The *Thermosynechococcus* genus: wide environmental distribution, but a highly conserved genomic core

A. Paulina Prondzinsky[1,2]*, Sarah J. Berkemer[3,4], Lewis M. Ward[2,5], Shawn E. McGlynn[2]*

[1]Department of Chemical Science and Engineering, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8550 Japan

[2]Earth-Life Science Institute, Ota-ku, Tokyo 145-0061, Japan

[3]Bioinformatics Group, Department of Computer Science, University Leipzig, Leipzig, Germany

[4]Competence Center for Scalable Data Services and Solutions, Dresden/Leipzig, Germany

[5]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA

*prondzinsky@elsi.jp / mcglynn@elsi.jp

Abstract

Cyanobacteria thrive in very diverse environments. In Earth history however, delayed oxygenation has raised questions of growth limitation in ancient environmental conditions. As a single genus, the *Thermosynechococcus* are known to be cosmopolitan and live in chemically diverse habitats. To understand the genetic basis for this, we compared the protein coding component of *Thermosynechococcus* genomes. Supplementing the known genetic diversity of *Thermosynechococcus*, we report draft metagenome-assembled genomes of two *Thermosynechococcus* recovered from ferrous carbonate hot springs in Japan. We find that as a genus, *Thermosynechococcus* is genomically conserved, having a small pan-genome with few accessory genes per individual strain and only 18 protein clusters appearing in all *Thermosynechococcus* but not in any other cyanobacteria in our analysis. Furthermore, by comparing orthologous protein groups, including an analysis of genes encoding proteins with an iron related function (uptake, storage or utilization), no clear differences in genetic content, or adaptive mechanisms could be detected between genus members, despite the range of environments they inhabit. Overall, our results highlight a seemingly innate ability for *Thermosynechococcus* to inhabit diverse habitats without having undergone substantial genomic adaptation to accommodate this. The finding of *Thermosynechococcus* in both hot and high iron

environments without adaptation recognizable from the perspective of protein coding genes has implications for understanding the basis of thermophily within this clade, and also suggests that ferrous iron in ancient oceans may not have inhibited the proliferation of Cyanobacteria on Earth. The conserved core genome may be indicative of an allopatric lifestyle – or reduced genetic complexity of hot spring habitats relative to other environments.

Keywords: cyanobacteria, great oxygenation event, hot springs, comparative genomics, pan-genome, *Thermosynechococcus*

Introduction

Water oxidizing cyanobacteria fundamentally altered the distribution of carbon and electrons on Earth (Canfield, 1998; Raven, 2009; Ward and Shih, 2019). A marker of this reorganization is in the Great Oxygenation Event (GOE), which marks a major transition in the evolution of life on Earth ~2.3 billion years ago (Lyons *et al.*, 2014; Fischer *et al.*, 2016). While the GOE is widely accepted to have been driven by the production of O_2 by oxygenic photosynthesis performed by members of the Cyanobacteria, the timing and proximal trigger for the GOE is debated**.** At least six hypotheses for the timing of the GOE can be considered i) the evolution of oxygenic photosynthesis by Cyanobacteria just before the GOE (Fischer *et al.*, 2016; Shih *et al.*, 2017), ii) an earlier evolution of Cyanobacteria with O_2 accumulation delayed due to the transition of cyanobacteria from small-scale freshwater to large-scale marine environments (Sánchez-Baracaldo, 2015), iii) the transition from unicellular to multicellular organisms for increased evolutionary success (Schirrmeister *et al.*, 2015), iv) the inhibition of early cyanobacteria due to high iron concentrations (Swanner, Mloszewska, *et al.*, 2015; Swanner, Wu, *et al.*, 2015), v) a possible nitrogen throttle on cyanobacterial growth (Fennel *et al.*, 2005; Shi and Falkowski, 2008; Allen *et al.*, 2019), vi) or depressed Archaean productivity due to phosphate availability (Hao *et al.* 2020*)*. Although cyanobacterial taxonomy continues to be refined (Knoll 2006; Tomitani *et al.*, 2006; Shih *et al.*, 2013; Soo *et al.*, 2019; Parks *et al.*, 2020), understanding the ecological distribution of cyanobacteria may help us generate, or reject hypotheses about their evolutionary trajectories, and the factors which led to the observed timing of the GOE.

Name	Strain name (Thermosynechococcus)	Max. Temperature [°C]	pH	Sulfate [mM]	Iron [mM]	References
Kangding Geothermal Field (Sichuan, China)	*T. elongatus* PKUAC-SCTE54	94	6.35 – 8.84	1.2	0.0104	Mei-hua *et al.*, 2015; Guo *et al.*, 2017; Tang *et al.*, 2018; Wang *et al.*, 2019
Okuoku-hachikurou Hot Spring (Akita, Japan)	*T. nakabusensis* OHK	44	6.8	6.5	0.114	Ward *et al.*, 2017
Yunomine Hot Spring (Wakayama, Japan)	*T. vulcanus* NIES2134	91	8	0.06 – 0.229	< 0.0018	Ichikuni and Kikuchi, 1972 Onsen information sheet*
Nakabusa Hot Spring (Nagano, Japan)	*T. nakabusensis* NK55a	76	8.5 – 9	0.218 – 0.246	0.0004	Nakagawa and Fukui, 2002; Nakamura *et al.*, 2002; Stolyar *et al.*, 2014 Onsen information sheet*
Jinata Hot Spring (Shikinejima,	*T. nakabusensis* Jinata	63	5.4	17.4	0.261	Ward *et al.*, 2019

Tokyo, Japan)						
Chung-Lun Hot Spring (Taiwan)	*T.* CL1	62	9.3	1.35 – 1.39	0.0006	Yeh *et al.*, 2005; Hsueh *et al.*, 2007; Maity *et al.*, 2011; Cheng *et al.*, 2020
Beppu Hot Spring Kamegawa Shinoyu (Oita, Japan)	*T. elongatus* BP1	78	6.8	1.09	0.0036	Onsen information sheet*

Table 1: Geochemical parameters for each hot spring. References are related to the first publication of the strains or the geochemistry of the hot spring. *information available online through local governments, last accessed in 11/2019.

Cyanobacteria are found in a wide range of environments – the knowledge of which continues to expand (Whitton, 1992; Puente-Sánchez *et al.*, 2018; Callieri *et al.*, 2019). As a single phylogenetic group, members of the *Thermosynechococcus* genus have been documented to inhabit a range of chemical environments, some of which might be similar to places and times on the early Earth. Those *Thermosynechococcus* with genomes available are from hot springs which vary in temperature from 44 – 94 °C, pH ranging from 5.4 – 9.3, sulfate concentrations between 0.06 mM and 17.4 mM and iron concentrations between 0.0004 mM and 0.261 mM (Table 1) at their source. It is noteworthy that for two of the genomes - those from Okuoku-hachikurou and Jinata hot springs - the ferrous iron concentrations at the source by far exceed those at the other springs with 114 μM at OHK and 261 μM at Jinata (Ward *et al.*, 2019; Cheng *et al.*, 2020).

