Revisiting the early evolution of Cyanobacteria with a new thylakoid-less and deeply diverged isolate from a hornwort

Nasim Rahmatpour¹, Duncan A. Hauser¹, Jessica M. Nelson¹, Pa Yu Chen², Juan Carlos Villarreal A.^{3,4}, Ming-Yang Ho^{2,5,*}, Fay-Wei Li^{1,6,*}

¹Boyce Thompson Institute, Ithaca, New York, USA

²Department of Life Science, National Taiwan University, Taipei, Taiwan

³Department of Biology, Laval University, Quebec City, Quebec, Canada

⁴Smithsonian Tropical Research Institute, Panama City, Panama

⁵Institute of Plant Biology, National Taiwan University, Taipei, Taiwan

⁶Plant Biology Section, Cornell University, Ithaca, New York, USA

*Authors of correspondence: Fay-Wei Li (fl329@cornell.edu) and Ming-Yang Ho (mingyang@ntu.edu.tw)

Summary

Cyanobacteria have played pivotal roles in Earth's geological history especially during the rise of atmospheric oxygen. However, our ability to infer the early transitions in Cyanobacteria evolution has been limited by their extremely lopsided tree of life—the vast majority of extant diversity belongs to Phycobacteria (or “crown Cyanobacteria”), while its sister lineage, Gloeobacteria, is depauperate and contains only two closely related species of *Gloeobacter* and a metagenome-assembled genome. Here we describe a new culturable member of Gloeobacteria, *Anthocero bacter panamensis*, isolated from a tropical hornwort. *Anthocero bacter* diverged from *Gloeobacter* over 1.4 billion years ago and has low 16S identities with environmental samples. Our ultrastructural, physiological, and genomic analyses revealed that this species possesses a unique combination of traits that are exclusively shared with either Gloeobacteria or Phycobacteria. For example, similar to *Gloeobacter*, it lacks thylakoids and circadian clock genes, but the carotenoid biosynthesis pathway is typical of Phycobacteria. Furthermore, *Anthocero bacter* has one of the most reduced gene sets for photosystems and phycobilisomes among Cyanobacteria. Despite this, *Anthocero bacter* is capable of oxygenic photosynthesis under a wide range of light intensities, albeit with much less efficiency. Given its key phylogenetic position, distinct trait combination, and availability as a culture, *Anthocero bacter* opens a new window to further illuminate the dawn of oxygenic photosynthesis.

Keywords

Anthocero bacter panamensis, *Candidatus Aurora*, Cyanobacteria, *Gloeobacter*, photosynthesis

Introduction

The rise of atmospheric oxygen is undoubtedly one of the most transformative events in Earth's history. During the Paleoproterozoic era, oxygenic photosynthesis carried out by Cyanobacteria fueled the Great Oxygenation Event, which altered the biogeochemical cycles and fundamentally shifted the evolutionary trajectories of life on Earth^{1, 2}. None of the closest living relatives of Cyanobacteria, Vampirovibrionia and Sericytochromatia³, are capable of oxygenic photosynthesis nor carry any intermediate photosynthetic machinery. Therefore, to understand how photosynthesis evolved over time, it is imperative to identify and examine the lineages that diverged the deepest within the Cyanobacteria tree of life.

The Phylum Cyanobacteria (defined by Garcia-Pichel et al.⁴) is composed of two extant groups: Gloeobacteria and Phycobacteria that diverged around 2 billion years ago (BYA)^{5, 6}. Phycobacteria encompasses >99.9% of the known cyanobacterial diversity and is sometimes also referred to as the "crown Cyanobacteria". Gloeobacteria, on the other hand, is rather enigmatic and has only two species described thus far: *Gloeobacter violaceus* and *G. kilaueensis*. Little is known about the ecology and distribution of these two species; *G. kilaueensis* is only known from a lava cave in Hawaii⁷, and *G. violaceus* was originally isolated from a limestone rock in the Swiss alps⁸ and subsequently reported on waterfall walls in Europe and Mexico^{9, 10}. The diversity of Gloeobacteria is likely much greater than these two species, as evidenced by a few distinct 16S clades from environmental samples¹¹ and a recent metagenome-assembled genome (MAG), *Candidatus Aurora vandensis*, from Lake Vanda, Antarctica¹².

Gloeobacteria is a pivotal lineage to reconstruct the early evolution of Cyanobacteria, given their many distinct, and presumably plesiomorphic traits. In contrast to Phycobacteria, *Gloeobacter* lacks thylakoid membranes as well as a circadian clock^{8, 13}. Due to the absence of thylakoids, the photosynthesis and respiratory apparatus are both located on the cytoplasmic membrane¹⁴. The *Gloeobacter* photosystems are also atypical^{15–19}, and their phycobilisomes are bundle-shaped instead of hemidiscoidal as found in Phycobacteria^{20, 21}. In addition, the carotenoid biosynthesis pathway in *Gloeobacter* is of "bacterial-type"^{22, 23}, which is again different from other photosynthetic organisms. To date, a large body of research has been done on *Gloeobacter* and has significantly advanced our knowledge on Cyanobacteria biology and evolution.

However, the depauperate nature of Gloeobacteria makes it difficult to ascertain whether the idiosyncrasies of *Gloeobacter* can properly reflect the ancestral states or are instead a product of lineage-specific reductive evolution. The recent discovery of *Candidatus Aurora* MAG is an important step toward capturing the broader genomic diversity of Gloeobacteria¹², although the fragmented nature of MAG, as well as its absence as a culturable organism, prohibit further in-depth investigations.

Here we report a new culturable Gloeobacteria species that is distantly related to *Gloeobacter*. Its complete circular genome, ultrastructure, and basic physiological properties provided new insights into the early evolution of Cyanobacteria and associated oxygenic photosynthesis.

Figure 1. Morphology of *Anthocero bacter panamensis*. (A) Blue-green colonies on a solid agar plate. (B) Rod-shaped cells under light microscope with polyphosphate granules (P) often visible toward the cell poles. (C-E) TEM images of *Anthocero bacter* in longitudinal (C) and transverse (D-E) sections. No thylakoid is present. P: polyphosphate granule. C: carboxysome. CM: cytoplasmic membrane. CL: electron-dense cytoplasmic layer.

Results and Discussion

Isolation of Anthocero bacter panamensis from a tropical hornwort

A new cyanobacterial culture was obtained from a surface-sterilized thallus of the hornwort (Bryophyta) *Leiosporoceros dussii* from Panama (Fig. 1). Our 16S phylogeny places this isolate (named *Anthocero bacter panamensis*) sister to a clade of uncultured environmental samples from the Arctic and Antarctic regions (e.g. Canadian tundra, Iceland, Antarctic lakes, and Patagonia) (Fig. 2A). This “polar clade” also includes *Candidatus Aurora vandensis* MAGs from Antarctica. However, *Anthocero bacter* is the only culturable strain in this clade. The 16S sequence identity between *Anthocero bacter* is around 96%, 88% and 70% when compared to polar clade members, *Gloeobacter*, and *Synechocystis* sp. PCC 6803, respectively (Fig. 2D). The equatorial origin of *Anthocero bacter* is in stark contrast to its sister polar clade, which speaks to the uniqueness of this new isolate.

Because all hornwort species have symbiotic associations with filamentous nitrogen-fixing Cyanobacteria^{24, 25}, it raises the question of whether *Anthocero bacter* could be a major symbiont. We suggest this is unlikely to be the case as *Anthocero bacter* is not filamentous (Fig. 1B), and its genome lacks nitrogenase genes while nitrogen fixation is the selective advantage in plant-Cyanobacteria symbiosis. Recently, the microbiome of *L. dussii* was profiled using 16S amplicon sequencing²⁶, but no trace of *Anthocero bacter* could be found (though interestingly, a unique clade of *Gloeobacter* was uncovered therein: ASV172, ASV719, and ASV661 in Fig. 2A). We therefore speculate that *Anthocero bacter* could be a soil bacterium that happened to survive our surface sterilization, an endophyte that inhabits hornwort intercellular space, or a freeloader inside symbiotic cavities with other functional nitrogen-fixing cyanobionts. Future studies are needed to clarify the ecology and natural history of *Anthocero bacter*.

