Insights into the evolution of oxygenic photosynthesis from a phylogenetically novel, low-light cyanobacterium

Insights into oxygenic photosynthesis from Aurora

Christen L. Grettenberger[1*], Dawn Y. Sumner[1], Kate Wall[1], C. Titus Brown[2,3], Jonathan Eisen[2], Tyler J. Mackey[1,4], Ian Hawes[5], Anne D. Jungblut [6]

1. University of California Davis, Department of Earth and Planetary Sciences, Davis, CA, USA
2. University of California Davis Genome Center, Davis, CA, USA
3. University of California Davis, Veterinary Medicine Population Health and Reproduction, Davis, CA, USA
4. Massachusetts Institute of Technology, Department of Earth, Atmospheric, and Planetary Sciences, Cambridge, MA
5. University of Waikato, Tauranga, New Zealand
6. The Natural History Museum, London, Life Sciences Department, Cromwell Road, SW7 5BD, London, United Kingdom

*Christen L Grettenberger

Earth and Planetary Sciences, 1 Shields Avenue

University of California Davis

Davis, CA 95616

(208) 869-1271

christygrettenberger@gmail.com

Abstract

Atmospheric oxygen level rose dramatically around 2.4 billion years ago due to oxygenic photosynthesis by the Cyanobacteria. The oxidation of surface environments permanently changed the future of life on Earth, yet the evolutionary processes leading to oxygen production are poorly constrained. Partial records of these evolutionary steps are preserved in the genomes of organisms phylogenetically placed between non-photosynthetic Melainabacteria, crown-group Cyanobacteria, and *Gloeobacter*, representing the earliest-branching Cyanobacteria capable of oxygenic photosynthesis. Here, we describe nearly complete, metagenome assembled genomes of an uncultured organism phylogenetically placed between the Melainabacteria and crown-group Cyanobacteria, for which we propose the name Candidatus *Aurora vandensis {au.rora* Latin noun *dawn* and *vand.ensis*, originating from Vanda}.

The metagenome assembled genome of *A. vandensis* contains homologs of most genes necessary for oxygenic photosynthesis including key reaction center proteins. Many extrinsic proteins associated with the photosystems in other species are, however, missing or poorly conserved. The assembled genome also lacks homologs of genes associated with the pigments phycocyanoerethrin, phycoeretherin and several structural parts of the phycobilisome. Based on the content of the genome, we propose an evolutionary model for increasing efficiency of oxygenic photosynthesis through the evolution of extrinsic proteins to stabilize photosystem II and I reaction centers and improve photon capture. This model suggests that the evolution of oxygenic photosynthesis may have significantly preceded oxidation of Earth's atmosphere due to low net oxygen production by early Cyanobacteria.

1. Introduction

Around 2.4 billion years ago, Earth's surface environments changed dramatically. Atmospheric oxygen rose from $<10^{-5}$ times present atmospheric level (PAL) to >1% PAL [1-4]. This Great Oxygenation Event (GOE) permanently changed Earth's surface geochemistry, fundamentally reshaped the cycling of key elements [5] and altered the evolutionary path of life by allowing widespread oxygen respiration [6]. The GOE was enabled by the evolution of oxygenic photosynthesis in the Cyanobacteria, making this one of the most important innovations in Earth's history [4,7,8]. However, the evolutionary processes leading to oxygenic photosynthesis are poorly constrained [9-14]. In one hypothesis, Cyanobacteria acquired photosynthetic genes for both photosystems I and II (PSI and PSII, respectively) via horizontal gene transfer and then combined and refined them to form the photosystems that drive oxygenic photosynthesis in crown-group Cyanobacteria [15,16]. In another hypothesis, the common ancestor of all phototrophic bacteria contained the genes necessary for photosynthesis, which diversified through time and were selectively lost in non-phototrophic portions of those lineages [17-21]. In either scenario, early branching Cyanobacteria will be important to elucidating the evolution of oxygenic photosynthesis.

Due to the importance of oxygenic photosynthesis, many have attempted to extract evolutionary information by studying the genus *Gloeobacter*, the earliest branching Cyanobacteria capable of this process [22,23]1. *Gloeobacter* lack traits common in photosynthetic, non-*Gloeobacter* (crown-group Cyanobacteria) indicating that they may lack traits derived within the crown-group Cyanobacteria. For example, the *Gloeobacter* do not contain thylakoid membranes, which host photosynthesis enzymes in crown-group Cyanobacteria [24,25]. In *Gloeobacter*, photosynthesis

and respiration occur in the cytoplasmic membrane [26]. *Gloeobacter* also contain a uniquely structured phycobilisome, the protein complex responsible for absorbing photons and transferring energy to the PSII reaction center. The six rods of the *Gloeobacter* phycobilisome form a single bundle whereas they are hemidiscoidal in the other crown-group Cyanobacteria [27]. Additionally, *Gloeobacter* lack PSII proteins including PsbY, PsbZ and Psb27, whereas others, including PsbO, PsbU, and PsbV, are poorly conserved [28]. As a result, *Gloeobacter* only grows slowly (23) and in low irradiance environments [29,30]. The absence of the thylakoid membrane, differences in light harvesting, and missing photosynthesis proteins help contextualize the evolution of oxygenic photosynthesis and the ecology and photochemistry of ancestral Cyanobacteria.

The Melainabacteria are an early branching sister group to the *Gloeobacter* and crown-group Cyanobacteria [10,11,31], and researchers have also interrogated their genomes for insight into the evolution of oxygenic photosynthesis [10-12,31]. Unlike the *Gloeobacter*, no known Melainabacteria have the potential for photosynthesis [10,11,31]. Therefore, the genes necessary for photosynthesis were either present in the common ancestor of Melainabacteria and Cyanobacteria and then lost in Melainabacteria and related lineages [32] or oxygenic photosynthesis evolved after the divergence of Melainabacteria and crown-group Cyanobacteria [10-12,31]. The phylogenetic space between Melainabacteria and crown-group Cyanobacteria contains an undescribed group of organisms known only from 16S rRNA gene surveys [33-36] which are either a sister group or basal to the *Gloeobacter*.

We recovered two nearly complete metagenome assembled genomes (MAGs) of a taxon within this early-diverging group from microbial mats in Lake Vanda, McMurdo Dry Valleys,

Antarctica. Here, we report on the MAGs of this organism, which we have named Candidatus *Aurora vandensis*. Based on reduced photosynthetic complex within the MAG, we propose a model that sheds light on evolutionary processes that led to increased photosynthetic efficiency through stabilization of the reaction centers and better photon harvesting systems.

2. Methods

Site Description

Lake Vanda is a perennially ice-covered lake located within Wright Valley, McMurdo Dry Valleys, Antarctica. Lake Vanda has a perennial ice cover of 3.5-4.0 m. The ice cover transmits 15-20% of incident photosynthetically active radiation [37]. Wavelengths shorter than 550 nm dominate the light spectrum because ice transmits little red light and water is particularly transparent to blue-green light [38]. Nutrient concentrations are low, and therefore there is little biomass in the water column [39]. However, benthic mats are abundant [38,40], covering the lake bottom from the base of the ice to >50 m [41]. The microbial mats are prostrate with abundant 0.1-30 cm tall pinnacles (41). They incorporate annual mud laminae. Mat surfaces have brown-purple coloration due to trapped sediment and pigments. The underlaying layers are characterized by green and purple pigmentation. The inner sections of large pinnacles are comprised of beige decomposing biomass. The dominant cyanobacterial genera based on morphological and 16S rRNA gene surveys are *Leptolynbya, Pseudanabaena, Wilmottia, Phormidium, Oscillatoria* and some unicellular morphotypes [42,43]. The microbial mats also contain diverse algae and other bacteria and archaea [40,44]. Incident irradiance penetrates millimeters into the mats, and most of the samples analyzed here were exposed to low irradiance in their natural environment [38].

Sampling and DNA extraction

To obtain samples, SCUBA divers collected benthic microbial mats and brought them to the surface in sterilized plastic containers. Pinnacles were dissected in the field using sterile technique. Subsamples were placed in Zymo Xpedition buffer (Zymo Research, Irvine, CA), and cells were lysed via bead beating in the field. The stabilized samples were then frozen on dry ice and maintained frozen in the field. Samples were transported at -80 °C to UC Davis. DNA was extracted at UC Davis using the QuickDNA Fecal/Soil Microbe kit using the manufacturer's instructions (Zymo Research, Irvine, CA, USA). The extracted DNAs were quantified using Qubit (Life Technologies) and were concentrated via evaporation until the concentration was ≥ 10 ng/uL. One bulk mat and one purple subsample were sequenced at the Joint Genome Institute (JGI).