Seminal work by Papke et al. on phylotype:geographical relationships of *Thermosynechococcus* posed questions as to how these organisms could be so widely distributed: in their analysis, the distribution of *Thermosynechococcus* could not be explained by measured geochemical parameters. *Thermosynechococcus* thus appear to be cosmopolitan, but the basis for this remains unresolved. Motivated by the finding of *Thermosynechococcus* members in ferrous iron carbonate hot springs that we have been studying (Ward *et al.*, 2017, 2019), and in an attempt to find a genetic basis for the geochemically wide distribution of the organisms, we took a comparative genomics approach to search for resolvable traits which may underlie environmental adaptations of *Thermosynechococcus* members.

80

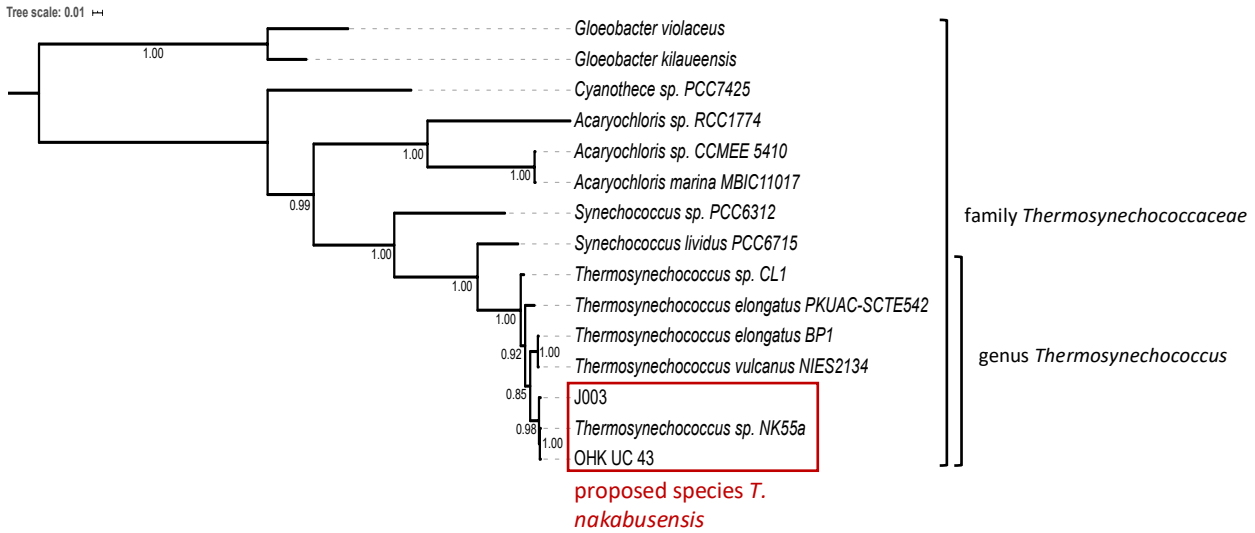

Figure 1*: Thermosynechococcaceae* phylogeny built with concatenated ribosomal proteins. Branch supports are derived from bootstrapping with BOOSTER and the tree scale bar indicates substitutions per nucleotide site.

81

82 **Methods**

83 **Genome recovery**

84 The OHK43 genome was recovered from genome-resolved metagenomic sequencing of samples from
85 Okuoku-hachikurou Onsen (OHK) in Akita Prefecture, Japan, following methods described previously
86 (Ward *et al.*, 2019, 2020) and described briefly here. Samples of thin biofilms in the outflow of the hot
87 spring were sampled for metagenomic sequencing in September 2016. DNA was preserved in the field
88 with a Zymo Terralyzer BashingBead Matrix and Xpedition Lysis Buffer (Zymo Research, Irvine, CA)
89 after disruption of cells in polyethylene sample tubes via attachment to and operation of a cordless
90 reciprocating saw (Makita JR101DZ). Microbial DNA was extracted and purified after return to the lab
91 with a Zymo Soil/Fecal DNA extraction kit (Zymo Research, Irvine, CA Quantification of DNA was
92 performed with a Qubit 3.0 fluorimeter (Life Technologies, Carlsbad, CA). DNA was submitted to
93 SeqMatic LLC (Fremont, CA) for library preparation using an Illumina Nextera XT DNA Library
94 Preparation Kit prior to 2x100bp paired-end sequencing via Illumina HiSeq 4000 technology. Raw
95 sequence reads were quality controlled with BBTools (Bushnell 2014) and assembled with MegaHit v.
96 1.02 (Li *et al.*, 2016). The OHK43 genome bin was recovered via differential coverage binning with
97 MetaBAT (Kang *et al.*, 2015). Completeness and contamination/redundancy were determined with
98 CheckM v1.1.2 (Parks *et al.*, 2015). The genome was uploaded to RAST v2.0 for annotation and
99 characterization (Aziz *et al.*, 2008). Presence or absence of metabolic pathways of interest was

100 predicted using MetaPOAP v1.0 (Ward *et al.*, 2018). Taxonomic assignment was determined with
101 GTDB-Tk v1.2 (Parks *et al.*, 2018, 2020; Chaumeil *et al.*, 2019). Genome assembly and quality statistics
102 for OHK43 and other *Thermosynechoccus* genomes are reported in Supplemental Table 4.

103 **Organismal Phylogenies**

104 Concatenated ribosomal phylogenies were constructed following methods from Hug *et al.* (2016).
105 Members of the *Thermosynechococcaceae* and outgroups were identified using GTDB (Chaumeil *et al.*,
106 2019) and their genomes downloaded from the NCBI WGS and Genbank databases. Ribosomal protein
107 sequences were extracted using the tblastn function of BLAST+ (Camacho *et al.*, 2009) and aligned
108 with MUSCLE (Edgar, 2004). Trees were built with RAxML v.8.2.12 (Stamatakis, 2014) on the Cipres
109 science gateway (Miller *et al.*, 2010). Transfer bootstrap support values were determined with
110 BOOSTER (Lemoine *et al.*, 2018). Visualization of trees was performed with the Interactive Tree of Life
111 Viewer (Letunic and Bork, 2016).

112 **Genome comparison**

113 We compared the core- and pangenomes of *Thermosynechococcus* at the genus level, family level and
114 with a sub-sample of Cyanobacteria across the GTDB defined class Cyanobacteriia. ProteinOrtho
115 (Lechner *et al.*, 2011) was used for the identification of Conserved Likely Orthologous Groups (CLOG)
116 and analysis.

117 At the genus level 7 *Thermosynechococcus* strains from varying hot spring environments (Table 1 and
118 2) were compared. The genomic data from five available sequences of *Thermosynechococcus* strains
119 *T.* sp. CL1/1-2178 (CL1), *T. elongatus* BP1/1-2178 (*elongatus* BP1), *T. vulcanus* NIES2134/1-2178
120 (*vulcanus*), *T.* sp. NK55a/1-2022 (NK55a), *T. elongatus* PKUAC-SCTE542 (*elongatus* PKUAC) and two
121 unnamed strains from Jinata Onsen (Jinata) and Oku-Okuhachikurou Onsen (OHK) in Japan were
122 compared using ProteinOrtho (Lechner *et al.*, 2011), BLAST (Altschul *et al.*, 1990), and FeGenie (Garber
123 *et al.*, 2019). Phylogenetic relationships between these strains were established using concatenated
124 ribosomal protein phylogenies following Hug *et al.* (2016), taxonomic classifications with GTDB-tk
125 (Chaumeil *et al.*, 2019) and average nucleotide identities (Rodriguez-R and Konstantinidis, 2017).