***Anthocero bacter* represents a novel and deeply branched cyanobacterial lineage**

To corroborate the phylogenetic placement of *Anthocero bacter* and to estimate the timing of divergence, we compiled a 12-gene matrix based on the dataset of Shih *et al.*⁶. These genes are slow-evolving, and are shared between plastid and mitochondria genomes, thereby allowing for cross-validations to reuse and link the same calibration across multiple nodes. As in the 16S phylogeny, we recovered the same branching order among *Anthocero bacter*, *Candidatus Aurora*, *Gloeobacter*, and Phycobacteria with strong support (Fig. 2B). It should be noted that our topology is different from the 16S tree of Grett enberger *et al.*¹², which resolved *Candidatus Aurora* sister to *Gloeobacter* + Phycobacteria, but is consistent with their 37-gene phylogeny. Given the high branch support from both of our datasets, we believe our result, (((*Anthocero bacter*, *Candidatus Aurora*), *Gloeobacter*), Phycobacteria), represents the most plausible relationship.

The chronogram from BEAST suggested that *Anthocero bacter* diverged from *Candidatus Aurora* at around 374.74 million years ago (MYA) (95% highest posterior density (HPD): 155.86–694.79 MYA), and from *Gloeobacter* around 1.45 BYA (95% HPD: 857.64–1,975.35 MYA) (Fig. 2B). Because molecular dating analyses can be strongly influenced by calibrations, we further explored the robustness of our time estimates by adopting a suite of different calibration schemes, models, and data matrices made by Baracaldo *et al.*²⁷. Comparable but generally deeper divergences were obtained, and the results are summarized in Fig. 2C. Taken together, our phylogenetic and molecular clock analyses indicate that *Anthocero bacter* is currently the sole culturable representative of a novel cyanobacterial lineage, which split from *Gloeobacter* over 1.4 billion years ago.

***Anthocero bacter* genome**

Nanopore sequencing resulted in one circular genome assembly of 4,187,797bp and a GC content of 55.51%. The Average Amino-acid Identity (AAI) is around 57% when compared to *Gloeobacter* and 69% to *Candidatus Aurora* (Fig. 2D). A total of 4,166 protein-coding genes were predicted, 1,805 of which can be functionally annotated. Protein family circumscription based on 103 cyanobacterial species identified a total of 18,419 orthogroups. Of these, 2,818 gene families are present in *Anthocero bacter*, 35 are species-specific, and 140 are unique to *Anthocero bacter* + *Candidatus Aurora* (Table S1; Fig. S1). A total of 454 predicted genes in *Anthocero bacter* were singletons that could not be assigned to any orthogroup. Furthermore, we found very little synteny between *Anthocero bacter* and *Gloeobacter spp.*, consistent with the deep divergence between the two genera (Fig. 2E).

***Anthocero bacter* lacks thylakoids**

One of the most distinctive features of *Gloeobacter* is its lack of thylakoids^{8,9}. While the ultrastructure of *Candidatus Aurora* is not available, analysis of its metagenome found a lack of the *VIPP1* (vesicle inducing protein in plastid 1) gene¹², which is essential for biogenesis of thylakoid membrane²⁸. The availability of *Anthocero bacter* as a pure culture allowed us to directly examine it using transmission electron microscopy (TEM). We found no trace of thylakoids (Fig. 1C-E), nor a *VIPP1* homolog in the genome. This result reaffirmed that the lack of thylakoid is likely a general feature of *Gloeobacteria*, and thylakoids were an innovation of Phycobacteria.

Spectroscopy reveals unique features of *Anthocero*

In a stark contrast to the purple to brown color of *Gloeobacter*^{7,9}, *Anthocero* are blue-green and has no apparent red hue (Fig. 1A). To better understand the pigment composition, we characterized and compared *Anthocero*'s absorption and fluorescence emission spectra with those of *Gloeobacter violaceus*, and *Synechocystis* sp. PCC 6803 (a model unicellular species from Phycobacteria; hereafter Syn6803). In the whole cell absorption spectra (Fig. 3A), *Anthocero* lacks clear peaks representing phycoerythrin (PE), which is different from *G. violaceus*²⁹ but consistent with the absence of PE genes in the *Anthocero* genome (see below). Compared to Syn6803, which shows predominant Soret and Q_y bands of chlorophyll (Chl) *a* (440 nm and 682 nm), the absorption at 682 nm in *Anthocero* is lower, not showing a pronounced peak (Fig. 3A). The peak at 635 nm from phycocyanin (PC) is higher in *Anthocero* than in Syn6803 at 629 nm, though it is common that the absorption peaks of the same phycobiliproteins differ from species to species^{21, 30, 31}. The absorption feature at 495 nm is from carotenoids. Because *Anthocero* has no PE nor phycoerythrocyanin (PEC) identified in the genome, the 582 nm feature, which is absent in Syn6803, is from an unknown source, possibly from carotenoid-binding proteins. In *G. violaceus*, this wavelength region is overlaid with the absorption of PE in the whole cell absorption spectra and was therefore not discussed previously²⁹.

The absorption spectra from methanol-extracted pigments (Fig. 3B) shows that *Anthocero* contain only Chl *a* (665 nm and 436 nm) and no other chlorophylls, and the Q_y band (665 nm) matches perfectly with the Q_y band from Chl *a* in Syn6803. However, although both strains show features of carotenoids at 475 nm, *Anthocero* have stronger absorption at 436 nm and an additional absorption feature around 530 nm, indicating it has one or more carotenoids with long-wavelength absorption. A long-wavelength absorbing carotenoid was also identified in *G. violaceus*, and its absorption maximum is similar to oscilloxanthin^{14, 29}. At high-light condition (100 μmol photons m⁻²s⁻¹), *Anthocero* produced more carotenoids (absorption peaks at wavelengths 443 nm, 474 nm, and 530 nm) per unit of Chl *a* (Fig. 3C and 5B). Further analyses are required to characterize the species and quantities of carotenoids in *Anthocero*.

Figure 3. Spectral analyses of *Anthocero* in comparison with other Cyanobacteria. Absorption spectra of (A) whole cells, (B) pigment extraction in methanol, and (C) pigment extraction in methanol of *Anthocero* in high and low light (100 and 10 μmol photons m⁻²s⁻¹) (D) Fluorescence spectroscopy was taken at 77K with excitation wavelength at 440 nm using whole cells of *Anthocero*, Syn6803, and *Chlorogloeopsis fritschii* PCC 9212 (Chloro9212). The absorbance spectra were normalized at 750 nm in (A) and at 665 nm in (B) and (C). Specific wavelengths were labeled above the peaks.

Finally, 77K fluorescence spectroscopy was used to measure fluorescence emission from chromophore-bound protein complexes such as phycobilisomes, photosystem I (PSI) and II (PSII). In *Anthocero bacter*, phycobiliproteins are the predominant light-harvesting complexes, resulting fluorescence emission largely coming from phycobiliproteins even though the excitation wavelength (at 440 nm) is targeted for chlorophylls. Three emission peaks are observed in *Anthocero bacter* and assigned as PC (646 nm), AP (664 nm), and PSII (684 nm) (Fig. 3D). The peak at 684 nm is asymmetric, indicating that the second peak of PSII is likely present but not strong enough forming a separate peak, a pattern also found in *G. violaceus*^{15, 29}. The long-wavelength emission from PSI is absent in *Anthocero bacter*, which is different from Syn6803 (and other Phycobacteria), but similar to *G. violaceus* (Figure 3D)^{15, 29}. Previous studies in *G. violaceus* concluded that the absence of long-wavelength emission beyond 700 nm is due to the lack of long-wavelength Chl *a* in PSI²⁹, which is possibly the situation in *Anthocero bacter* as well. These measurements indicate that the spectral properties of *Anthocero bacter* are distinct from Syn6803 but largely similar to *G. violaceus* with the exception that *Anthocero bacter* lacks PE.