DNA sequencing

The JGI generated sequence data using Illumina technology. An Illumina library was constructed and sequenced 2x151 bp using the Illumina HiSeq-2500 1TB platform. BBDuk (version 37.36) was used to remove contaminants, trim reads that contained adapter sequence and right quality trim reads where quality drops to 0. BBDuk was also used to remove reads that contained 4 or more 'N' bases, had an average quality score across the read less than 3 or had a minimum length ≤ 51 bp or 33% of the full read length. Reads mapped to masked human, cat, dog and mouse references at 93% identity were removed. Reads aligned to common microbial contaminants were also removed.

135 Bioinformatic analysis

136 Quality controlled, filtered raw data were retrieved from IMG Gold (JGI Gold ID GP0191362
137 and Gp0191371). Metagenomes were individually assembled using MEGAHIT [45] using a
138 minimum contig length of 500 bp and the paired end setting. Reads were mapped back to the
139 assembly using Bowtie2 [46]. A depth file was created using jgi_summarize_bam_contig_depths
140 and the assemblies were binned using MetaBAT [47]. CheckM assessed the quality of the bins
141 [48], and bins of interest were identified based on phylogenetic placement. Average nucleotide
142 identity (ANI) was calculated using the OrthoANI algorithm [49]. Protein coding regions were
143 identified by prodigal [50] within CheckM. GhostKOALA and Prokka were used to annotate
144 translated protein sequences [51,52].
145
146 When homologs of genes from the KEGG photosynthesis module were not present in the bin,
147 they were searched for in assembled, unbinned data by performing a BLASTX search with an E-
148 value cutoff of 1E-5. BLASTP was used to find the best hit for the retrieved sequences and to
149 exclude those that were not the target gene. Any sequences phylogenetically similar to *A.*
150 *vandensis* were identified based on their position in a phylogenetic gene tree constructed using
151 the methodology described below.

152 Phylogenetic inference

153 Aligned, nearly full length 16S rRNA gene sequences were collected from the Silva database
154 (v123; [53]). The recovered 16S sequence from the bulk mat was added to this alignment using
155 MAFFT [54]. A maximum likelihood tree was constructed in RAxML-HPC2 on XSEDE [55] in
156 the CIPRES Science Gateway [56]. Non-full-length sequences were added to the tree using the
157 evolutionary placement algorithm in in RAxML-HPC2 on XSEDE. Trees were rooted and

158 visualized in the interactive tree of life [57]. Maximum likelihood trees based on 16S rRNA gene
159 trees were separately constructed in MEGA7 [58]. For these trees, sequences were aligned with
160 Muscle and a maximum likelihood tree was constructed using 100 bootstrap replicates.
161
162 Concatenated marker genes from Campbell et al. [59] were retrieved as described in the anvi'o
163 workflow for phylogenomics [60]. The alignments were concatenated, and a maximum
164 likelihood tree was constructed as described above. A maximum likelihood tree was also
165 constructed for each individual ribosomal protein set. A genome tree was constructed in KBASE
166 by inserting the MAGs and published Melainabacteria genomes into a species tree using the
167 species tree builder (0.0.7; [61]).Trees were rooted and visualized in the interactive tree of life
168 [57].

169 **3. Results**

170 Assembled metagenomes contained 313-1306 Mbp in 228837-861358 contigs with a mean
171 sequence length of 1301-1669 bp. 49.6 and 53.3% of unassembled reads mapped back to the
172 assembly for the bulk and purple samples, respectively. We recovered two MAGs of a taxon
173 most closely related to *Gloeobacter,* one from each sample. The bins were 3.07 and 2.96 Mbp,
174 had a GC content of 55.4% and 55.3%, and contained 3,025 and 3,123 protein coding sequences.
175 Bins were 90.1 and 93.2% complete with 1.7 and 0.85% contamination based on marker gene
176 analysis in CheckM. GhostKOALA annotated 41.1 and 41.7% of the predicted protein coding
177 sequences. Marker gene sequences and key photosynthetic gene sequences from the bins were
178 identical or nearly identical and the genomes were 99.96% similar based on ANI.
179

The MAG is most similar to *G. violaceous* with which it had 66.8% ANI across the genome. The KBASE genome tree placed the MAGs as a sister group to the *Gloeobacter* (Figure S1a). The individual marker gene trees differed in their topologies, and the concatenated tree placed *A. vandensis* as a sister group to the *Gloeobacter* (Figures 1a and S1b). The 16S rRNA gene from the MAG was >99% similar to clones from moss pillars in an Antarctic lake (AB630682) and tundra soil (JG307085) and was 91% similar to *G. violaceous* strain PCC 7421 (NR_074282; Figure 1b). Phylogenies based on 16S rRNA gene sequences varied and placed *A. vandensis* either as branching before or sister to the *Gloeobacter* dependent on which groups were included in the analysis (Figures 1 and S1c, d). The genome-based phylogeny placed *A. vandensis* as a sister group to the *Gloeobacter*.

Based on KEGG annotations, the MAG contained homologs of all the genes necessary for carbon fixation via the Calvin Cycle. It also contained many of the genes necessary for glycolysis via the Embden-Meyerhof-Parnas pathway (EMP; missing *pfkABC*) and citrate cycle.

The MAG contained homologs of many genes associated with oxygenic photosynthesis, but *psbJ*, *psbM*, *psbT*, *psbZ*, *psbY*, *psb27*, or *psbU* from photosystem II (PSII) were missing. Similarly, homologs of *psbA* were absent from the bin, but a BLASTX search of assembled, unbinned data located a *psbA* that branches before the a Gloeobacter D1 group 4 sequence and likely belongs to the MAG. PSII genes *psbP*, *psbO,* and *psbV* conserved (Table S1). The MAG lacked homologs of genes encoding phycobilisome proteins *apcD*, *apcF*, *cpcD*, *rpcG*, *cpcG*, and any genes associated with phycoerythrocyanin (PEC) or phycoerythrin (PE) (Table 1). The PSI genes *psaI*, *psaJ*, *psaK*, and *psaX*, and the photosynthetic electron transport gene *petJ* (cytochrome c6) were

203 also absent. For each missing photosynthesis gene, no homologs were found in the assembled,
204 unbinned data that had similar phylogenetic placements to other genes in the MAG, except *psbA*.

205 **4. Discussion**

206 Genus and Species Description

207 We propose that our MAG is the first genome within a new genus. Compared to the most similar
208 genome available, *G. violaceous* strain PCC 7421, it has a 66.8% average nucleotide identity
209 (ANI) and a 91% similarity for its 16S rRNA gene. On average, genera contain taxa that are
210 96.5% similar based on 16S rRNA genes Therefore, we propose the creation of a new genus,
211 *Aurora*, which includes our MAG, *Aurora vandensis,* and numerous representatives in 16S
212 rRNA gene sequence databases. The candidate genus is named after Aurora, the goddess of the
213 dawn, to reflect its divergence from other photosynthetic Cyanobacteria near the dawn of
214 oxygenic photosyntheis and its presence in low light environments. Aurora also refers to the
215 northern and southern lights aurora borealis and aurora australis, so the name also mirrors
216 *Aurora*'s apparent preference for high latitude locations. The species, *A. vandensis,* is named
217 after Lake Vanda where the samples originated. Lake Vanda was named after a sled dog used in
218 the British North Greenland Expedition [62].
219
220 The phylogenetic placement of *A. vandensis* varies based on the genes or proteins used to
221 construct the phylogeny, the taxa included in the analysis, and the tree building algorithm (e.g.
222 Figures 1 and S1). However, it nearly always appears as sister or immediately basal to the
223 *Gloeobacter*. Aurora's family-level classification requires additional genomes to resolve.
224

To date, *Aurora* is composed of taxa from high altitude or high latitude regions including Arctic microbial mats [63], Patagonian Andes [35], Nunavut, Canada [34], The French Alps [64], and perennially ice-covered lakes in Antarctica [33] and current study; Figure 1b) and a single taxon from stromatolites in Tasmania [65]. Based on this geographic distribution, *Aurora* may be a cold adapted clade [63,66].

Metabolic Characterization of the uncultured *Aurora* genome

Aurora vandensis contains homologs for the complete complement of genes necessary for carbon fixation via the Calvin Cycle and a nearly complete pathway for glycolysis via the EMP. Many Cyanobacteria contain the genes for the EMP pathway [67] and use it to ferment glycogen under dark conditions [68,69]. *Aurora vandensis* may use this pathway to ferment glycogen during the 6 months of darkness over the Antarctic winter.