126 For the analysis at the family level, we included 6 more species that appear as representative strains at
127 the *Thermosynochoccaceae* family level in the Genome Taxonomy Data Base (GTDB). The
128 representative GTDB strains include *Acaryochloris marina* MBIC11017, *Cyanothece* sp. PCC 7425,

129 *Acaryochloris* sp. CCMEE 5410, *Synechococcus* sp. PCC 6312, *Synechococcus lividus* PCC 6715 and
130 *Acaryochloris* sp. RCC1774. This resulted in a total of 13 strains for family level analysis.

131 At the class level we used the 7 genus level strains and additionally 16 species, which do not cover all
132 of the family level species, 15 after (Beck *et al.*, 2012) and one more *Gloeobacter* species (*G. kilaueensis*
133 JS-1). Here the goal was to test the coherence of our analysis parameters when compared to previous
134 studies of cyanobacterial core- and pangenomes (Beck *et al.*, 2012, 2018). We acknowledge that the
135 distribution of species within the families of the class is not even, however this comparison was used to
136 verify our methods and the stability of the class level core. This resulted in a total of 23 genomes for the
137 class level analysis.

138 It is important to note that although the spring source water properties differ, the source water is not
139 necessarily where the DNA originated. Thus there is some uncertainty in the precise
140 geochemistry:genome relationships discussed here. We also acknowledge that not all genomes in this
141 study reach 100 % completeness. We might therefore be missing some gene clusters and some
142 genomes might cluster differently if they were complete. We compared genomic data as of August 4th,
143 2020.

144 To compare the appearance of genes related to iron uptake and regulation we used FeGenie (Garber
145 *et al.*, 2019) with standard parameters. We used results from ProteinOrtho (Lechner *et al.*, 2011) for
146 further analysis of the core, shared, unique, TS core and TS shared clusters. Proteinortho was applied
147 such that the output included also singleton clusters (only containing a single protein) and with an
148 algebraic connectivity of 0 as a measure for the structure of the orthologous clusters. We did not obtain
149 many differences when running ProteinOrtho on our data using a value of 0 or 0.1 (default) for the
150 algebraic connectivity, however, a value of 0 resulted in slightly larger clusters for the core and a few
151 less singletons. A comparison of results from FeGenie and ProteinOrtho resulted in similar gene clusters
152 for iron related genes such that our results here are based on both program outputs.

153

154

155

156

157

158

159 **Results and Discussion**

Genome Size	*T. nakabusensis* Jinata	*T. nakabusensis* OHK	*T. elongatus* BP-1	*T. elongatus PKUAC-SCTE542*	*T. sp CL1*	*T. nakabusensis* NK55a	*T. vulcanus* NIES2134
T. nakabusensis Jinata	2.31 MB	99.70	92.65	86.82	87.53	99.68	92.59
T. nakabusensis OHK		2.25 MB	92.58	86.85	87.52	99.84	92.55
T. elongatus BP-1			2.59 MB	86.65	87.34	92.53	99.14
T. elongatus PKUAC-SCTE542				2.64 MB	89.95	86.81	86.637
T. sp CL1					2.64 MB	87.50	87.32
T. nakabusensis NK55a						2.51 MB	92.45
T. vulcanus NIES2134							2.57 MB

Genome Sizes for GTDB representative family level species:	
Acaryochloris marina MBIC11017	8.36 MB
Acaryochloris sp. CCMEE 5410	7.87 MB
Cyanothece sp. PCC 7425	5.78 MB
Synechococcus lividus PCC 6715	2.65 MB
Synechococcus sp. PCC 6312	3.72 MB
Acarychloris RCC 1774	5.93 MB

160 Table 2: Genus level average nucleotide identities (ANI) and genome sizes (diagonal). Additionally, genomes sizes for
161 family level species are added in the lower left.

162 ***Phylogeny of the Thermosynechococcus and the species proposal T. nakabusensis***

163 The *Thermosynechococcus* genus is phylogenetically coherent within the Cyanobacteria (Figure 1) and
164 the genome sizes of genus members are similar to one another (Table 2). Based on similarity observed
165 with ANI and GTDB-tk (Table 2, supplementary Table 4), 4 species are present within the
166 *Thermosynechococcus* genus: *T. elongatus* BP1 and *T. vulcanus* belonging to one species, *T.* NK55a,
167 the Jinata and OHK genomes belonging to a second species, and *T.* CL1 and *T. elongatus* PKUAC-
168 SCTE542 as one species each. For the species including the *T.* NK55a, Jinata, and OHK genomes we
169 propose the name *Thermosynechococcus nakabusensis* as the first and so far only isolated organism
170 which originates from Nakabusa hot spring in Nagano Prefecture, Japan.

171 **Genus and family level comparison of *Thermosynechococcus* and**
172 ***Thermosynochococcaceae***

173 Comparing the conserved likely orthologous groups (CLOGS), we analyzed i) the core-genomes: those
174 CLOGS shared by all genomes in an analysis, ii) the shared CLOGs: those shared by at least 2 but not
175 all of the genomes in the analysis, and iii) unique CLOGs: those CLOGs that are unique to a single

176 genome (Table 3). The *Thermosynechococcus* genus specific core (**core TS)** comprises CLOGs shared
177 by all 7 genus level genomes that are not present in any other species, and the *Thermosynechococcus*
178 genus-specific shared CLOGs (**shared TS)** corresponds to CLOGs that are shared by at least 2 and at
179 most 6 genus level genomes.

	Genus (7 genomes)	Family (13 genomes)	Class (23 genomes)
Core of all genomes analyzed	1878	1283	742
Core of 5 Thermosynechococcus (not including Jinata/OHK)	27	3	3
Jinata and OHK only	28	25	24
Protein Clusters Unique to Thermosynechococcus genus	-	18	79*

180 Table 3 - Numbers of CLOGs per grouping and phylogenetic level. *note that the number of genus level
181 core CLOGs increases due to the exclusion of some family level genomes at the class level.

182 Comparing the genomes of the seven genus members, the protein core is made up of 1878 CLOGs and
183 contains 80 to 89 % of the putative protein coding genes in a genome. This is higher than in other
184 comparisons of organisms, for example in Reno *et al.* (2009) 69 – 79 % of genes from *S. islandicus*
185 strains made up the core. Wu *et al.* (2018) found that the core genes of species belonging to the genus
186 *Comamonas* account for 18 – 33 % of all genes, and Barajas *et al.* (2019) investigated the core genome
187 within the genus *Streptococcus* which ranges in size from 9.6 % to 24 %. The high proportion of CLOGs
188 which make up the core leaves few unique CLOGs for each genome and between genus members, and
189 only slight variations in genome content between genome are observed: for example 27 CLOGs are
190 specific to the 5 genomes that do not include Jinata and OHK and 28 CLOGs are specific to Jinata and
191 OHK (Figure 2, supplementary Figure 1).