***Anthocero bacter* lacks cyanobacterial circadian clock genes**

Circadian clocks enable organisms to adjust their physiological activities according to extrinsic daily changes such as light, temperature, and humidity. In Cyanobacteria, environmental signals are transmitted through redox sensitive proteins to core oscillators (KaiA, KaiB and KaiC), followed by output pathways to regulate physiological activities³². We found that the *Anthocero bacter* genome is missing the core oscillator KaiABC genes, similar to what was reported in *Gloeobacter*¹³. This implies that the cyanobacterial circadian clock most likely evolved in the common ancestor of Phycobacteria.

The carotenoid biosynthesis pathways of Anthocero bacter and Gloeobacter differ

Carotenoids are an integral part of light harvesting complexes and play key roles in photoprotection in conjunction with orange carotenoid proteins (OCP). While *Anthocero bacter* OCP is in the OCPX clade (defined by Bao *et al.*³³) like *Gloeobacter* (Fig. S2, Table S2), the pathway through which carotenoids are made differs between the two lineages. The first step of carotenoid biosynthesis is the condensation of two geranylgeranyl pyrophosphate into phytoene by phytoene synthase (CrtB)(Fig. S3). Following this, phytoene is converted to lycopene through two different routes depending on the organisms. The “plant-type” pathway, which can be found in all Cyanobacteria (except *Gloeobacter*) and plants, relies on three separate enzymes: phytoene desaturase (CrtP), ζ -carotene desaturase (CrtQ), and cis-carotene isomerase (CrtH)³⁴. By contrast, the “bacteria-type”, found in *Gloeobacter*, anoxygenic photosynthetic bacteria, and fungi, uses only a single phytoene desaturase (CrtI)^{22, 23}. Interestingly, we found that both *Anthocero bacter* and *Candidatus Aurora* have the genetic chassis for the plant-type carotenoid biosynthesis, instead of the bacteria-type reported in *Gloeobacter* (Fig. S3). This implies that the origin of the plant-type carotenoid biosynthesis pathway should be pushed back to the ancestor of all Cyanobacteria, with *Gloeobacter* later substituting it with the bacterial pathway. While we cannot entirely rule out that *Anthocero bacter* horizontally acquired the plant-type pathway from Phycobacteria, the phylogenetic positions of *Anthocero bacter crtP*, *crtQ*, and *crtH* are consistent with the species phylogeny (Fig. S4 and S5).

Anthoceroobacter has a unique phycobilisome composition

Cyanobacteria use phycobilisomes as the light harvesting antennae to capture and transfer light energy to photosystems. Typically, phycobilisomes are fan-shaped and composed of the allophycocyanin core (AP) and several peripheral proteins: phycocyanin (PC) and phycoerythrin (PE) or phycoerythrocyanin (PEC)^{21, 30}. One notable exception is *Gloeobacter*, whose phycobilisomes are bundle-shaped and consist of AP, PC, and PE^{20, 21}. Compared to *Gloeobacter*, *Anthoceroobacter* further lacks homologs for PE (and PEC), a situation also reported from *Candidatus Aurora* MAG (Fig. 4)¹². On the other hand, most of the AP and PC subunits are present in *Anthoceroobacter*, which is consistent with our absorption and fluorescence emission spectral data (Fig. 3A and D). The missing AP and PC subunits are ApcD, ApcF, and CpcD, although the mutations of which did not appear to affect the phycobilisome function in *Synechococcus* sp. PCC 7002^{35, 36} or in Syn6803^{37, 38}.

Another peculiar feature of *Anthoceroobacter* phycobilisomes is the presence of *cpcG*, which encodes for a rod-core linker protein. This protein is critical in phycobilisome assembly and energy transfer from PC to AP but is absent in *Gloeobacter* and *Candidatus Aurora*. In *Gloeobacter*, a different linker protein, CpcJ, was used instead of CpcG³⁹. Interestingly, this *Gloeobacter*-specific *cpcJ* gene is also present in *Anthoceroobacter*, suggesting that *Anthoceroobacter* has two different types of rod-core linker (CpcG and CpcJ) for connecting PC and AP (Table S2). In other words, *Anthoceroobacter* may have a considerably different phycobilisome from that of *Gloeobacter*, which is already distinct among Cyanobacteria. Furthermore, our phylogenetic analysis of phycobilisome linker proteins placed *Anthoceroobacter cpcG* sister to the

Figure 4. Comparison of phycobilisome and photosystem subunits among *Anthoceroobacter*, *Candidatus Aurora*, *Gloeobacter*, and other Cyanobacteria. Filled and empty boxes indicate gene presence and absence respectively. *Candidatus Aurora* “green” and “purple” are MAG from separate sediment layers. **psbA* was not found in Aurora MAG but on separate contigs that could not be properly binned.

rest of cyanobacterial homologs (Fig. S6), suggesting that the ancestor of Cyanobacteria very likely had the CpcG linker.

Anthocero bacter has a reduced set of photosystem subunits

With respect to photosystems, *Anthocero bacter* is similar to *Gloeobacter* in that they both lack several highly conserved genes encoding PsbY, PsbZ, and Psb27 at photosystem II (PSII), and PsaI, PsaJ, PsaK, PsaX at photosystem I (PSI)(Fig. 4). These subunits are involved in electron transfer (e.g. PsbZ)⁴⁰, or the formation and stabilization of photosystems (e.g. Psb27, PsaI, and PsaJ)^{41–45}. *Anthocero bacter* also lacks the known pathway genes for synthesizing phyloquinone/menaquinone (KEGG M00116), which is a secondary electron acceptor of PSI; such absence is also shared with *Gloeobacter*¹⁸.

Interestingly, we found even more genes encoding photosystem subunits are missing in *Anthocero bacter* (and to some extent *Candidatus Aurora* as well) compared to *Gloeobacter*, including PsbM, PsbU, PetM, and PetN. PsbM is localized in the PSII center and plays a key role in PSII dimer formation and stability⁴⁶. PsbU is needed for making a thermally stable oxygen-evolving complex, and aids electron transport and energy transfer from phycobilisome to PSII^{47, 48}. At the electron transport chain, PetN is necessary for the proper function and assembly of the cytochrome b₆f complex at least in Syn6803 and tobacco⁴⁹. While PetM mutations in Syn6803 did not impair the cytochrome b₆f function, the amounts of phycobilisomes and PSI decreased⁵⁰. In all, *Anthocero bacter* has one of the most reduced sets of photosynthesis subunits among Cyanobacteria, which prompted us to further investigate the physiological properties of *Anthocero bacter* and compare them with existing cyanobacterial strains.

Anthocero bacter is able to grow photoautotrophically and evolve oxygen, but slowly

The growth of *Anthocero bacter* was measured under low light (LL) and high light (HL) conditions (10 and 100 $\mu\text{mol photons m}^{-2}\text{s}^{-1}$, respectively) in the medium without supplement of carbon sources. *Anthocero bacter* can grow photoautotrophically under both light intensities, with a faster rate under HL (Fig. 5A). The pigment analyses show that under HL, the Chl *a*:phycobiliprotein ratio is about half in *Anthocero bacter* as in Syn6803 (Fig. 5B-C). *Anthocero bacter* produces less Chl *a* and phycobiliproteins but more carotenoids in HL than in LL, probably because photoprotection is turned on. A similar change from lower to higher light intensity was also observed in *Gloeobacter*¹⁵.

