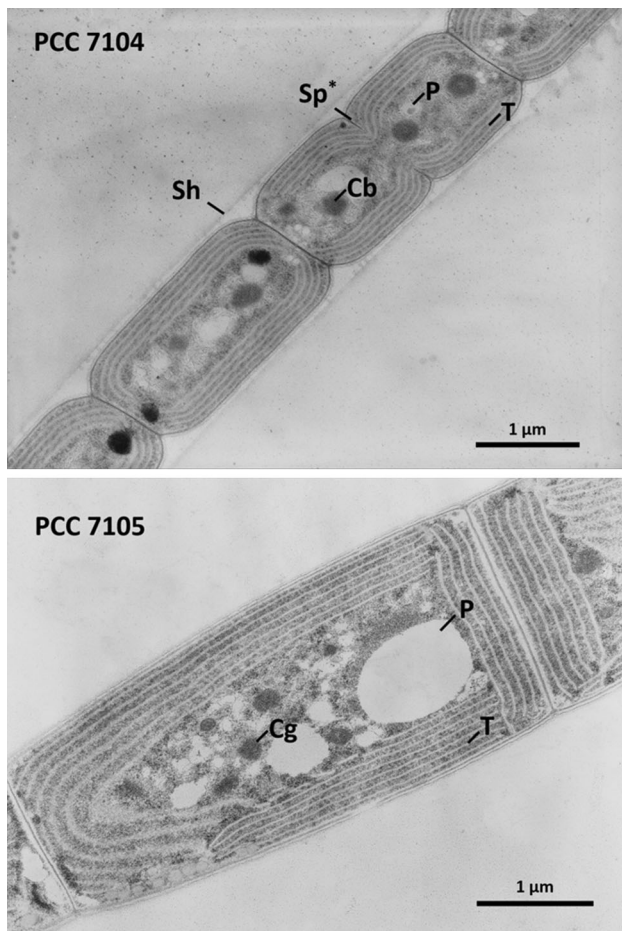
Fig. 3 Micrographs of select cyanobacterial strains from Morphotype III. *Top*: *N. nodulosa* PCC 7104 (Fixation: GLA) and *bottom*: *Geitlerinema* sp. PCC 7105 (Fixation: GLA). *Cb* carboxysome, *Cg* cyanophycin granule, *P* polyphosphate body, *Sh* sheath, *Sp* septum, *Sp** forming septum, *T* thylakoid

Fig. 4 Micrographs of select cyanobacteria from Morphotype IV. *Top*: *Calothrix* sp. PCC 7103 (Fixation: KEL) and *bottom*: *Nostoc* sp. PCC 7107 (Fixation: KEL). *Cn* cyanophycin granule, *P* polyphosphate body, *Sh* sheath, *Sp* septum, *T* thylakoid

