



HAL
open science

Diel changes in the expression of a marker gene and candidate genes for intracellular amorphous CaCO₃ biomineralization in *Microcystis*

Apolline Bruley, Juliette Gaëtan, Muriel Gugger, Claire Pancrace, Maxime Millet, Geoffroy Gaschignard, Manuela Dezi, Jean-Francois Humbert, Julie Leloup, Fériel Skouri-Panet, et al.

► To cite this version:

Apolline Bruley, Juliette Gaëtan, Muriel Gugger, Claire Pancrace, Maxime Millet, et al.. Diel changes in the expression of a marker gene and candidate genes for intracellular amorphous CaCO₃ biomineralization in *Microcystis*. 2024. hal-04855399

HAL Id: hal-04855399

<https://hal.science/hal-04855399v1>

Preprint submitted on 24 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1 Diel changes in the expression of a marker gene 2 and candidate genes for intracellular 3 amorphous CaCO₃ biomineralization in 4 *Microcystis*

5
6 Apolline Bruley^{1,#}, Juliette Gaëtan^{1,#}, Muriel Gugger², Claire
7 Pancrace^{2,3}, Maxime Millet¹, Geoffroy Gaschignard¹, Manuela
8 Dezi¹, Jean-François Humbert³, Julie Leloup³, Fériel Skouri-Panet¹,
9 Isabelle Callebaut¹, Karim Benzerara*¹, Elodie Duprat*¹

10
11 ¹ Sorbonne Université, Muséum National d'Histoire Naturelle, UMR CNRS 7590, Institut de Minéralogie, de Physique
12 des Matériaux et de Cosmochimie, IMPMC – Paris, France

13 ² Institut Pasteur, Université Paris Cité, Collection of Cyanobacteria – Paris, France

14 ³ Sorbonne Université, CNRS, INRAE, IRD, Université Paris Est Créteil, Université Paris Cité, UMR 7618. Institut
15 d'Ecologie et des Sciences de l'Environnement de Paris (iEES-Paris) – Paris, France

16
17 #Equal contributions

18 *Corresponding authors

19 Correspondence: karim.benzerara@sorbonne-universite.fr & elodie.duprat@sorbonne-universite.fr

20 21 **ABSTRACT**

22 Phylogenetically diverse cyanobacteria biomineralize intracellular amorphous calcium carbonate
23 (iACC) inclusions. This includes several genotypes of the *Microcystis* genus, a potentially toxic,
24 bloom-forming cyanobacterium found worldwide in freshwater ecosystems. While we ignore the
25 biological function of iACC and the molecular mechanisms driving their formation, this process
26 may impact local geochemical cycles and/or be used for bioremediation strategies. Recently, a
27 marker gene of this biomineralization pathway, named *ccyA*, was discovered. However, the
28 function of the calcyanin protein encoded by *ccyA* remains unknown. Here, based on an RNA-
29 Seq approach, we assess the expression of the *ccyA* gene in *Microcystis aeruginosa* PCC 7806
30 during a 24 h day/night cycle. The *ccyA* gene shows a clear day/night expression pattern with
31 maximum transcript abundances during the second half of the night. This is consistent with the
32 assumption that iACC biomineralization is related with photosynthesis and may therefore follow
33 a day/night cycle as well. Moreover, several genes directly co-localized upstream and
34 downstream of *ccyA*, on the same DNA strand show a similar expression pattern, including a *cax*
35 gene encoding a calcium/proton exchanger and a gene encoding a protein with a domain also
36 present in the N-terminal region of calcyanins in many iACC-forming cyanobacteria. This suggests
37 that they all could be part of an operon, and may play a concerted role in iACC formation. Last,
38 several other genes involved in carbon concentrating mechanisms and calcium transport show
39 an expression pattern similar to that of *ccyA*. Overall, this study provides a list of candidate genes
40 that may be involved in the biomineralization of iACC by cyanobacteria and whose role could be,
41 in the future, analyzed by biochemistry and genetics approaches.
42

43 **Keywords:** Biomineralization, amorphous calcium carbonate, cyanobacteria, RNA-seq, *cax* gene, CCM

44 Introduction

45 Biomineralization is a process through which some living organisms form mineral phases. It is found in all
46 the domains of life (Boskey, 2003). In some cases, biomineralization is genetically controlled, *i.e.* it is tightly
47 controlled by specific genes, as documented in diverse eukaryotic organisms (Brownlee et al., 2015; Gilbert et
48 al., 2022; Knoll, 2003; Marin & Luquet, 2004; Monteiro et al., 2016). For example, key genes encoding, *e.g.*,
49 carbonic anhydrases (Chen et al., 2018; Le Roy et al., 2014), Ca^{2+} , H^+ and HCO_3^- transporters (Mackinder et al.,
50 2011), and/or molecules inhibiting mineralization (Marin et al., 2000, 1996), play pivotal roles in the control of
51 calcium carbonate (CaCO_3) precipitation by eukaryotic organisms. However, there are much fewer
52 documented cases of genetically controlled biomineralization by prokaryotes (Cosmidis & Benzerara, 2022;
53 Görden et al., 2020). For a long time, the only well documented case was the intracellular formation of
54 magnetites (Fe_3O_4) and/or greigites (Fe_3S_4) by magnetotactic bacteria (Monteil et al., 2018). This process is
55 regulated by a suite of more than 40 genes involved in the control of the size, shape and crystallochemistry of
56 the mineral byproducts (Liu et al., 2023; Taoka et al., 2023; Uebe & Schüler, 2016). More recently, the
57 formation of intracellular amorphous calcium carbonates (iACC) by several species of cyanobacteria has been
58 evidenced as an additional case of genetically controlled biomineralization in bacteria (Benzerara et al., 2014;
59 Couradeau et al., 2012). Before this discovery, cyanobacteria were usually thought to induce the precipitation
60 of CaCO_3 as an indirect by-product of photosynthesis, with no apparent involvement of specific genes
61 (Altermann et al., 2006). Cyanobacteria biomineralizing iACC are cosmopolitan (Ragon et al., 2014) and can
62 sequester alkaline earth elements such as Ca, Sr, Ba and Ra, including their radioactive isotopes (Blondeau et
63 al., 2018a; Cam et al., 2016; Mehta et al., 2022a). As a result, these bacteria may be interesting candidates for
64 bioremediating some radioactive pollutions (Mehta et al., 2022a). Moreover, the biomineralization of iACC has
65 been evidenced in phylogenetically diverse strains within the phylum of Cyanobacteria as well as in other
66 bacterial phyla (Liu et al., 2021; Monteil et al., 2021). It is also found in several strains (but not all) of the well-
67 studied *Microcystis* genus (Gaëtan et al., 2023). This cyanobacterium is particularly interesting since it can form
68 blooms, with some genotypes being toxic for human and wildlife (Harke et al., 2016), causing numerous
69 environmental issues in eutrophic freshwater ecosystems (Dick et al., 2021). Whether the formation of iACC
70 by some *Microcystis* strains may impact their ecological success under certain conditions remains to be
71 deciphered. Some putative biological functions have been suggested for iACC such as intracellular storage of
72 inorganic carbon, intracellular pH buffering or their use as a ballast modifying cell buoyancy (Cosmidis &
73 Benzerara, 2022). About the later hypothesis, we note that buoyancy is not just controlled by cell density but
74 also by additional parameters such as EPS adhering to the cells, enhancing cell aggregation. For example, Gu
75 et al. (2020) showed that Ca induces EPS production by *M. aeruginosa*, which can increase buoyancy through
76 the increased size of cell aggregates. The relative contribution of these opposing parameters on the buoyancy
77 should be assessed in the future. Moreover, we can speculate that the fulfilment of these functions may
78 involve some kind of homeostasis which could require both iACC precipitation and dissolution processes,
79 although iACC dissolution has not been evidenced yet in *Cyanobacteriota*. Last, the formation of iACC, a
80 concentrated intracellular sink of Ca, by the large biomasses of *Microcystis* during blooms may impact the local
81 geochemical cycle of Ca in these environments, although future studies will have to further document this
82 possibility (Gaëtan et al., 2023). Indeed, the fraction of a natural *Microcystis* bloom with Ca-concentrating
83 capabilities remains to be determined. Moreover, the dependence of Ca-sequestration by cells on extracellular
84 Ca concentrations needs to be specified. While Ca concentrations may be lower in many lakes than in the BG11
85 medium (less than 50% of the lakes with a pH higher than 7.4 in the database investigated by Weyhenmeyer
86 et al. 2009), we note, however, that concentrations in lakes vs laboratory cultures cannot be simply compared.
87 Indeed, cultures in the laboratory of iACC-forming cyanobacteria have been conducted so far under batch
88 conditions with a Ca supply limited by the volume of the culture, a condition that may be very different from
89 those encountered in an open system such as a lake.

90 Despite the potential environmental importance of this process, the molecular mechanisms driving the
91 biomineralization of iACC in cyanobacteria remain poorly known. It has been shown that it costs some energy
92 to the cells (Cam et al., 2018; De Wever et al., 2019). Moreover, a genetic control has been proposed
93 (Benzerara et al., 2022). Indeed, based on a comparative genomics approach, one family of genes, named *ccyA*,

94 was found in phylogenetically diverse iACC-forming cyanobacteria and was absent from cyanobacteria not
95 biomineralizing iACC. The *ccyA* gene was found in about one third of all publicly released genomes of
96 *Microcystis i.e.*, 93 out of 282 genomes assemblies available at the time of publication (Gaëtan et al., 2023).
97 The protein encoded by *ccyA*, called calcyanin, lacks detectable full-length homologs with a known function in
98 public databases. Therefore, its function remains unknown. Several clues suggest that calcyanin may play a
99 role in calcium homeostasis (Benzerara et al., 2023, 2022). However, since the expression of the *ccyA* gene in
100 mutants of cyanobacteria not forming iACC did not result in the formation of iACC, this biomineralization
101 process likely involves additional genes (Benzerara et al., 2022). For example, some of these genes may play a
102 role in the fact that iACC inclusions form within an intracellular compartment, the envelope of which remains
103 of unknown chemical composition (Blondeau et al., 2018b). Such a compartmentalization may serve to achieve
104 chemical conditions allowing the precipitation and the stabilization of ACC. Additional examples of molecular
105 processes that may favor iACC precipitation are carbon-concentrating mechanisms (CCMs) (Li et al., 2016;
106 Görgen et al., 2021), which have been abundantly described in cyanobacteria and involve diverse genes and
107 multiple mechanisms (Badger & Price, 2003; Kupriyanova et al., 2023). Last, calcium transport genes may also
108 be involved in the accumulation of high amounts of calcium within iACC (De Wever et al., 2019). Consistently,
109 the presence/absence of the *ccyA* gene in genomes was correlated with the presence/absence of genes
110 involved in the transport and homeostasis of calcium and inorganic carbon (Benzerara et al., 2022). Moreover,
111 in some genomes, *ccyA* was even colocalized with such genes.

112 Overall, a first step to gaining a better understanding about the molecular mechanisms involved in iACC
113 biomineralization is to look at the expression of the *ccyA* gene together with other genes, including those
114 neighboring *ccyA* in the genome and those mentioned above. The diel (day/night) cycle is an important
115 parameter driving large variations of gene expression in cyanobacteria (Stöckel et al., 2008; Welkie et al., 2019;
116 Zinser et al., 2009). For example, the diel cycle influences the expression of numerous genes involved in key
117 cellular processes in *Microcystis aeruginosa* such as photosynthesis, carbon fixation, nitrogen metabolism, and
118 circadian rhythms as shown, for example, by Huang et al. (2004) in *M. aeruginosa* PCC 7820. Similarly, the
119 genes involved in a circadian cycle were studied in *M. aeruginosa* PCC 7806 (Straub et al., 2011) under a 24-
120 hour light/dark cycle. They found that while the *kaiA* gene showed no significant variations during the cycle,
121 the transcription patterns of the *kaiB* and *kaiC* genes, as well as the *sasA* gene encoding the two-component
122 sensor histidine kinase, a KaiC-interacting protein, exhibited significant changes. These findings suggested that
123 light is not the sole factor triggering the transcription of genes involved in photosynthesis and respiration;
124 instead, their transcription may also be regulated by an endogenous circadian clock. Therefore, we have
125 investigated the expression of the whole genome of *Microcystis aeruginosa* PCC 7806, which contains the *ccyA*
126 gene and forms iACC under laboratory culture conditions, over a 24 h day/night cycle. In particular, we
127 assessed whether the expression of genes either neighboring *ccyA* in the genome and/or involved in CCM and
128 calcium transport might be correlated with that of *ccyA*.