Compared to other model Cyanobacteria, *Anthocero bacter* is very slow growing; its doubling time is estimated around 67 h under HL and over 8 days under LL (Fig. 5A). By comparison, Syn6803 doubles in around 8 hours under the same HL condition. The reported doubling time for *Gloeobacter* ranges from 73 h to 17 days^{14, 51}, depending on the culturing conditions. The slow growth of *Anthocero bacter* likely results from its inefficient photosynthetic apparatus and/or the inability to tolerate high light intensities, which could be caused by the fact that they lack accessory subunits of PSI and PSII (Fig. 4).

Finally, we demonstrated that *Anthocero bacter* does produce oxygen, and the level of which increased with light intensity (from 100, 500, to 2000 $\mu\text{mol photons m}^{-2}\text{s}^{-1}$) (Fig. 5D). In general, *Anthocero bacter* has about half of the oxygen evolution rates compared to Syn6803 under the same conditions. While *Gloeobacter* was not included in our experiments, we could cross-reference the results with those of Koyama *et al.*⁵², which reported a similar difference in oxygen evolution rates between Syn6803 and *G.*

violaceus. In other words, despite having even fewer photosynthesis subunits than *Gloeobacter*, *Anthocero bacter* can nevertheless carry out comparable oxygenic photosynthesis. These results in *Anthocero bacter* may contribute to determination of the minimal and essential photosynthetic subunits required for photoautotrophic growth in nature.

Figure 5. Measurements of growth, pigment analyses, and oxygen evolution. *Anthocero bacter* was grown in high light (HL, 100 μmol photon m⁻²s⁻¹) and low light (LL, 10 μmol photon m⁻²s⁻¹) conditions for 8 days. (A) The OD₇₅₀ values at the beginning of inoculation (Day 0) and the end of cultivation (Day 8) were recorded for estimation of doubling time. Quantifications of (B) Chl a and carotenoids, and (C) phycobiliproteins were performed using *Anthocero bacter* (Ap) cells grown in LL and HL on Day 8. *Synechocystis* PCC6803 (Syn6803) grown under HL was used as a comparison. All the calculations were based on three biological replicates. (D) Rates of oxygen evolution (Y-axis) over different light intensities (μmol photons m⁻²s⁻¹). Three biological replicates were used for all the measurements.

Comparison with Candidatus Aurora

Very recently, *Candidatus Aurora* was described from a metagenome assembly¹¹, which based on our phylogenetic analysis belongs to the same lineage as *Anthocero bacter* (Fig. 2). There are however several key differences between the two. First, the assembly of *Candidatus Aurora* MAG is highly fragmented and incomplete, making it difficult to confidently conclude gene presence/absence and discuss the broader implications on cyanobacterial evolution. In fact, *psbA* (which encodes PSII reaction center D1 protein) was not present in Aurora MAG, but instead found on a separate small contig that was not binned properly. Here, with a complete circular genome of *Anthocero bacter*, we are able to fully describe the gene content. A case in point is the phycobilisome linker protein CpcG. Because it is missing in *Gloeobacter* genomes as well as Aurora MAG, *cpcG* gene would be inferred to originate in Phycobacteria. However, we found a clear *cpcG* homolog in *Anthocero bacter* that is phylogenetically sister to the rest of cyanobacterial *cpcG* (Fig. S6). This suggests that *cpcG* is probably ancestral to all Cyanobacteria (but subsequently lost in *Gloeobacter*) and raises the possibility that its absence in Aurora MAG might be due to incomplete assembly.

Second, the existence of *Candidatus Aurora* had been hinted by earlier environmental samples from Arctic and Antarctic regions, which share highly similar 16S sequences (>99% identity; Fig. 2)¹¹. *Anthocero bacter*, on the hand, was not known from any environmental sample that we are aware of. In

addition, as an isolate from a tropical hornwort, *Anthocero bacter* shows a contrasting distribution pattern from its sister polar clade (Fig. 2).

Most importantly, *Candidatus Aurora* is not an isolated culture, rendering further studies difficult if not impossible. The fact that *Anthocero bacter* is the only culturable organism in this lineage allowed us to characterize the morphology and physiology of this representing strain in detail. The lack of thylakoid in *Anthocero bacter* was observed by TEM (Fig. 1), which is much more convincing than basing on the absence of *VIPP1* homolog alone in *Candidatus Aurora*¹¹, especially considering gene absence could also be due to incomplete MAG assembly. Our spectral and growth experiments also provided important insights into the biology of *Anthocero bacter* that are otherwise impossible to obtain for *Candidatus Aurora*. For example, *Anthocero bacter* lacks long-wavelength emission from PSI similar to *Gloeobacter*, suggesting that this feature is not an idiosyncrasy of *Gloeobacter* but deeply shared across *Gloeobacteria*. In addition, *Anthocero bacter* was experimentally demonstrated capable of oxygen evolution and photoautotrophic growth, which are not possible to be verified in *Candidatus Aurora*. While there is no doubt that the discovery of *Candidatus Aurora* is significant, *Anthocero bacter*, owing to its culturable nature, is the one that can advance our understandings of this key cyanobacterial lineage.

Conclusion

Anthocero bacter and *Candidatus Aurora* together represent a novel and ancient lineage of Cyanobacteria, which diverged from *Gloeobacter* over 1.4 billion years ago. The deep divergence and availability as a culture enable us to use *Anthocero bacter* to revisit the early evolution of Cyanobacteria. One key question we were able to address is whether the many unique features in *Gloeobacter* are indeed characteristic of the entire *Gloeobacteria* lineage. On one hand, we found that similar to *Gloeobacter*, *Anthocero bacter* lacks thylakoids, circadian clock oscillators, nitrogenase, canonical pathway for phyloquinone synthesis, long-wavelength PSI fluorescence, as well as many conserved photosynthesis subunits (summarized in Fig. S7). The most parsimonious interpretation is that the origins of these components took place in the most recent common ancestor (MRCA) of Phycobacteria, and the absence of which likely define *Gloeobacteria*. Conversely, the carotenoid biosynthesis pathway in *Anthocero bacter* is not the bacterial-type as found in *Gloeobacter*, but typical of the rest Cyanobacteria, implying that the plant-type pathway likely evolved in the MRCA of Cyanobacteria and not of Phycobacteria (Fig. S7). *Anthocero bacter* also possesses several unique features, such as its phycobilisome composition and an even further reduced collection of photosystem subunits. Nevertheless, we were able to demonstrate that *Anthocero bacter* can indeed grow photoautotrophically and evolve oxygen across a range of light intensities, although less efficiently compared to other cyanobacterial strains. In summary, *Anthocero bacter* is an important addition to the depauperate *Gloeobacteria* and will facilitate future investigations into the origin of Cyanobacteria and oxygenic photosynthesis.

366

367 ***Taxonomic treatment***

368 ***Anthocerbacter*** F.-W. Li gen. nov. (Fig. 1)

369 Differ from genera of Chroococcales, Chroococcidiopsidales, Nostocales, Oscillatoriales, Pleurocapsales,
370 Spirulinales, and Synechococcales in lacking thylakoid membranes. Most similar to *Gloeobacter* in
371 Gloeobacterales but lacks phycoerythrin.

372 **Etymology:** An.tho.cero.bact'er. N.L. neut. n. Anthocerotophyta, hornworts; N.L. masc. n. bacter, a small
373 rod; N.L. masc. n. Anthocerbacter, a rod from hornworts.

374 **Type species:** *Anthocerbacter panamensis* F.-W. Li sp. nov.