Aurora vandensis contains homologs of many many core genes necessary for oxygenic photosynthesis, but it lacks homologs encoding several extrinsic proteins in the photosystems. As such, it is likely capable of performing oxygenic photosynthesis, but at lower efficiency than the crown-groups with more diverse extrinsic proteins.

Photosystem II

Phycobilisomes harvest photons for use in PSII. These structures contain stacks of pigment proteins (biliproteins) connected by linker proteins and are anchored in to the thylakoid membrane in crown-group Cyanobacteria or into the cell membrane in the *Gloeobacter*. The pigments in the phycobilisome include a core of allophycocyanin (AP) which best captures photons at ~650 nm, surrounded by rods of phycoycanin (PC; ~620 nm), phycoerethrin (PE; maxima between 495-560 nm) and phycoerethrocyanin (PEC; 575 nm). Not all Cyanobacteria

247 use all four pigment types, instead adapting the composition of their phycobilisomes to available
248 irradiance [70].

249 *Aurora vandensis* contains homologs of the genes necessary to construct the AP core and PC
250 rods but does not contain homologs of any biliproteins associated with PE, PEC, or many of the
251 linker proteins associated with these pigments (Table S1). Therefore, we infer that *Aurora's*
252 phycobilisomes do not contain pigments that best capture energy from yellow and yellow-green
253 photons, even though the majority of irradiance available in Lake Vanda is at wavelengths at 550
254 nm or below. In contrast, less than 5% of the irradiance in the AP and PC spectral ranges, which
255 *A. vandensis* can capture, is transmitted though the ice at Lake Vanda [38].

256 We consider two possible hypotheses for the absence of PE and PEC related genes in *A.*
257 *vandensis*: 1) presence of these genes in the common ancestor of *A. vandensis* and *Gloeobacter*
258 and adaptive gene loss in *A. vandensis* or 2) absence in the common ancestor and addition only
259 in the branch containing Gloeobacter and crown-group Cyanobacteria. Gene loss would limit the
260 ability of *A. vandensis* to harvest light energy from its environment but may provide two
261 advantages. First, because other organisms in the mat contain PE, those wavelengths are
262 absorbed in the top few millimeters of the mat [38]. Thus, *A. vandensis* may use AP and PC to
263 avoid competition for light with other organisms. Second, loss of PE might protect *A. vandensis*
264 from photoinbibition. Alternately, the absence of PE in *A. vandensis* might reflect an ancestral
265 character state of oxygenic photosynthesis with limited ability to capture photon energy. Apt *et*
266 *al.,* [71] suggested that the biliproteins originated from a common ancestor, with AP being the
267 earliest branching lineage followed by the divergence of PC and PE, and finally PEC from PC.
268 They propose that the ancestor of all Cyanobacteria contained AP, PC, and PE biliproteins but
269 did not contain PEC related proteins. *Aurora vandensis* partially fits this model with the absence

of PEC. However, it also lacks PE. Thus, we propose an alternative model in which PE diverges after PC rather than simultaneously.

Aurora vandensis also lacks homologs of *apcD*, *acpF*, *cpcD*, and *rpcG*, which are structurally important to the phycobilisome and facilitate energy transfer from the antenna proteins to PSII and PSI. Knockouts of these genes in other Cyanobacteria demonstrate that they are not essential to oxygenic photosynthesis, but mutants often operate less efficiently than wildtype strains [72]. *Aurora vandensis* likely has lower effectiveness of energy transfer between the light-harvesting complex and the reaction centers relative to crown-group Cyanobacteria due to the absence of homologs of these genes. Like *Gloeobacter*, *A. vandensis* lacks homologs of *cpcG*, which encodes a phyobilisome rod-core linker protein. *Gloeobacter* also lacks this gene and instead uses *cpcJ* (Glr2806), which connects PC and AP, and *cpeG* (Glr1268*),* which connects PC and PE. These genes allow energy transfer from PC and AP to the reaction center [28,73]. *Aurora vandensis* contains sequences ~43-58% similar to these genes, but we cannot determine if they serve the same function.

Overall, *Aurora vandensis* can likely capture irradiance for growth, but does so less efficiently than crown-group Cyanobacteria. The absence of homologs of PE creates a mismatch between available irradiance and photo capture optima, which likely limits energy transfer between the antennae proteins and the reaction centers in *A. vandensis*.

Energy flows from phocobilisomes to PSII reaction centers and excites P680, which contains the D1 and D2 reaction center dimers (*psbA* and *psbD*). This process oxidizes water and releases oxygen at the oxygen evolving complex (OEC). The reaction center also contains homologs of chlorophyll apoproteins CP43 and CP47 (*psbC* and *psbB*) and two subunits of cytochrome b559

(*psbE* and *psbF*). Other common subunits support the OEC (e.g. *psbO*, *psbV*, *psbU*) or facilitate electron flow through the reaction center.

The *A. vandensis* MAG contains homologs of all the main subunits for the PSII reaction center including the D1 and D2 proteins (Table S1). It contains homologs of *psbA* and *psbD* genes that are 91% similar to those of *G. violaceus* (WP_023172020 and WP_011142319). The translated *psbA* sequence produces a D1 protein within Group 4 [74]. Group 4 D1 proteins include all the "functional," non-rogue D1 proteins, and all phototrophic Cyanobacteria possess a protein within this group [74].

The *A. vandensis* genome lacks a homolog of *psbM*, which helps stabilize the PSII D1/D2 dimer. However, the D1/D2 dimer still forms in the absence of PsbM in crown-group Cyanobacteria [75]. Therefore, it is unlikely that the lack of this protein prevents *A. vandensis* from forming a stable PSII reaction center. It also lacks a homolog of *psbJ*, which regulates the number of PSII reaction centers in the thylakoid membrane [76]. Mutants missing *psbJ* have less stable D1/D2 dimers and lower rates of oxygen production than wildtype strains [77]. Although *A. vandensis* may be less efficient without these genes, their absence is unlikely to prevent it from performing oxygenic photosynthesis.

When P680 reduces pheophytin a, it triggers water to donate an electron to P680 and return it to its ground redox state. Repeated four times, this process splits water into O_2 and H^+ at the OEC. The OEC is composed of a Mn_4CaO_5 cluster bound to D1, D2, CP47 and CP43 proteins. It also contains extrinsic proteins, including PsbO, PsbU, and PsbV, which help to support the OEC and provide a geochemical environment that is conducive to water oxidation [78].

313 The translated D1 and D2 proteins from *A. vandensis* contain all of the D1 amino acid Mn_4CaO_5
314 ligands described previously (Asp170, Glu333, Glu189, Asp342, Ala344, His332, His 337, and
315 Ala344; [79] and the D2 Glu69 ligand [80]. The gene encoding PsbO is poorly conserved in *A.*
316 *vandensis* and is only 46% similar similar to PsbO in *Gloeobacter* and 36% or less similar to
317 those in other crown-group Cyanobacteria compared with ~55% or greater similarity among
318 crown-group Cyanobacteria. Despite this, PsbO in *A. vandensis* contains all the features
319 necessary to interact with other PSII proteins and the D1, D2, CP43 and C47 subunits [81].
320 Therefore, the *A. vandensis* PsbO likely helps stabilize the Mn_4CaO_5 cluster and support the OEC
321 despite the lack of sequence similarity. Similarly, PsbV in *A. vandensis* is dissimilar to that in
322 crown-group Cyanobacteria, *Synechocystis* sp. PCC 6803 mutants that lack this gene are capable
323 of evolving oxygen [82,83]. *Aurora vandensis* appears to be missing homologs of a gene
324 encoding PsbU which stabilizes the OEC [84]. Cyanobacterial mutants missing *psbU* have
325 decreased energy transfer between AP and PSII [85], are highly susceptible to photoinhibition,
326 have decreased light utilization under low-light conditions, and have lowered oxygen evolution
327 and electron donation rates than the wildtype [86]. In addition, the OEC becomes significantly
328 more labile [86].

329 PsbO, PsbU, PsbV, a region of the D1, and other extrinsic proteins help control the concentration
330 of Cl^-, Ca^{2+}, and H^+ and create an environment that is amenable to water oxidation [87-89].
331 Specifically, chloride may be involved in removing protons from the OEC [90]. Although PsbU
332 is missing in *A. vandensis*, the other proteins conserve important residues. For example, the D1
333 chloride ligand site Asn338 is conserved in the translated *psbA*, but the sequence is not long
334 enough to determine if Glu354 is also conserved. Similarly, the translated *psbO* contains Glu54,

335 Glu114, and His231 residues that bind with Ca^{2+} [91], suggesting some Cl^- and Ca^{2+} regulation
336 capabilities in *A. vandensis*.