129

130

Methods

131 Cultures and sampling

132

133 The study was conducted using the cyanobacterial strain *Microcystis aeruginosa* PCC 7806 available at the
134 Pasteur culture of cyanobacteria (PCC) collection at the Institut Pasteur. The culture, sampling, nucleic acid
135 extraction, and RNA sequencing were carried out at the the Institut Pasteur in 2017. The photosynthetic
136 photon flux density provided during the day (50 $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$; cool white OSRAM L 18/640) was
137 measured using a LICOR LI-185B quantum/radiometer/photometer equipped with a LICOR LI-193SB spherical
138 sensor. A preculture of *M. aeruginosa* strain PCC 7806 was grown at 22 °C in 40 mL of the BG11 growth medium
139 (Rippka et al., 1979) amended with 10 mM NaHCO_3 under a 13/11 h day/night cycle. A 13-hour light period
140 corresponds to what is observed in September at midlatitudes in the Northern hemisphere, a month during
141 which *Microcystis* proliferates in numerous ecosystems. Then, the culture was transferred into three separate
142 flasks (named A, B and C). The triplicate cultures of 40 mL were progressively transferred to a final set of three
143 cultures with a volume of 1250 mL each, in fresh BG11 medium supplemented with 20 mM NaHCO_3 and
144 synchronized in an INFORS incubator with 1% CO_2 and agitation at 50-100 rpm under the same day/night cycle

145 and temperature conditions. Synchronization involved three consecutive transfers, aiming for an initial optical
146 density (OD) at 750 nm of 0.05 and growing for a month. Sampling was performed during the mid-exponential
147 phase (OD of 0.6-0.7) for each biological replicate, over a period of 24 h at eight time points: (i) 7 am, 60 min
148 before the transition from night to day (t1_N); (ii) 9 am, 60 min after the transition from night to day (t2_D);
149 (iii) 12 am (t3-D); (iv) 4 pm (t4_D); (v) 8 pm, 1 h before the transition from day to night (t5_D); (vi) 10 pm, 60
150 min after the transition from day to night (t6_N); (vii) 2 am, *i.e.* the middle of the night (t7_N), and (viii) in the
151 end of the night period (t8_N), 24 h after the first sampling point. The sampling procedure involved the removal
152 of 150 mL of culture, which were centrifuged, rinsed with sterile water, and centrifuged again. The resulting
153 pellets were flash-frozen in liquid nitrogen and stored at -80 °C until further processing. To minimize batch
154 effects, the samples were randomized before subsequent processing. We note that relatively high phosphate
155 concentrations such as those found in BG11 (~180 $\mu\text{mol}\cdot\text{L}^{-1}$) can inhibit the precipitation of extracellular
156 carbonates and/or induce that of Ca-phosphates as shown by, *e.g.*, Rivadeneyra et al. (2006; 2010). However,
157 by contrast, it has been shown that this does not prevent at all the precipitation of intracellular carbonates,
158 which has been experimentally studied in BG11 before (Cam et al., 2017; De Wever et al., 2019).

159

160 **Nucleic acid extraction**

161 RNA extraction was carried out using a Trizol reagent (Life Technologies) and the NucleoSpin miRNA kit
162 (Macherey-Nagel). Firstly, 1 mL of Trizol preheated at 65 °C was added to the frozen pellets and incubated at
163 65 °C for 12 min with vortexing three times. After centrifugation (10 min at 4 °C and 12,000 rcf), the upper
164 phase was transferred to a new tube and mixed with 500 μL of Trizol preheated at 65 °C. RNA extraction was
165 achieved by adding 400 μL of chloroform, and incubating samples for 3 min at room temperature before
166 centrifugation (10 min at 4°C and 12,000 rcf). The upper phase was again transferred to a new tube for
167 extraction with 400 μL of Trizol and 400 μL of chloroform with similar incubation and centrifugation conditions.
168 Then, the upper phase obtained from this step was further processed using the NucleoSpin miRNA kit
169 according to the manufacturer's instructions. The final purified RNA was eluted in 20 μL of RNase/DNase-free
170 sterile water. The absence of genomic DNA contamination was confirmed by PCR targeting the 16S rRNA gene
171 using specific primers. The quality and quantity of RNA were assessed using a Nanodrop spectrophotometer
172 (ThermoFisher Scientific) and a Bioanalyzer (Agilent) with RNA Nano chips.

173

174 **Library Preparation and RNA Sequencing**

175 Library preparation was conducted using the TruSeq Stranded mRNA sample preparation kit (Illumina, San
176 Diego, California) following the manufacturer's instructions. Ribosomal RNA (rRNA) depletion was performed
177 using the Ribo-Zero rRNA removal kit (bacteria, #MRZB12424, Illumina) with 5 μg of total RNA. The rRNA-
178 depleted RNA was then fragmented using divalent ions at 94°C for 8 min. The resulting fragmented RNA
179 samples were reverse-transcribed using random primers, followed by complementary-strand synthesis to
180 generate double-stranded cDNA fragments. There was no need for an end repair step in this process. A 3'-end
181 adenine was added, and specific Illumina adapters were ligated to the cDNA fragments. The ligated products
182 were amplified by PCR. The quality of the resulting oriented libraries was assessed using Bioanalyzer DNA1000
183 Chips (Agilent, # 5067-1504) and quantified using spectrofluorimetry (Quant-iT™ High-Sensitivity DNA Assay
184 Kit, #Q33120, Invitrogen). Sequencing was performed on the Illumina HiSeq2500 platform of the Institut
185 Pasteur, generating single-end 65 bp reads with strand specificity. Each lane of sequencing accommodated a
186 mixture of 18 multiplexed samples.

187

188 **Read processing**

189 Raw RNA-seq reads (*available online, see section Data, scripts, code, and supplementary information*
190 *availability*) were cleaned from adapter sequences and low-quality sequences using an in-house program
191 (https://github.com/baj12/clean_ngs). Only sequences of at least 25 nt in length were considered for further
192 analysis. Transcript abundance was quantified from the cleaned read dataset using the Salmon software (Patro
193 et al., 2017). Pseudo-mapping of reads was done considering the coding sequences (CDS, including
194 pseudogenes) of the complete RefSeq genome assembly of *M. aeruginosa* PCC 7806 as a reference
195 (GCF_002095975.1). For each replicate, the estimated transcript abundances were normalized by Salmon
196 according to the transcript size, genome size (number of CDS) and sample size (number of reads).

197

198 **Differential expression analysis**

199 Total abundance data were further processed using the DiCoExpress R-script-based tool (Lambert et al.,
200 2020) in order to detect genes expressed differentially between different time points. First, normalization was
201 achieved using Trimmed Mean of the M-value (Robinson and Oshlack, 2010). During this procedure, 220 low-
202 count transcripts were discarded, resulting in a total of 4,440 genes with normalized counts. Principal
203 component analysis (PCA) was performed on this normalized dataset using the R packages FactoMineR and
204 Factoshiny (Lê et al., 2008). Then, the differential expression between time points was tested using a
205 generalized linear model. For comparison between each time point, statistics of differential gene expression
206 are provided as Fold Change (FC) and p-values.

207 **Gene functional annotation**

208 In addition to the annotations provided by the NCBI prokaryotic genome annotation pipeline for the
209 genome assembly of *M. aeruginosa* PCC 7806
210 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_002095975.1), protein sequences translated from CDS
211 were processed using DeepNOG protein orthologous groups assignment (Feldbauer et al., 2021) as
212 implemented in the Genovi tool (Cumsille et al., 2023). An orthologous group with a COG accession number is
213 assigned to each gene, as well as a confidence score ranging within [0, 1]. The stringent threshold was set at
214 0.8. Orthologs of genes involved in carbon-concentrating mechanisms (CCM) and calcium transport were
215 specifically detected as bidirectional best BLASTp hits using protein sequence queries described in (Tang et al.,
216 2022) (15 genes from *Synechocystis* sp. PCC 6803, and the *ecaA* gene from *Anabaena* PCC 7120) and (De Wever
217 et al., 2019) (13 genes from *Chroococcidiopsis thermalis* PCC 7203), respectively. The following thresholds were
218 used: E-value cut-off of 1E-6, $\geq 30\%$ identity and 70% coverage.

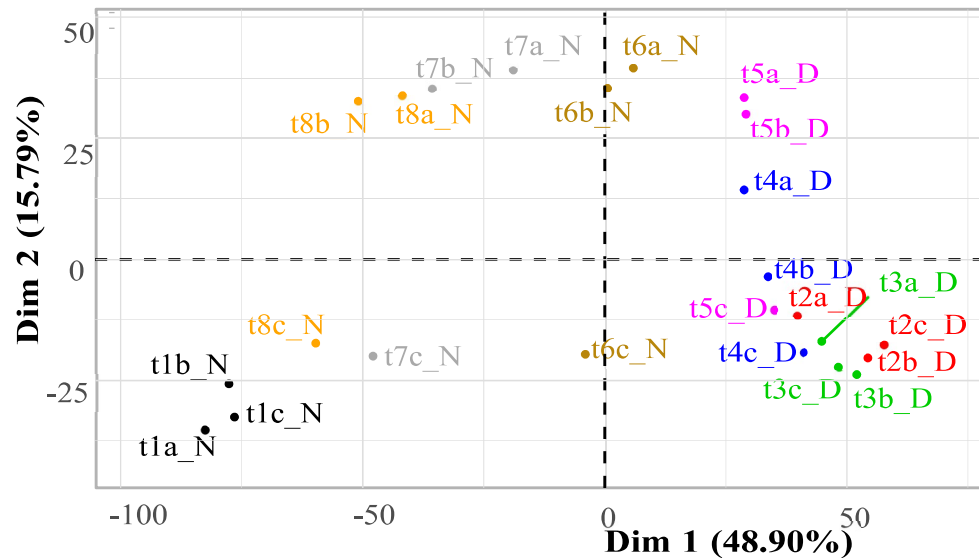
219 **Sequence and structure analysis of unknown proteins**

220 General information on the three domains of unknown function (DUF) encoded by genes located in the
221 genomic neighborhood of the *ccyA* gene was extracted from their respective InterPro entries (Paysan-Lafosse
222 et al., 2023; DUF5132: IPR033456; DUF454: IPR00740; DUF1269: IPR009200). AlphaFold2 (AF2) 3D structure
223 models of the three DUF-containing proteins and two hypothetical proteins neighboring *ccyA* were build using
224 ColabFold (Mirdita et al., 2022). Foldseek (van Kempen et al., 2023) was used to search for structural
225 similarities against the Protein Data Bank (PDB) and AlphaFold Database (AFDB). 3D structures were
226 manipulated using Chimera (Pettersen et al., 2004). Hydrophobic cluster analysis (HCA) (Callebaut et al., 1997)
227 was used to analyze protein secondary structure features.

228 **Results**

229 **Overall expression pattern**

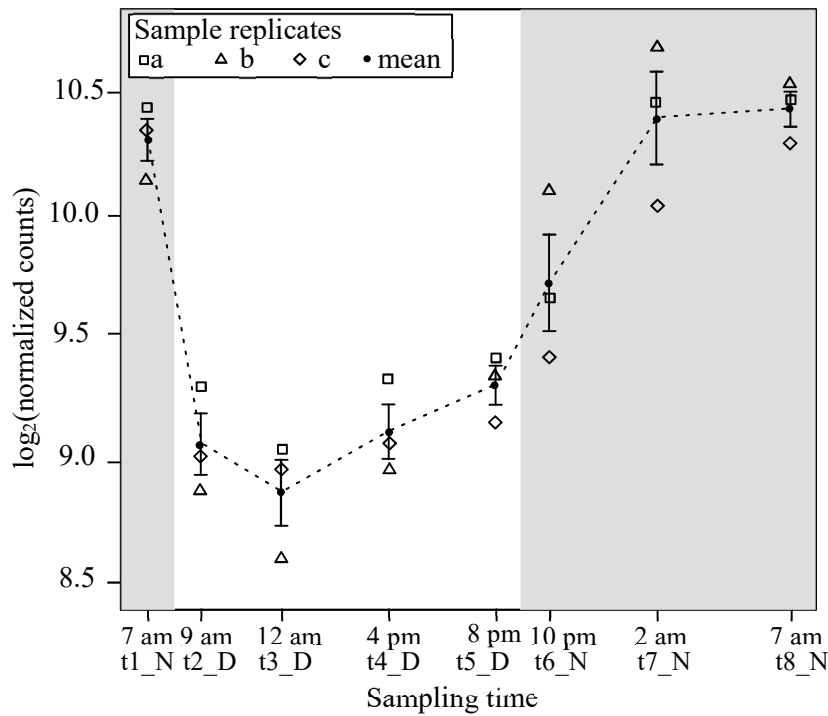
230 A principal component analysis (PCA) was performed on the transcript abundances of the 4,440 genes, for
231 each replicate at each sampling time (Figure 1). All the data points were projected onto a two-dimensional
232 plane defined by the first two principal components, which accounted for 64.69% of the total variance (Figure
233 1). Among the triplicates (i.e. three independent cultures for each time-point) only minor variations were
234 observed, at least along axis 1, and their values were clearly separated from the samplings at other time points.
235 Furthermore, the resulting plot exhibited a clear separation along axis 1 between night (N) and day (D) points.
236
237
238
239



240
241 **Figure 1 - Principal component analysis of the transcript abundances of 4440 *M. aeruginosa* genes during**
242 **a day/night cycle.** Abundances are expressed as log base 2 of the normalized counts (see Methods section for
243 normalization procedure), for each replicate (a-c) at each sampling time (t1-t8). The 24 samples are projected
244 onto the first two principal components (Dim. 1 and Dim. 2), explaining 48.9% and 15.8% of the variance of
245 the dataset, respectively. Samples/replicates collected at different times over the day/night cycle appear with
246 a different color code. The “_N” and “_D” codes in the sample names refer to night and day, respectively.