375 ***Anthocerbacter panamensis*** F.-W. Li sp. nov. (Fig. 1)

376 The cells are rod-shaped, 0.5-1 μm wide and 2-10 μm long. Thylakoid membranes are absent.
377 Carboxysomes are present, and polyphosphate granules often appear toward the cell poles. The major
378 pigments are allophycocyanin, phycocyanin, and chlorophyll *a*. No phycoerythrin nor
379 phycoerythrocyanin is present. Colonies are generally light to dark green in both liquid and solid BG11
380 medium, and never purple, orange, or brown like in *Gloeobacter*. Growth can occur under 10 - 100 μmol
381 photons $\text{m}^{-2}\text{s}^{-1}$ light intensities. No nitrogenase genes are present and cells do not grow without
382 supplementing combined nitrogen. The type strain (C109) was originally isolated from a surface-
383 sterilized hornwort *Leiosporoceros dussii* from Coclé, Panama (8°37'34.6" N, 80°08'13.7" W). GC content
384 of the type strain is 55.51%.

385 **Etymology:** pa.na.men'sis. N.L. masc./fem. adj. panamensis, of or pertaining to Panama where the type
386 strain was isolated.

387 **Holotype:** The type strain C109 was deposited at Bailey Hortorium (BH) under the accession xxx, and
388 American Type Culture Collection (no. xxx).

Materials and Methods

Culturing

The hornwort *Leiosporoceros dussii* (Steph.) Hässel was collected in Coclé, Panama (8°37'34.6" N, 80°08'13.7" W). Sterilization of the plant thalli and initial culturing of the isolate (type strain C109) followed the protocol of Nelson *et al.*⁵³. *Synechocystis* sp. PCC 6803 was a gift from Dr. Hsiu-An Chu at Institute of Plant and Microbial Biology, Academia Sinica, Taiwan, and *Chlorogloeopsis fritschii* PCC 9212 was obtained from the Pasteur Culture Collection⁵⁴. *Anthoceroobacter*, *Chlorogloeopsis fritschii* PCC 9212, and *Synechocystis* sp. PCC 6803 were grown in the B-HEPES growth medium⁵⁵, a modified BG11 medium containing 1.1 g L⁻¹ HEPES (final concentration) with the pH adjusted to 8.0 with 2.0 M KOH. Cool white LED light provided continuous illumination at either 10 or 100 μmol photons m⁻² s⁻¹ in a 30°C growth chamber supplemented with 1% (v/v) CO₂ in air. The doubling time is estimated based on the change of OD₇₅₀ absorption over time.

DNA extraction

Cyanobacteria colonies from either liquid cultures or solid BG11 plates were centrifuged at 17,000 x g for 2 mins to pellet the cells. The cells were ground with a pestle in liquid nitrogen and resuspended in a prewarmed 2X CTAB solution with 1% beta-mercaptoethanol. The mixture was incubated at 65°C for 1h, with gentle mixing every 15 min. Equal volume of 24:1 chloroform:isoamyl alcohol was added to the samples twice, each time with 5 min mixing on a Labnet Mini Labroller and 5 min centrifugation at 17,000 x g. The supernatant was transferred to a new tube by wide bore pipette tips to maintain DNA integrity. DNA was precipitated with an equal volume of isopropanol and pelleted for 30 min at 17,000 x g and 4°C. The DNA pellets were washed twice with 70% ethanol, then air dried in a sterile air hood and resuspended in Tris-HCl pH 7.5, before treated with RNase for 1 h at 37°C.

Nanopore and Illumina sequencing

Genomic DNA was sequenced on both Oxford Nanopore MinION and Illumina platforms. Nanopore libraries were prepared using the Ligation Sequencing kit (SQK-LSK109) and sequenced on two MinION R9 flowcells (FLO-MIN106D) for 60 hours. In one MinION run, the focal sample was multiplexed with another cyanobacteria sample and barcoded by Native Barcoding Expansion Kit 1-12, while the other run contains only the focal sample. Basecalling was carried out using Guppy v3.1.5 with the high accuracy flip-flop mode. The Illumina library was prepared and sequenced on NextSeq500 by MiGS (Microbial Genome Sequencing Center). The raw nanopore and Illumina reads are deposited at NCBI SRA under the accessions SRR12713672 and SRR12713672.

Genome assembly and annotation

The basecalled nanopore reads were first filtered to remove reads shorter than 10kb, and then *de novo* assembled using Flye v2.4.1⁵⁶ with the --plasmids flag on, followed by polishing by medaka

(<https://github.com/nanoporetech/medaka>). The resulting assembly contained three circular contigs. Based on the PATRIC taxonomic assignments⁵⁷, one has the closest affinity to *Gloeobacter*, while the other two belong to *Pseudomonas fluorescens*, indicating contaminations were introduced during liquid culture (which was later confirmed by light microscopy). We therefore started a new culture for generating Illumina data. After trimming by TRIMMOMATIC (min score=25, min length=36)⁵⁸, the resulting Illumina reads had 91.74% mapping rate against the cyanobacterial contig (based on the bwa mem aligner⁵⁹), suggesting most (if not all) contaminations were removed. Genome polishing with Illumina reads was done by pilon⁶⁰ for 4 iterations. Structural and functional annotation was performed using PATRIC⁵⁷, with additional KEGG annotation by BlastKOALA⁶¹. The genome assembly is deposited at NCBI under BioProject PRJNA665722.

Genome comparisons

Pairwise 16S sequence divergence and Average Amino-acid Identity (AAI) were calculated by seqinr v3.6-1⁶² and CompareM v0.1.1 (<https://github.com/dparks1134/CompareM>), respectively. We used Orthofinder v2.3.12⁶³ to classify families of orthologous genes in the genomes of *Anthoceroobacter* and 102 other cyanobacterial species. UpSetR⁶⁴ was used to summarize the shared and unique orthogroups found in *Anthoceroobacter*. To infer synteny between *Anthoceroobacter* and the two *Gloeobacter* species, MCScanX⁶⁵ was used with Diamond⁶⁶ as the search engine. The syntenic relationship was then visualized in AccuSyn (<https://accusyn.usask.ca>).

Phylogenetic analyses and divergence time estimates

We first compiled a 16S data matrix that includes major Cyanobacteria lineages, outgroups⁶, and available environmental sequences with high sequence similarities to *Gloeobacter* and *Anthoceroobacter*^{11, 12}. Alignment was done by PASTA v3⁶⁷, and maximum-likelihood phylogeny was inferred using IQTREE v1.6.11 with 1,000 ultrafast bootstrap replicates⁶⁸.

Next we integrated our data with previous molecular clock analyses on Cyanobacteria: Shih *et al.*⁶ and Baracaldo *et al.*²⁷. The dataset from Shih *et al.*⁶ is composed of 11 concatenated protein sequences (*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*, *rpl2*, *rpl16*, *rps3*, *rps12*, and elongation factor Tu) and 16S nucleotide sequences. We used the same age calibrations and substitution models (CpREV for proteins and GTR+Gamma for 16S) as in Shih *et al.*⁶. Phylogeny and divergence time were inferred using BEAST v.1.8.3⁶⁹ on CIPRES Science Gateway (v3.3)⁷⁰. Five MCMC were run in total, and for each chain samples were taken every 10,000 generations. The posterior distributions were inspected in Tracer v.1.7.1 to ensure proper mixing and convergence. The resulting trees from five runs were combined by LogCombiner and summarized by TreeAnnotator⁶⁹.