The micrographs from *Nodosilinea nodulosa* PCC 7104 and *Geitlerinema* sp. PCC 7105 (Fig. 3) are representatives of Morphotype III (filamentous). The trichomes are formed of chain of vegetative cells, which divide by intercalary cell divisions on a single plane perpendicular to the long axis of the trichome (Rippka et al. 1979) (i.e., Figure 3— PCC 7104). Morphotype III strains are represented by micrographs from *Oscillatoria acuminata* PCC 6304, *Leptolyngbya boryana* PCC 6306, *Spirulina major* PCC 6313, *Leptolyngbya* sp. PCC 6406, *Oscillatoria formosa* PCC 6407, *Oscillatoria* sp. PCC 6412, *Leptolyngbya* sp. PCC 6703, *Pseudanabaena* sp. PCC 6802, *Oscillatoria nigro-viridis* PCC 7112, *Microcoleus* sp. PCC 7113, *Leptolyngbya* sp. PCC 7124, *Pseudanabaena* sp. PCC 7367, *Leptolyngbya* sp. PCC 7375, *Leptolyngbya* sp. PCC 7376, *Oscillatoria sancta* PCC 7515, and *Crinalium epipsammum* PCC 9333 (Online Resource 3). The filamentous strains from Morphotype IV (heterocystous) also are composed of
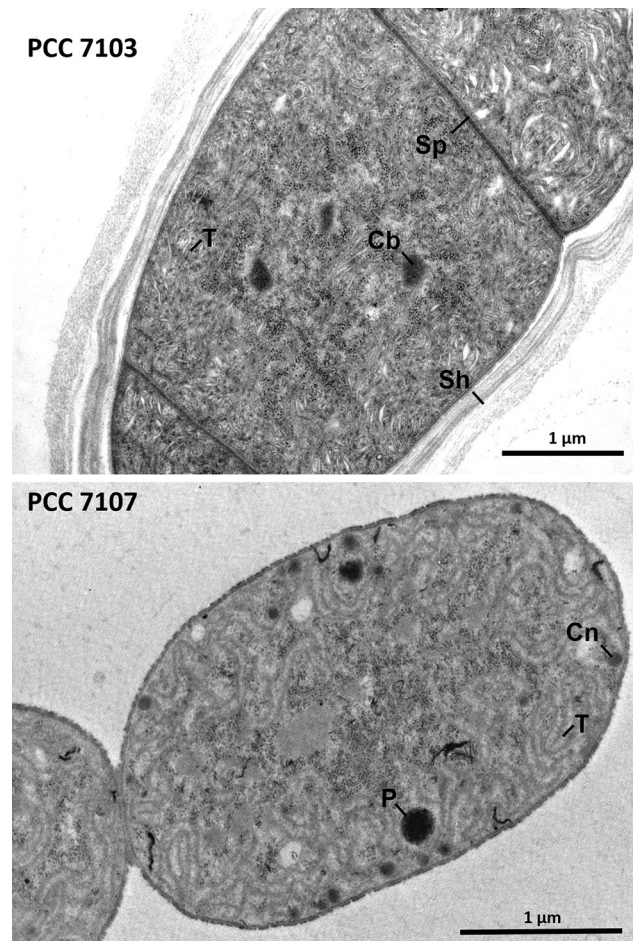
trichomes that divide on a single plane (Rippka et al. 1979). The heterocystous cyanobacteria are able to differentiate two types of cells: heterocysts to perform nitrogen fixation and akinetes to survive unfavorable environmental conditions and to develop new trichomes. *Calothrix* sp. PCC 7103 and *Nostoc* sp. PCC 7107 as well as *Calothrix* sp. PCC 6303, *Nostoc* sp. PCC 7524, and *Tolypothrix* sp. PCC 9009 illustrate some of the diversity found among the heterocystous cyanobacteria (Fig. 4 and Online Resource 4). Finally, the strains from Morphotype V (ramified) are also filamentous and able to differentiate cells into heterocysts and akinetes and can form true ramifications. The strains from Morphotype V can divide in multiple planes, resulting in branched trichomes (Rippka et al. 1979), as in *Fischerella* sp. PCC 9431 and *Mastigocladopsis repens* PCC 10914, with clear septations at different planes of division (Fig. 5). As in phylogenomic studies (Shih et al., 2013), we also include the unclassified cyanobacterium PCC 7702 as part of this group (Online Resource 5), as this
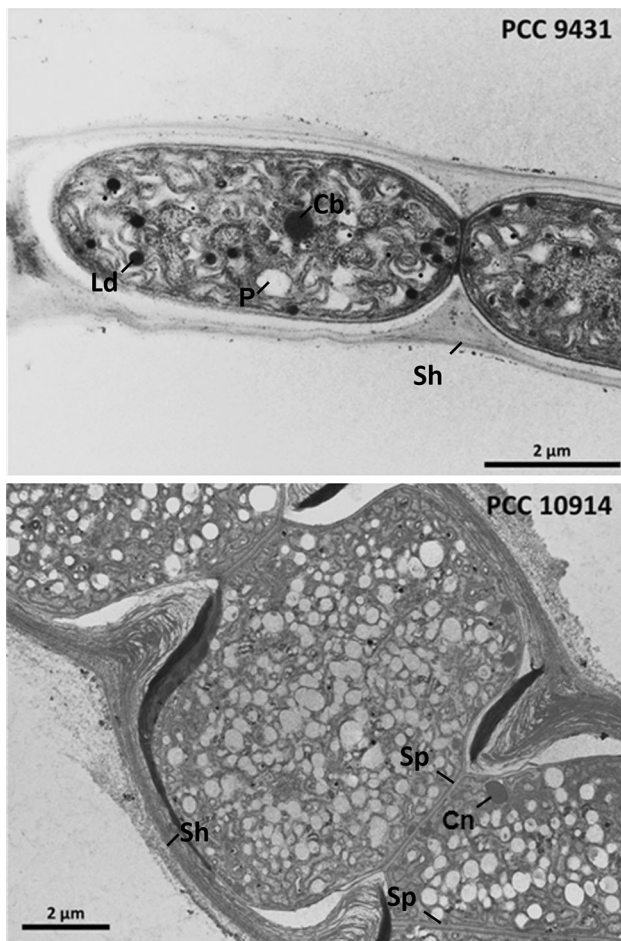
Fig. 5 Micrographs of select cyanobacterial strains from Morphotype V. *Top*: *Fischerella* sp. PCC 9431 (Fixation: GLA) and *bottom*: *M. repens* PCC 10914 (Fixation: GLA). *Cb* carboxysome, *Cn* cyanophycin granule, *Ld* lipid droplet, *P* polyphosphate body, *Sh* sheath, *Sp* septum

strain, although atypical, groups with members of *Chlorogloeopsis* that are known to lose, during development, their filamentous morphology to appear unicellular (Rippka et al. 1979).