337 *Cytochrome b6f*

338 Once through PSII, the electrons move through an electron transport chain, and pass through the
339 cytochrome b6f complex, which pumps protons across the membrane. This process creates a
340 proton gradient that is used to generate ATP. The cytochrome b6f complex is composed of eight
341 subunits. The *A. vandensis* genome contains homologs of genes encoding five of these subunits,
342 including the four large subunits, PetA, PetB, PetC, PetD, PetM and the small subunit PetG.
343 However, it appears to be missing *petL* and *petN*. A *Synechocystis* mutant was able to grow
344 photoautotrophcally without *petL* but the rate of oxygen evolution was reduced [92]. Deletion of
345 *petN* prevents plants from photosynthesizing [93,94]. These results have been interpreted to
346 mean that *petN* is necessary for photosynthesis in plants and Cyanobacteria [92,95] but attempts
347 to delete *petN* in Cyanobacteria have been unsuccessful [92] so it is not possible to determine
348 what effect its absence may have on electron transport in *A. vandensis*. Overall, the absence of
349 these genes may cause *A. vandensis* to transfer energy less efficiently than other Cyanobacteria
350 but likely does not prohibit it from performing oxygenic photosynthesis or aerobic respiration.

351 Cytochrome b6f is restricted to crown-group Cyanobacteria, *Gloeobacter*, and *Aurora*. The
352 Melainabacteria and Sericytochromatia contain multiple aerobic respiratory pathways, but do not
353 contain cytochrome b6f. This has been interpreted as evidence that these three classes
354 independently acquired aerobic respiration [12]. Based on the presence of cytochrome b6f in
355 *Aurora* we infer that aerobic respiration evolved before the divergence of *Aurora* from
356 *Gloeobacter,* and thus the ability to perform oxygenic photosynthesis also predated this
357 divergence.

Photosystem I

The end of the electron transport chain is either plastocyanin or cytochrome c6, which donate electrons to P700 in PSI. *Aurora vandensis* contains homologs of genes necessary to produce plastocyanin, but lacks homologs of *petJ,* which codes for cytochrome c6, so plastocyanain is the final electron carrier delivering electrons to PSI in *A. vandensis*.

Photosystem I in *A. vandensis* is similar to that in *Gloeobacter*. Both contain all the main subunits for PSI, but lack homologs of several genes including *psaI*, *psaJ*, *psaK*, and *psaL* that are present in crown-group Cyanobacteria. In addition, both contain homologs of many genes involved in chlorophyll biosynthesis. Therefore, PSI in *A. vandensis* likely functions similarly to PSI in *Gloeobacter*

Photoprotection

Cyanobacteria can experience photoinhibition under high light conditions when photon absorption outstrips the ability to dissipate electrons through photochemical pathways, and reactive oxygen species accumulate at the PSII reaction center. These reactive species damage photosynthetic machinery, especially the D1 protein, which requires reassembly proteins (96-98)**.** Cyanobacteria protect themselves from photoinhibition in two key ways. First, they use orange carotenoid proteins (OCP) as receptors to reduce the amount of energy transferred from the phycobilisome to PSII and PSI [96]. The *A. vandensis* genome contains two copies of a gene coding for a protein 68% similar to the OCP in *G. violaceous.* The OCP interacts directly with the phycobilisome [96]. Thus, the sequence differences may reflect structural differences in the phycobilisomes of *A. vandensis* and *G. violaceous.*

379 Cyanobacteria also protect themselves from photoinhibition using high light inducible proteins
380 (HLIP) to dissipate energy. *Aurora vandensis* contains homologs of genes for three proteins that
381 are 69-85% similar to HLIP in *G. violaceous.* We hypothesize that these genes act as HLIP and
382 protect *A. vandensis* against photoinhibition.

383 Despite containing mechanisms for photoprotection, *A. vandensis* occupies a low-irradiance
384 environment in Lake Vanda, particularly in the wavelengths absorbed by its biliproteins.
385 Similarly, many other *Aurora* taxa originated from low irradiance environments. For example,
386 one was collected from Hotoke-Ike where only 20-30% of incident PAR reaches the lake bed
387 [97]. 16S rRNA gene sequences were found at 1 cm depth in sediments [98] where they were
388 protected from light. Additionally, biomass may shield *Aurora* from irradiance in soil crusts in
389 Greenland [36]. *Gloeobacter* are also sensitive to high irradiance [24] and if both *Gloeobacter*
390 and *A. vandensis* are low-light adapted, this may be an ancestral trait of the Cyanobacteria.

391 Conceptual model of the evolution of Cyanobacteria and photosynthesis
392 The exact phylogenetic placement of *Aurora* is uncertain and diverged before the divergence of
393 *Gleobacter* and crown-group Cyanobacteria or is a sister group to the *Gloeobacter*. *Aurora*
394 *vandensis* lacks many of the photosynthetic genes present in photosynthetic Cyanobacteria which
395 may resemble the gene content of the ancestor of it and other Cyanobacteria. Based on these
396 traits, we propose a model for progressive evolutionary stabilization of early oxygenic
397 photosynthesis. Alternative models calling on gene loss or horizontal gene transfer (HGT) can
398 also explain differences among *Aurora*, *Gloeobacter* and crown-group Cyanobacteria (Figure 2b,
399 c).

For the progressive evolutionary stabilization model core photosynthetic domains were present in Cyanobacteria prior to the divergence of *Aurora* and *Gloeobacter* and were stabilized and became more efficient through the course of evolutionary time in some lineages (Figure 2a). This model predicts that the common ancestor of *Aurora*, *Gloeobacter*, and crown-group Cyanobacteria contained genes encoding core photosynthetic proteins including PsbA, PsbD, PsaA, PsaB, extrinsic proteins including PsbO, PsbM, and PsbV, and the AP and PC biliproteins (Figure 2a). Many of these genes appear to be essential for photosynthesis and were likely present in the common ancestor of all oxygenic phototrophs, possibly before PSII and PSI were linked to perform oxygenic photosynthesis. After the divergence of *Aurora* from *Gloeobacter* and crown-group Cyanobacteria, extrinsic proteins evolved to stabilize the reaction centers, improve water splitting, improve the flow of electrons through the reaction centers, and aid in the assembly of the reaction center. The lineage also expanded its ability to capture photons with the evolution of PE (Figure 2a). Finally, between the divergence of *Gloeobacter* and diversification of crown-group Cyanobacteria, additional extrinsic proteins were added to PSII, PEC was added to the phycobilisome, and PsaIJK and PsaX were added to PSI (Figure 2a). These reflect continued stabilization, and many may have been associated with the evolution and stabilization of the thylakoid membrane. In this model, each protein addition is predicted to increase the efficiency of oxygenic photosynthesis and be driven by selection processes.

Many alternative evolutionary models exist that rely on gene loss or HGT to explain the distribution of photosynthetic genes in *Aurora*, *Gloeobacter*, and crown-group Cyanobacteria. End-members models include one that relies exclusively on gene loss and another that relies on HGT (Figures 2b, c). In both models, core and extrinsic photosystems genes and much of the ETC and aerobic respiratory pathways were present in the common ancestor of the crown-group

Cyanobacteria, *Aurora*, and *Gloeobacter* (Figure 2b, c). This organism also possessed AP, PC, and linker proteins for the phycobilisome. In the gene loss model, the common ancestor also contained the genes for additional extrinsic proteins in PSII, PsaIJK and PsaX in PSI, and PE. These genes were then lost in *Aurora* (Figure 2b). In the HGT model, this suite of genes evolved independently either within the *Gloeobacter* or between the divergence of *Gloeobacter* and the diversification of crown-group Cyanobacteria. The genes were then transferred between these two groups, but not into *Aurora*. Horizontal transfer appears more parsimonious than gene loss because a single HGT event can transfer multiple photosynthetic genes [99,100] and the transfer of beneficial traits between *Gloeobacter* and crown-group Cyanobacteria seems more likely than their loss.

Aurora branches before the divergence of *Gloeobacter* and crown-group Cyanobacteria (Figure 2a) is most parsimonious with the emergence of oxygenic photosynthesis, a new metabolism capable of generating large amounts of chemical energy from light energy but at the expense of significant metabolic machinery damage. Through time, evolutionary pressures led to progressive increases in stability and productivity in some lineages, which allowed the expansion of early Cyanobacteria into environments with greater irradiance. Based on this model, we predict that ancestral lineages that emerged prior to the GOE may have needed to occupy low irradiance habitats due to photoinhibition, and high UV doses that would have accompanied other wavelengths in the pre-oxygenated atmosphere. As the photosystems stabilized, photon capture efficiency improved, and oxygenic phototrophs expanded to higher-light environments. Both would have resulted in significantly higher primary productivity and rates of oxygen production.