247 **Expression of the *ccyA* gene**

248 In order to analyze the expression profile of the *ccyA* gene during the 24 h day/night cycle, we plotted the
249 normalized expression of *ccyA* in all replicates at the different time points (Figure 2). A clear day/night pattern
250 can be observed in the expression of *ccyA*, with maximum transcript abundances during the second half of the
251 night (t7_N and t8_N, Table 1) and a minimum transcript abundance at 12 am during daytime (t3_D, Table 1).
252 Pairwise comparisons of *ccyA* transcript abundance at different time points show that the expression of *ccyA*
253 at daytime points (t2_D to t5_D) did not significantly differ between each other, but were significantly different
254 (p-value < 0.01) from some of the nighttime points, t7_N and t8_N, respectively (Table S1). The 10 pm nighttime
255 point (t6_N) did not significantly differ from any others except from the 12 am daytime point (t3_D; p-value <
256 0.05) and can therefore be considered as a transition point between day and night. When compared with the
257 mean expression of the 4,440 genes of the whole transcriptome, the mean expression of the *ccyA* gene was
258 significantly higher (p-value < 0.05) during the second half of the night, *i.e.* at the two-time points 2 am (t7_N)
259 and 7 am (t8_N) (Table 1).



260
261
262
263
264
265

Figure 2 - Time course of the abundance of *ccyA* transcripts during a day/night cycle. Abundances are expressed as log base 2 of the normalized counts. The abundance mean for the three replicates is shown by a black filled circle. Replicates a, b and c are represented by different symbols. For each sampling time (t1-t8), error bar represents the standard error on the mean of the three replicates. Grey shaded areas outline the night periods.

	t1_N(*) 7 am	t2_D 9 am	t3_D 12 am	t4_D 4 pm	t5_D(*) 8 pm	t6_N(*) 10 pm	t7_N(**) 2 am	t8_N(**) 7 am
Mean (<i>ccyA</i> gene, n=3 replicates)	10.31	9.07	8.88	9.12	9.31	9.73	10.39	10.43
Standard deviation (<i>ccyA</i> gene, n=3 replicates)	0.15	0.22	0.23	0.19	0.14	0.34	0.33	0.12
Mean (4440 genes, n=3 replicates)	6.87	7.35	7.39	7.44	7.44	7.36	7.23	7.18
Standard deviation (4440 genes, n=3 replicates)	3.79	2.45	2.57	2.67	2.61	2.87	3.21	3.33

266
267
268
269

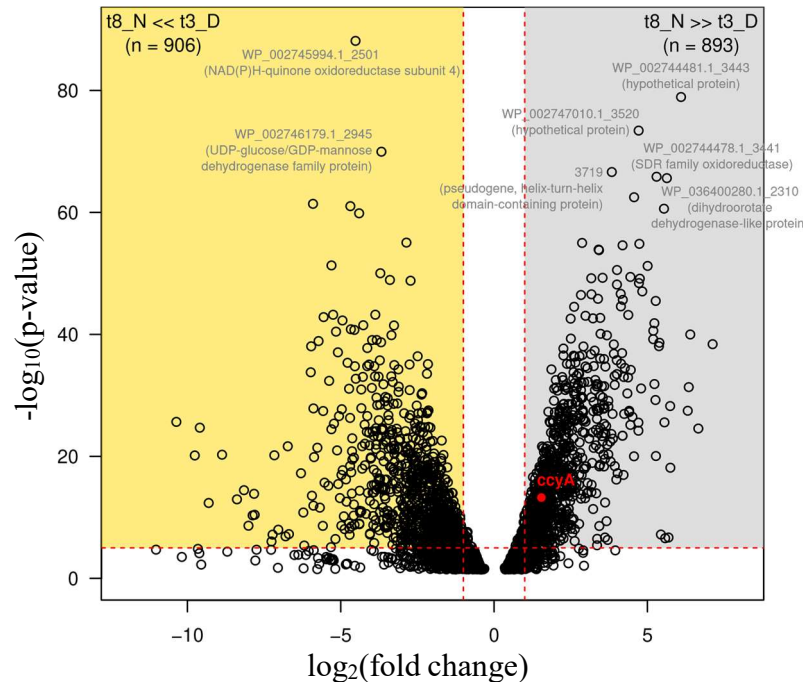
Table 1 - Comparison between the transcript abundances of *ccyA* and the whole transcriptome, for each time point. Abundances are expressed as log base 2 of the normalized counts. When means significantly differ, p-values of Wilcoxon signed rank test with continuity correction are depicted as follows: (**): 0.01 – 0.05, (*): 0.05 – 0.1. Night points are shaded in grey.

270
271
272
273
274
275

Search of differential gene expression during the day/night cycle

Subsequently, a comprehensive analysis was conducted throughout the genome to identify all genes exhibiting diel variations in their expression, using the minima and maxima of *ccyA* gene expression (t3_D and t8_N) as reference time points (Figure 3). Overall, 1,799 of the 4,440 total genes showed a significant difference in expression (at least 2-fold) between 12 am (during the day, t3_D) and 7 am (during the night, t8_N) (Figure

276 3 and Table S2). Among these 1,799 genes, 906 were significantly more expressed during the day (t3_D), and
 277 conversely, 893 were over-expressed during the night (t8_N).
 278 Figure 3 provides the annotations for the genes with the most substantial expression differences between
 279 these two time points ($-\log_{10}(p\text{-value}) > 60$): genes overexpressed during the day included a NAD(P)H-quinone
 280 oxidoreductase and a protein of the UDP-glucose/GDP-mannose dehydrogenase family. The genes
 281 overexpressed at night included: two genes encoding hypothetical proteins, a pseudo-gene encoding a protein
 282 containing a helix-turn-helix domain, a gene encoding a protein from the SDR oxidoreductase family, and a
 283 gene encoding a dihydroorotate dehydrogenase-like protein.
 284

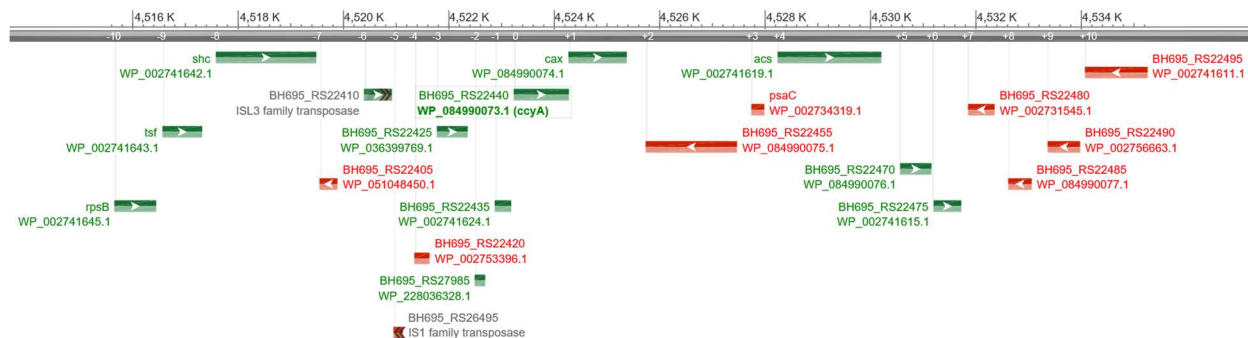


285

286 **Figure 3 - Differential expression of the whole transcriptome between sampling time t3_D (12 am,**
 287 **during the day period) and t8_N (7 am, end of the night period).** Log base 2 of the fold change (FC) values
 288 (x-axis) are defined for 3116 CDS. CDS with a negative log₂(FC) value correspond to transcripts
 289 significantly more abundant at 12 am (t3_D) compared with 7 am (t8_N) (gold area; 906 CDS). CDS with
 290 a positive log₂(FC) value correspond to transcripts significantly more abundant at 7 am (t8_N) compared
 291 with 12 am (t3_D) (grey area; 893 CDS). The *ccyA* gene is indicated by a red dot. The horizontal red dashed
 292 line indicates the significance p-value threshold after Bonferroni correction ($-\log_{10}(P\text{-value})=5$). The
 293 vertical red dashed lines indicate FC=2 and FC=0.5, e.g., 2-fold abundance ratio between the two sampling
 294 times. RefSeq accessions and annotations are indicated for CDS with $P < 2.0e-66$. See Suppl. Table S2 for
 295 details.

296 Genomic neighborhood of *ccyA*

297 Figure 4 displays the map of the 20 genes (18 genes and 2 pseudo-genes) located upstream (10 genes) and
 298 downstream (10 genes) to *ccyA*. This number of 20 genes was set arbitrarily and considered as a close
 299 neighborhood. When available, the functional annotations were extracted from the NCBI RefSeq database
 300 (Table 2). In this database, the *ccyA* gene is incorrectly annotated as a gene coding for the cell envelope
 301 biogenesis protein *OmpA* (Benzerara et al., 2022). Interestingly, three transposition elements were found near
 302 the *ccyA* gene: an ISL3 family and an IS1 family transposases, both pseudo-genes located at -6 and -5 genes
 303 compared to *ccyA*, and an IS1634 family transposase located at +2 genes compared with *ccyA*. The gene located
 304 directly upstream of *ccyA* was annotated as encoding a domain-of-unknown-function (DUF5132)-containing
 305 protein, and the one directly downstream a calcium/proton exchanger (*cax* gene), previously characterized in
 306 eukaryotes (Hirschi et al., 1996; Waight et al., 2013). At position -3 (upstream of *ccyA*), another DUF-containing
 307 protein (DUF454) was detected.