The dataset from Baracaldo *et al.*²⁷ is composed of 8 protein-coding genes (*atpA*, *atpB*, *petB*, *psaC*, *psbA*, *psbD*, *rbcl*, and S12). Following the protocol of Baracaldo *et al.*²⁷, PhyloBayes⁷¹ was used to estimate divergence time, with CAT-GTR+Gamma substitution model, three different relaxed molecular clock models: log normal (ln), Cox–Ingersoll–Ross model (cir) and uncorrelated gamma multiplier (ugam), and two different root priors: (1) 95% of the prior distribution falls between 2.32 and 2.7 BYA and (2)

between 2.3 and 3 BYA. The sequence alignments, input files for BEAST and PhyloBayes, as well as the resulting tree files can be found at <https://gitlab.com/NasimR/cyanobacteria>.

Phylogenetic inferences were also done for selected orthogroups using IQTREE v1.6.11 with 1,000 ultrafast bootstrap replicates and automatic model selection⁶⁸.

Phylogeny of orange carotenoid proteins (OCP)

OCP is composed of two domains, N-terminal domain (NTD) and C-terminal domain (CTD), connected by a flexible loop linker⁷². Both domains have their own paralogs with either stand-alone NTD or CTD; the former is called helical carotenoid protein (HCP) and the latter CTD-like proteins (CTDHs)⁷².

Anthocero bacter has one OCP and one NTD, but no CTDH. While Grett enberger et al¹² did report on the presence of OCP in *Candidatus Aurora*, they did not distinguish OCP, HCP, and CTDH. Our reanalysis showed that *Candidatus Aurora* has one OCP, one HCP, and one putative CTDH. We then incorporated the OCP sequences from *Anthocero bacter* and *Candidatus Aurora* into the dataset compiled by Bao et al³³, which was used to define the major OCP clades³³. Sequence alignment was done by PASTA v3⁶⁷ followed by phylogenetic inference using IQTREE v2.0.3 with automatic model selection and 1,000 ultrafast bootstrap replicates⁶⁸.

Phylogeny of phycobilisome linker proteins

The phycobilisome linkers from *Anthocero bacter* and *Gloeobacter violaceus* were added to the protein sequences compiled by Watanabe and Ikeuchi³⁸. Sequence alignment was done by PASTA v3⁶⁷, and phylogeny was reconstructed using IQTREE v2.0.3 with automatic model selection and 1,000 ultrafast bootstrap replicates⁶⁸.

Microscopy

Pure cultures were processed for transmission electron microscopy as described previously⁷³ with slight modifications. Cultures were fixed in 3% glutaraldehyde, 1% fresh formaldehyde, and 0.75 % tannic acid in 0.05 M Na-cacodylate buffer, pH 7, for 3 h at room temperature. After several rinses in 0.1M buffer, the samples were post-fixed in buffered (0.1M, pH 6.8) 1% osmium tetroxide overnight at 4°C, dehydrated in an ethanol series and embedded in Spurr's resin via ethanol. Thin sections were cut with a diamond histo-knife, stained with methanolic uranyl acetate for 15 min and in Reynolds' lead citrate for 10 min, and observed with a Hitachi H-7100 transmission electron microscope at the Imaging-Microscopy Platform of the IBIS, Universite Laval. Images of colonies on an agar plate were taken by a SMZ-171TLED stereomicroscope (Bio Pioneer Tech Co, Taiwan) equipped with a TrueChrome AF digital camera (TUCSEN, China). Light microscopy images were taken under a Leica DMR microscope (Leica, Germany) equipped with a Leica MC 170HD digital camera (Leica, Germany).

Absorption and fluorescence spectral analyses

The spectroscopy measurements were performed as previously described with modifications⁷⁴. The OD₇₅₀ values of the cyanobacterial cultures were adjusted to 0.25 in 1 mL of B-HEPES medium for absorption spectroscopy, and 0.25 in 1 mL of PEG buffer (500 µl of B-HEPES medium and 500 µl of 30% (w/v) PEG4000) for fluorescence spectroscopy. The measurement of absorption spectra (wavelength from 400 to 800 nm) was performed by MAPADA UV1800 with an equipped PC software (Datech Instruments Ltd., Taiwan). Fluorescence spectroscopy was performed by the F-4500 fluorescence spectrophotometer (Hitachi, Japan). Fluorescence emission was recorded from 600 to 800 nm by excitation at 440 nm for chlorophyll *a*.

Measurements of pigment composition and oxygen evolution

The quantifications of chlorophyll *a*, carotenoids, and phycobiliproteins were performed as previously described⁷⁵. Steady-state rates of oxygen evolution were measured as previously described with minor modifications⁷⁶. An Oxytherm System oxygen electrode (Model V-560, Hansatech Instruments Ltd, UK) was used for the measurement. Cyanobacterial cells were diluted in B-HEPES medium to the final concentration as 3.5 µg (Chl) ml⁻¹ in a stirred, water-jacketed cell chamber at 25°C. Potassium ferricyanide and 2,6-dichloro-p-benzoquinone (DCBQ) were added sequentially into the cell chamber as artificial electron acceptors to the final concentration as 2 mM. Different strengths of light intensities were provided from both sides of the cell chamber by two fiber-optic illuminators (Dolan-Jenner model MI 150, USA). Data recording and processing were performed with the manufacture software O₂ view (Hansatech Instruments Ltd, UK).

Acknowledgments

We thank Dr. Noris Salazar Allen for providing fieldwork support to J.C.V.A., Dr. Hsiu-An Chu and Keng-Min Lin for assisting oxygen evolution measurement, Dr. Patrick Shih for sharing alignment matrices and early discussions, and Drs. Kathleen Pryer and John Meeks for comments. This work was funded by National Science Foundation (NSF) grant no. DEB1831428 to F.-W.L., Ministry of Science and Technology (Taiwan) Young Scholar Fellowship Einstein Program (108C5368 and 109C5369) and Ministry of Education (Taiwan) Yushan Young Scholar Program (108V1102 and 109V1102) to M.-Y.H., and Chaires de Recherche du Canada (950-232698), CRNSG (RGPIN-2016-05967) and Earl S. Tupper Fellowship (STRI) to J.C.V.A.

Author Contributions

F.-W.L. and M.-Y.H. designed research; N.R., D.A.H., J.M.N., P.Y.C., J.C.V.A., M.-Y.H., F.-W.L. performed research; N.R., M.-Y.H., and F.-W.L. analyzed data; and N.R., M.-Y.H., and F.-W.L. wrote the paper.

Declaration of interests

The authors declare no competing interest.