445 Importance of *Aurora vandensis*

446 The crown-group Cyanobacteria diversified between 2.3 and 1.9 billion years ago [14],
447 approximately 600 to 900 million years after the divergence of the phototrophic Cyanobacteria
448 and the Melainabacteria [14]. The only characterized lineages that diverged within this interval
449 are *G. violaceous*, *G. kilauensis*, which diverged 2.2 to 2.6 billion years ago, and now *A.*
450 *vandensis,* with *A. vandensis* potentially diverging between the Melainabacteria and *Gloeobacter*.
451 If basal to the *Gloeobacter,* this new genome provides key insight into the evolutionary
452 processes occurring over the 300-650 million years [14,104] spanning the invention of the most
453 transformative metabolism on Earth, oxygenic photosynthesis. Thus, the genome of *A. vandensis*
454 is particularly important for contextualizing this innovation. Specifically, an evolutionary model
455 in which *Aurora* is basal to *Gloeobacter* (Figure 2a) is parsimonious with the emergence of
456 oxygenic photosynthesis as a new metabolism capable of generating substantial chemical energy
457 from light but at the expense of significant metabolic machinery damage. Thus, early
458 cyanobacterial lineages may have inhabited only low irradiance habitats due to photoinhibition.
459 Through time, evolutionary selection led to progressive increases in stability and productivity,
460 which allowed expansion of Cyanobacteria into environments with greater irradiance. As
461 photosystems stabilized, photon capture efficiency also improved, increasing primary
462 productivity. Eventually, habitat expansion and improvements in efficiency allowed
463 Cyanobacteria to produce enough oxygen to cause oxidative weathering [101] and finally trigger
464 the GOE [3].

465 Low photosynthetic efficiency in early Cyanobacteria can reconcile models that predict rapid
466 oxidation of Earth's surface [102] with the geological record, which shows whiffs of oxygen

467 before the GOE [4,103]. If our evolutionary model is correct, cyanobacterial oxygen production
468 could have initiated long before oxygen accumulated in the oceans and environment.

469 **Acknowledgements**

470 Sequencing was provided by the U.S. Department of Energy Joint Genome Institute, a DOE
471 Office of Science User Facility, and is supported under Contract No. CSP502867. Samples
472 used in this project were collected during a field season supported by the New Zealand
473 Foundation for Research, Science and Technology (grant number CO1X0306) with field
474 logistics provided by Antarctica New Zealand (project K-081). Salary support for CG was
475 provided by the Massachusetts Institute of Technology node of the NASA Astrobiology
476 Institute.

477

478 **Competing Interests**

479 The authors declare that they have no competing financial interests.

480

481 **References**

482 1. Karhu JA, Holland HD. Carbon isotopes and the rise of atmospheric oxygen. Geology
483 1996; **24**:867–870.

484 2. Kump LR. The rise of atmospheric oxygen. Nature 2008; **451**:277–278.

485 3. Lyons TW, Reinhard CT, Planavsky NJ. The rise of oxygen in Earth's early ocean and
486 atmosphere. Nature 2014; **506**:307–315.

4. Planavsky NJ, Reinhard CT, Wang X, Thomson D, McGoldrick P, Rainbird RH, et al. Earth history. Low mid-Proterozoic atmospheric oxygen levels and the delayed rise of animals. Science 2014; **346**:635–8.

5. Reinhard CT, Planavsky NJ, Robbins LJ, Partin CA, Gill BC, Lalonde SV, et al. Proterozoic ocean redox and biogeochemical stasis. Proc. Natl. Acad. Sci. U.S.A. 2013; **110**:5357–5362.

6. Summons RE, Bradley AS, Jahnke LL, Waldbauer JR. Steroids, triterpenoids and molecular oxygen. Philos Trans R Soc London B Biol Sci 2006; **361**:951–968.

7. Blankenship RE. Early evolution of photosynthesis. Plant Physiol. 2010; **154**:434–438.

8. Crowe SA, Døssing LN, Beukes NJ, Bau M, Kruger SJ, Frei R, et al. Atmospheric oxygenation three billion years ago. Nature 2013; **501**:535–538.

9. Mulkidjanian AY, Koonin EV, Makarova KS, Mekhedov SL, Sorokin A, Wolf YI, et al. The cyanobacterial genome core and the origin of photosynthesis. Proc Natl Acad Sci USA 2006; **103**:13126–13131.

10. Di Rienzi SC, Sharon I, Wrighton KC, Koren O, Hug LA, Thomas BC, et al. The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. eLife 2013; **2**:611–625.

11. Soo RM, Skennerton CT, Sekiguchi Y, Imelfort M, Paech SJ, Dennis PG, et al. An Expanded Genomic Representation of the Phylum Cyanobacteria. Genome Biology and Evolution 2014; **6**:1031–1045.

12. Soo RM, Hemp J, Parks DH, Fischer WW, Hugenholtz P. On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. Science 2017; **355**:1436–1440.

13. Cardona T. Evolution of Photosynthesis. Chichester, UK: John Wiley & Sons, Ltd; 2001.

14. Magnabosco C, Moore KR, Wolfe JM, Fournier GP. Dating phototrophic microbial lineages with reticulate gene histories. Geobiology 2018; **16**:179–189.

15. Mathis P. Compared structure of plant and bacterial photosynthetic reaction centers. Evolutionary implications. Biochim et Biophysica Acta 1990; **1018**:163–167.

16. Hohmann-Marriott MF, Blankenship RE. Evolution of photosynthesis. Annu. Rev. Plant Biol. 2011; **62**:515–548.

17. Olson JM. The Evolution of Photosynthesis. Science 1970; **168**:438–446.

18. Olson JM. Evolution of photosynthetic reaction centers. BioSystems 1981; **14**:89–94.

19. Olson JM. Evolution of Photosynthesis' (1970), re-examined thirty years later. Photosynth Res 2001; **68**:95–112.

20. Sousa FL, Shavit-Grievink L, Allen JF, Martin WF. Chlorophyll Biosynthesis Gene Evolution Indicates Photosystem Gene Duplication, Not Photosystem Merger, at the Origin of Oxygenic Photosynthesis. Genome Biol and Evol 2012; **5**:200–216.

21. Cardona T. A fresh look at the evolution and diversification of photochemical reaction centers. Photosynth Res 2015; **126**:1–24.

22. Turner S, Pryer KM, Miao VP, Palmer JD. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. J. Eukaryot. Microbiol. 1999; **46**:327–338.

23. Gupta RS. Protein signatures (molecular synapomorphies) that are distinctive characteristics of the major cyanobacterial clades. Int J Syst Evol Microbiol 2009; **59**:2510–2526.

24. Rippka R, Waterbury J, Cohen-Bazire G. A cyanobacterium which lacks thylakoids. Archives of Microbiology 1974; **100**:419–436.

25. Saw JHW, Schatz M, Brown MV, Kunkel DD, Foster JS, Shick H, et al. Cultivation and Complete Genome Sequencing of Gloeobacter kilaueensis sp. nov., from a Lava Cave in Kīlauea Caldera, Hawai'i. PLoS ONE 2013; 8:e76376–12.

26. Rexroth S, Mullineaux CW, Ellinger D, Sendtko E, Rögner M, Koenig F. The Plasma Membrane of the Cyanobacterium Gloeobacter violaceusContains Segregated Bioenergetic Domains. Plant Cell 2012; **23**:2379–2390.

27. Guglielmi G, Cohen-Bazire G, Bryant DA. The structure of Gloeobacter violaceus and its phycobilisomes. Archives of Microbiology 1981; **129**:181–189.

28. Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, et al. Complete genome structure of Gloeobacter violaceus PCC 7421, a cyanobacterium that lacks thylakoids. DNA Res. 2003; **10**:137–145.

29. Koenig F, Schmidt M. Gloeobacter violaceus - investigation of an unusual photosynthetic apparatus. Absence of the long wavelength emission of photosystem I in 77 K fluorescence spectra. Physiologia Plantarum 1995; **94**:621–628.

30. Mimuro M, Tomo T, Tsuchiya T. Two unique cyanobacteria lead to a traceable approach of the first appearance of oxygenic photosynthesis. Photosynth Res 2008; **97**:167–176.

31. Soo RM, Woodcroft BJ, Parks DH, Tyson GW, Hugenholtz P. Back from the dead; the curious tale of the predatory cyanobacterium Vampirovibrio chlorellavorus. PeerJ 2015; **3**:e968–22.