308
309
310
311
312
313

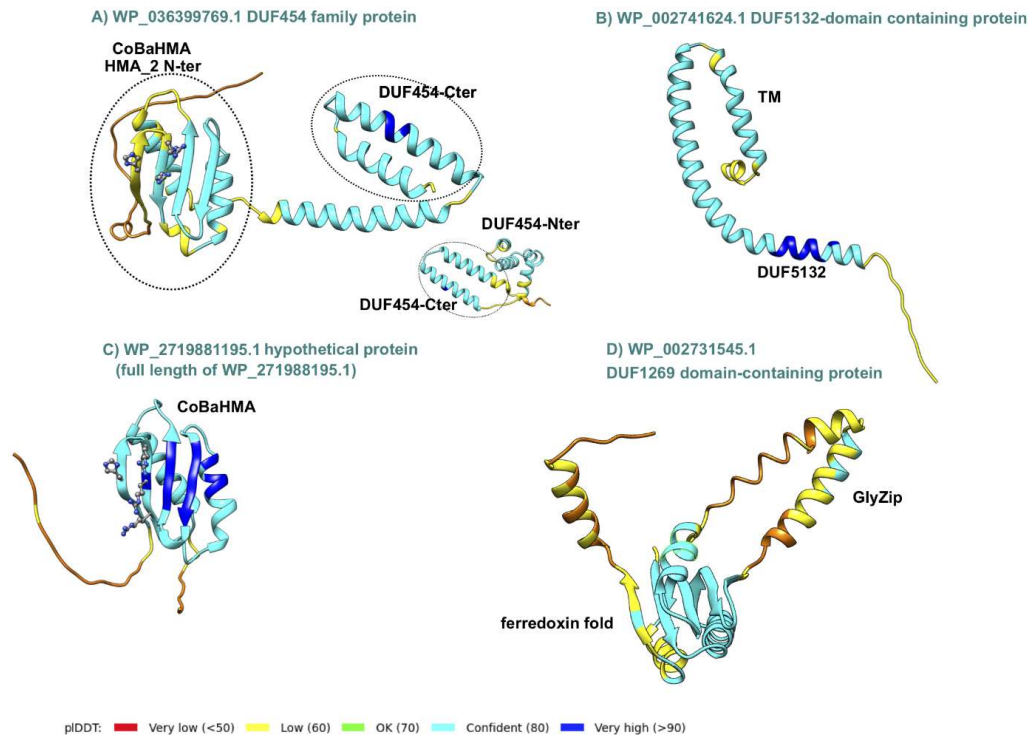
314 **Figure 4 - Schematic representation of the *ccyA* gene neighborhood in *M. aeruginosa* PCC 7806 complete genome assembly (NCBI genome browser display).** The *ccyA* gene is outlined by a
315 green box. RefSeq protein accession, gene name (if defined), gene orientation (red for minus; green for plus) and locus tag are indicated for the 10 CDS located upstream (-10 to -1) and the 10
316 CDS located downstream (+1 to +10) with respect to *ccyA*. Protein functional annotation is provided in Table 2. This genomic region spans within the nucleotide position interval [4515650,
317 4535259] in the NCBI genome assembly NZ_CP020771.1. CDS boundaries and features correspond to RefSeq annotations (GCF_002095975.1). Gene symbols stand for *shc*: squalene--hopene
318 cyclase; *tsf*: translation elongation factor Ts; *rpsB*: 30S ribosomal protein S2 ; *cax*: calcium/proton exchanger; *acs*: acetate--CoA ligase; *psaC*: photosystem I iron-sulfur center protein PsuC.

319 **Table 2 - Neighborhood of the *ccyA* gene in *M. aeruginosa* PCC 7806 complete genome assembly.** General features and annotations are listed for the ten CDS located upstream (-10 to -1)
320 and the ten CDS located downstream (+1 to +10) of the position of *ccyA* (in bold, protein improperly annotated in RefSeq as OmpA). RefSeq annotations correspond to the assembly accession
321 GCF_002095975.1 (NZ_CP020771.1).

Protein accession (NCBI RefSeq)	CDS position		Gene orientation	Start (nt)	End (nt)	Gene name	Protein functional annotation (NCBI RefSeq)	COG (DeepNOG annotation)		
	NCBI RefSeq	relative to <i>ccyA</i>						Accession	Confidence	Category
WP_002741645.1	4262	-10	+	4515650	4516447	<i>rpsB</i>	30S ribosomal protein S2	COG0052	1.0	J
WP_002741643.1	4263	-9	+	4516564	4517319	<i>tsf</i>	translation elongation factor Ts	COG0264	1.0	J
WP_002741642.1	4264	-8	+	4517577	4519484	<i>shc</i>	squalene--hopene cyclase	COG1657	1.0	I
WP_051048450.1	4265	-7	-	4519545	4519886		transcriptional regulator	COG5499	0.99	V
pseudo (partial=3')	4266	-6	+	4520389	>4520916		ISL3 family transposase			
pseudo (partial=3')	4267	-5	-	<4520947	4521127		IS1 family transposase			
WP_002753396.1	4268	-4	-	4521343	4521633		DUF5132 domain-containing protein	COG4980	0.99	R
WP_036399769.1	4269	-3	+	4521771	4522358		DUF454 family protein	COG2205	0.28	T
WP_228036328.1	4270	-2	+	4522489	4522689		hypothetical protein	COG2161	0.20	V
WP_002741624.1	4271	-1	+	4522868	4523182		DUF5132 domain-containing protein	COG1826	0.96	U
WP_084990073.1	4272	0	+	4523229	4524278	<i>ccyA</i>	cell envelope biogenesis protein OmpA	COG3134	0.95	S
WP_084990074.1	4273	+1	+	4524268	4525377	<i>cax</i>	calcium/proton exchanger	COG0387	1.0	P
WP_084990075.1	4274	+2	-	4525733	4527469		IS1634 family transposase	COG5421	0.99	X
WP_002734319.1	4275	+3	-	4527740	4527985	<i>psaC</i>	photosystem I iron-sulfur center protein PsuC	COG5703	1.0	C
WP_002741619.1	4276	+4	+	4528237	4530207	<i>acs</i>	acetate--CoA ligase	COG0365	1.0	I
WP_084990076.1	4277	+5	+	4530562	4531158		hypothetical protein	COG1196	0.40	D
WP_002741615.1	4278	+6	+	4531198	4531722		GNAT family N-acetyltransferase	COG1670	0.99	J,O
WP_002731545.1	4279	+7	-	4531850	4532356		DUF1269 domain-containing protein	COG4803	1.0	S
WP_084990077.1	4280	+8	-	4532613	4533059		DUF1269 domain-containing protein	COG4803	0.99	S
WP_002756663.1	4281	+9	-	4533361	4533978		CBS domain-containing protein	COG5792	1.0	C,T
WP_002741611.1	4282	+10	-	4534069	4535259		glycosyltransferase family 4 protein	COG0438	1.0	M

322

323 AlphaFold 2 (AF2) 3D structure models of two hypothetical proteins, as well as five proteins containing
324 DUFs present in the *ccyA* genomic neighborhood were analyzed (Figure 5) and compared with experimental
325 3D structures using Foldseek, in order to highlight possible structural and functional relationships. (i)
326 Analysis of the DUF454 family protein (WP_036399769.1; position -3 relatively to *ccyA*) indicates that its
327 C-terminal end matches the C-terminal end of the DUF454 IPR profile (inner membrane protein YbaN from
328 *E.coli*). Meanwhile, its N-terminal part matches the Pfam HMA_2 profile (PF19991), corresponding to a
329 CoBaHMA domain. CoBaHMA (after Conserved Basic residues HMA) is the name of a recently identified
330 family of domains belonging to the HMA superfamily (Gaschignard et al., 2024). Within this superfamily,
331 the CoBaHMA domain family shows distinct features such as the presence of an additional strand β_0 at the
332 N-terminus of the domain and the presence of a charged patch on one side of the β -sheet. The AF2 3D
333 structure model also consistently highlights this two-domain architecture, with the C-terminal, DUF454-
334 like domain folding as a helical hairpin, while the CoBaHMA domain possesses the basic conserved amino
335 acids characteristic of this family (Figure 5-A). (ii) The DUF5132-containing protein (WP_002741624.1;
336 position -1 relatively to *ccyA*) is covered by the DUF5132 profile (aa 48-89) over a little less than half of its
337 length and contains a predicted transmembrane segment in its N-terminal part (aa 12-32). The AF2 3D
338 structure model, characterized overall by good pLDDT values, indicates a putative assembly, in a hairpin-
339 like conformation, of the helical transmembrane segment and another alpha-helix preceding a long,
340 soluble helix matching the DUF5132 profile (Figure 5-B). (iii) A second member of this family, with similar
341 structural features is also present in the *ccyA* neighborhood (WP_002753396.1; position -4 relatively to
342 *ccyA*). No significant relationship with any known 3D structures could be detected by FoldSeek. (iv) A
343 hypothetical protein (WP_228036328.1; position -2 relatively to *ccyA*) corresponds to a truncated part of
344 a hypothetical protein referenced under RefSeq WP_271988195.1, which is also composed of a CoBaHMA
345 domain sharing the basic signature of this family (Figure 5-C). (v) Another hypothetical protein
346 (WP_084990076.1; position +5 relatively to *ccyA*) is predicted as a long coiled-coil (data not shown). (vi)
347 and (vii) Finally, two DUF1269-containing proteins (WP_002731545.1 and WP_084990077.1; genes at +7
348 and +8 relatively to *ccyA*), corresponding to two isoforms differing by their C-terminal end, match the
349 DUF1269 IPR profile over their whole lengths and include a long glycine-zipper motif embedded in a soluble
350 domain, which based on Foldseek searches, corresponds to a ferredoxin fold, as found for instance in the
351 copper tolerance CutA1 protein (pdb 1V6H, Bagautdinov, 2014) and in the nickel-responsive transcription
352 factor NikR (pdb 2BJ3, Chivers and Tahirov 2003) (Figure 5-D). This glycine-zipper motif is similar, albeit
353 more hydrophobic, to the GlyZip motif in the C-ter domain of calcyanins, composed of two parts separated
354 by a proline preceded by a small amino acid (here a serine, while it is a glycine in calcyanins) (Figure 5-D).
355 Overall, this analysis highlighted some features in the neighborhood of the *ccyA* gene, which were not
356 captured by automatic annotations, including motifs that are found in calcyanins of all cyanobacteria
357 (GlyZip and hairpin-like helices) as well as calcyanins of some species (CoBaHMA domains).



358

359 **Figure 5 - AlphaFold2 3D structure models of some proteins encoded by the genomic neighbors of *ccyA*.**

360 The 3D structure models are represented as ribbons colored according to the pLDDT values (color code given at
361 bottom). In panels A and C, the basic amino acids constituting the signature of the CoBaHMA family are shown
362 as balls and sticks. (A) The CoBaHMA_DUF454-Cter architecture (encoding gene in position -3 relatively to *ccyA*)
363 is compared with a protein entirely covered by the DUF454 IPR profile (inner membrane protein YbaN
364 (YABN_ECOLI, UniProt P0AAR5); insert). (B) DUF5132 protein (encoding gene in position -1 relatively to *ccyA*).
365 TM stands for TransMembrane. (C) Hypothetical protein (encoding gene in position -2 relatively to *ccyA*) with a
366 typical CoBaHMA domain. (D) DUF1269 domain-containing protein (encoding gene in position +7 relatively to
367 *ccyA*) containing a GlyZip motif, predicted to fold as a helical hairpin (see HCA plot and sequence alignment in
368 Figure S1).
369

370 **Correlated expression of other genes with the *ccyA* gene**

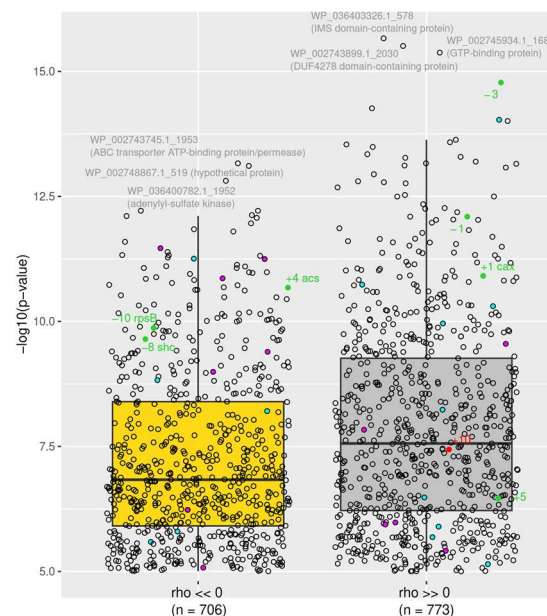
371 The expression patterns of genes exhibiting a significant difference in their circadian expression
372 between daytime (t3_D, 12 am) and nighttime (t8_N, 7 am) were compared with that of *ccyA* to figure out
373 which ones may have a maximum expression at the same time (Figure 6; Table S3). We note that such gene
374 expression variations over a day/night cycle are expected in an autotrophic organisms and at least some
375 correlations are likely not causal. However, their identification offers one possibility to better understand
376 which other pathways may be activated at the same time as *ccyA* is expressed. As a result, two subsets
377 were identified: on the one hand, 773 genes showed a highly positive correlation with the expression
378 pattern of *ccyA* (with a Pearson correlation coefficient $\rho > 0.75$, this threshold being arbitrarily fixed; grey
379 subset in Figure 6; see also Figure S2 and Table S3); on the other hand, 706 displayed a negative correlation
380 with the expression pattern of *ccyA* ($\rho < -0.75$, yellow subset in Figure 6; see also Figure S2 and Table S3).
381 Figure 6 displays the significance level (p-value) of the correlations between the expression profiles of these
382 1479 genes and that of *ccyA*, and provides annotation for the genes with the most significant correlation.

383 The genes for which the expression patterns were the most significantly positively correlated with that
384 of the *ccyA* gene were: (i) an intermembrane mitochondrial space (IMS) domain-containing protein, (ii) a
385 GTP-binding protein, and (iii) a DUF-containing protein (DUF4278). The genes the most significantly
386 anticorrelated with *ccyA* expression profile were annotated as (i) an ABC transporter ATP-binding
387 protein/permease, (ii) an adenylyl-sulfate kinase, and (iii) a hypothetical protein.