192 Golicz *et al.* (2020) suggested that the size of a pangenome is related to organismal lifestyles, with
193 sympatric organisms having open pangenomes with many accessory genes and allopatric (isolated from
194 other organisms) organisms having more closed and conserved pangenomes. The
195 *Thermosynechococcus* genus could thus be considered as allopatric as they have a comparatively large
196 core and few shared and unique genes. Chen *et al.* (2020) also noted that differences in horizontal gene
197 transfer (HGT) are related to genome size with smaller genomes showing less HGT and bigger genomes
198 having a larger probability that HGT occurred. They also found that hot spring cyanobacteria specifically
199 have smaller genome sizes and less HGT into the genome. Excluding gene loss, we suggest that the
200 conserved core might be indicative of a more ancient gene repertoire in hot spring cyanobacteria, with

201 other cyanobacteria gaining more functionality through HGT over evolutionary timescales. Furthermore
202 thermotolerance is phylogenetically scattered across the cyanobacterial tree of life, mostly in organisms
203 comprising smaller genomes. Since all *Thermosynechococcus* in this study are found in hot springs –
204 which typically have reduced microbial diversity in comparison to other environments such as soils
205 (Ward *et al.*, 1998) – this finding provides support for the still tentative hypothesis that hot spring
206 environments may provide more limited opportunity for lateral gene transfer, which in turn could lead to
207 less opportunity for lateral gene transfer and smaller genomes.

208 In contrast to the genus level where genome size varies from 2.25 MB to 2.64 MB, at the family level it
209 varies to 8.36 MB (*Acaryocholoris marina* MBIC11017, Table 2). Running proteinortho analyses with the
210 seven genus level *Thermosynechococcus* and the six family level sequences, the overall core is reduced

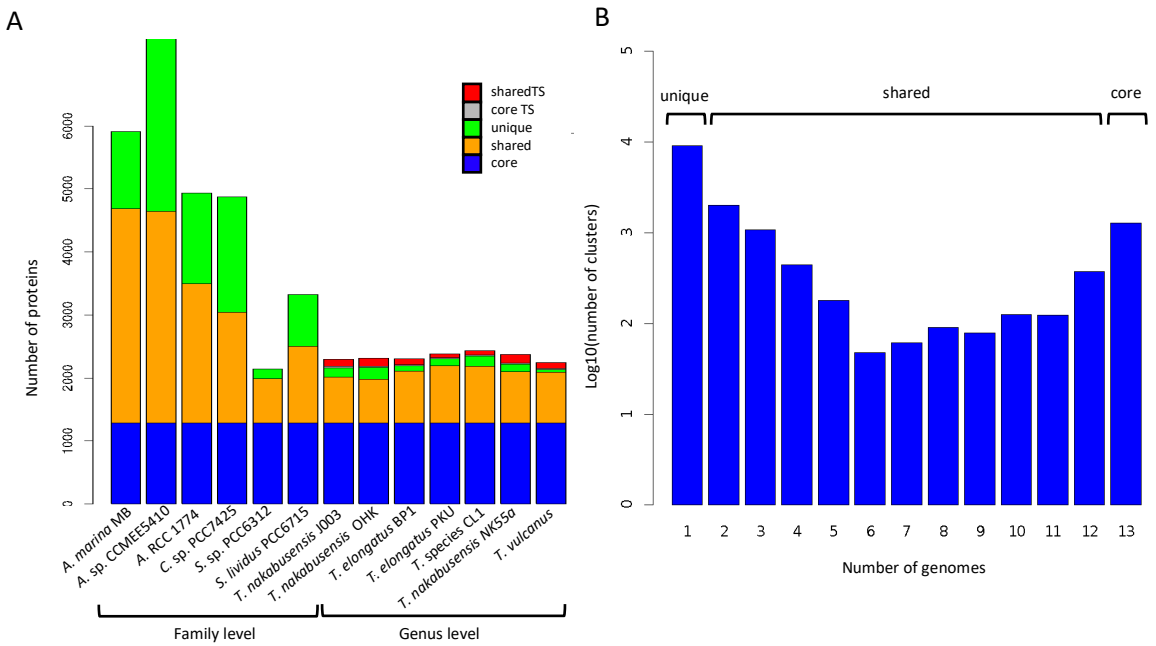

Figure 2: (A) Thermosynechococcaceae family level comparison of core- and pangenomes, (B) number of CLOGs observed at the family level in relationship to the number of genomes in the analysis. coreTS indicates CLOGs found in all genus level genomes, sharedTS are CLOGs found in at least 2 but not all genus level genomes and no other genomes. Full species names as mentioned above.

211 by just over 30%, from 1878 to 1283 CLOGs. This number of ~1900 core CLOGs within the genus is
212 different from a recent analysis (Cheng *et al.*, 2020), with the higher number observed here being due
213 to the inclusion of the newly available and revised *T. elongatus* PKUAC-SCTE542 genome. Shared
214 genes, not core genes, comprise a larger percentage of the genomes for smaller genomes, while unique
215 genes are abundant in larger genomes (Figure 2a).

216 Including the added diversity at the family level, the number of CLOGs shared between the 5
217 *Thermosynechococcus* genus members that do not include Jinata and OHK is reduced from 27 to 3,
218 showing that these CLOGs are found in other closely related members at the family level. However, the
219 number of CLOGs found only in Jinata and OHK changes from 28 to 25 CLOGs, highlighting that these
220 are unique. Also observed when comparing the *Thermosynechococcus* genus across the
221 *Thermosynochococcaceae* family, the specific CLOGs that appear in only, and all of the 7 genus level
222 genomes is made up of 18 CLOGs, showing a low number of conserved *Thermosynechococcus* genus
223 only CLOGs. From the CLOG perspective, the seven members share much in common with their closest
224 relatives, but they appear to lack major distinguishing qualities.

225 **Genomic differentiation of the *Thermosynechococcus* genus from other cyanobacteria**

226 At the class level the number of *Thermosynechococcus* genus level core CLOGs increases when compared
227 to the family level due to the exclusion of some family level species at this level (see Methods and Table 3).
228 The overall core decreases by almost 60% when class level representatives are added (1878 core
229 CLOGs at the genus level and 742 core CLOGs at the class level). This is expected as Beck *et al.* (2018)
230 suggested that the clustering of CLOGs depends on the variability in genome size as well as
231 phylogenetic distance between the analyzed genomes as well as the total amount of genomes analyzed .
232 Moreover this is also in line with the larger analyses of microbial genomes which have shown that the
233 continued addition of taxonomic diversity in an analysis in an analysis leads to increasingly smaller cores
234 (Charlebois and Doolittle, 2004; Lapierre and Gogarten, 2009).

235 Overall these results conform with Beck *et al.* (2012), with variations attributed to the different methods
236 and parameters used for each analysis. Additionally, we confirm that the class level core is stable when
237 adding a higher diversity of organisms, similar to Beck *et al.* (2018), as the number of core family CLOGs
238 only slightly decreases when adding more species at the class level.