Bioinformatic analysis of genomic patterns influencing cell structure

The micrographs presented here represent approximately one-third of the strains in the cyanobacterial phylogenetic tree after updating with the CyanoGEBA dataset (Shih et al. 2013). While these micrographs capture the morphological diversity across the cyanobacterial phylum, bioinformatic analysis of the represented strains enhances the understanding of the ultrastructural differences. Previous studies have correlated morphological information with genomic sequence data. Schirrmeister et al. (2015)

demonstrated that multicellularity precedes the diversification of all modern cyanobacterial morphologies. Shih et al. (2013) searched for genes responsible for morphological transitions between subsections but were unable to detect a specific gene set underlying any of the transitions. In the context of cyanobacterial inclusions, a survey of the genomes of 16 sequenced strains found that glycogen-related genes were prevalent, with fewer observations of cyanophycin-related and PHA-related genes (Beck et al. 2012).

We searched for differences in the genomes of the strains used for the phylogenetic tree from Shih et al. (2013) to identify contrasts at the genome level between clades. We focused on protein domains; the appearance of new domains or domain combinations frequently provides information on evolution of function and structure. For the domain search, we used Pfams, which are a manually curated compilation of protein domains based on multiple sequence alignments and hidden Markov models (Finn et al. 2014). On average, 71.4 ± 5.3 % of the genes in any given strain's genome contained at least one Pfam domain (Online Resource 8). We assembled a list of 11912 Pfams available in the 126 strains used by Shih et al. (2013) and annotated the occurrence of genes for each Pfam per strain (Online Resource 9). For a given Pfam, we normalized to the strain's total gene number to obtain the relative abundance of each Pfam (Online Resource 10). We then converted the Pfam abundances to proportions by dividing the relative abundance by the overall average per Pfam and to the average of each Pfam per clade (Online Resource 10) and plotted it as a graph to depict the overrepresentation and underrepresentation of Pfams per strain and per morphological type (Fig. 6). Interestingly, the Pfam distribution pattern of the picocyanobacteria (clade E) appears to be above the overall average (Fig. 6A), even though they have comparatively smaller genomes (Online Resource 8). This overabundance of domains can be explained by the evolutionary phenomenon of recombining existing domains in proteins to generate enzymes with novel functions (therefore, the same Pfam domain is present on different proteins). Differences in Pfam distribution patterns when analyzed by clade show differences between morphological types; for example, in clade C, the multicellular strains appear to contain more overrepresented Pfams than the unicellular (Fig. 6B), which could suggest a genomic complexity that correlates with morphological complexity. While the information from Fig. 6 provides an overview of the underlying differences between morphological types, it is difficult to quantify these differences.

To facilitate a numerical comparison between morphological groups, we reassigned genomes to clades, collated the relative abundances of every Pfam and tallied the Pfams conserved in at least 99 % of the strains belonging
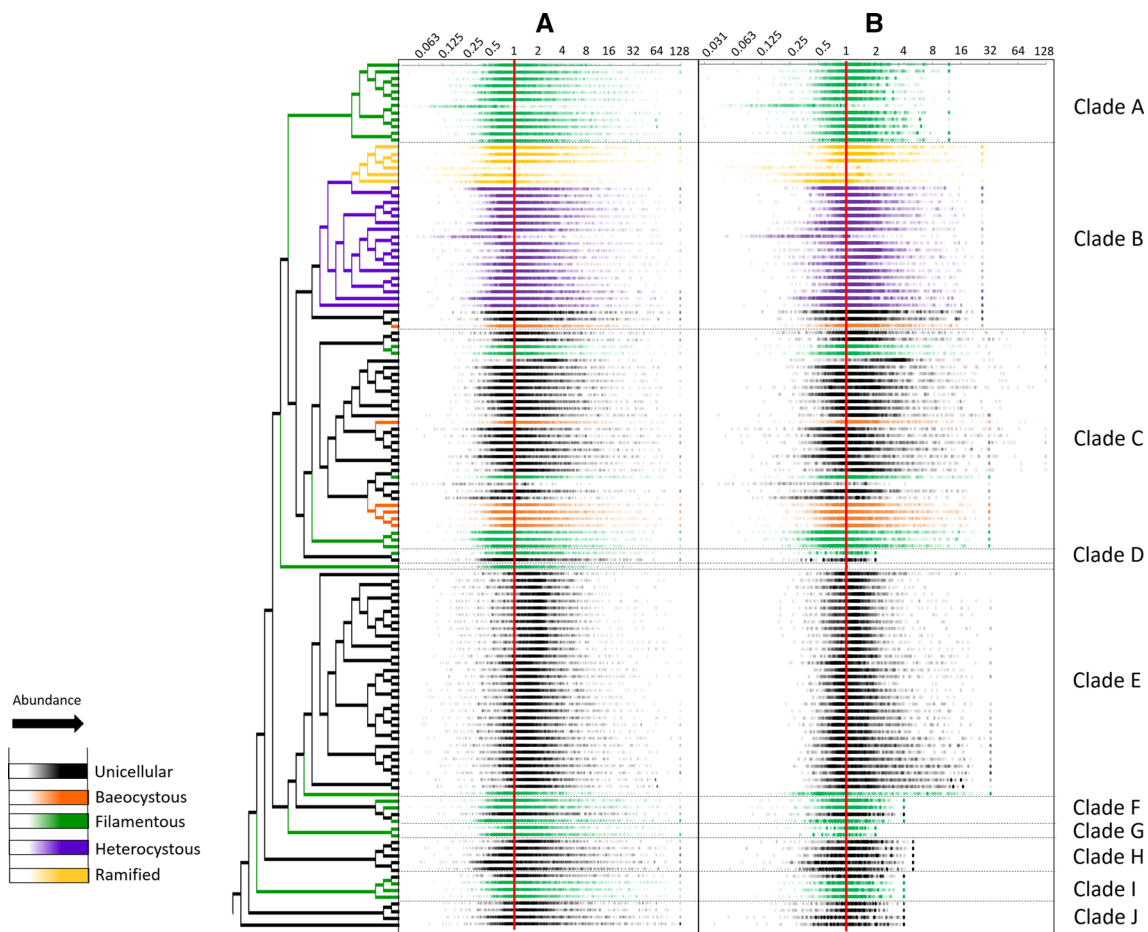
Fig. 6 Pfam domain distribution in a modified phylogenetic tree from Shih et al. (2013). Pfams were divided by the overall average of each Pfam (A) or to the clade average for each Pfam (B), and each proportion corresponds to a *dot plotted* in the *graph*. *Red line* The