32. Cardona Londono T. Origin of water oxidation at the divergence of Type I and Type II photochemical reaction centres. 2017; **4**:1–11.

33. Nakai R, Abe T, Baba T, Imura S, Kagoshima H, Kanda H, et al. Microflorae of aquatic moss pillars in a freshwater lake, East Antarctica, based on fatty acid and 16S rRNA gene analyses. Polar Biol 2011; **35**:425–433.

34. Lynch MDJ, Bartram AK, Neufeld JD. Targeted recovery of novel phylogenetic diversity from next-generation sequence data. ISME J 2012; **6**:2067–2077.

35. Elser JJ, Bastidas Navarro M, Corman JR, Emick H, Kellom M, Laspoumaderes C, et al. Community Structure and Biogeochemical Impacts of Microbial Life on Floating Pumice. Appl Environ Microbiol 2015; **81**:1542–1549.

563 36. Pushkareva E, Pessi IS, Wilmotte A, Elster J. Cyanobacterial community composition in
564 Arctic soil crusts at different stages of development. FEMS Microbiol Ecol 2015;
565 **91**:fiv143.

566 37. Howard-Williams C, Schwarz A-M, Hawes I, Priscu JC. Optical Properties of the
567 Mcmurdo Dry Valley Lakes, Antarctica. In: Ecosystem Dynamics in a Polar Desert: the
568 Mcmurdo Dry Valleys, Antarctica. Washington, D. C.: American Geophysical Union;
569 2013. pp 189–203.

570 38. Hawes I, Schwarz A-M. Absorption and utilization of irradiance by cyanobacterial mats
571 in two ice-covered antarctic lakes with contrasting light climates. J Phycol 2001; **37**:5-15.

572 39. Vincent WF, Vincent CL. Factors Controlling Phytoplankton Production in Lake Vanda
573 (77°S). Can J Fish Aquat Sci 1982; **39**:1602–1609.

574 40. Love FG, Simmons GM Jr., Parker BC, Wharton RA Jr., Seaburg KG. Modern
575 conophyton-like microbial mats discovered in Lake Vanda, Antarctica. Geomicrobiol J
576 2009; **3**:33–48.

577 41. Mackey TJ, Sumner DY, Hawes I, Jungblut AD. Morphological signatures of microbial
578 activity across sediment and light microenvironments of Lake Vanda, Antarctica.
579 Sediment Geol 2017; **361**:82–92.

580 42. Zhang L, Jungblut AD, Hawes I, Andersen DT, Sumner DY, Mackey TJ. Cyanobacterial
581 diversity in benthic mats of the McMurdo Dry Valley lakes, Antarctica. Polar Biol 2015;
582 **38**:1097–110.

583 43. Sumner DY, Jungblut AD, Hawes I, Andersen DT, Mackey TJ, Wall K. Growth of
584 elaborate microbial pinnacles in Lake Vanda, Antarctica. Geobiology 2016; **14**:556–574.

585 44. Kaspar M, Simmons GM, Parker BC, Seaburg KG, Wharton RA, Smith RIL. Bryum
586 Hedw. Collected from Lake Vanda, Antarctica. The Bryologist 1982; **85**:424-430.

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

590 46. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods
591 2012; **9**:357–359.

592 47. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately
593 reconstructing single genomes from complex microbial communities. PeerJ 2015;
594 **3**:e1165.

595 48. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing
596 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
597 Genome Res. 2015; **25**:1043–1055.

598 49. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to
599 calculate average nucleotide identity. Antonie van Leeuwenhoek 2017; **110**:1281–1286.

600 50. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal:
601 prokaryotic gene recognition and translation initiation site identification. BMC
602 Bioinformatics 2010; **11**:119.

603 51. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for
604 Functional Characterization of Genome and Metagenome Sequences. J. Mol. Biol. 2016;
605 **428**:726–731.

606 52. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;
607 **30**:2068–2069.

608 53. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA
609 ribosomal RNA gene database project: improved data processing and web-based tools.
610 Nucl. Acids Res. 2013; **41**:D590–596.

611 54. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
612 improvements in performance and usability. Mol Biol and Evol 2013; **30**:772–780.

613 55. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
614 thousands of taxa and mixed models. Bioinformatics 2006; **22**:2688–2690.

615 56. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for
616 inference of large phylogenetic trees. IEEE; 2010. pp 1–8.

617 57. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree
618 display and annotation. Bioinformatics 2007; **23**:127–8.

619 58. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis
620 Version 7.0 for Bigger Datasets. Mol Bioland Evol 2016; **33**:1870–1874.

621 59. Campbell BJ, Yu L, Heidelberg JF, Kirchman DL. Activity of abundant and rare
622 bacteria in a coastal ocean. Proc Natl Acad Sci USA 2011; **108**:12776–81271.

623 60. Eren AM, Esen ÖC, Quince C, Vineis JH, Sogin ML, Delmont TO. Anvi'o: An
624 advanced analysis and visualization platform for 'omics data. Peer J 2015;
625 10.7717/peerj.1319

626 61. Allen B, Drake M, Harris N, Sullivan T. Using KBase to Assemble and Annotate
627 Prokaryotic Genomes. Curr Protoc Microbiol 2017; **46**:1E.13.1–1E.13.18.

628 62. Chinn T, Mason P. The first 25 years of the hydrology of the Onyx River, Wright Valley,
629 Dry Valleys, Antarctica. Polar Record 2015; **52**:16–65.

630 63. Jungblut AD, Lovejoy C, Vincent WF. Global distribution of cyanobacterial
631 ecotypes in the cold biosphere. ISME J 2010; **4**:191–202.

632 64. Billard E, Domaizon I, Tissot N, Arnaud F, Lyautey E. Multi-scale phylogenetic
633 heterogeneity of archaea, bacteria, methanogens and methanotrophs in lake sediments.
634 Hydrobiologia 2015; **751**:159–73.

635 65. Proemse BC, Eberhard RS, Sharples C, Bowman JP, Richards K, Comfort M, et al.
636 Stromatolites on the rise in peat-bound karstic wetlands. Sci Rep 2017; **7**:15384.

637 66. Chrismas NAM, Anesio AM, Sánchez-Baracaldo P. Multiple adaptations to polar and
638 alpine environments within cyanobacteria: a phylogenomic and Bayesian approach.
639 Front. Microbiol. 2015; **6**:3041–3010.

640 67. Chen X, Schreiber K, Appel J, Makowka A, Fähnrich B, Roettger M, et al. The Entner-
641 Doudoroff pathway is an overlooked glycolytic route in cyanobacteria and plants. Proc.
642 Natl. Acad. Sci. U.S.A. 2016; **113**:5441–5446.

643 68. Moezelaar R, Stal LJ. Fermentation in the unicellular cyanobacterium Microcystis
644 PCC7806. Archives of Microbiology 1994; **162**:63–69.

645 69. Moezelaar R, Bijvank SM, Stal LJ. Fermentation and Sulfur Reduction in the Mat-
646 Building Cyanobacterium Microcoleus chthonoplastes. Appl EnvironMicrobiol 1996;
647 **62**:1752–1758.

648 70. Bogorad L. Phycobiliproteins and Complementary Chromatic Adaptation. Annu Rev
649 Plant Physiol 1975; **26**:369–401.

650 71. Apt KE, Collier JL, Grossman AR. Evolution of the phycobiliproteins. J. Mol. Biol.
651 1995; **248**:79–96.

652 72. Ashby MK, Mullineaux CW. The role of ApcD and ApcF in energy transfer from
653 phycobilisomes to PS I and PS II in a cyanobacterium. Photosynth Res 1999; **61**:169–
654 179.

655 73. Koyama K, Tsuchiya T, Akimoto S, Yokono M, Miyashita H, Mimuro M. New linker
656 proteins in phycobilisomes isolated from the cyanobacterium Gloeobacter violaceus
657 PCC 7421. FEBS Letters 2006; **580**:3457–3461.

658 74. Cardona T. A fresh look at the evolution and diversification of photochemical reaction
659 centers. Photosynth Res 2015; **126**:1–24.

660 75. Kawakami K, Umena Y, Iwai M, Kawabata Y, Ikeuchi M, Kamiya N, et al. Roles of
661 PsbI and PsbM in photosystem II dimer formation and stability studied by deletion
662 mutagenesis and X-ray crystallography. Biochim. Biophys. Acta 2011; **1807**:319–325.

663 76. Lind LK, Shukla VK, Nyhus KJ, Pakrasi HB. Genetic and immunological analyses of
664 the cyanobacterium Synechocystis sp. PCC 6803 show that the protein encoded by the
665 psbJ gene regulates the number of photosystem II centers in thylakoid membranes. J.
666 Biol. Chem. 1993; **268**:1575–9.