388 Among the 20 neighboring genes of *ccyA*, eight showed an expression profile either positively or
389 negatively correlated with that of *ccyA* (see Figure S3 for a time plot of expression of all 20 genes).

390 Interestingly, the expression profile of the two genes adjacent to *ccyA*, located on the same DNA strand,
391 *i.e.* encoding for a calcium/proton exchanger (CAX) and DUF5132 (Table 2), displayed a strong positive
392 correlation with the expression profile of *ccyA* (Figure 6 and 7). The expression profile of the gene coding
393 for a DUF454 family protein (position -3 relatively to *ccyA*; Table 2), exhibited one of the strongest positive
394 correlations with the expression profile of *ccyA* (Figure 6), with a nearly identical expression profile (Figure
395 7). Finally, the expression profiles of two other genes, located at +10 and +5 genes with respect to *ccyA*
396 showed a weaker but significant positive correlation with the expression profile of *ccyA*. The gene at +10
397 was annotated as a glycosyltransferase family 4 protein. The gene at +5 codes for a hypothetical protein
398 predicted as a long coiled-coil (Table 2).

399 Moreover, proteins involved in calcium transport and in CCM were searched in the genome of PCC
400 7806 (Table S4). In total, 24 genes involved in CCM and/or their homologs and 24 genes involved in calcium
401 passive or active transport and their homologs were identified. The expression patterns of seven out of the
402 24 genes involved in CCM were correlated negatively with the expression pattern of *ccyA*, whereas the
403 expression patterns of five of CCM genes were positively correlated with the expression pattern of *ccyA*
404 (Figure 6, in magenta; see Figure S4 for a time plot of the expression of all CCM genes). Among the 24 genes
405 involved in calcium transport, the expression patterns of five of them were negatively correlated with that
406 of *ccyA*, while the expression patterns of ten of them were positively correlated with that of *ccyA* (Figure
407 6, in cyan except the *cax* gene in green; see Figure S5 for a plot of time expression of all “Ca genes”).



408

409

410

411

412

413

414

415

416

417

418

419

Figure 6 - CDS with a day/night expression pattern strongly (anti-)correlated with that of *ccyA*. The Pearson correlation coefficient (ρ) is highly negative for 706 CDS (left). The Pearson correlation coefficient is highly positive for 773 CDS (right). The distribution of correlation coefficients over the 4440 CDS is shown in Figure S2. RefSeq accessions and annotations are indicated for the 3 CDS with the strongest negative and positive correlations (see Table S3 for details). CDS belonging to the carbon concentrating mechanism or the calcium transport system are represented as magenta and cyan dots, respectively. CDS located in the [-10, +10] neighborhood of *ccyA* (Table 2 and Figure 4) are represented by green and red dots (for positive and negative DNA strands, respectively). The *cax* gene located at a genomic position of +1 from the *ccyA* gene also belongs to the calcium transport system and appears in green (see Table S4 for details).

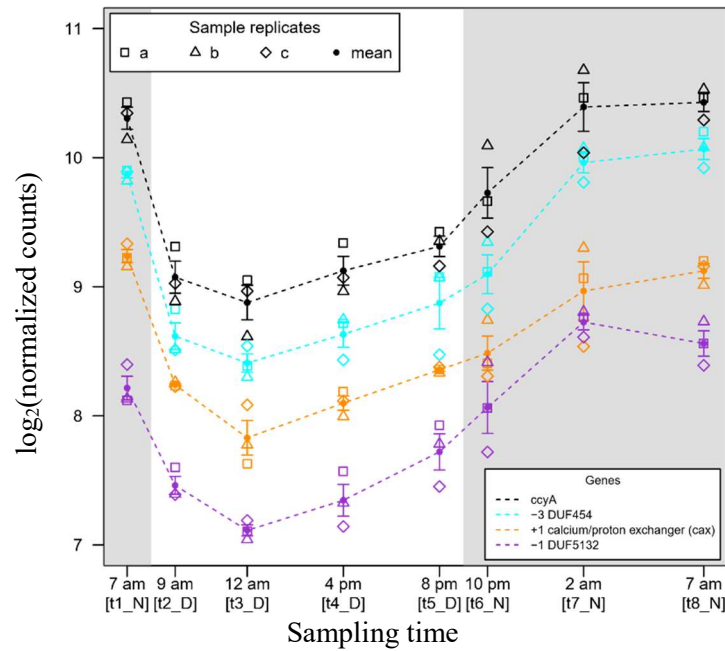


Figure 7 – Time course of the abundances of three genomic neighbors of the *ccyA* gene during a day/night cycle. The CDS located at positions -3 (blue), -1 (purple) and +1 (orange) correspond to the *ccyA* neighbors with transcript abundance profiles the most correlated to *ccyA* transcript abundance profile: Pearson correlation coefficient of 0.97 (p-value=1.7e-15), 0.95 (p-value=8.0e-13), 0.94 (p-value=1.2e-11), respectively. See Methods section for more details about how counts were normalized.

Discussion

The identification of a marker gene of iACC biomineralisation in cyanobacteria was a first step towards understanding the molecular mechanisms controlling the formation of these biomineral phases (Benzerara et al., 2022). However, up to now, only the presence/absence of the gene and the structural features of the encoded protein have been investigated (Benzerara et al., 2022; Gaëtan et al., 2023; Gaschignard et al., 2024). Here, we explored whether this marker gene is transcribed in an organism capable of forming iACC, how this may be influenced by day/night cycles and how the expression of neighboring genes compares.

Several tools can be used for assessing the transcription of a large number of genes within a genome, including DNA microarray and high-throughput RNA sequencing (RNAseq) (Wang et al., 2009). Microarrays are limited by the use of probes: the number of studied genes remains limited by the available knowledge of the targeted genome. By contrast, RNAseq allows targeting all the mRNAs in a cell and has the advantage of measuring the level of transcriptional activity in the organisms (Mantione et al., 2014). Moreover, RNAseq can achieve higher resolution (*i.e.* higher accuracy in the estimation of gene expression) than microarrays. Because this technique does not suffer from microarray-based limitations such as background noise and saturation, it can also achieve a lower detection limit, and thus detect very low gene expression (Zhao et al., 2014). Straub et al. (2011) already studied the expression of the whole genome of *M. aeruginosa* PCC 7806 over a 24-h day/night cycle, using DNA microarray. Using their publicly available dataset, we determined that the expression level of the *ccyA* gene (MIC_2500) was not significantly different from the background noise. Here, the use of RNA-seq on the same organism enabled the detection and analysis of the *ccyA* gene expression dynamics. Indeed, the expression level of *ccyA* in our dataset was above the mean expression of the 4,440 genes of *Microcystis* at all time and this was statistically significant for several time points during the night. We can thus conclude that the *ccyA* gene is clearly expressed in an iACC-forming strain.

In line with the study by Straub et al. (2011) on the dynamics of gene expression in *M. aeruginosa* PCC 7806, we found, as expected, that the day/night cycle is an important factor driving variations in global gene expression in PCC 7806. Cyanobacteria have a highly defined circadian clock, and this day/night gene expression pattern has been demonstrated in several strains (Pattanayak & Rust, 2014; Stöckel et al., 2008;

456 Zinser et al., 2009). Straub et al. (2011) showed that the metabolism of *M. aeruginosa* undergoes significant
457 changes between the light and dark periods. Specifically, during the light period, processes including
458 carbon uptake, photosynthesis, and the reductive pentose phosphate pathway result in glycogen synthesis.
459 Conversely, during the dark period, glycogen degradation, the oxidative pentose phosphate pathway, the
460 tricarboxylic acid (TCA) branched pathway, and ammonium uptake promote amino acid biosynthesis.
461 Furthermore, the biosynthesis of secondary metabolites, such as microcystins, aeruginosin, and
462 cyanopeptolin, predominantly occurs during the light period.

463 A clear day-night expression pattern was also observed for the *ccyA* gene: the expression was
464 significantly higher than the average genome expression from the middle (2 am, t7_N) to the end of the
465 night (7 am, t8_N). In all other time points of the 24-h cycle, the expression level of this gene did not
466 statistically differ from the average expression of all genes. The gene is therefore overtranscribed during
467 the second half of the night. While mRNA abundance provides valuable information about gene expression,
468 it does not always directly correlate with the protein levels in the cell (Ingolia et al., 2009; Riba et al., 2019;
469 Schwanhäusser et al., 2011). Several factors, including translation, initiation and elongation rate (Riba et
470 al., 2019), post-translational modifications, protein stability, and degradation rates (Christiano et al., 2014;
471 Gancedo et al., 1982) may impact the relationship between mRNA abundance and protein content. As
472 such, it is difficult to infer the abundance of calcyanin at a given time from the transcript expression level
473 only. However, one can speculate that the protein may be more produced during the second part of the
474 night, and remain at an optimal abundance during the day when it may most function. In the future, direct
475 measurements of the protein abundance should be performed using proteomics in order to quantify the
476 calcyanin content in cells. Last, Straub et al. (2011) mentioned that the metabolism of *M. aeruginosa*
477 strongly changes between the light and the dark period. For example, during the latter, amino acid
478 biosynthesis is promoted by glycogen degradation, the oxidative pentose phosphate pathway, the TCA
479 branched pathway and ammonium uptake. This illustrates that seeing *ccyA* expression at its strongest at
480 the end of the dark period is not particularly surprising energetically-wise.

481 Interestingly, we showed that a $\text{Ca}^{2+}/\text{H}^{+}$ antiporter (*cax* gene) and two DUF-containing proteins were
482 not only directly co-localized upstream and downstream of *ccyA*, on the same DNA strand but they were
483 also co-expressed with *ccyA*. This suggests that those three genes could be part of a functioning unit of
484 DNA that could be transcribed together into a single mRNA strand, *i.e.* an operon, and may play a concerted
485 role in iACC formation.

486 The process of iACC formation necessitates the accumulation of calcium ions at concentrations higher
487 than the cytosolic concentrations typically measured in cells (Cam et al., 2015). Therefore, it may be
488 suggested that the protein encoded by the *cax* gene is involved in the build-up of these accumulations.
489 However, the functioning of CAX proteins raises questions. Indeed, the single $\text{Ca}^{2+}/\text{H}^{+}$ exchanger found in
490 *Synechocystis* PCC 6803 has been shown to be located in the plasma membrane and to catalyze Ca^{2+} efflux
491 out of the cells (Waditee et al., 2004). By contrast, it has been suggested that CAX proteins in plants may
492 have different locations and efflux Ca^{2+} from the cytoplasm either to the extracellular space or to
493 intracellular vacuoles (Shigaki et al., 2006). Investigating further the similarities/dissimilarities of the CAX
494 protein of *M. aeruginosa* PCC 7806 with already known CAX proteins and locating these transporters within
495 *M. aeruginosa* PCC 7806 cells would therefore be crucial to better understand their potential involvement
496 in iACC formation.

497 The gene encoding a DUF454 protein that co-localizes (position -3 relatively to *ccyA*) and co-expresses
498 with *ccyA* is particularly interesting. Indeed, we show here that this DUF454 protein bears a CoBaHMA
499 (after Conserved Basic residues HMA) domain (Gaschignard et al., 2024). Strikingly, the calcyanins of many
500 cyanobacteria are similarly composed of a CoBaHMA domain in their N-terminal region and a C-terminal
501 three-fold repeat of a long glycine zipper motif conserved in all iACC-forming cyanobacteria (Benzerara et
502 al 2022). This is not the case for the calcyanins of *Microcystis*, which does not contain a CoBaHMA domain
503 but instead displays an unknown domain in their N-terminal region. Since the discovery of the calcyanin
504 family, it has been surprising to observe such a broad diversity in their N-terminal domains despite an
505 expected common function. Here, although we still ignore the function of calcyanin, the discovery of a
506 CoBaHMA-encoding gene neighboring the *ccyA* gene suggests that *Microcystis* calcyanin may work in
507 concert with other proteins, including a DUF454 protein and resulting in a protein complex with a possibly
508 similar architecture in *Microcystis* and other iACC-forming cyanobacteria. Moreover, the presence in the
509 close genomic neighborhood of *ccyA* of genes encoding for another protein of the CoBaHMA family (-2

510 position relative to *ccyA*), as well as proteins possessing a GlyZip-like motif (+7 and +8 positions relative to
511 *ccyA*) relatively similar to the GlyZip motif in the C-ter domain of calcyanins additionally suggests the
512 possibility of complementary roles of these proteins for a function to identify.