individual Pfam relative abundance divided by the average abundance = 1 (no over or under representation). *Color* density depicts the number of Pfams at a given abundance

to each morphological group of a given clade (Online Resource 11). The Venn diagrams (Online Resource 11) show the distribution of the 99 % conserved Pfams in each of the morphological types belonging to a clade. The renamed clades (see Fig. 8) containing two morphologies (E, J and G) show a higher Pfam conservation in the unicellular strains compared to the filamentous strains (280/222, 330/93, and 251/84, respectively). In clade C, with three different morphologies (unicellular, baeocystous, and heterocystous), the heterocystous forms have the smallest number of unique Pfams (53), and interestingly, the baeocystous forms share a large number of Pfams with both the unicellular (186) and the heterocystous members (178) (Online Resource 11). In clade D, the majority (707) of conserved Pfams belong to the filamentous forms, and 414 Pfams are conserved between the baeocystous and filamentous morphologies (Online Resource 11). A similar analysis of overall distribution of Pfams (75 % conserved) per morphological section is compiled in Online Resource
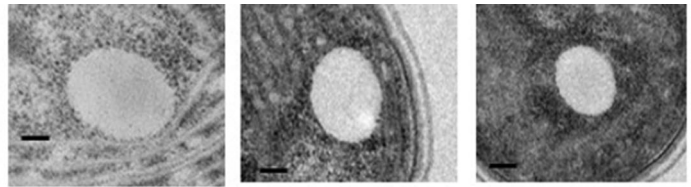
12. As expected, there is no protein signature characteristic of a morphological transition. Interestingly, in both the overall analysis and the per-clade analysis, the baeocystous group contains a larger number of abundant Pfams in common with the filamentous group relative to the unicellular group, which is consistent with the evolution of cyanobacteria from unicellular toward a higher morphological complexity.

To connect inclusions visible in the micrographs to genomic sequence data, we performed a survey of genes containing Pfam domains associated with the production of inclusions (carboxysomes, cyanophycin, glycogen and polyhydroxybutyrate granules, polyphosphate bodies, gas vesicles, and lipid droplets; Fig. 7) using the IMG JGI database (Markowitz et al. 2012; Fig. 8 and Online Resource 8). A detailed description of cyanobacterial inclusions and associated TEM images can be found in the reviews by Lang (1968), Gantt and Conti (1969), and Allen (1984).
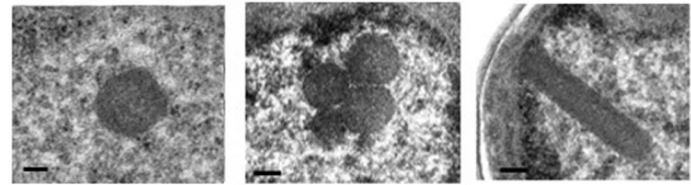
Fig. 7 Close-up of cyanobacterial inclusions showing their diversity in appearance. PHA bodies appear as electron clear inclusions as they are frequently lost during sample preparation. Carboxysomes are generally identified by their sharp edges; a "rod" carboxysome ocassionally seen in wildtype strains is shown in the right panel. Glycogen granules are small clear inclusions near thylakoids. Cyanophycin granules are generally round and can be identified by a radiating pattern (or as in this case, by modifying nutrient conditions). Polyphosphate granules are clear inclusions, distinguished from PHA bodies in this case by confirming the absence of PHA synthesis genes. Fixation: GLA. *Scale bar* 100 nm
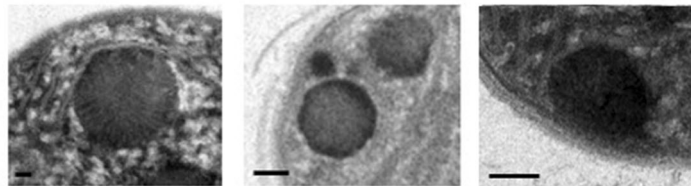
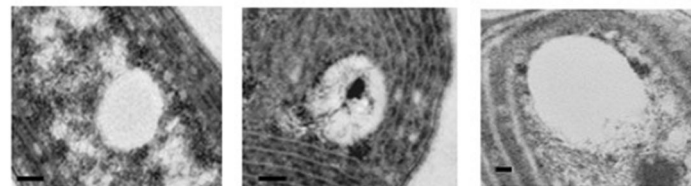Carboxysomes are polyhedral bodies that typically appear electron dense in micrographs (e.g., as in Fig. 3—PCC 7104); they allow for efficient CO_2 fixation by concentrating the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) in a protein shell (Kerfeld and Melnicki 2016). Morphologically, carboxysomes are visible only in cells that carry out CO_2 fixation—not in heterocysts (Allen 1984). Carboxysomes with a diameter of 90 nm have been reported in *Prochlorococcus* MED4 (Ting et al. 2007) and up to 600 nm in *Cyanothece* sp.