667 77. Sugiura M, Iwai E, Hayashi H, Boussac A. Differences in the interactions between the
668 subunits of photosystem II dependent on D1 protein variants in the thermophilic
669 cyanobacterium Thermosynechococcus elongatus. J. Biol. Chem. 2010; **285**:30008–
670 30018.

671 78. Umena Y, Kawakami K, Shen J-R, Kamiya N. Crystal structure of oxygen-evolving
672 photosystem II at atomic resolution. Acta Cryst 2010; **66**:s124–125.

673 79. Murray JW. Sequence variation at the oxygen-evolving centre of photosystem II: a new
674 class of "rogue" cyanobacterial D1 proteins. Photosynth Res 2011; **110**:177–184.

675 80. Vermaas W, Charité J, Shen GZ. Glu-69 of the D2 protein in photosystem II is a
676 potential ligand to Mn involved in photosynthetic oxygen evolution. Biochemistry 1990;
677 **29**:5325–5332.

678 81. Koyama K, Suzuki H, Noguchi T, Akimoto S, Tsuchiya T, Mimuro M. Oxygen
679 evolution in the thylakoid-lacking cyanobacterium Gloeobacter violaceus PCC 7421.
680 Biochimica et Biophysica Acta (BBA) - Bioenergetics 2008; **1777**:369–378.

681 82. Shen JR, Burnap RL, Inoue Y. An independent role of cytochrome c-550 in
682 cyanobacterial photosystem II as revealed by double-deletion mutagenesis of the psbO
683 and psbV genes in Synechocystis sp. PCC 6803. Biochemistry 1995; **34**:12661–12668.

684 83. Shen JR, Qian M, Inoue Y, Burnap RL. Functional characterization of Synechocystis sp.
685 PCC 6803 delta psbU and delta psbV mutants reveals important roles of cytochrome c-
686 550 in cyanobacterial oxygen evolution. Biochemistry 1998; **37**:1551–1558.

687 84. Nishiyama Y, Los DA, Hayashi H, Murata N. Thermal protection of the oxygen-
688 evolving machinery by PsbU, an extrinsic protein of photosystem II, in Synechococcus
689 species PCC 7002. Plant Physiol. 1997; **115**:1473–1480.

690 85. Veerman J, Bentley FK, Eaton-Rye JJ, Mullineaux CW, Vasil'ev S, Bruce D. The PsbU
691 subunit of photosystem II stabilizes energy transfer and primary photochemistry in the
692 phycobilisome-photosystem II assembly of Synechocystis sp. PCC 6803. Biochemistry
693 2005; **44**:16939–16948.

694 86. Inoue-Kashino N, Kashino Y, Satoh K, Terashima I, Pakrasi HB. PsbU Provides a
695 Stable Architecture for the Oxygen-Evolving System in Cyanobacterial Photosystem II †.
696 Biochemistry 2005; **44**:12214–12228.

697 87. Chu HA, Nguyen AP, Debus RJ. Amino acid residues that influence the binding of
698 manganese or calcium to photosystem II. 2. The carboxy-terminal domain of the D1
699 polypeptide. Biochemistry 1995; **34**:5859–5882.

700 88. Guskov A, Gabdulkhakov A, Broser M, Glöckner C, Hellmich J, Kern J, et al. Recent
701 progress in the crystallographic studies of photosystem II. Chem physchem 2010;
702 **11**:1160–71.

703 89. Umena Y, Kawakami K, Shen JR, Kamiya N. Crystal structure of oxygen-evolving
704 Photosystem II at 1.9 angstrom resolution. Nature 2011; **473**:55-60

90. Yocum C. The calcium and chloride requirements of the O2 evolving complex. Coordination Chemistry Reviews 2008; **252**:296–305.

91. Murray JW, Barber J. Identification of a Calcium-Binding Site in the PsbO Protein of Photosystem II. Biochemistry 2006; **45**:4128–4130.

92. Schneider D, Volkmer T, Rögner M. PetG and PetN, but not PetL, are essential subunits of the cytochrome b6f complex from Synechocystis PCC 6803. Res. Microbiol. 2007; **158**:45–50.

93. Hager M, Biehler K, Illerhaus J, Ruf S, Bock R. Targeted inactivation of the smallest plastid genome-encoded open reading frame reveals a novel and essential subunit of the cytochrome b(6)f complex. EMBO J. 1999; **18**:5834–5842.

94. Schwenkert S, Legen J, Takami T, Shikanai T, Herrmann RG, Meurer J. Role of the low-molecular-weight subunits PetL, PetG, and PetN in assembly, stability, and dimerization of the cytochrome b6f complex in tobacco. Plant Physiol. 2007; **144**:1924–1935.

95. Bernát G, Rögner M. Center of the Cyanobacterial Electron Transport Network: The Cytochrome b 6 f Complex. In: Bioenergetic Processes of Cyanobacteria. Dordrecht: Springer Netherlands; 2011. pp 573–606.

96. Kirilovsky D. Photoprotection in cyanobacteria: the orange carotenoid protein (OCP)-related non-photochemical-quenching mechanism. Photosynth Res 2007; **93**:7–16.

724 97. Tanabe Y, Ohtani S, Kasamatsu N, Fukuchi M, Kudoh S. Photophysiological responses
725 of phytobenthic communities to the strong light and UV in Antarctic shallow lakes.
726 Polar Biol 2009; **33**:85–100.

727 98. Stoeva MK, Aris-Brosou S, Chételat J, Hintelmann H, Pelletier P, Poulain AJ. Microbial
728 community structure in lake and wetland sediments from a high Arctic polar desert
729 revealed by targeted transcriptomics. PLoS ONE 2014; **9**:e89531.

730 99. Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. Transfer of
731 photosynthesis genes to and from Prochlorococcus viruses. Proc Natl Acad Sci USA
732 2004; **101**:11013–11018.

733 100. Monier A, Pagarete A, de Vargas C, Allen MJ, Read B, Claverie J-M, et al. Horizontal
734 gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA
735 virus. Genome Res. 2009; **19**:1441–1449.

736 101. Lalonde SV, Konhauser KO. Benthic perspective on Earth's oldest evidence for
737 oxygenic photosynthesis. Proc Natl Acad Sci USA 2015; **112**:995–1000.

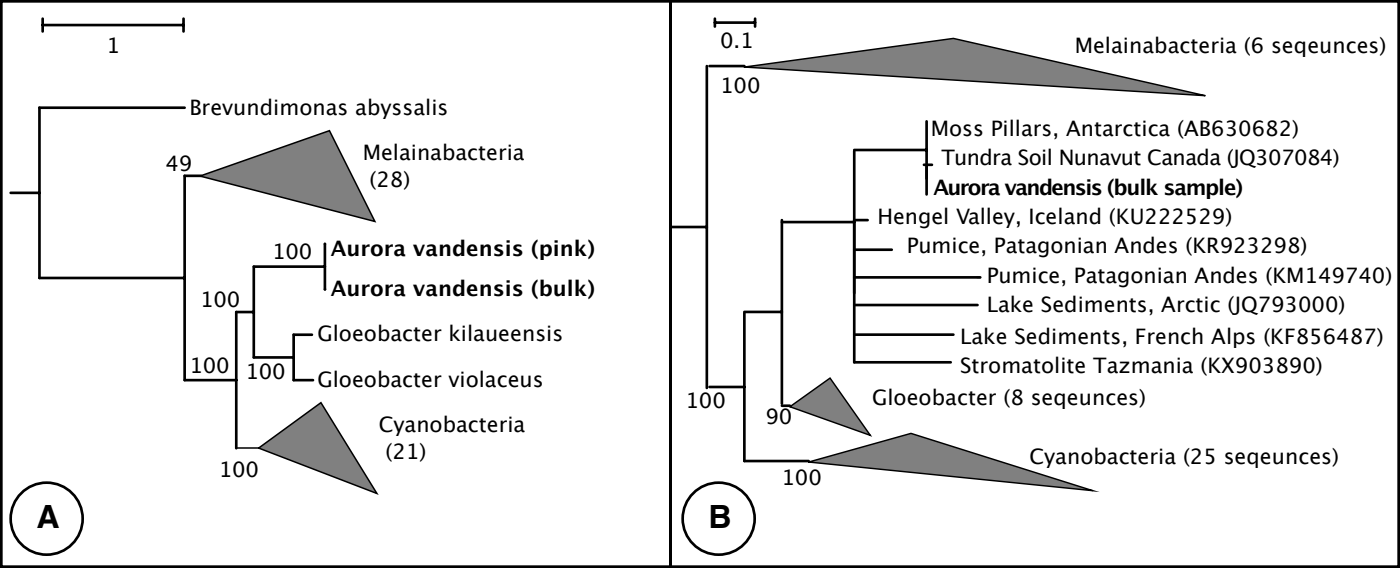
738 102. Ward LM, Kirschvink JL, Fischer WW. Timescales of Oxygenation Following the
739 Evolution of Oxygenic Photosynthesis. Orig Life Evol Biosph 2016; **46**:51–65.

