



HAL
open science

Horizontal Gene Transfer and Recombination in Cyanobacteriota

Devaki Bhaya, Gabriel Birzu, Eduardo P.C. Rocha

► **To cite this version:**

Devaki Bhaya, Gabriel Birzu, Eduardo P.C. Rocha. Horizontal Gene Transfer and Recombination in Cyanobacteriota. Annual Review of Microbiology, In press, <10.1146/annurev-micro-041522-100420>. <pasteur-05262034>

HAL Id: pasteur-05262034

<https://pasteur.hal.science/pasteur-05262034v1>

Submitted on 15 Sep 2025

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 CC BY-NC 4.0 - Attribution - Non-commercial use - International License

1 **HORIZONTAL GENE TRANSFER AND RECOMBINATION IN THE CYANOBACTERIOTA**

2 Authors: Devaki Bhaya^{1,2}, Gabriel Birzu^{2,3}, Eduardo P. C. Rocha⁴

3 ¹ Biosphere Sciences and Engineering, Carnegie Institution for Science, Stanford CA 94305 USA, ²
4 Stanford University, CA 94305 USA, ³ University of Florida, USA, ⁴ Institut Pasteur Université Sur, Paris
5 Cité, CNRS UMR3525, Microbial Evolutionary Genomics, Paris, 75015, France.

6 Email: dbhaya@carnegiescience.edu; gbirzu@ufl.edu; eduardo.rocha@pasteur.fr

7 Corresponding author contact information: dbhaya@carnegiescience.edu

8 ORCID # Devaki Bhaya: 0000-0001-7965-4258

9 ORCID # Gabriel Birzu 0000-0003-3561-024X

10 ORCID # Eduardo P. C. Rocha 0000-0001-7704-822X

11
12 **KEYWORDS** Gene flow, photosynthesis, auxiliary metabolic genes, adaptive evolution, *Microcystis*,
13 Thermophiles

14
15
16 **ABSTRACT**

17
18
19 Cyanobacteria played a pivotal role in shaping Earth's early history and are key players in many
20 important ecosystems. As versatile prokaryotic phototrophs, they are models for the evolution of
21 oxygenic photosynthesis, nitrogen fixation, circadian rhythms and symbiosis. Cyanobacterial genomes
22 and metagenomes exhibit high levels of genomic diversity within and across species, partly caused by
23 gene flow in the form of recombination and horizontal gene transfer. We review current knowledge
24 about these processes in terms of molecular mechanisms, barriers to gene flow, and selection for novel
25 traits, with a focus on auxiliary metabolic genes. We evaluate the effect of gene flow at
26 macroevolutionary scales when key functions emerged, and at micro-evolutionary scales where
27 populations adapt to local conditions. Unraveling the dynamics and mechanisms of gene flow is needed
28 to understand key functional innovations in the phylum, to leverage them in synthetic biology and to
29 predict their future adaptation to climate change.

30
31
32

1. INTRODUCTION

1.1 A brief history of the role of Cyanobacteria in Earth's History

Somewhere between 2.1 and 2.5 billion years ago the Earth's atmosphere began changing irreversibly from a reducing to an oxidizing atmosphere. This change was caused by oxygen released by the water-splitting photosynthetic activity of the ancestors of modern-day Cyanobacteriota (61, 122) and had a profound effect on the natural history of the planet (87, 88). Cyanobacteriota, in addition to being the only known phylum with oxygenic photosynthetic activity, are one of the few bacterial phyla which have left fossil records and biosignatures of their activity, going back ~3 billion years. This data has been used to develop scenarios of the emergence of the major cyanobacterial groups and their potential impact on biogeochemical cycles, although inconsistencies between geological records and genomic data remain (118).