ATCC 5114 (Liberton et al. 2011). Occasionally, rod carboxysomes can be observed in wild-type cells (Fig. 7—Right panel and Online Resource 3—PCC 6802). For the bioinformatics survey, we used as the search template the Pfam03319 (the conserved domain found in the homopentamers that cap the vertices of the shell) and verified the presence of this gene within all of the genomes (with the exception of a symbiont cyanobacterium; see Thompson et al. 2012) (Fig. 8). The genomes of clade F each contain two Pfam03319-associated genes, which is the

Fig. 8 Bioinformatic survey of genes related to ultrastructure features. Phylogenetic tree modified from Shih et al. (2013). Genes counted are essential for the production of the specified inclusion and contain the specific Pfams: PHA bodies (Pfam09712), carboxysomes (Pfam03319), glycogen granules (Pfam08323 and Pfam00534), cyanophycin granules (Pfam08245, Pfam13535, and Pfam02875), polyphosphate bodies (Pfam02503), and gas vesicles (Pfam00741). Branches from the phylogenetic tree are colored according to Morphotype (*black*: unicellular; *orange*: baeocystous; *green*: filamentous; *pink*: heterocystous; *yellow*: ramified)
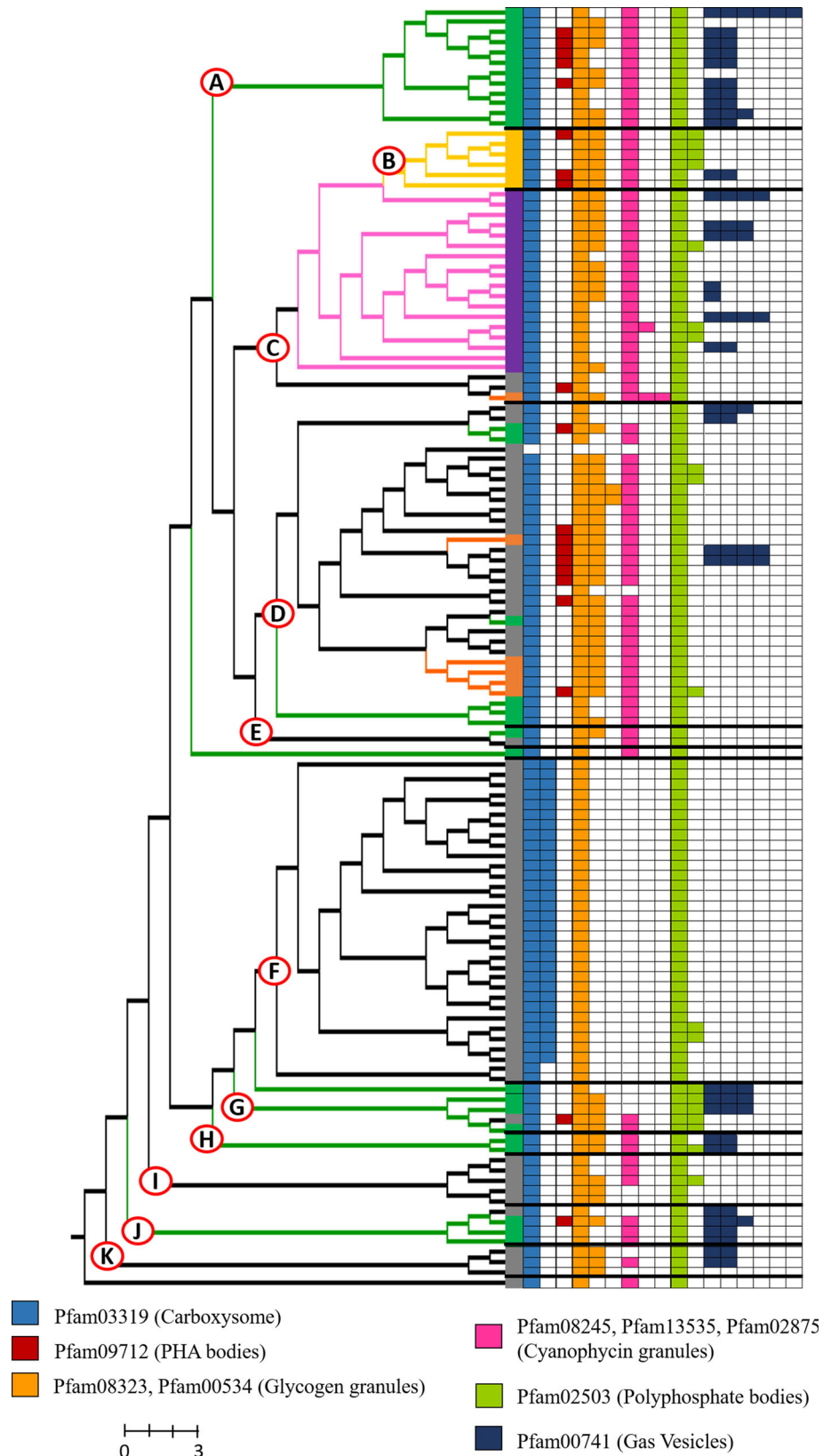
characteristic of the α-carboxysomes produced by the majority of the picocyanobacteria (Fig. 8).