239 At the class level, we looked at specific cases among the shared, but not unique CLOGs, to investigate
240 gene content which may differentiate the *Thermosynechococcus* genus from other cyanobacteria. In the
241 case of CLOGs which appear in all *Thermosynechococcus* but not in any other cyanobacteria in our
242 analysis, 79 CLOGs were found, 47 of which are hypothetical proteins of unknown functions and could
243 be of interest to further biochemical studies to understand the adaptations of the genus. CLOGs unique
244 to the genus appear to be of known functionalities account for a variety of processes with genes involved
245 in transcriptional regulation, transporters and membrane proteins (e.g. acetyltransferases, glycosidases

246 and ATPases; more information in supplementary Table 2). Since all the *Thermosynechococcus*
247 analyzed are from hot springs, these genes potentially provide some basis for that lifestyle.

248 **Adaptation of Thermosynechococcus to their respective environments**

249 We were especially interested in those CLOGs that are shared between the strains from environments
250 with elevated iron concentrations (Jinata and OHK) but which are not present in any other cyanobacteria.
251 Previous studies have shown that some cyanobacteria express higher levels of genes involved in iron
252 ion homeostasis which are expressed in iron limiting conditions (Cheng and He, 2014), and here we
253 looked at the presence or absence of iron related gene products in *Thermosynechococcus* compared
254 to other cyanobacteria using FeGenie and BLAST comparisons. 24 CLOGs are specific to Jinata and
255 OHK at the class level, some of which show high partial identity but low coverage matches with genes
256 from other *Thermosynechococcus* (supplementary Table 4). It is notable that these 24 CLOGs do not
257 appear in the third strain of the same species (NK55a). With our current understandings after
258 considering results from BLAST and FeGenie, none of the 24 CLOGs comprises genes that could
259 explain the organisms adaptation to elevated iron concentrations. 3 unique CLOGs are found in the 5
260 *Thermosynechococcus* that do not include Jinata and OHK, and we confirmed that none of those
261 CLOGs are related to iron regulation as best as can be assessed from the sequence, similar to the
262 Jinata and OHK CLOGs (supplementary Table 4). There is no sequence resolvable genomic signature
263 specific to Jinata and OHK related to iron tolerance.

264 These genomes lack genes coding ferrous iron transport and uptake proteins EfeB, EfeO and EfeU, the
265 metal transport gene ZupT, cellular iron storage proteins Bfr and the iron regulator under iron limiting
266 conditions PfsR (Table 4). Genes encoding these proteins are found in other species of cyanobacteria,
267 but not in the *Thermosynechococcus* genus members with currently available genomes. Within the
268 genus there are no differences with regard to genes encoding proteins involved in iron regulation (PfsR),
269 and in all cases the same genes encoding proteins related to ferrous iron uptake (FeoA, FeoB, YfeA
270 and YfeB), ferric iron uptake or transport (ExbD, FutA, FutB and FutC), siderophore iron acquisition
271 (FpvD), metal ion binding (Ho1 and Ho2) and iron starvation acclimation (IsiA) are present (Table 4).

272

273

274

Protein	Function	BP-1	PKUAC	CL1	NK55a	vulcanus	Jinata	OHK	Present in Themosyn	Reference
FeoA	ferrous iron uptake	X	X	X	X	X	X	X	All	Lau *et al.*, 2016
FeoB	ferrous iron uptake	X	X	X	X	X	X	X	All	Lau *et al.*, 2016
YfeA	ferrous iron uptake	X	X	X	X	X	X	X	All	Toulza *et al.*, 2012
YfeB	ferrous iron uptake	X	X	X	X	X	X	X	All	Toulza *et al.*, 2012
FpvD	siderophore iron acquisition	X	X	X	X	X	X	X	All	Brillet *et al.*, 2012
ExbD	ferric iron uptake	X	X	X	X	X	X	X	All	Jiang *et al.*, 2015
PfsR	iron regulator under iron limiting conditions	Not found when Blasting Synechocystis sequence							Synechocystis and others	Cheng and He, 2014
FutA	ABC-type ferric iron transport	X	X	X	X	X	X	X	All	Katoh *et al.*, 2001
FutB	ABC-type ferric iron transport	X	X	X	X	X	X	X	All	Katoh et al., 2001
FutC	ABC-type ferric iron transport	X	X	X	X	X	X	X	All	Katoh *et al.*, 2001
EfeB	ferrous iron transport	Not found when Blasting Synechococcus sequence							Synechococcus and others	Lau *et al.*, 2016
EfeO	ferrous iron transport	Not found when Blasting Synechococcus sequence							Synechococcus and others	Lau *et al.*, 2016
EfeU	ferrous iron uptake	Not found when Blasting Microcystis sequence							Microcystis aeruginosa	Lau *et al.*, 2016
ZupT	metal transport (including ferrous iron)	Not found when Blasting with Synechococcus sequence as query							Synechococcus, J055 and J083 and others	Lau *et al.*, 2016
Bfr	cellular iron storage	Not found when Blasting with Synechococcus sequence as query							Synechococcus and others	Keren *et al.*, 2004; Cheng and He, 2014
IsiA	iron starvation acclimation	X	X	X	X	X	X	X	Some	Cheng and He, 2014
Ho1	metal ion binding	X	X	X	X	X	X	X	Some	Cheng and He, 2014
Ho2	metal ion binding	X	X	X	X	X	X	X	Some	Cheng and He, 2014

275 Table 4: Genes known to be involved in iron regulation within the class Cyanobacteriia and their presence in the

276 *Thermosynechococcus* genus. Presence or absence of genes was confirmed with BLAST searches.

277

The conserved genomic core of *Thermosynechococcus* in relationship to environmental distribution is unique

In addition to resolving a genetic basis for widespread environmental distribution, our work is also relevant to historical proliferation of cyanobacteria, since some modern-day hot springs and their biogeochemistries can be used as historical process analogues (Brown *et al.*, 2005, 2007; Ward *et al.*, 2019). Considering contemporary environments, the analysis of *Thermosynechococcus* also provide insight into island biogeography of microbes. Ionescu *et al.* (2010) observed that the speciation patterns of microorganisms are shaped by local community structures and environmental influences, and Bahl *et al.* (2011) additionally suggest a positive correlation between geographic and genetic distance. Papke *et al.* (2003) found that isolated environments such as geothermal springs may lead to evolutionary divergence of closely related *Thermosynechococcus* strains due to an island effect. Our analysis suggests that geographically widespread organisms belonging to the genus inhabit hot springs with varying geochemistries without genomically recognizable adaptations specific to their site of origin. Instead, the finding of highly conserved genomes within the genus, and furthermore, that the genetic content of the genus is not markedly different from other cyanobacteria, implies that the genus is inherently flexible and viable in the geochemical regimes studied. The large portion of shared genes within the genus provides a genetic basis for the lack of correlation between geographic and genetic distances within the genus found by Papke *et al.* (2003).

This genomic coherence at the genus level is in contrast to other studies, such as Reno *et al.* (2009; Wu *et al.* (2018) or Barajas *et al.* (2019)as mentioned above. Furthermore the organisms tested by Reno *et al.* (2009) are obligate thermoacidophilic archaea at the species level which are environmentally restricted similar to organisms here. They found that the core and pan genomes of these organisms are shaped by their geographical distribution and relatedness within and across different environments. Apparently, *Thermosynechococcus* is environmentally promiscuous, and have fewer restrictive requirements concerning their distribution.