Another pivotal event in Earth's history was the integration of a cyanobacterial endosymbiont into a eukaryotic host about a billion years ago, giving rise to plastids in the green lineage and diverse algal groups which resulted in the emergence of modern day plants (111, 127). Today, the plant kingdom accounts for more than eighty percent of the planet's biomass (7). As a consequence, Cyanobacteriota have been referred to as "Earth's microbial heroes" (75) and "among the most important organisms ever to have evolved on our planet" (118). Arguably, Cyanobacteriota had an outsized impact on the natural history of the Earth. Yet, our understanding of how genomic diversity is generated and maintained is fragmentary. We present current knowledge about horizontal gene transfer (HGT) and recombination in terms of molecular mechanisms, barriers to gene flow, selection for novel traits, with a focus on auxiliary metabolic genes. Challenges in evaluating the role of gene flow in evolutionary history and diversification in different ecosystems are presented.

1.2 Key features of Cyanobacteriota

The phylum Cyanobacteriota comprises a large, monophyletic, morphologically diverse group of Gram negative bacteria (124, 133, 141). Their photoautotrophic lifestyle and ability to withstand environmental extremes has enabled them to thrive in almost every terrestrial and marine environment where light is available for photosynthesis. For instance, marine cyanobacteria may be responsible for as much as 25% of the net primary productivity in the oceans (41). In addition, several species (both unicellular and filamentous genera) fix atmospheric nitrogen in diverse environments (162). Thus, Cyanobacteriota play an important role in both carbon and nitrogen biogeochemical cycles. They are symbiotic partners with plant or animal hosts, providing both fixed carbon and nitrogen (28, 42, 96). It has been estimated that between 40 and 80% of bacterial cells exist within biofilms, where the likelihood of HGT and recombination events may be greater due to the high density of cells within an extracellular matrix (40). It is important to emphasize that in many environments, Cyanobacteriota evolve and interact as keystone members of microbial communities or biofilms with the ability to donate or take up exogenous DNA. These biofilms are present in diverse environments such as microbial mats in thermal springs, freshwater and hypersaline lakes, polar habitats, desert crusts (46, 64, 126, 134).

75 1.3 Assessing the role of gene flow in Cyanobacteriota in a post-genomic era

76 Gene flow plays a crucial, but sometimes underappreciated role, in adaptive evolution of many bacterial
77 phyla (5, 15). Until recently, only a few genomes of model organisms from phylogenetically widely
78 separated cyanobacterial genera were available (34, 124). To set the stage, we used standard tools and
79 a subset of available complete genomes of Cyanobacteriota to highlight features such as genome size
80 and molecular/structural features that are relevant to gene flow, such as Type IV pili (T4P), plasmids,
81 prophages, and conjugative and mobility systems (**Figure 1**).

82
83 The phylum Cyanobacteriota exhibits a wide range of genome sizes from the reduced genome of the
84 symbiont Candidatus *Atelocyanobacterium thalassa* (1.4 Mb) to the 12.4Mb genome of the filamentous,
85 heterocyst forming *Calothrix* sp. PCC 7716. This represents a difference in gene repertoires of almost an
86 order of magnitude. In Bacteriota, genome size variation is usually attributed to a balance between gene
87 gain by HGT or gene duplication, subfunctionalization and gene loss. HGT is considered to be far more
88 important than gene duplication in shaping bacterial gene repertoires (143, 144). In Cyanobacteriota,
89 certain features are almost universal such as the presence of T4P (missing in the highly reduced
90 genomes of symbionts). In contrast, other characteristics, such as the presence or absence of plasmids
91 are patchy, which may be a function of incompletely curated datasets. Figure 1 also highlights that
92 certain clades and genera are overrepresented in current datasets, e.g. the marine *Synechococcus* and
93 *Prochlorococcus* clades or multiple re-sequenced variants of a model organism such as *Synechocystis*
94 PCC6803. In contrast, close variants of species that are co-evolving in the natural environment are less-
95 well represented. However, we anticipate a rapid increase in the number of complete cyanobacterial
96 genomes, single cell amplified genomes (SAGs), metagenomes, and metagenome-assembled genomes
97 (MAGs) of microbial communities from diverse habitats. In conjunction with improved sequencing
98 platforms, development of refined annotation and computational pipelines, visualization tools (e.g.
99 *anvi'o*) these datasets will be crucial to identify gene flow in natural populations (3, 37, 147).