513 Moreover, building upon the approach followed by De Wever et al. (2019), we conducted a
514 comprehensive search for annotated genes involved in CCM and calcium transport, as well as their
515 homologous counterparts in the genome of *M. aeruginosa* PCC 7806. Firstly, a greater number of genes
516 associated with CCM or their homologs appeared to be expressed preferentially during the day instead of
517 night (nine vs four, respectively), compared with genes associated with calcium transport (three vs nine,
518 respectively) (Table S4). Similarly, Straub et al. (2011) showed that the transcript abundance of CCM genes
519 showed a peak in expression just after the night/day transition and then a rapid decrease after the
520 day/night transition. The elevated concentration of CO₂ in carboxysomes resulting from the CCM is a way
521 to improve carbon fixation (Price and Howitt, 2011). The expression of CCM genes during the day is
522 therefore expected. Yet, among the 24 genes identified as involved or potentially involved in CCM, 5 had
523 an expression pattern positively correlated with that of *ccyA*, *i.e.* maximum at night (Table S4). Notably,
524 two of these genes (*cmpC* and *cmpD*) encode two subunits of the inorganic carbon transporter BCT1
525 involved in the cyanobacterial CCM (Price & Howitt, 2011). BCT1 is a high-affinity but low-flux HCO₃⁻
526 transporter. It is composed of four subunits, and among them, CmpC and CmpD are extrinsic cytoplasmic
527 proteins that possess binding sites for ATP (Price & Howitt, 2011). It has been shown that the expression
528 of these genes in *M. aeruginosa* PCC 7806 depends on the ambient pCO₂ (Sandrini et al., 2015). Notably,
529 under high pCO₂ conditions (1450 ppm in their experiments), the expression of high-affinity bicarbonate
530 uptake systems (notably *cmpA*, *cmpB*, and *cmpC* genes) is down-regulated, and cells shift towards CO₂ and
531 low-affinity bicarbonate uptake compared with cells grown under low pCO₂ conditions (200 ppm) that the
532 authors considered as C_i-limited. For the production of the transcriptomes that we studied here, pCO₂ was
533 not regulated nor monitored, and the change of expression in CCM-involved genes such as *cmpD* and *cmpC*
534 could be due to changes in pCO₂ conditions during day/night transitions, even if there are expected to be
535 lower than in the Sandrini et al. (2015)'s study. Moreover, we note that *cmpA* and *cmpB* were more
536 expressed during the day. Therefore, the potential variation of the BCT1 activity over a diel cycle remains
537 to be ascertained. In future studies, exploring whether the co-expression of the five genes involved in CCM
538 with *ccyA* may play a role in iACC biomineralization would be interesting, and should include a special focus
539 on the pCO₂ variations during the experiments. Last, Walter et al. (2016) demonstrated that in *Anabaena*
540 sp. PCC 7120, the concentration of extracellular Ca²⁺ can also impact the expression of genes, including
541 those involved in bicarbonate uptake. While we do not expect a dramatic change of the extracellular Ca²⁺
542 concentration over a 24-hour period based on previous studies (Cam et al., 2018; De Wever et al., 2019),
543 this parameter should also be considered in future studies. In contrast to the bicarbonate transporters and
544 their homologs, the number of genes possibly linked with calcium transport (24 genes) and having an
545 expression pattern positively correlated with that of *ccyA* (10 genes) was higher than genes with an
546 expression pattern negatively correlated with *ccyA* (5 genes) (Table S4). The transport of Ca²⁺ ions into the
547 cytosol is commonly considered as a passive process facilitated by channels that exhibit low ionic specificity
548 (Domínguez et al., 2015). Here, only one out of the ten genes whose expression was positively correlated
549 with that of *ccyA* was identified as a passive transporter (hB1-1). The other nine genes include the *cax* gene,
550 neighboring *ccyA*, two *apnhaP* genes (Na⁺/H⁺ antiporter acting as a Ca²⁺/H⁺ antiporter at alkaline pH as
551 shown by Waditee et al. (2001)) and three of their homologs, a UPF0016 gene (putative Ca²⁺/H⁺
552 transporter) and a Ca²⁺-ATPase and one of their homologs. They could thus also be candidate genes
553 contributing to the intracellular accumulation of calcium with the same question arising about the polarity
554 of the transport as for *cax* (towards the exterior of the cells or inside compartments containing iACC).

555 In a comparative genomics approach performed at the scale of the whole *Cyanobacteriota* phylum,
556 *ccyA* was the only gene identified as shared by iACC+ strains and absent in iACC- strains (Benzerara et al.,
557 2022). However, some genes from both phenotypes (iACC- and iACC+), and therefore not highlighted by
558 the above-mentioned comparative genomics approach, could also be involved in iACC formation. Indeed,
559 these genes may be recruited in iACC+ strains for iACC formation, whereas they would have a different
560 function in iACC- strains. Moreover, there could also be genes specific to *Microcystis* involved in iACC
561 formation, with other non-homologous genes but of similar functions involved in other cyanobacterial
562 genera. The present study is a step forward towards identifying possible partners in iACC formation.
563 However, only future studies specifically targeting these genes will be able to provide definitive answers

564 about their implications in the iACC formation process. Furthermore, investigations into the
565 comprehensive response of *Microcystis* to fluctuations in calcium and/or pCO₂ levels could provide
566 valuable insights into the molecular mechanisms underlying the biomineralization of iACC. These studies
567 should encompass monitoring of the physicochemical parameters of the environment and the
568 quantification of iACC formation. Notably, certain semi-quantitative techniques, such as Fourier transform
569 infrared spectroscopy (FTIR) (Mehta et al., 2022b), scanning transmission x-ray microscopy (STXM)
570 (Benzerara et al., 2023) and X-ray absorption near edge structure (XANES) spectroscopy (Mehta et al.,
571 2023), now offer the capability to detect cyanobacterial iACC and assess their relative abundance. In the
572 future, studying control strains not hosting the *ccyA* gene might be an additional interesting perspective.
573 However, for this purpose, one cannot consider any *ccyA*- strain as a robust control since its genome may
574 differ from the one of PCC 7806 in several genes. A more reliable control would consist in a yet-unavailable
575 mutant of PCC 7806 with a deactivated/deleted *ccyA* gene. Last, one may speculate that cells may
576 sometimes, also direct iACC dissolution to regulate their abundance. While this process has been shown in
577 *Achromatium* (e.g., Yang et al., 2019), it still needs to be unambiguously evidenced in *Cyanobacteriota*
578 before assessing which genes might be involved.

579 Finally, it can be noted that the correlation between the *ccyA* gene and genes involved in calcium
580 transport and the CCM is not restricted to *Microcystis*. For example, the combined presence of homologs
581 of a Ca²⁺/H⁺ antiporter gene and the Na⁺-dependent bicarbonate *bicA* gene was shown to be significantly
582 correlated with that of *ccyA* in the *Cyanobacteriota* phylum (Benzerara et al., 2022). Moreover, the *ccyA*
583 gene was co-localized with a Ca²⁺/H⁺ antiporter gene in several other cyanobacteria (*C. fritschii* PCC 9212
584 and PCC 6912, and *Fischerella* sp. NIES-4106), mirroring the situation observed in *M. aeruginosa* PCC 7806.
585 Overall, this offers some target genes for future genetic studies to better depict the molecular mechanisms
586 of iACC biomineralization.
587

588

Acknowledgements

589 We thank Varet H, Sismeiro O, Dillies MA, Legendre R and Coppe JY who run the RNA platform of the
590 Institut Pasteur and provided us with the processed transcriptomics data. We also thank two reviewers for
591 comments that significantly improved the manuscript.

592

Data, scripts, code, and supplementary information availability

593 Raw transcriptomics data are available online on NCBI, Gene expression Omnibus, GEO accession
594 number: GSE255450. Supplementary information, supplementary data (output data from Salmon,
595 DICOEXPRESS and DeepNOG analyses as well as gene expression correlation) and supplementary figures
596 and tables are available online on bioRxiv, BIORXIV/2024/602159 - Version 3,
597 <https://doi.org/10.1101/2024.07.07.602159>. Links to the tools used in this study are noted in the Material
598 and methods section.

599

Conflict of interest disclosure

600 The authors declare that they comply with the PCI rule of having no financial conflicts of interest in
601 relation to the content of the article.

602

Funding

603 This work was supported by the Agence Nationale de la Recherche (ANR Harley, ANR-19-CE44-0017-
604 01; ANR PHOSTORE, ANR-19-CE01-0005). The PCC collection is funded by the Institut Pasteur. The PhD of
605 C.P. was supported by the Ile-de-France ARDoC Grant. Juliette Gaëtan PhD grant was funded by the
606 Learning Planet Institute and University Paris Cité grant Frontiers of Innovation in Research and Education.

References

- 608 Raw transcriptomics data are available on NCBI, Gene expression Omnibus, GEO accession number:
609 GSE255450. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE255450>
- 610 Supplementary information and supplementary data, figures and tables can be found on bioRxiv, at the
611 following address: <https://doi.org/10.1101/2024.07.07.602159>
- 612 Altermann, W., Kazmierczak, J., Oren, A., Wright, D.T., 2006. Cyanobacterial calcification and its
613 rock-building potential during 3.5 billion years of Earth history. *Geobiology* **4**, 147–166.
614 <https://doi.org/10.1111/j.1472-4669.2006.00076.x>
- 615 Badger, M.R., Price, G.D., 2003. CO₂ concentrating mechanisms in cyanobacteria: molecular
616 components, their diversity and evolution. *Journal of Experimental Botany* **54**, 609–622.
617 <https://doi.org/10.1093/jxb/erg076>
- 618 Benzerara, K., Duprat, E., Bitard-Feildel, T., Caumes, G., Cassier-Chauvat, C., Chauvat, F., Dezi, M.,
619 Diop, S.I., Gaschignard, G., Görden, S., Gugger, M., López-García, P., Millet, M., Skouri-
620 Panet, F., Moreira, D., Callebaut, I., 2022. A new gene family diagnostic for intracellular
621 biomineralization of amorphous Ca carbonates by cyanobacteria. *Genome Biology and
622 Evolution* **14**, evac026. <https://doi.org/10.1093/gbe/evac026>
- 623 Benzerara, K., Görden, S., Athar, K.M., Chauvat, F., March, K., Menguy, N., Mehta, N., Skouri-Panet,
624 F., Swaraj, S., Travert, C., Cassier-Chauvat, C., Duprat, E., 2023. Quantitative mapping of
625 calcium cell reservoirs in cyanobacteria at the submicrometer scale. *Journal of Electron
626 Spectroscopy and Related Phenomena* **267**, 147369.
627 <https://doi.org/10.1016/j.elspec.2023.147369>
- 628 Benzerara, K., Skouri-Panet, F., Li, J., Féraud, C., Gugger, M., Laurent, T., Couradeau, E., Ragon, M.,
629 Cosmidis, J., Menguy, N., Margaret-Oliver, I., Tavera, R., López-García, P., Moreira, D., 2014.
630 Intracellular Ca-carbonate biomineralization is widespread in cyanobacteria. *Proceedings of
631 the National Academy of Sciences* **111**, 10933–10938.
632 <https://doi.org/10.1073/pnas.1403510111>
- 633 Blondeau, M., Benzerara, K., Ferard, C., Guigner, J.-M., Poinot, M., Coutaud, M., Tharaud, M.,
634 Cordier, L., Skouri-Panet, F., 2018a. Impact of the cyanobacterium *Gloeomargarita lithophora*
635 on the geochemical cycles of Sr and Ba. *Chemical Geology* **483**, 88–97.
636 <https://doi.org/10.1016/j.chemgeo.2018.02.029>
- 637 Blondeau, M., Sachse, M., Boulogne, C., Gillet, C., Guigner, J.-M., Skouri-Panet, F., Poinot, M.,
638 Ferard, C., Miot, J., Benzerara, K., 2018b. Amorphous calcium carbonate granules form within
639 an intracellular compartment in calcifying cyanobacteria. *Front. Microbiol.* **9**, 1768.
640 <https://doi.org/10.3389/fmicb.2018.01768>
- 641 Boskey, A.L., 2003. Biomineralization: an overview. *Connective Tissue Research* **44**, 5–9.
642 <https://doi.org/10.1080/03008200390152007>
- 643 Brownlee, C., Wheeler, G.L., Taylor, A.R., 2015. Coccolithophore biomineralization: New questions,
644 new answers. *Seminars in Cell & Developmental Biology*, **46**, 11–16.
645 <https://doi.org/10.1016/j.semcdb.2015.10.027>
- 646 Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B., Mornon,
647 J.P., 1997. Deciphering protein sequence information through hydrophobic cluster analysis
648 (HCA): current status and perspectives. *Cell Mol Life Sci*, **53**, 621–645.
649 <https://doi.org/10.1007/s000180050082>
- 650 Cam, N., Benzerara, K., Georgelin, T., Jaber, M., Lambert, J.-F., Poinot, M., Skouri-Panet, F., Cordier,
651 L., 2016. Selective uptake of alkaline earth metals by cyanobacteria forming intracellular
652 carbonates. *Environ. Sci. Technol.* **50**, 11654–11662. <https://doi.org/10.1021/acs.est.6b02872>
- 653 Cam, N., Benzerara, K., Georgelin, T., Jaber, M., Lambert, J.-F., Poinot, M., Skouri-Panet, F., Moreira,
654 D., Lopez-Garcia, P., Raimbault, E., Cordier, L., Jezequel, D., 2018. Cyanobacterial formation
655 of intracellular Ca-carbonates in undersaturated solutions. *Geobiology* **16**, 49–61.
656 <https://doi.org/10.1111/gbi.12261>
- 657 Cam, N., Georgelin, T., Jaber, M., Lambert, J.-F., Benzerara, K., 2015. In vitro synthesis of amorphous
658 Mg-, Ca-, Sr- and Ba-carbonates: What do we learn about intracellular calcification by