Outlook

One aim of this study was to identify those proteins that make hot-spring inhabiting *Thermosynechococcus* genus members unique. Although the genus *Thermosynechococcus* is rather genomically conserved, 47 out of the 79 CLOGs unique to these genomes are identified as hypothetical proteins of unknown function and thus biochemical studies on these proteins is warranted.

308 Based on the genome comparisons presented here, a viability test of isolates in environments other
309 than those of their origin is suggested as future work. For example, since *Thermosynechococcus* is
310 lacking genes that are involved in iron regulation under limiting conditions, genus level iron tolerance
311 experiments are proposed to test if strains from low-iron environments can also withstand elevated iron
312 without any adaptations in their genome. In a similar way thermotolerance of these organisms could
313 also be investigated. This could help us understand if *Thermosynechococcus* is indeed less restrictive
314 or to identify mechanisms unresolvable by the CLOG approach that account for the geographical
315 distribution.

316 Traits which are unresolvable with an orthology comparison approach like that employed here include
317 amino acid substitutions or sequence changes that lead to variations in enzyme activity, structure, and
318 regulation. Furthermore our analysis is not sensitive to differences in gene regulation or gene copy
319 number. Access to solvents, solubility, active side peptides and differential folding can influence a
320 protein's functionality and therefor may have implications for adaptability or infer evolutionary history
321 (Lesk and Chothia, 1986; Pandey and Braun, 2020). If the orthology approach employed here is correct,
322 the geochemical parameters that are often thought to be important in microbial selection might simply
323 be not as important as we assume them to be in the environment. Here, orthology between genes was
324 assessed based on sequence comparison (proteinortho, BLAST) and sequence and structure
325 comparison (FeGenie, HMMer). However, there exists a diverse set of methods and approaches to
326 determine orthology between genes (Lechner *et al.*, 2011; Forslund *et al.*, 2018), and as orthology
327 definition techniques change, our results could be interpreted.

328 From an Earth history perspective, the presence of cyanobacteria in high-iron environments today could
329 indicate that ancient cyanobacteria may not have been limited by elevated iron concentrations on early
330 Earth, whereas previous studies have reported on the toxicity of high iron concentrations to
331 cyanobacteria (Swanner, Mloszewska, *et al.*, 2015; Swanner, Wu, *et al.*, 2015). We could not find
332 striking adaptive mechanisms on a genome level that can explain the tolerance of strains found in
333 environments with elevated iron concentrations, however the presence and perseverance of
334 cyanobacteria in these environments have important implications when considering the earliest
335 organisms capable of oxygenic photosynthesis on early Earth. Ionescu et al. proposed that simply by
336 increasing photosynthetic rate and oxygen production, cyanobacteria might protect themselves from
337 ferrous iron by promoting its precipitation at some distance from the cell (Ionescu *et al.*, 2014). In line
338 with this observation, is worthwhile to note that the biomass accumulation in high iron environments like

339 Jinata hot spring is appreciable, with co-occurring visible biomass and super saturated dissolved oxygen
340 concentrations from cyanobacterial activity (Ward *et al.*, 2019).

341 **Data availability for newly described strains**

342 The Whole Genome Shotgun project for OHK43 has been deposited at DDBJ/ENA/GenBank under the
343 accession JACOMP000000000. The version described in this paper is version JACOMP010000000.

344 **Supplemental Material**

345 SI Figure 1 – class and genus level comparison of core- and pan-genomes; distribution of number of
346 CLOGs with number of genomes at class and genus level. Core TS corresponds to the CLOGs found
347 in all *Thermosynechococcus* genus members, and shared TS corresponds to CLOGs that are shared
348 by at least 2 and at most 6 genus level genomes.

349 SI Table 1 – *Thermosynechococcus* core CLOGs at class level and their hits in the BLAST protein
350 database.

351 SI Table 2 – Jinata and OHK specific CLOGs at class level with coverage and identity scores, 5 TS
352 specific CLOGs at class level

353 SI Table 3 – genome information for all family level genomes.

354 **Acknowledgements**

355 LMW was supported by a Simons Foundation Postdoctoral Fellowship in Marine Microbial Ecology.
356 Metagenomic sequencing of OHK was supported by NSF grant #OISE 1639454. SEM acknowledges
357 support from the Astrobiology Center Program of the National Institutes of Natural Sciences (grant no.
358 AB311013). We would also like to thank the group around M. Daroch for providing us with an updated
359 version for the *T. elongatus* PKUAC-SCTE542 genome which increased the strain's coherence
360 compared to previous versions of the genome.