Fixed carbon can accumulate in storage polymers, such as glycogen granules and polyhydroxyalkanoates (PHA) bodies. Glycogen granules are small elongated electron-clear inclusions commonly found between thylakoid membranes (i.e., Online Resource 3—PCC 7124); their length on the long axis can range between 65 and 300 nm (Jensen 1993). PHA bodies may appear as clear or lightly electron dense, round inclusions up to 800 nm (Jensen 1993) in TEM micrographs. Glycogen granules are produced by glycogen synthase, which contains the Pfam08323 and Pfam00534 domains (determined by analysis of known glycogen synthases using InterPro [Hunter et al. 2012]). These glycogen synthase homologs are conserved among the cyanobacterial strains analyzed here; the majority of strains contain two homologs except for the cyanobacterial genomes of clade F, which encode only one (Fig. 8). The two distinct homologs of glycogen synthases have been shown in *Synechocystis* sp. PCC 6803 to have different chain elongation properties (Yoo et al. 2014). Carbon and energy storage by glycogen is more common than by PHA, and genes encoding proteins that synthetize PHA bodies are less frequently observed; only 20 out of 126 genomes possess a Pfam domain associated with the poly (3-hydroxyalkanoate) synthase *phaE* (Pfam09712) and with no obvious pattern of clade distribution. Thus, it appears that PHA bodies have evolved as an ancillary storage body, possibly conferring a competitive advantage by allowing carbon storage in an alternate inclusion.

A less understood inclusion for carbon storage is the lipid droplet, frequently observed in micrographs from cultures at stationary phase (Peramuna and Summers 2014). Lipid droplets appear as small round electron-dense inclusions (e.g., see Fig. 5—PCC10914). We searched for proteins containing Pfam01277 (oleosin) and Pfam05042 (Caleosin). These Pfams are found in the proteins from plant lipid droplets (Chapman et al. 2012); however, we could not detect an analogous domain signature for cyanobacterial lipid droplets.

Polyphosphate bodies are phosphorus and energy storage inclusions synthesized by polyphosphate kinases (containing Pfam02503, Pfam13089, and Pfam13090). In micrographs, these granules may appear electron clear (Fig. 5—PCC 9431) or as round dark inclusions (Fig. 4—PCC 7107), depending on the sample processing (Allen 1984). Polyphosphate granules range from 15 nm to several micrometers in diameter (Jensen 1993) and are frequently found in proximity to carboxysomes. The appearance of a polyphosphate body may be similar to a PHA body and so they may be misidentified (Tsang et al. 2013). The data shown in Fig. 8 and Online Resource 8 provide a benchmark

for interpretation; for example, the round clear inclusion in Online Resource 3—PCC 9333—is likely a polyphosphate body, as this strain does not contain proteins with Pfam09712 domains (PHA synthesis). However, in micrographs such as in Fig. 5—PCC 10914, which shows a cell full of clear bodies, an alternative identification method is required. Notably, the gene for polyphosphate body synthesis is the only fully conserved ultrastructure-related gene in all clades. It appears overrepresented in clades B and G (Fig. 8); the conservation of genes for polyphosphate synthesis is perhaps related to the frequency with which phosphate is a limiting nutrient in the aquatic environments inhabited by cyanobacteria (Tyrrell 1999).