100

101 2. MOBILE GENETIC ELEMENTS AND MECHANISMS DRIVING GENE FLOW

102

103 The set of mobile genetic elements (MGEs) in an organism that supports gene flow within the genome
104 or across species is called a 'mobilome'. Recognizable MGEs such as plasmids, transposable elements,
105 and bacteriophages, are present in almost all (cyano)bacteria. The exceptions include the reduced
106 genomes of endosymbiotic species and the minimal genome of *A. thalassa*. Gene flow in bacteria occurs
107 by a multitude of mechanisms of which the three most common are natural transformation, conjugation
108 and transduction (32). Exogenous DNA can be incorporated into the genome by two pathways. If the
109 DNA is homologous to the recipient genome, then genetic exchange occurs by the process of
110 homologous recombination and results in gene conversion and allelic exchange, which is usually
111 restricted to gene flow between closely related genomes. If the DNA entering is non-homologous to the
112 host genome, it can remain in the cell as an extrachromosomal element (plasmid), or be integrated into
113 the genome using a site-specific recombinase (usually integrases). Both processes may result in the
114 addition of novel genes that could alter host fitness or accelerate its ability to occupy new ecological
115 niches.

116

117 2.1 Natural transformation

118 Natural transformation is the best described process of gene flow in Cyanobacteriota. It requires
119 recipient bacteria to be in a 'competent state' i.e. express the ability to take up DNA. Natural
120 transformation in Cyanobacteriota was first reported in 1970 and has been experimentally observed in
121 several unicellular and filamentous genera (for reviews see (119, 154)). Transformation is a multi-step
122 process (32) that generally requires a Type IV pilus (T4P). The first step involves attachment of free
123 extracellular DNA to a retractile T4P that brings it closer to the cell surface. Most cyanobacterial
124 genomes encode components of a T4P (27) (**Figure 1**), with the exception of genera that have
125 streamlined or reduced genomes, such as endosymbionts or marine unicellular *Prochlorococcus* sp.
126 (119). T4P is composed of one major structural pilin and minor pilins, some of which might be involved
127 in binding DNA (See Table 1 in (119) for details). In Cyanobacteriota, T4P are multifunctional entities
128 required for natural transformation in addition to surface-dependent motility, phototaxis, cell adhesion,
129 biofilm formation and secretion (23, 95). How these multiple roles of T4P interact and affect
130 transformation efficiency has not been explored extensively.

131
132 The second step of transformation typically requires the DNA transport across the cell envelope and
133 involves a periplasmic DNA binding protein (ComEA) with nuclease activity and a membrane channel
134 (ComEC). Subsequently, a membrane-associated protein, ComF, presents single stranded DNA to the
135 recombination machinery (32). Competent cyanobacteria encode all the three key proteins (119). In the
136 final step, DNA is targeted by the homologous recombination machinery for incorporation into the
137 genome. Most cyanobacteria encode the minimal set of functions required for homologous
138 recombination and alternative paths have been identified (8) (99) but so far, detailed biochemical
139 characterizations are limited. In the fast-growing non-transformable strain *Synechococcus elongatus*
140 UTEX 2973, introduction of an exogenous *pilN* gene restored natural competence (155). In *Synechocystis*
141 sp., transformation can contribute to the transfer of small cryptic plasmids (100). In *S. elongatus* PCC
142 7942, the active prokaryotic Argonaute nuclease, SeAgo, was found to reduce natural transformation
143 and prevent maintenance of RSF1010 replicons, suggesting that it may act as a defense system to
144 prevent HGT (136). Natural competence in filamentous cyanobacteria is less well documented and it was
145 suggested to be less common than in unicellular cyanobacteria but no systematic study of this group has
146 yet been carried out (98, 99, 119). In summary, the machinery of natural transformation in
147 Cyanobacteriota has many characteristics in common with other competent bacteria but features that
148 may be specific to Cyanobacteriota or how transformation impacts genetic diversification and the
149 acquisition of novel genes is not well-documented.