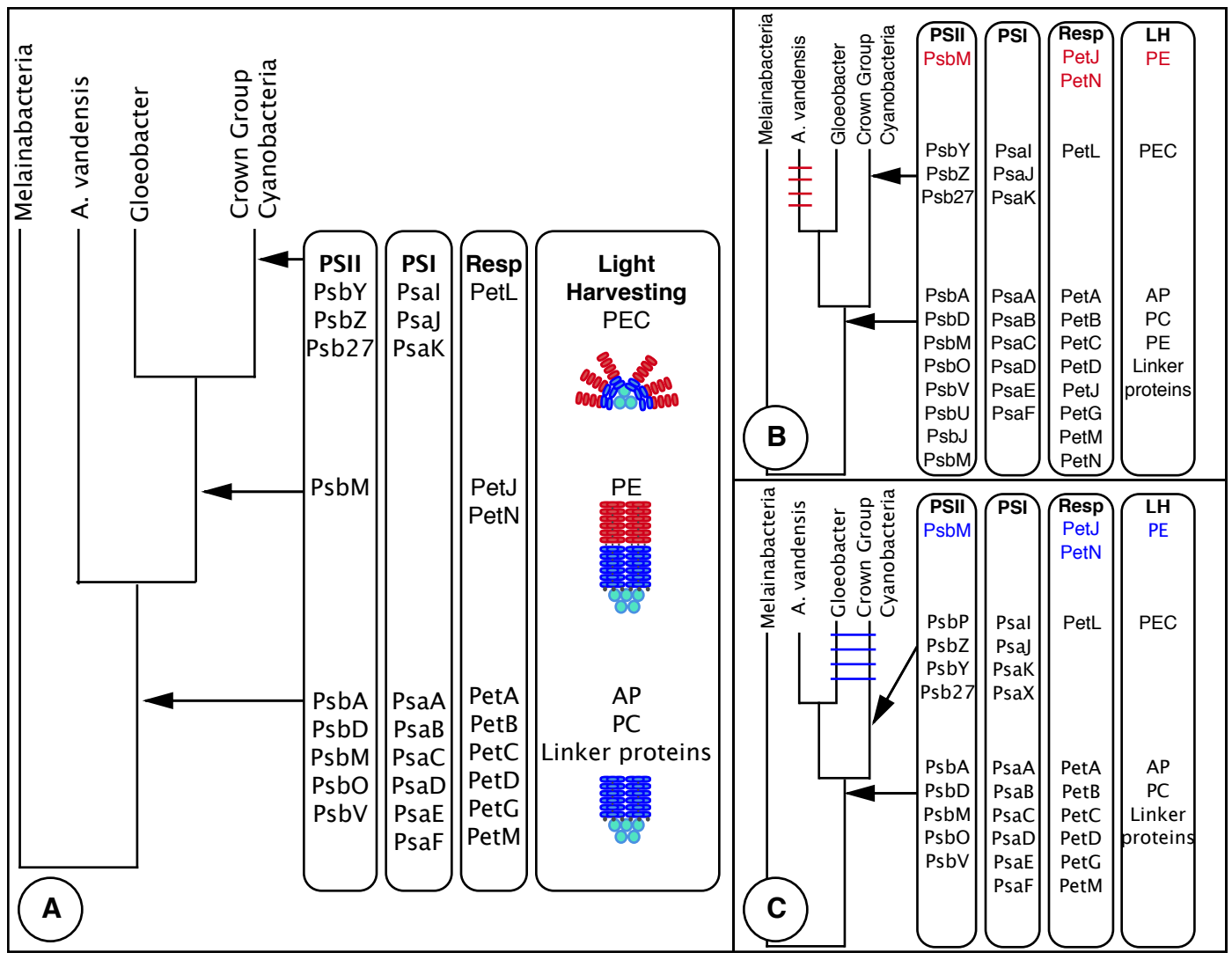
740 103. Anbar AD, Duan Y, Lyons TW, Arnold GL, Kendall B, Creaser RA, et al. A whiff of
741 oxygen before the great oxidation event? Science 2007; **317**:1903–6.

742

743 **Figure and Table Captions**

744 **Figure 1.** Phylogenetic placement of *A. vandensis*. A) Phylogeny constructed by inserting the
745 *A. vandensis* bin and 38 complete or nearly complete Melainabacteria and Sericytochromatia
746 draft genomes into a species tree containing 98 Cyanobacterial genomes. B) 16S rRNA gene
747 phylogeny. The genus *Aurora* is indicated by the dotted line. Bootstrap values are from the
748 original backbone tree.
749
750 **Figure 2.** Evolutionary model of oxygenic photosynthesis. A) Our preferred model showing
751 the progressive stabilization of oxygenic photosynthesis through time with *Aurora* basal to the
752 *Gloeobacter*. B) Model showing gene loss in the genus *Aurora*. C) Model showing horizontal
753 gene transfer between the ancestor of *Gloeobacter* and the ancestor of crown-group
754 Cyanobacteria. Models B and C show *Aurora* as a sister clade to *Gloeobacter*.
755
756 **Figure S1.** A) Genome phylogeny from KBASE showing *A. vandensis* as a sister group to the
757 *Gloeobacter.* B) 16S rRNA gene phylogeny showing the genus *Aurora* as basal to the
758 *Gloeobacter*. C) Ribosomal protein L2 phylogeny with *A. vandensis* sister to the *Gloeobacter*
759 and D) IF3 C phylogeny showing *A. vandensis* diverging before the divergence of the
760 *Gloeobacter*.
761
762 **Table S1.** Photosynthetic genes present in *A. vandensis*, *Gloeobacter*, and crown-group
763 *Cyanobacteria*. Differences between the early branching Aurora and Gloeobacter and the crown-
764 group Cyanobacteria are indicated in green. Difference between Aurora and Gloeobacter are
765 indicated in blue. Modified from ref 8.
766


1
Brevundimonas abyssalis
49
Melainabacteria
(28)
100
Aurora vandensis (pink)
Aurora vandensis (bulk)
100
Gloeobacter kilaueensis
100
100
Gloeobacter violaceus
Cyanobacteria
(21)
100
A
0.1
Melainabacteria (6 seqeunces)
100
Moss Pillars, Antarctica (AB630682)
Tundra Soil Nunavut Canada (JQ307084)
Aurora vandensis (bulk sample)
Hengel Valley, Iceland (KU222529)
Pumice, Patagonian Andes (KR923298)
Pumice, Patagonian Andes (KM149740)
Lake Sediments, Arctic (JQ793000)
Lake Sediments, French Alps (KF856487)
Stromatolite Tazmania (KX903890)
100
90
Gloeobacter (8 seqeunces)
Cyanobacteria (25 seqeunces)
100
B


Melainabacteria
A. vandensis
Gloeobacter
Crown Group Cyanobacteria
PSII
PsbY
PsbZ
Psb27
PsbM
PsbA
PsbD
PsbM
PsbO
PsbV
PSI
PsaI
PsaJ
PsaK
PsaA
PsaB
PsaC
PsaD
PsaE
PsaF
Resp
PetL
PetJ
PetN
PetA
PetB
PetC
PetD
PetG
PetM
Light Harvesting
PEC
PE
AP
PC
Linker proteins
A
B
PSII
PsbM
PsbY
PsbZ
Psb27
PsbA
PsbD
PsbM
PsbO
PsbV
PSI
PsaI
PsaJ
PsaK
PsaA
PsaB
PsaC
PsaD
PsaE
PsaF
Resp
PetJ
PetN
PetL
PetA
PetB
PetC
PetD
PetG
PetM
LH
PE
PEC
AP
PC
PE
Linker proteins
C
PSII
PsbM
PsbP
PsbZ
PsbY
Psb27
PsbA
PsbD
PsbM
PsbO
PsbV
PSI
PsaI
PsaJ
PsaK
PsaX
PsaA
PsaB
PsaC
PsaD
PsaE
PsaF
Resp
PetJ
PetN
PetL
PetA
PetB
PetC
PetD
PetG
PetM
LH
PE
PEC
AP
PC
Linker proteins