- 659 cyanobacteria? *Geochimica et Cosmochimica Acta* **161**, 36–49.
660 <https://doi.org/10.1016/j.gca.2015.04.003>
- 661 Chen, S., Gagnon, A.C., Adkins, J.F., 2018. Carbonic anhydrase, coral calcification and a new model
662 of stable isotope vital effects. *Geochimica et Cosmochimica Acta* **236**, 179–197.
663 <https://doi.org/10.1016/j.gca.2018.02.032>
- 664 Christiano, R., Nagaraj, N., Fröhlich, F., Walther, T.C., 2014. Global proteome turnover analyses of the
665 yeasts *S. cerevisiae* and *S. pombe*. *Cell Reports* **9**, 1959–1965.
666 <https://doi.org/10.1016/j.celrep.2014.10.065>
- 667 Cosmidis, J., Benzerara, K., 2022. Why do microbes make minerals? *Comptes Rendus. Géoscience* **354**,
668 1–39. <https://doi.org/10.5802/crgeos.107>
- 669 Couradeau, E., Benzerara, K., Gerard, E., Moreira, D., Bernard, S., Brown, G.E., Lopez-Garcia, P.,
670 2012. An early-branching microbialite cyanobacterium forms intracellular carbonates. *Science*
671 **336**, 459–462. <https://doi.org/10.1126/science.1216171>
- 672 Cumsille, A., Durán, R.E., Rodríguez-Delherbe, A., Saona-Urmeneta, V., Cámara, B., Seeger, M.,
673 Araya, M., Jara, N., Buil-Aranda, C., 2023. GenoVi, an open-source automated circular genome
674 visualizer for bacteria and archaea. *PLOS Computational Biology* **19**, e1010998.
675 <https://doi.org/10.1371/journal.pcbi.1010998>
- 676 De Wever, A., Benzerara, K., Coutaud, M., Caumes, G., Poinso, M., Skouri-Panet, F., Laurent, T.,
677 Duprat, E., Gugger, M., 2019. Evidence of high Ca uptake by cyanobacteria forming
678 intracellular CaCO₃ and impact on their growth. *Geobiology* **17**, 676–690.
679 <https://doi.org/10.1111/gbi.12358>
- 680 Dick, G.J., Duhaime, M.B., Evans, J.T., Errera, R.M., Godwin, C.M., Kharbush, J.J., Nitschky, H.S.,
681 Powers, M.A., Vanderploeg, H.A., Schmidt, K.C., Smith, D.J., Yancey, C.E., Zwiers, C.C.,
682 Denef, V.J., 2021. The genetic and ecophysiological diversity of *Microcystis*. *Environmental*
683 *Microbiology* **23**, 7278–7313. <https://doi.org/10.1111/1462-2920.15615>
- 684 Domínguez, D.C., Guragain, M., Patrauchan, M., 2015. Calcium binding proteins and calcium signaling
685 in prokaryotes. *Cell Calcium, Evolution of Calcium Signaling* **57**, 151–165.
686 <https://doi.org/10.1016/j.ceca.2014.12.006>
- 687 Feldbauer, R., Gosch, L., Lüftinger, L., Hyden, P., Flexer, A., Rattei, T., 2021. DeepNOG: fast and
688 accurate protein orthologous group assignment. *Bioinformatics* **36**, 5304–5312.
689 <https://doi.org/10.1093/bioinformatics/btaa1051>
- 690 Gaëtan, J., Halary, S., Millet, M., Bernard, C., Duval, C., Hamlaoui, S., Hecquet, A., Gugger, M., Marie,
691 B., Mehta, N., Moreira, D., Skouri-Panet, F., Travert, C., Duprat, E., Leloup, J., Benzerara, K.,
692 2023. Widespread formation of intracellular calcium carbonates by the bloom-forming
693 cyanobacterium *Microcystis*. *Environmental Microbiology* **25**, 751–765.
694 <https://doi.org/10.1111/1462-2920.16322>
- 695 Gancedo, J.M., López, S., Ballesteros, F., 1982. Calculation of half-lives of proteins in vivo.
696 Heterogeneity in the rate of degradation of yeast proteins. *Mol Cell Biochem* **43**, 89–95.
697 <https://doi.org/10.1007/BF00423096>
- 698 Gaschignard, G., Millet, M., Bruley, A., Benzerara, K., Dezi, M., Skouri-Panet, F., Duprat, E.,
699 Callebaut, I., 2024. AlphaFold2-guided description of CoBaHMA, a novel family of bacterial
700 domains within the heavy-metal-associated superfamily. *Proteins*, 1-19. <https://doi.org/10.1002/prot.26668>
- 701
- 702 Gilbert, P.U.P.A., Bergmann, K.D., Boekelheide, N., Tambutté, S., Mass, T., Marin, F., Adkins, J.F.,
703 Erez, J., Gilbert, B., Knutson, V., Cantine, M., Hernández, J.O., Knoll, A.H., 2022.
704 Biomineralization: Integrating mechanism and evolutionary history. *Science Advances* **8**,
705 eabl9653. <https://doi.org/10.1126/sciadv.abl9653>
- 706 Görgen, S., Benzerara, K., Skouri-Panet, F., Gugger, M., Chauvat, F., Cassier-Chauvat, C., 2020. The
707 diversity of molecular mechanisms of carbonate biomineralization by bacteria. *Discov Mater*
708 **1**, 2. <https://doi.org/10.1007/s43939-020-00001-9>
- 709 Gu, P., Li, Q., Zhang, W., Zheng, Z., Luo, X., 2020. Effects of different metal ions (Ca, Cu, Pb, Cd) on
710 formation of cyanobacterial blooms. *Ecotoxicol. Environ. Saf.* **189**, 109976.
711 <https://doi.org/10.1016/j.ecoenv.2019.109976>

- 712 Harke, M.J., Steffen, M.M., Gobler, C.J., Otten, T.G., Wilhelm, S.W., Wood, S.A., Paerl, H.W., 2016.
713 A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium,
714 *Microcystis* spp. *Harmful Algae* **54**, 4–20. <https://doi.org/10.1016/j.hal.2015.12.007>
- 715 Hirschi, K.D., Zhen, R.G., Cunningham, K.W., Rea, P.A., Fink, G.R., 1996. CAX1, an H⁺/Ca²⁺
716 antiporter from *Arabidopsis*. *Proceedings of the National Academy of Sciences* **93**, 8782–8786.
717 <https://doi.org/10.1073/pnas.93.16.8782>
- 718 Huang, J., Wang, J., Xu, H., 2014. The circadian rhythms of photosynthesis, ATP content and cell
719 division in *Microcystis aeruginosa* PCC7820. *Acta Physiol Plant* **36**, 3315–3323.
720 <https://doi.org/10.1007/s11738-014-1699-1>
- 721 Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., Weissman, J.S., 2009. Genome-wide analysis in
722 vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223.
723 <https://doi.org/10.1126/science.1168978>
- 724 Knoll, A.H., 2003. Biomineralization and evolutionary history. *Reviews in Mineralogy and*
725 *Geochemistry* **54**, 329–356. <https://doi.org/10.2113/0540329>
- 726 Kupriyanova, E.V., Pronina, N.A., Los, D.A., 2023. Adapting from low to high: an update to CO₂-
727 concentrating mechanisms of cyanobacteria and microalgae. *Plants* **12**, 1569. doi:
728 [10.3390/plants12071569](https://doi.org/10.3390/plants12071569).
- 729 Lambert, I., Paysant-Le Roux, C., Colella, S., Martin-Magniette, M.-L., 2020. DiCoExpress: a tool to
730 process multifactorial RNAseq experiments from quality controls to co-expression analysis
731 through differential analysis based on contrasts inside GLM models. *Plant Methods* **16**, 68.
732 <https://doi.org/10.1186/s13007-020-00611-7>
- 733 Le Roy, N., Jackson, D.J., Marie, B., Ramos-Silva, P., Marin, F., 2014. The evolution of metazoan α-
734 carbonic anhydrases and their roles in calcium carbonate biomineralization. *Frontiers in*
735 *Zoology* **11**, 75. <https://doi.org/10.1186/s12983-014-0075-8>
- 736 Lê, S., Josse, J., Husson, F., 2008. FactoMineR: an R package for multivariate analysis. *Journal of*
737 *Statistical Software* **25**, 1–18. <https://doi.org/10.18637/jss.v025.i01>
- 738 Li, J., Margaret Oliver, I., Cam, N., Boudier, T., Blondeau, M., Leroy, E., Cosmidis, J., Skouri-Panet,
739 F., Guigner, J.-M., Féraud, C., Poinot, M., Moreira, D., Lopez-Garcia, P., Cassier-Chauvat, C.,
740 Chauvat, F., Benzerara, K., 2016. Biomineralization patterns of intracellular carbonatogenesis
741 in cyanobacteria: molecular hypotheses. *Minerals* **6**, 10. <https://doi.org/10.3390/min6010010>
- 742 Liu, P., Liu, Y., Ren, X., Zhang, Z., Zhao, X., Roberts, A.P., Pan, Y., Li, J., 2021. A novel magnetotactic
743 alphaproteobacterium producing intracellular magnetite and calcium-bearing minerals. *Applied*
744 *and Environmental Microbiology* **87**, e0155621. <https://doi.org/10.1128/AEM.01556-21>
- 745 Liu, P., Zheng, Y., Zhang, R., Bai, J., Zhu, K., Benzerara, K., Menguy, N., Zhao, X., Roberts, A.P.,
746 Pan, Y., Li, J., 2023. Key gene networks that control magnetosome biomineralization in
747 magnetotactic bacteria. *National Science Review* **10**, nwac238.
748 <https://doi.org/10.1093/nsr/nwac238>
- 749 Mackinder, L., Wheeler, G., Schroeder, D., von Dassow, P., Riebesell, U., Brownlee, C., 2011.
750 Expression of biomineralization-related ion transport genes in *Emiliania huxleyi*.
751 *Environmental Microbiology* **13**, 3250–3265. <https://doi.org/10.1111/j.1462-2920.2011.02561.x>
- 752
- 753 Mantione, K.J., Kream, R.M., Kuzelova, H., Ptacek, R., Raboch, J., Samuel, J.M., Stefano, G.B., 2014.
754 Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. *Med*
755 *Sci Monit Basic Res* **20**, 138–141. <https://doi.org/10.12659/MSMBR.892101>
- 756 Marin, F., Corstjens, P., de Gaulejac, B., de Vrind-De Jong, E., Westbroek, P., 2000. Mucins and
757 molluscan calcification: molecular characterization of mucoperlin, a novel mucin-like protein
758 from the nacreous shell layer of the fan mussel *Pinna nobilis* (Bivalvia, pteriomorphia). *Journal*
759 *of Biological Chemistry* **275**, 20667–20675. <https://doi.org/10.1074/jbc.M003006200>
- 760 Marin, F., Luquet, G., 2004. Molluscan shell proteins. *Comptes Rendus Palevol* **3**, 469–492.
761 <https://doi.org/10.1016/j.crpv.2004.07.009>
- 762 Marin, F., Smith, M., Isa, Y., Muyzer, G., Westbroek, P., 1996. Skeletal matrices, mucins, and the origin
763 of invertebrate calcification. *Proceedings of the National Academy of Sciences* **93**, 1554–1559.
764 <https://doi.org/10.1073/pnas.93.4.1554>