150 151 2.2 Conjugation

152 Conjugation is driven by two types of MGEs: conjugative plasmids and integrative conjugative elements
153 (ICEs), also called conjugative transposons. Conjugation requires an origin of transfer (*oriT*), a relaxase
154 and a Type IV secretion system (T4SS) to transfer ssDNA between cells that are in direct contact. The
155 process starts when the relaxase interacts with the *oriT*, making a single strand nick and attaching to the
156 nicked strand. Next, the nucleo-protein complex is transferred by the T4SS to the recipient cell. At the
157 end of the transfer, DNA ligation and replication reconstitute a copy of the element in both the donor
158 and recipient cells. In Cyanobacteriota, early studies reported the presence of plasmids that were of

159 various lengths (some as large as ~400 kb), and some carried several genes such as toxin-antitoxin
160 systems or CRISPR-Cas modules (59, 114). Conjugation systems of Cyanobacteriota were identified
161 almost forty years ago (78) and several families of relaxases (47) and key components of the T4SS (53)
162 were identified in the genomes of Cyanobacteriota. Cyanobacteriota are unusual in that they have many
163 more predicted conjugative plasmids than ICEs (54). Despite this, to the best of our knowledge, genetic
164 engineering of Cyanobacteriota is generally carried out using conjugation systems derived from
165 Proteobacteria (76)(36). Phylogenetic studies indicate that cyanobacterial conjugation systems are
166 evolutionarily very ancient, being amongst the first to diverge from those of other phyla (52). This
167 implies that our knowledge of cyanobacterial systems, e.g. whether they have narrow or wide host
168 range, whether they can mate in liquid or only on solid surfaces, and whether they require receptors on
169 recipient cells, requires further investigation.

170
171 The requirement for direct cell-cell contact for conjugation and the fact that conjugative pili formed
172 between cells can be easily disrupted by agitation suggest this mechanism may be more efficient for
173 DNA transfer within stable structured communities, such as biofilms. Interestingly, almost a third of the
174 Cyanobacteriota genomes carry mobile genetic elements that encode a relaxase and are thus potentially
175 mobilizable by conjugation (**Figure 1**). A recent survey suggests that Cyanobacteriota may be a
176 significant environmental reservoir for antimicrobial resistance which could have a potentially negative
177 effect on the environment. It remains to be determined if native cyanobacterial conjugation systems can
178 mobilize these genes into other phyla. In Proteobacteria and Firmicutes, conjugation has been shown to
179 play a key role in the transfer of traits such as those that confer antibiotic resistance (140, 157).

180 181 *2.3 Transduction and the role of cyanophages*

182 Bacteriophages are viruses that infect bacteria and are ubiquitous in nature. Phages that infect
183 cyanobacteria are sometimes referred to as 'cyanophages'. Cyanophages are frequently from the Class
184 Caudoviricetes, and have one of the three most common morphologies: cyanomyovirus (contractile
185 tails), cyanosiphovirus (long non-contractile tails), and cyanopodovirus (short non-contractile tails) (for a
186 recent review see (50) Cyanobacteria can be infected by virulent or temperate phages. Both types
187 undergo a lytic cycle causing cell lysis and release of phage progeny. However, temperate phages can
188 integrate their genome ('prophage') into the bacterial genome ('lysogen'). Most prophage functions are
189 silenced until a physiological or environmental signal induces the lytic cycle resulting in phage replication
190 and cell host lysis. Prophages have been identified in diverse Cyanobacteriota, but there has not been a
191 recent systematic analysis of their frequency in available cyanobacterial genomes (90). Our exploratory
192 analyses of cyanobacterial genomes reveals that ~25% of genomes carry prophages (**Figure 1**), a fraction
193 that might increase as more environmental isolates or metagenomes are characterized and novel types
194 of phages are discovered. Hence, the impact of lysogeny in natural environments can be significant, but
195 remains unexplored. In other phyla, prophages have been shown to endow bacteria with traits that have
196 adaptive value such as virulence factors, defense systems and metabolic functions (22).