References

1. Soo, R.M., Hemp, J., Parks, D.H., Fischer, W.W., and Hugenholtz, P. (2017). On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science* 355, 1436–1440.
2. Planavsky, N.J., Reinhard, C.T., Wang, X., Thomson, D., McGoldrick, P., Rainbird, R.H., Johnson, T., Fischer, W.W., and Lyons, T.W. (2014). Low Mid-Proterozoic atmospheric oxygen levels and the delayed rise of animals. *Science* 346, 635–638.
3. Soo, R.M., Hemp, J., and Hugenholtz, P. (2019). Evolution of photosynthesis and aerobic respiration in Cyanobacteria. *Free Radic. Biol. Med.* 140, 200–205.
4. Garcia-Pichel, F., Zehr, J.P., Bhattacharya, D., and Pakrasi, H.B. (2020). What's in a name? The case of cyanobacteria. *J. Phycol.* 56, 1–5.
5. Cavalier-Smith, T. (2002). The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.* 52, 7–76.
6. Shih, P.M., Hemp, J., Ward, L.M., Matzke, N.J., and Fischer, W.W. (2016). Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobiology* 15, 19–29.
7. Saw, J.H.W., Schatz, M., Brown, M.V., Kunkel, D.D., Foster, J.S., Shick, H., Christensen, S., Hou, S., Wan, X., and Donachie, S.P. (2013). Cultivation and complete genome sequencing of *Gloeobacter kilaueensis* sp. nov., from a Lava Cave in Kīlauea Caldera, Hawai'i. *PLoS ONE* 8, e76376.
8. Rippka, R., Waterbury, J., and Cohen-Bazire, G. (1974). A cyanobacterium which lacks thylakoids. *Arch. Microbiol.* 100, 419–436.
9. Mareš, J., Hrouzek, P., Kaňa, R., Ventura, S., Strunecký, O., and Komárek, J. (2013). The primitive thylakoid-less cyanobacterium *Gloeobacter* is a common rock-dwelling organism. *PLoS ONE* 8, e66323.
10. Montejano, G., Becerra-Absalón, I., Gold-Morgan, M., and Osorio-Santos, K. (2018). *Gloeobacter violaceus*: primitive reproductive scheme and its significance. *Plant Syst. Evol.* 304, 1221–1229.
11. Lynch, M.D.J., Bartram, A.K., and Neufeld, J.D. (2012). Targeted recovery of novel phylogenetic diversity from next-generation sequence data. *ISME J.* 6, 2067–2077.
12. Grettenberger, C.L., Sumner, D.Y., Wall, K., Brown, C.T., Eisen, J.A., Mackey, T. J., Hawes, I., Jospin, G., and Jungblut, A.D. (2020). A phylogenetically novel cyanobacterium most closely related to *Gloeobacter*. *ISME J.* 14, 2142–2152.
13. Nakamura, Y., Kaneko, T., Sato, S., Mimuro, M., Miyashita, H., Tsuchiya, T., Sasamoto, S., Watanabe, A., Kawashima, K., Kishida, Y., *et al.*, (2003). Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. *DNA Res.* 10, 137–145.
14. Rexroth, S., Mullineaux, C.W., Ellinger, D., Sendtko, E., Rögner, M., and Koenig, F. (2011). The plasma membrane of the cyanobacterium *Gloeobacter violaceus* contains segregated bioenergetic domains. *Plant Cell* 23, 2379–2390.
15. Koenig, F., and Schmidt, M. (1995). *Gloeobacter violaceus*— investigation of an unusual photosynthetic apparatus. Absence of the long wavelength emission of photosystem I in 77 K fluorescence spectra. *Physiol. Plant.* 94, 621–628.
16. Mangels, D., Kruij, J., Berry, S., Rögner, M., Boekema, E.J., and Koenig, F. (2002) Photosystem I from the unusual cyanobacterium *Gloeobacter violaceus*. *Photosynth. Res.* 72, 307–319.

17. Inoue, H., Tsuchiya, T., Satoh, S., Miyashita, H., Kaneko, T., Tabata, S., Tanaka, A., and Mimuro, M. (2004). Unique constitution of photosystem I with a novel subunit in the cyanobacterium *Gloeobacter violaceus* PCC 7421. *FEBS Lett.* **578**, 275–279.
18. Mimuro, M., Tsuchiya, T., Inoue, H., Sakuragi, Y., Itoh, Y., Gotoh, T., Miyashita, H., Bryant, D.A., and Kobayashi, M. (2005). The secondary electron acceptor of photosystem I in *Gloeobacter violaceus* PCC 7421 is menaquinone-4 that is synthesized by a unique but unknown pathway. *FEBS Lett.* **579**, 3493 – 3496.
19. Dreher, C., Hielscher, R., Prodöhl, A., Hellwig, P., and Schneider, D. (2010) Characterization of two cytochrome b_6 proteins from the cyanobacterium *Gloeobacter violaceus* PCC 7421. *J. Bioenerg. Biomembr.* **42**, 517–526.
20. Guglielmi, G., Cohen-bazire, G., and Bryant, D.A. (1981). The structure of *Gloeobacter violaceus* and its phycobilisomes. *Arch. Microbiol.* **129**, 181–189.
21. Bryant, D.A., and Canniffe, D.P. (2018). How nature designs light-harvesting antenna systems: design principles and functional realization in chlorophototrophic prokaryotes. *J. Phys. B.* **51**, 033001.
22. Tsuchiya, T., Takaichi, S., Misawa, N., Maoka, T., Miyashita, H., and Mimuro, M. (2005). The cyanobacterium *Gloeobacter violaceus* PCC 7421 uses bacterial-type phytoene desaturase in carotenoid biosynthesis. *FEBS Lett.* **579**, 2125–2129.
23. Steiger, S., Jackisch, Y., and Sandmann, G. (2005). Carotenoid biosynthesis in *Gloeobacter violaceus* PCC4721 involves a single crtI-type phytoene desaturase instead of typical cyanobacterial enzymes. *Arch. Microbiol.* **184**, 207–214.
24. Meeks, J.C. (1998). Symbiosis between nitrogen-fixing cyanobacteria and plants. *Bioscience* **48**, 266–276.
25. Duckett, J.G., Prasad, A.K.S.K., Davis, D.A., and Walker, S. (1977). A cytological analysis of the *Nostoc*-Bryophyte relationship. *New Phytol.* **79**, 349–362.
26. Bouchard, R., Peñaloza-Bojacá, G., Toupin, S., Guadalupe, Y., Gudiño, J., Allen, N.S., Li, F.W., and Villarreal A, J.C. (2020). Contrasting bacteriome of the hornwort *Leiosporoceros dussii* in two nearby sites with emphasis on the hornwort-cyanobacterial symbiosis. *Symbiosis* **81**, 39–52.
27. Sánchez-Baracaldo, P., Raven, J.A., Pisani, D., and Knoll, A.H. (2017). Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7737–E7745.
28. Kroll, D., Meierhoff, K., Bechtold, N., Kinoshita, M., Westphal, S., Vothknecht, U.C., Soll, J., and Westhoff, P. (2001). *VIPP1*, a nuclear gene of *Arabidopsis thaliana* essential for thylakoid membrane formation. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4238–4242.
29. Mimuro, M., Ookubo, T., Takahashi, D., Sakawa, T., Akimoto, S., Yamazaki, I., and Miyashita, H. (2002). Unique fluorescence properties of a cyanobacterium *Gloeobacter violaceus* PCC 7421: Reasons for absence of the long-wavelength PSI Chl a fluorescence at –196°C. *Plant Cell Physiol.* **43**, 587–594.
30. Bryant, D.A., Guglielmi, G., Tandeau de Marsac, N., Castets, A-M., and Cohen-Bazire, G. (1979). The structure of cyanobacterial phycobilisomes: a model. *Arch. Microbiol.* **123**, 113–127.
31. Bryant, D.A. (1982). Phycoerythrocyanin and phycoerythrin: properties and occurrence in cyanobacteria. *J. G. Microbiol.* **128**, 835–844.
32. Cohen, S.E., and Golden, S.S. (2015). Circadian rhythms in Cyanobacteria. *Microbiol. Mol. Biol. Rev.*

619 79, 373–385.

620 33. Bao, H., Melnicki, M.R., Pawlowski, E.G., Sutter, M., Agostoni, M., Lechno-Yossef, S., Cai, F.,
621 Montgomery, B.L., and Kerfeld, C.A. (2017). Additional families of orange carotenoid proteins in the
622 photoprotective system of cyanobacteria. *Nat. Plants* 3, 17089.

623 34. Takaichi, S. (2011). Carotenoids in algae: distributions, biosynthesis and functions. *Mar. Drugs* 9,
624 1101–1118.

625 35. Ashby, M.K., and Mullineaux, C.W. (1999). The role of ApcD and ApcF in energy transfer from
626 phycobilisomes to PS I and PS II in a cyanobacterium. *Photosynth. Res.* 61, 169–179.

627 36. Gindt, Y.M., Zhou, J., Bryant, D.A., and Sauer, K. (1992). Core mutations of *Synechococcus* sp. PCC
628 7002 phycobilisomes: A spectroscopic study. *J. Photochem. Photobiol. B. Biol.* 15, 75–89.

629 37. Ughy, B., and Ajlani, G. (2004). Phycobilisome rod mutants in *Synechocystis* sp. Strain PCC6803 strain
630 PCC6803. *Microbiol.* 150, 4147–4156.