361

362

363

364

365

366 **References**

367 Allen, J.F., Thake, B., and Martin, W.F. (2019) Nitrogenase Inhibition Limited Oxygenation of Earth's Proterozoic
368 Atmosphere. *Trends in Plant Science* **24**: 1022–1031.
369 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool.
370 *Journal of Molecular Biology* **215**: 403–410.
371 Aziz, R.K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R.A., et al. (2008) The RAST Server: Rapid
372 annotations using subsystems technology. *BMC Genomics* **9**: 75.
373 Bahl, J., Lau, M.C.Y., Smith, G.J.D., Vijaykrishna, D., Cary, S.C., Lacap, D.C., et al. (2011) Ancient origins
374 determine global biogeography of hot and cold desert cyanobacteria. *Nature Communications* **2**: 1–6.
375 Barajas, H.R., Romero, M.F., Martínez-Sánchez, S., and Alcaraz, L.D. (2019) Global genomic similarity and core
376 genome sequence diversity of the Streptococcus genus as a toolkit to identify closely related bacterial
377 species in complex environments. *PeerJ* **2019**: e6233.
378 Beck, C., Knoop, H., Axmann, I.M., and Steuer, R. (2012) The diversity of cyanobacterial metabolism: genome
379 analysis of multiple phototrophic microorganisms. *BMC genomics* **13**: 56.
380 Beck, C., Knoop, H., and Steuer, R. (2018) Modules of co-occurrence in the cyanobacterial pan-genome reveal
381 functional associations between groups of ortholog genes. *PLOS Genetics* **14**: e1007239.
382 Brillet, K., Ruffenach, F., Adams, H., Journet, L., Gasser, V., Hoegy, F., et al. (2012) An ABC Transporter with Two
383 Periplasmic Binding Proteins Involved in Iron Acquisition in Pseudomonas aeruginosa. *ACS Chemical*
384 *Biology* **7**: 2036–2045.
385 Callieri, C., Slabakova, V., Dzhembekova, N., Slabakova, N., Peneva, E., Cabello-Yeves, P.J., et al. (2019) The
386 mesopelagic anoxic Black Sea as an unexpected habitat for Synechococcus challenges our understanding
387 of global "deep red fluorescence." *ISME Journal* **13**: 1676–1687.
388 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L.
389 (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* **10**: 1–9.
390 Canfield, D.E. (1998) A new model for Proterozoic ocean chemistry. *Nature* **396**: 450.
391 Charlebois, R.L. and Doolittle, W.F. (2004) Computing prokaryotic gene ubiquity: Rescuing the core from
392 extinction. *Genome Research* **14**: 2469–2477.
393 Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019) GTDB-Tk: a toolkit to classify genomes with
394 the Genome Taxonomy Database. *Bioinformatics* **36**: 1925–1927.
395 Chen, M.Y., Teng, W.K., Zhao, L., Hu, C.X., Zhou, Y.K., Han, B.P., et al. (2020) Comparative
396 genomics reveals insights into cyanobacterial evolution and habitat adaptation. *ISME J* 1–17.
397 Cheng, D. and He, Q. (2014) PfsR Is a Key Regulator of Iron Homeostasis in Synechocystis PCC 6803. *PLoS ONE*
398 **9**: e101743.
399 Cheng, Y.-I., Chou, L., Chiu, Y.-F., Hsueh, H.-T., Kuo, C.-H., and Chu, H.-A. (2020) Comparative Genomic Analysis
400 of a Novel Strain of Taiwan Hot-Spring Cyanobacterium Thermosynechococcus sp. CL-1. *Frontiers in*
401 *Microbiology* **11**: 82.
402 Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity.
403 *BMC Bioinformatics* **5**:113
404 Fennel, K., Follows, M., and Falkowski, P.G. (2005) The co-evolution of the nitrogen, carbon and oxygen cycles
405 in the proterozoic ocean. *American Journal of Science* **305**: 526–545.
406 Fischer, W.W., Hemp, J., and Valentine, J.S. (2016) How did life survive Earth's great oxygenation? *Current*
407 *Opinion in Chemical Biology* **31**: 166–178.
408 Forslund, K., Pereira, C., Capella-Gutierrez, S., da Silva, A.S., Altenhoff, A., Huerta-Cepas, J., et al. (2018) Gearing
409 up to handle the mosaic nature of life in the quest for orthologs. *Bioinformatics* **34**: 323–329.
410 Garber, A.I., Nealson, K.H., Okamoto, A., McAllister, S.M., Chan, C.S., Barco, R.A., and Merino, N. (2019)
411 FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in
412 genomes and metagenome assemblies. *bioRxiv* 777656.
413 Golicz, A.A., Bayer, P.E., Bhalla, P.L., Batley, J., and Edwards, D. (2020) Pangenomics Comes of Age: From
414 Bacteria to Plant and Animal Applications. *Trends in Genetics* **36**: 132–145.
415 Guo, Q., Pang, Z., Wang, Y., and Tian, J. (2017) Fluid geochemistry and geothermometry applications of the
416 Kangding high-temperature geothermal system in eastern Himalayas. *Applied Geochemistry* **81**: 63–75.
417 Hao, J., Knoll, A.H., Hazen, R.M., Daniel, I., Huang, F., and Schieber, J. (2020) Cycling Phosphorus on the Archean
418 Earth: Part II. Phosphorus Limitation on Primary Production in Archean Ecosystems Pattern-forming
419 abiotic reactions during organic diagenesis View project Deep Time Data Infrastructure View project
420 Cycling phosphorus on the Archean Earth: Part II. Phosphorus limitation on primary production in
421 Archean ecosystems ScienceDirect.

Hsueh, H.T., Chu, H., and Chang, C.C. (2007) Identification and Characteristics of a Cyanobacterium Isolated from a Hot Spring with Dissolved Inorganic Carbon. *Environmental Science & Technology* **41**: 1909–1914.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., et al. (2016) A new view of the tree of life. *Nature Microbiology* **1**: 1–6.

Ichikuni, M. and Kikuchi, K. (1972) Retention of boron by travertines. *Chemical Geology* **9**: 13–21.

Ionescu, D., Buchmann, B., Heim, C., HÃ¤usler, S., de Beer, D., and Polerecky, L. (2014) Oxygenic photosynthesis as a protection mechanism for cyanobacteria against iron-encrustation in environments with high Fe2+ concentrations. *Frontiers in Microbiology* **5**: 459.

Jiang, H.-B., Lou, W.-J., Ke, W.-T., Song, W.-Y., Price, N.M., and Qiu, B.-S. (2015) New insights into iron acquisition by cyanobacteria: an essential role for ExbB-ExbD complex in inorganic iron uptake. *The ISME journal* **9**: 297.

Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **2015**: e1165.

Katoh, H., Hagino, N., and Ogawa, T. (2001) Iron-Binding Activity of FutA1 Subunit of an ABC-type Iron Transporter in the Cyanobacterium Synechocystis sp. Strain PCC 6803, Oxford Academic.

Keren, N., Aurora, R., and Pakrasi, H.B. (2004) Critical roles of bacterioferritins in iron storage and proliferation of cyanobacteria. *Plant Physiology* **135**: 1666–1673.

Knoll, A.H. (2006) Cyanobacteria and Earth History Diverse aspects of testate amoebae evolution View project The Co-Evolution of the Geo-and Biospheres: An Integrated Program for Data-Driven Abductive Discovery in Earth Sciences View project.

Lapierre, P. and Gogarten, J.P. (2009) Estimating the size of the bacterial pan-genome. *Trends in Genetics* **25**: 107–110.

Lau, C.K.Y., Krewulak, K.D., and Vogel, H.J. (2016) Bacterial ferrous iron transport: the Feo system. *FEMS microbiology reviews* **40**: 273–298.

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P.F., and Prohaska, S.J. (2011) Proteinortho: detection of (co-) orthologs in large-scale analysis. *BMC bioinformatics* **12**: 124.

Lemoine, F., Entfellner, J.-B.D., Wilkinson, E., Correia, D., Felipe, M.D., De Oliveira, T., and Gascuel, O. (2018) Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**: 452–456.

Lesk, A.M. and Chothia, C.H. (1986) The response of protein structures to amino-acid sequence changes. *Philosophical Transactions of the Royal Society of London Series A, Mathematical and Physical Sciences* **317**: 345–356.

Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**: W242–W245.

Li, D., Luo, R., Liu, C.M., Leung, C.M., Ting, H.F., Sadakane, K., et al. (2016) MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**: 3–11.

Lyons, T.W., Reinhard, C.T., and Planavsky, N.J. (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**: 307.

Maity, J.P., Liu, C.-C., Nath, B., Bundschuh, J., Kar, S., Jean, J.-S., et al. (2011) Biogeochemical characteristics of Kuan-Tzu-Ling, Chung-Lun and Bao-Lai hot springs in southern Taiwan. *Journal of Environmental Science and Health, Part A* **46**: 1207–1217.

Mei-hua, W., Junhao, W., and Tingshan, T. (2015) Study of the Scaling Trend of Thermal Groundwater in Kangding County of Sichuan Province. *World Geothermal Congress 2015* 11.

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 gateway computing environments workshop (GCE)*. Ieee, pp. 1–8.

Nakagawa, T. and Fukui, M. (2002) Phylogenetic characterization of microbial mats and streamers from a Japanese alkaline hot spring with a thermal gradient. *The Journal of General and Applied Microbiology* **48**: 211–222.

Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S., et al. (2002) Complete genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1. *DNA research* **9**: 123–130.

Pandey, A. and Braun, E.L. (2020) Phylogenetic Analyses of Sites in Different Protein Structural Environments Result in Distinct Placements of the Metazoan Root. *Biology* **9**: 64.

Papke, R.T., Ramsing, N.B., Bateson, M.M., and Ward, D.M. (2003) Geographical isolation in hot spring cyanobacteria. *Environmental Microbiology* **5**: 650–659.