- 765 Mehta, N., Bougoure, J., Kocar, B.D., Duprat, E., Benzerara, K., 2022a. Cyanobacteria accumulate
766 radium (^{226}Ra) within intracellular amorphous calcium carbonate inclusions. *ACS EST Water*
767 **2**, 616–623. <https://doi.org/10.1021/acsestwater.1c00473>
- 768 Mehta, N., Gaëtan, J., Giura, P., Azais, T., Benzerara, K., 2022b. Detection of biogenic amorphous
769 calcium carbonate (ACC) formed by bacteria using FTIR spectroscopy. *Spectrochimica Acta*
770 *Part A: Molecular and Biomolecular Spectroscopy* **278**, 121262.
771 <https://doi.org/10.1016/j.saa.2022.121262>
- 772 Mehta, N., Vantelon, D., Gaëtan, J., Fernandez-Martinez, A., Delbes, L., Travert, C., Benzerara, K.,
773 2023. Calcium speciation and coordination environment in intracellular amorphous calcium
774 carbonate (ACC) formed by cyanobacteria. *Chemical Geology* **641**, 121765.
775 <https://doi.org/10.1016/j.chemgeo.2023.121765>
- 776 Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold:
777 making protein folding accessible to all. *Nat Methods* **19**, 679–682.
778 <https://doi.org/10.1038/s41592-022-01488-1>
- 779 Monteil, C.L., Benzerara, K., Menguy, N., Bidaud, C.C., Michot-Achdjian, E., Bolzoni, R., Mathon,
780 F.P., Coutaud, M., Alonso, B., Garau, C., Jézéquel, D., Viollier, E., Ginet, N., Floriani, M.,
781 Swaraj, S., Sachse, M., Busigny, V., Duprat, E., Guyot, F., Lefevre, C.T., 2021. Intracellular
782 amorphous Ca-carbonate and magnetite biomineralization by a magnetotactic bacterium
783 affiliated to the Alphaproteobacteria. *ISME J* **15**, 1–18. [https://doi.org/10.1038/s41396-020-](https://doi.org/10.1038/s41396-020-00747-3)
784 [00747-3](https://doi.org/10.1038/s41396-020-00747-3)
- 785 Monteil, C.L., Perrière, G., Menguy, N., Ginet, N., Alonso, B., Waisbord, N., Cruveiller, S., Pignol, D.,
786 Lefèvre, C.T., 2018. Genomic study of a novel magnetotactic Alphaproteobacteria uncovers
787 the multiple ancestry of magnetotaxis. *Environmental Microbiology* **20**, 4415–4430.
788 <https://doi.org/10.1111/1462-2920.14364>
- 789 Monteiro, F.M., Bach, L.T., Brownlee, C., Bown, P., Rickaby, R.E.M., Poulton, A.J., Tyrrell, T.,
790 Beaufort, L., Dutkiewicz, S., Gibbs, S., Gutowska, M.A., Lee, R., Riebesell, U., Young, J.,
791 Ridgwell, A., 2016. Why marine phytoplankton calcify. *Science Advances* **2**, e1501822.
792 <https://doi.org/10.1126/sciadv.1501822>
- 793 Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-
794 aware quantification of transcript expression. *Nat Methods* **14**, 417–419.
795 <https://doi.org/10.1038/nmeth.4197>
- 796 Pattanayak, G., Rust, M.J., 2014. The cyanobacterial clock and metabolism. *Current Opinion in*
797 *Microbiology* **18**, 90–95. <https://doi.org/10.1016/j.mib.2014.02.010>
- 798 Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L.,
799 Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D.H., Letunić, I., Marchler-Bauer, A., Mi,
800 H., Natale, D.A., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I.,
801 Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A., 2023. InterPro in 2022.
802 *Nucleic Acids Research* **51**, D418–D427. <https://doi.org/10.1093/nar/gkac993>
- 803 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E.,
804 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput*
805 *Chem* **25**, 1605–1612. <https://doi.org/10.1002/jcc.20084>
- 806 Price, G.D., Howitt, S.M., 2011. The cyanobacterial bicarbonate transporter *BicA*: its physiological role
807 and the implications of structural similarities with human SLC26 transporters. *Biochem. Cell*
808 *Biol.* **89**, 178–188. <https://doi.org/10.1139/O10-136>
- 809 Ragon, M., Benzerara, K., Moreira, D., Tavera, R., Lopez-Garcia, P., 2014. 16S rDNA-based analysis
810 reveals cosmopolitan occurrence but limited diversity of two cyanobacterial lineages with
811 contrasted patterns of intracellular carbonate mineralization. *Front. Microbiol.* **5**, 331.
812 <https://doi.org/10.3389/fmicb.2014.00331>
- 813 Riba, A., Di Nanni, N., Mittal, N., Arhné, E., Schmidt, A., Zavolan, M., 2019. Protein synthesis rates
814 and ribosome occupancies reveal determinants of translation elongation rates. *Proceedings of*
815 *the National Academy of Sciences* **116**, 15023–15032.
816 <https://doi.org/10.1073/pnas.1817299116>
- 817 Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., Stanier, R.Y., 1979. Generic assignments,
818 strain histories and properties of pure cultures of cyanobacteria. *Microbiology* **111**, 1–61.
819 <https://doi.org/10.1099/00221287-111-1-1>

- 820 Rivadeneyra, M.A., Martín-Algarra, A., Sánchez-Navas, A., Martín-Ramos, D. 2006. Carbonate and
821 phosphate precipitation by *Chromohalobacter marismortui*, *Geomicrobiology Journal*, **23**, 89-
822 101. <https://doi.org/10.1080/01490450500533882>
- 823 Rivadeneyra, M.A., Martín-Algarra, A., Sánchez-Román, M., Sánchez-Navas, A., Martín-Ramos, J.D.,
824 2010. Amorphous Ca-phosphate precursors for Ca-carbonate biominerals mediated by
825 *Chromohalobacter marismortui*. *ISME J*, **4**, 922–932. <https://doi.org/10.1038/ismej.2010.17>
- 826 Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis
827 of RNA-seq data. *Genome Biology* **11**, R25. <https://doi.org/10.1186/gb-2010-11-3-r25>
- 828 Sandrini, G., Cunsolo, S., Schuurmans, J.M., Matthijs, H.C.P., Huisman, J., 2015. Changes in gene
829 expression, cell physiology and toxicity of the harmful cyanobacterium *Microcystis aeruginosa*
830 at elevated CO₂. *Frontiers in Microbiology* **6**, 401. <https://doi.org/10.3389/fmicb.2015.00401>
- 831 Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., Selbach, M.,
832 2011. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342.
833 <https://doi.org/10.1038/nature10098>
- 834 Shigaki, T., Rees, I., Nakhleh, L., Hirschi, K.D., 2006. Identification of three distinct phylogenetic
835 groups of CAX cation/proton antiporters. *J Mol Evol* **63**, 815–825.
836 <https://doi.org/10.1007/s00239-006-0048-4>
- 837 Stöckel, J., Welsh, E.A., Liberton, M., Kunnvakkam, R., Aurora, R., Pakrasi, H.B., 2008. Global
838 transcriptomic analysis of *Cyanothece* 51142 reveals robust diurnal oscillation of central
839 metabolic processes. *Proceedings of the National Academy of Sciences* **105**, 6156–6161.
840 <https://doi.org/10.1073/pnas.0711068105>
- 841 Straub, C., Quillardet, P., Vergalli, J., Tandeau de Marsac, N., Humbert, J.-F., 2011. A day in the life
842 of *Microcystis aeruginosa* strain PCC 7806 as revealed by a transcriptomic analysis. *PLOS*
843 *ONE* **6**, e16208. <https://doi.org/10.1371/journal.pone.0016208>
- 844 Tang, J., Zhou, H., Yao, D., Riaz, S., You, D., Klepacz-Smółka, A., Daroch, M., 2022. Comparative
845 genomic analysis revealed distinct molecular components and organization of CO₂-
846 concentrating mechanism in thermophilic cyanobacteria. *Frontiers in Microbiology* **13**,
847 876272. <https://doi.org/10.3389/fmicb.2022.876272>
- 848 Taoka, A., Eguchi, Y., Shimoshige, R., Fukumori, Y., 2023. Recent advances in studies on
849 magnetosome-associated proteins composing the bacterial geomagnetic sensor organelle.
850 *Microbiology and Immunology* **67**, 228–238. <https://doi.org/10.1111/1348-0421.13062>
- 851 Uebe, R., Schüler, D., 2016. Magnetosome biogenesis in magnetotactic bacteria. *Nat Rev Microbiol* **14**,
852 621–637. <https://doi.org/10.1038/nrmicro.2016.99>
- 853 van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J.,
854 Steinegger, M., 2024. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol*
855 **42**, 243–246. <https://doi.org/10.1038/s41587-023-01773-0>
- 856 Waditee, R., Hibino, T., Tanaka, Y., Nakamura, T., Incharoensakdi, A., Takabe, T., 2001. Halotolerant
857 cyanobacterium *Aphanothece halophytica* contains an Na⁽⁺⁾/H⁽⁺⁾ antiporter, homologous to
858 eukaryotic ones, with novel ion specificity affected by C-terminal tail. *J Biol Chem*. **276**, 36931-
859 36938. <https://doi.org/10.1074/jbc.M10365020>
- 860 Waditee, R., Hossain, G.S., Tanaka, Y., Nakamura, T., Shikata, M., Takano, J., Takabe, Tetsuko,
861 Takabe, Teruhiro, 2004. Isolation and functional characterization of Ca²⁺/H⁺ antiporters from
862 Cyanobacteria. *J. Biol. Chem.* **279**, 4330–4338. <https://doi.org/10.1074/jbc.M310282200>
- 863 Waight, A.B., Pedersen, B.P., Schlessinger, A., Bonomi, M., Chau, B.H., Roe-Zurz, Z., Risenmay, A.J.,
864 Sali, A., Stroud, R.M., 2013. Structural basis for alternating access of a eukaryotic
865 calcium/proton exchanger. *Nature* **499**, 107–110. <https://doi.org/10.1038/nature12233>
- 866 Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev*
867 *Genet* **10**, 57–63. <https://doi.org/10.1038/nrg2484>
- 868 Welkie, D.G., Rubin, B.E., Diamond, S., Hood, R.D., Savage, D.F., Golden, S.S., 2019. A Hard Day's
869 Night: cyanobacteria in diel cycles. *Trends in Microbiology* **27**, 231–242.
870 <https://doi.org/10.1016/j.tim.2018.11.002>
- 871 Weyhenmeyer, G.A., Hartmann, J., Hessen, D.O., Kopáček, J., Hejzlar, J., Jacquet, S., Hamilton, S.K.,
872 Verburg, P., Leach, T.H., Schmid, M., Flaim, G., Nöges, T., Nöges, P., Wentzky, V.C., Rogora,
873 M., Rusak, J.A., Kosten, S., Paterson, A.M., Teubner, L., Higgins, S.N., Lawrence, G., Kangur,
874 K., Kokorite, I., Cerasino, L., Funk, C., Harvey, R., Moatar, F., de Wit, H.A., Zechmeister, T.,

875 2019. Widespread diminishing anthropogenic effects on calcium in freshwaters. *Sci Rep* **9**,
876 10450. <https://doi.org/10.1038/s41598-019-46838-w>
877 Yang, T., Teske, A., Ambrose, W., Salman-Carvalho, V., Bagnell, R., & Nielsen, L. P., 2019.
878 Intracellular calcite and sulfur dynamics of *Achromatium* cells observed in a lab-based
879 enrichment and aerobic incubation experiment. *Antonie van Leeuwenhoek*, **112**, 263-274.
880 <https://doi.org/10.1007/s10482-018-1153-2>
881 Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., Liu, X., 2014. Comparison of RNA-Seq and
882 microarray in transcriptome profiling of activated T cells. *PLOS ONE* **9**, e78644.
883 <https://doi.org/10.1371/journal.pone.0078644>
884 Zinser, E.R., Lindell, D., Johnson, Z.I., Futschik, M.E., Steglich, C., Coleman, M.L., Wright, M.A.,
885 Rector, T., Steen, R., McNulty, N., Thompson, L.R., Chisholm, S.W., 2009. Choreography of
886 the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph,
887 *Prochlorococcus*. *PLOS ONE* **4**, e5135. <https://doi.org/10.1371/journal.pone.0005135>
888
889
890