Cyanophycin granules consist of a nitrogen-storage polymer. They appear as electron-dense inclusions (i.e., Fig. 5—PCC 10914) (500 nm diameter and larger granules have been reported; see Lang et al. 1968), typically in radiating patterns (as in Fig. 7—Left panel) (Lang et al. 1972). They are produced by proteins that contain Pfam08245, Pfam13535, and Pfam02875 domains. Even though it has been suggested that nitrogen storage as cyanophycin plays a larger role in nitrogen-fixing strains in contrast to its storage in phycobilisomes (proposed to be the main nitrogen reservoir in non-N_2-fixing strains) (Li et al. 2001), the genes for cyanophycin synthases are widespread throughout the cyanobacterial genomes; they are notably absent among the marine picocyanobacteria in clade F (Fig. 8 and Online Resource 8).

A less widespread ultrastructural feature is the gas vesicle (Online Resource 3—PCC 7367). Gas vesicles are cylindrical inclusions that are used for buoyancy, with a diameter of about 45—117 nm and length from 100 to more than 800 nm (Walsby 1994). We searched for proteins containing Pfam00741, the Pfam found in the GvpA protein (the main protein component of the gas vesicle shell). Proteins with Pfam00741 were overrepresented among the filamentous strains, especially in clades A, G, H, and J (Fig. 8). Although the number of proteins containing Pfam00741 domains is highly abundant in strains of the filamentous Morphotype III, this is in many cases due to the existence of other less-conserved protein subunits of the gas vesicle (GvpJ, GvpK) that are fusions to the Pfam00741 domain.

Summary

With the genomic information gathered CyanoGEBA dataset (Shih et al. 2013) and with the compilation of metadata and electron micrographs from PCC strains, we have generated a compilation of cyanobacterial ultrastructure data that should be useful for the study of genotype and phenotype in a variety of experimental contexts.

References

Allen MM (1984) Cyanobacterial cell inclusions. Annu Rev Microbiol 38:1–25. doi:10.1146/annurev.mi.38.100184.000245

Beck C, Knoop H, Axmann I, Steuer R (2012) The diversity of cyanobacterial metabolism: genome analysis of multiple phototrophic microorganisms. BMC Genom 13:56. doi:10.1186/1471-2164-13-56

Benzerara K et al (2014) Intracellular Ca-carbonate biomineralization is widespread in cyanobacteria. Proc Natl Acad Sci 111:10933–10938. doi:10.1073/pnas.1403510111

Canfield DE (2005) The early history of atmospheric oxygen: Homage to Robert M. Garrels. Annu Rev Earth Planet Sci 33:1–36. doi:10.1146/annurev.earth.33.092203.122711

Chapman KD, Dyer JD, Mullen RT (2012) Biogenesis and functions of lipid droplets in plants. J Lipid Res 53:215–226. doi:10.1194/jlr.R021436

Dittmann E, Gugger M, Sivonen K, Fewer DP (2015) Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. Trends Microbiol 23:642–652. doi:10.1016/j.tim.2015.07.008

Finn RD et al (2014) Pfam: the protein families database. Nucleic Acids Res 42:D222–D230. doi:10.1093/nar/gkt1223

Gantt E, Conti SF (1969) Ultrastructure of blue-green algae. J Bacteriol 97:1486–1493

Hess WR (2011) Cyanobacterial genomics for ecology and biotechnology. Curr Opin Microbiol 14:608–614. doi:10.1016/j.mib.2011.07.024

Hoffmann L, Komárek J, Kaŝtovský J (2005) System of cyanoprokaryotes (cyanobacteria)-state in 2004. Algol Stud 117:95–115. doi:10.1127/1864-1318/2005/0117-0095

Hunter S et al (2012) InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res 40:D306–D312. doi:10.1093/nar/gkr948

Jensen TE (1984) Cell inclusions in the cyanobacteria. Ann N Y Acad Sci 435:279–282. doi:10.1111/j.1749-6632.1984.tb13793.x

Jensen TE (1993) Cyanobacterial ultrastructure. In: Berne T (ed) Ultrastructure of microalgae, 3rd edn. CRC Press, Florida, pp 8–51

Kaneko T et al (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res 3:109–136. doi:10.1093/dnares/3.3.109

Kaneko T et al (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. DNA Res 8:205–213. doi:10.1093/dnares/8.5.205

Kellenberger E, Ryter A, Sechaud J (1958) Electron microscope study of DNA-containing plasms. II. Vegetative and mature phage DNA as compared with normal bacterial nucleoids in different physiological states. J Biophys Biochem Cytol 4(6):671–678

Kerfeld CA, Melnicki MR (2016) Assembly, function and evolution of cyanobacterial carboxysomes. Curr Opin Plant Biol 31:66–75

Komárek J, Kaŝtovský J (2003) Coincidences of structural and molecular characters in evolutionary lines of cyanobacteria. Algol Stud 109:305–325

Komárek J, Kaŝtovský J, Mareš J, Johansen JR (2014) Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. Preslia 86:295–335