480 Parks, D.H., Chuvochina, M., Chaumeil, P.A., Rinke, C., Mussig, A.J., and Hugenholtz, P. (2020) A complete
481 domain-to-species taxonomy for Bacteria and Archaea. *Nature Biotechnology* 1–8.
482 Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., and Hugenholtz, P. (2018)
483 A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.
484 *Nature Biotechnology* **36**: 996.
485 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: Assessing the
486 quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*
487 **25**: 1043–1055.
488 Puente-Sánchez, F., Arce-Rodríguez, A., Oggerin, M., García-Villadangos, M., Moreno-Paz, M., Blanco, Y., et al.
489 (2018) Viable cyanobacteria in the deep continental subsurface. *Proceedings of the National Academy of*
490 *Sciences of the United States of America* **115**: 10702–10707.
491 Raven, J. (2009) Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and
492 oxygen fluxes in aquatic environments. *Aquatic Microbial Ecology* **56**: 177–192.
493 Reno, M.L., Held, N.L., Fields, C.J., Burke, P. v., and Whitaker, R.J. (2009) Biogeography of the Sulfolobus
494 islandicus pan-genome. *Proceedings of the National Academy of Sciences of the United States of America*
495 **106**: 8605–8610.
496 Rodriguez-R, L.M. and Konstantinidis, K.T. (2017) Bypassing Cultivation To Identify Bacterial Species Culture-
497 independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide
498 true insights into microbial species.
499 Sánchez-Baracaldo, P. (2015) Origin of marine planktonic cyanobacteria. *Scientific Reports* **5**:.
500 Schirrmeister, B.E., Gugger, M., and Donoghue, P.C.J. (2015) Cyanobacteria and the Great Oxidation Event:
501 evidence from genes and fossils. *Palaeontology* **58**: 769–785.
502 Shi, T. and Falkowski, P.G. (2008) Genome evolution in cyanobacteria: The stable core and the variable shell.
503 *Proceedings of the National Academy of Sciences of the United States of America* **105**: 2510–2515.
504 Shih, P.M., Hemp, J., Ward, L.M., Matzke, N.J., and Fischer, W.W. (2017) Crown group Oxyphotobacteria
505 postdate the rise of oxygen. *Geobiology* **15**: 19–29.
506 Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., et al. (2013) Improving the coverage of the
507 cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy*
508 *of Sciences of the United States of America* **110**: 1053–1058.
509 Soo, R.M., Hemp, J., and Hugenholtz, P. (2019) Evolution of photosynthesis and aerobic respiration in the
510 cyanobacteria. *Free Radical Biology and Medicine* **140**: 200–205.
511 Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
512 phylogenies. *Bioinformatics* **30**: 1312–1313.
513 Stolyar, S., Liu, Z., Thiel, V., Tomsho, L.P., Pinel, N., Nelson, W.C., et al. (2014) Genome sequence of the
514 thermophilic cyanobacterium Thermosynechococcus sp. strain NK55a. *Genome Announc* **2**: e01060-13.
515 Swanner, E.D., Mloszewska, A.M., Cirpka, O.A., Schoenberg, R., Konhauser, K.O., and Kappler, A. (2015)
516 Modulation of oxygen production in Archaean oceans by episodes of Fe(II) toxicity. *Nature Geoscience* **8**:
517 126–130.
518 Swanner, E.D., Wu, W., Hao, L., Wüstner, M.L., Obst, M., Moran, D.M., et al. (2015) Physiology, Fe(II) oxidation,
519 and Fe mineral formation by a marine planktonic cyanobacterium grown under ferruginous conditions.
520 *Frontiers in Earth Science* **3**: 60.
521 Tang, J., Jiang, D., Luo, Y., Liang, Y., Li, L., Shah, Md.M.R., and Daroch, M. (2018) Potential new genera of
522 cyanobacterial strains isolated from thermal springs of western Sichuan, China. *Algal Research* **31**: 14–20.
523 Tomitani, A., Knoll, A.H., Cavanaugh, C.M., and Ohno, T. (2006) The evolutionary diversification of
524 cyanobacteria: Molecular-phylogenetic and paleontological perspectives. *Proceedings of the National*
525 *Academy of Sciences of the United States of America* **103**: 5442–5447.
526 Toulza, E., Tagliabue, A., Blain, S., and Piganeau, G. (2012) Analysis of the Global Ocean Sampling (GOS) Project
527 for Trends in Iron Uptake by Surface Ocean Microbes. *PLoS ONE* **7**: e30931.
528 Wang, L., Liu, Q., Wu, X., Huang, Y., Wise, M.J., Liu, Z., et al. (2019) Bioinformatics Analysis of Metabolism
529 Pathways of Archaeal Energy Reserves. *Scientific Reports* 1–12.
530 Ward, D.M., Ferris, M.J., Nold, S.C., and Bateson, M.M. (1998) A Natural View of Microbial
531 Biodiversity within Hot Spring Cyanobacterial Mat Communities. *Microbiol Mol Biol Rev* **62**:
532 1353–1370.
533 Ward, L.M., Fischer, W.W., and Mcglynn, S.E. Candidatus Anthektikosiphon siderophilum OHK22, a New
534 Member of the Chloroflexi Family Herpetosiphonaceae from Oku-okuhachikurou Onsen. *Microbes*
535 *Environ* **35**: 2020.
536 Ward, L.M., Idei, A., Nakagawa, M., Ueno, Y., Fischer, W.W., and McGlynn, S.E. (2019) Geochemical and
537 Metagenomic Characterization of Jinata Onsen, a Proterozoic-Analog Hot Spring, Reveals Novel Microbial

Diversity including Iron-Tolerant Phototrophs and Thermophilic Lithotrophs. *Microbes and environments* **34**: 278–292.

Ward, L.M., Idei, A., Terajima, S., Kakegawa, T., Fischer, W.W., and McGlynn, S.E. (2017) Microbial diversity and iron oxidation at Okuoku-hachikurou Onsen, a Japanese hot spring analog of Precambrian iron formations. *Geobiology* **15**: 817–835.

Ward, L.M. and Shih, P.M. (2019) The evolution and productivity of carbon fixation pathways in response to changes in oxygen concentration over geological time. *Free Radical Biology and Medicine* **140**: 188–199.

Ward, L.M., Shih, P.M., and Fischer, W.W. (2018) MetaPOAP: presence or absence of metabolic pathways in metagenome-assembled genomes. *Bioinformatics (Oxford, England)* **34**: 4284–4286.

Whitton, B.A. (1992) Diversity, ecology, and taxonomy of the cyanobacteria. In *Photosynthetic prokaryotes*. Springer, pp. 1–51.

Wu, Y., Zaiden, N., and Cao, B. (2018) The Core- and Pan-Genomic Analyses of the Genus Comamonas: From Environmental Adaptation to Potential Virulence. *Frontiers in Microbiology* **9**: 3096.

Yeh, G.-H., Yang, T.F., Chen, J.-C., Chen, Y.-G., and Song, S.-R. (2005) Fluid geochemistry of mud volcanoes in Taiwan. *Mud Volcanoes, Geodynamics and Seismicity Springer, Netherlands* 227–237.