631 38. Watanabe, M., and Ikeuchi, M. (2013). Phycobilisome: architecture of a light-harvesting
632 supercomplex. *Photosynth Res.* 116, 265–276.

633 39. Koyama, K., Tsuchiya, T., Akimoto, S., Yokono, M., Miyashita, H., and Mamoru, M. (2006). New linker
634 proteins in phycobilisomes isolated from the cyanobacterium *Gloeobacter violaceus* PCC 7421. *FEBS*
635 *Lett.* 580, 3457–3461.

636 40. Bishop, C.L., Ulas, S., Baena-Gonzalez, E., Aro, E-M., Purton, S., Nugent, J.H.A., and Mäenpää, P.
637 (2007). The PsbZ subunit of Photosystem II in *Synechocystis* sp. PCC 6803 modulates electron flow
638 through the photosynthetic electron transfer chain. *Photosynth Res.* 93, 139–147.

639 41. Shi, L-X., and Schröder, W.P. (2004). The low molecular mass subunits of the photosynthetic
640 supercomplex, photosystem II. *BBA - Bioenergetics.* 1608, 75–96.

641 42. Roose, J.L., and Pakrasi, H.B. (2008). The Psb27 protein facilitates Manganese cluster assembly in
642 photosystem II. *J. Biol. Chem.* 283, 4044–4050.

643 43. Becková, M., Gardian, Z., Yu, J., Konik, P., Nixon, P.J., and Komenda, J. (2017). Association of Psb28
644 and Psb27 proteins with PSII-PSI supercomplexes upon exposure of *Synechocystis* sp. PCC 6803 to high
645 light. *Mol. Plant* 10, 62–72.

646 44. Xu, Q., Odom, W.R., Guikema, J.A., Chitnis, V.P., and Chitnis, P.R. (1994). Targeted deletion of *psaJ*
647 from the cyanobacterium *Synechocystis* sp. PCC 6803 indicates structural interactions between the *PsaJ*
648 and *PsaF* subunits of photosystem I. *Plant Mol. Biol.* 26, 291–302.

649 45. Xu, Q., Yuo, L., Chitnis, V.P., and Chitnis, P.R. (1994). Function and organization of photosystem I in
650 cyanobacterial mutant strain that lacks *PsaF* and *PsaJ* subunits. *J. Biol. Chem.* 269, 3205–3211.

651 46. Kawakami, K., Umena, Y., Iwai, M., Kawabata, Y., Ikeuchi, M., Kamiya, N., and Shen, J-R. (2011). Roles
652 of PsbI and PsbM in photosystem II dimer formation and stability studied by deletion mutagenesis and
653 X-ray crystallography. *BBA - Bioenergetics* 1807, 319–325.

654 47. Inoue-Kashino, N., Kashino, Y., Satoh, K., Terashima, I., and Pakrasi, H.B. (2005). PsbU provides a
655 stable architecture for the oxygen-evolving system in cyanobacterial photosystem II. *Biochemistry* 44,
656 12214–12228.

657 48. Veerman, J., Bentley, F.K., Eaton-Rye, J.J., Mullineaux, C.W., Vasil'ev, S., and Bruce, D. (2005). The

PsbU subunit of photosystem II stabilizes energy transfer and primary photochemistry in the phycobilisome–photosystem II assembly of *Synechocystis* sp. PCC 6803. *Biochemistry* **44**, 16939–16948.

49. Schwenkert, S., Legen, J., Takami, T., Shikanai, T., Hermann, R.G., and Meurer, J. (2007). Role of the low-molecular-weight subunits PetL, PetG, and PetN in assembly, stability, and dimerization of the cytochrome b_6f complex in Tobacco. *Plant Physiol.* **144**, 1924–1935.

50. Schneider, D., Berry, S., Rich, P., Seidler, A., and Rögner, M. (2001). A Regulatory role of the PetM subunit in a cyanobacterial cytochrome b_6f complex. *J. Biol. Chem.* **276**, 16780–16785.

51. Castenholz, R.W. (2001). Oxygenic photosynthetic Bacteria. In *Bergey's manual of systematics Bacteriology*, Springer-Verlag, New York, NY, ed (Boone, D.R., Castenholz, R.W., Garrity, G.M), pp. 474–487.

52. Koyama K., Suzuki, H., Noguchi, T., Akimoto, S., Tsuchiya, T., and Mimuro, M. (2008). Oxygen evolution in the thylakoid-lacking cyanobacterium *Gloeobacter violaceus* PCC 7421. *BBA - Bioenergetics* **1777**, 369–378.

53. Nelson J.M., Hauser, D.A., Gudiño, J.A., Guadalupe, Y.A., Meeks, J.C., Allen, N.S. Villarreal, J.C and Li, F-W. (2019). Complete genomes of symbiotic cyanobacteria clarify the evolution of Vanadium-nitrogenase. *Genome Biol. Evol.* **11**, 1959–1964.

54. Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., and Stanier, R.Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Microbiol.* **111**, 1–61.

55. Dubbs, J.M., and Bryant, D.A. (1991). Molecular cloning and transcriptional analysis of the cpeBA operon of the cyanobacterium *Pseudanabaena* species PCC 7409. *Mol. Microbiol.* **5**, 3073–3085.

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

57. Davis, J.J., Wattam, A.R., Aziz, R.K., Brettin, T., Butler, R., Butler, R.M., Chlenski, P., Conard, N., Dickerman, A., Dietrich, E.M., *et al.*, (2020). The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res.* **48**, D606–D612.

58. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequences data. *Bioinformatics* **30**, 2114–2120.

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

60. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortmen, J., Young, S.K., *et al.*, (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963.

61. Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731.

62. Charif, D., and Lobry, J.R. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution: Molecules, networks, populations*, Springer Verlag, ed. (Bastolla, U., Porto, M., Roman, H.E., Vendruscolo, M) pp, 207–232.

63. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238.

64. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Genome analysis* 33, 2938–2940.
65. Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T-h., Jin, H., Marler, B., Guo, H., *et al.*, (2012). *MCScanX*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49.
66. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60.
67. Mirarab, S., Nguyen, N., Guo, S., Wang, L-S., Kim, J., and Warnow, T. (2015). PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *J. Comput. Biol.* 22, 377–386.
68. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Ecol.* 37, 1530–1534.
69. Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol* 7, 214 –221.
70. Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *In Gateway Computing Environments Workshop (GCE)*, 1–8.
71. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286–2288.
72. Muzzopappa, F., and Kirilovsky, D. (2020). Changing color for photoprotection: The orange carotenoid protein. *Trends Plant Sci.* 25, 92–104.
73. Villarreal, J.C., and Renzaglia, K.S. (2006). Structure and development of *Nostoc* strands in *Leiosporoceros dussii* (Anthocerotophyta): a novel symbiosis in land plants. *Am. J. Bot.* 93, 693–705.
74. Ho, M.-Y., Shen, G., Canniffe, D.P., Zhao, C., and Bryant, D.A. (2016). Light-dependent chlorophyll f synthase is a highly divergent paralog of PsbA of Photosystem II. *Science* 353, aaf9178.
75. Sakamoto, T., and Bryant, D.A. (1998). Growth at low temperature causes nitrogen limitation in the cyanobacterium *Synechococcus* sp. PCC 7002. *Arch. Microbiol.* 169, 10–19.
76. Huang, J-Y., Chiu, Y-F., Ortega, J.M., Wang, H-T., Tseng, T-S., Ke, S-C., Roncel, M., and Chu, H-A. (2016). Mutations of cytochrome b₅₅₉ and PsbJ on and near the Q_c Site in Photosystem II influence the regulation of short-term light response and photosynthetic growth of the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochemistry* 55, 22142–2226.