Lang NJ (1968) The fine structure of blue-green algae. Annu Rev Microbiol 22:15–46. doi:10.1146/annurev.mi.22.100168.000311

Lang N, Simon R, Wolk CP (1972) Correspondence of cyanophycin granules with structured granules in *Anabaena cylindrica*. Arch Mikrobiol 83:313–320. doi:10.1007/BF00425243

Li H, Sherman D, Bao S, Sherman L (2001) Pattern of cyanophycin accumulation in nitrogen-fixing and non-nitrogen-fixing cyanobacteria. Arch Microbiol 176:9–18. doi:10.1007/s002030100281

Liberton M, Austin JR, Berg RH, Pakrasi HB (2011) Unique thylakoid membrane architecture of a unicellular N_2-fixing cyanobacterium revealed by electron tomography. Plant Physiol 155:1656–1666. doi:10.1104/pp.110.165332

Machado IMP, Atsumi S (2012) Cyanobacterial biofuel production. J Biotechnol 162:50–56. doi:10.1016/j.jbiotec.2012.03.005

Markowitz VM et al (2012) IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res 40:D115–D122. doi:10.1093/nar/gkr1044

Peramuna A, Summers ML (2014) Composition and occurrence of lipid droplets in the cyanobacterium *Nostoc punctiforme*. Arch Microbiol 196:881–890. doi:10.1007/s00203-014-1027-6

Price GD et al (2013) The cyanobacterial CCM as a source of genes for improving photosynthetic CO_2 fixation in crop species. J Exp Bot 64:753–768. doi:10.1093/jxb/ers257

Raven J, Allen J (2003) Genomics and chloroplast evolution: what did cyanobacteria do for plants? Genome Biol 4:209

Rippka R, Waterbury J, Cohen-Bazire G (1974) A cyanobacterium which lacks thylakoids. Arch Microbiol 100:419–436. doi:10.1007/BF00446333

Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J Gen Microbiol 111:1–61

Ris H, Singh RN (1961) Electron microscope studies on blue-green algae. J Biophys Biochem Cytol 9:63–80

Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB (2010) Microbial biosynthesis of alkanes. Science 329:559–562. doi:10.1126/science.1187936

Schirrmeister BE, Antonelli A, Bagheri HC (2011) The origin of multicellularity in cyanobacteria. BMC Evol Biol 11:45–45. doi:10.1186/1471-2148-11-45

Schirrmeister BE, Gugger M, Donoghue PCJ (2015) Cyanobacteria and the great oxidation event: evidence from genes and fossils. Palaeontology 58:769–785. doi:10.1111/pala.12178

Shenoy SR, O'Keefe BR, Bolmstedt AJ, Cartner LK, Boyd MR (2001) Selective interactions of the human immunodeficiency virus-inactivating protein cyanovirin-N with high-mannose oligosaccharides on gp120 and other glycoproteins. J Pharmacol Exp Ther 297:704–710

Shih PM et al (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc Natl Acad Sci 110:1053–1058. doi:10.1073/pnas.1217107110

Stanier G (1988) Fine structure of cyanobacteria. In: LesterPacker ANG (ed) Methods in enzymology, vol 167. Academic Press, Cambridge, pp 157–172. doi:10.1016/0076-6879(88)67017-0

Stanier R, Bazine G (1977) Phototrophic prokaryotes: the cyanobacteria. Annu Rev Microbiol 31:225–274

Thompson AW et al (2012) Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. Science 337:1546–1550. doi:10.1126/science.1222700

Ting CS, Hsieh C, Sundararaman S, Mannella C, Marko M (2007) Cryo-electron tomography reveals the comparative three-dimensional architecture of *Prochlorococcus*, a globally important

marine cyanobacterium. J Bacteriol 189:4485–4493. doi:10. 1128/jb.01948-06

Tsang T, Roberson R, Vermaas WJ (2013) Polyhydroxybutyrate particles in *Synechocystis* sp. PCC 6803: facts and fiction. Photosynth Res 118:37–49. doi:10.1007/s11120-013-9923-1

Tyrrell T (1999) The relative influences of nitrogen and phosphorus on oceanic primary production. Nature 400:525–531. doi:10. 1038/22941

Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA (2004) Structure, function and evolution of multidomain proteins. Curr Opin Struct Biol 14:208–216. doi:10.1016/j.sbi.2004.03.011

Walsby AE (1994) Gas vesicles. Microbiol Rev 58:94–144

Yoo S-H, Lee B-H, Moon Y, Spalding MH, J-l Jane (2014) Glycogen synthase isoforms in *Synechocystis* sp. PCC6803: identification of different roles to produce glycogen by targeted mutagenesis. PLoS ONE 9:e91524. doi:10.1371/journal.pone.0091524

Zarzycki J, Axen SD, Kinney JN, Kerfeld CA (2013) Cyanobacterial-based approaches to improving photosynthesis in plants. J Exp Bot 64:787–798. doi:10.1093/jxb/ers294