197
198 Phages can also transfer (or 'transduce') fragments of host bacterial DNA into a recipient cell, by
199 homologous recombination (142). This may be the result of imprecise excision of the prophage from the
200 host ('specialized transduction'), or from prophage replication that occurs before excision from the

201 chromosome ('lateral transduction'), or from packaging random pieces of the bacterial genome
202 ('generalized transduction'). Systematic studies of transduction are mostly limited to marine
203 cyanobacteria where a model system for infection has been established (161). For example, in
204 *Prochlorococcus*, it was shown that myoviruses mispackaged host DNA at a low frequency, while growth
205 light intensity had a large effect on mispackaging frequency in podo- and siphoviruses (79). Recent
206 research has identified MGEs, called 'phage satellites' that can be packaged at high frequency into viral
207 particles to be transferred by transduction (at the cost of exclusion of the phage genome).
208 *Prochlorococcus* carries a large and diverse subgroup of phage satellites, called 'tycheposons' (55). They
209 do not seem to include cargo genes and their impact on the host remains to be determined. Of note,
210 other tycheposons do not seem to be phage satellites and could be novel types of MGEs. Transduction
211 allows the transfer of bacterial DNA in particles that can disperse across long distances, especially in
212 aquatic environments. This could be key to diversification of planktonic cyanobacteriota which are
213 usually at low concentrations in the ocean (11).

214

215 2.4 Mobile genetic elements

216 In addition to MGEs, such as phages and conjugative plasmids that are able to transfer autonomously
217 between cells, many other MGEs, such as transposons and insertion sequences (IS), drive intra-genomic
218 mobility. Their capacity to translocate or transpose genetic material within genomes can alter gene flow
219 by moving genes between MGEs or between these and the bacterial chromosome (for a recent review
220 see (153)). Transposition also results in gene duplication or gene inactivation that affects host fitness. IS
221 insertions can create mutations that have negative or positive effects by contributing to host adaptation
222 or by regulating neighboring genes by acting as a promoter (137). A comprehensive study of MGEs
223 across Cyanobacteriota would be instructive since our analysis suggests that they are frequent in
224 lineages with large genomes (**Figure 1**). Next, we highlight a few examples from the best studied
225 systems.

226

227 Early studies reported that the bloom forming colonial *Microcystis aeruginosa* NIES843 had a high
228 abundance of IS elements (~ 11% of the genome) (44, 82). Similarly, the colonial cyanobacterium
229 *Woronichinia naegeliana* WA131 from a freshwater lake that has periodic toxic cyanobacterial blooms,
230 has a genome size of 7.9 Mb, that is larger than other bloom-forming unicellular cyanobacteria, and
231 which has been attributed to its high number of ISs and pseudogenes (31). Because *Microcystis* and
232 other toxic bloom forming cyanobacteria are often health hazards in water bodies, they have been
233 extensively analyzed for the presence and acquisition of toxin producing gene clusters via HGT and the
234 role of recombination. Cyanobacterial blooms are increasing worldwide, so it is an area of active
235 research that we cannot review adequately here. We refer the interested reader to these excellent
236 reviews and references therein (29, 30, 159).

237

238 Another microbial community that has been intensively studied are the stratified biofilms in hot springs
239 where phototrophs such as cyanobacteria, chloroflexi and other heterotrophs adapt to extreme
240 conditions such as high temperatures and low nutrients (74, 134) (Shelton et al in review ISME). For
241 example, genomes of two related thermophilic unicellular genera, *Synechococcus* OS-A and
242 *Synechococcus* sp. OS-B' were shown to have a high level of genomic rearrangements and
243 recombination (10, 116). Consistent with this finding, they had a high abundance of ISs (between 71 and

244 82 full-length ISs). A comparison of the genomic locations of IS elements in these two genera with a
245 metagenomic dataset derived from the same community indicated that there was ongoing exchange of
246 the IS elements between the *Synechococcus* genera (97). A comparative analysis of the *Synechococcus*
247 OS-A and *Synechococcus* sp. OS-B' genomes also provided specific examples of HGT events. The
248 *Synechococcus* sp. OS-B' genome (but not *Synechococcus* OS-A) had acquired a ~8Kb DNA module, with
249 multiple genes that encoded a phosphonate lyase and a transporter (Phn operon). It was experimentally
250 demonstrated that acquisition of this module allowed them to grow on phosphonate as a sole P source
251 (2). The acquisition of the Phn operon by HGT has also been observed in the marine N-fixing
252 *Trichodesmium* sp. (35). Even though, in these cases, the donor bacteria and transfer mechanism were
253 not established, this study emphasizes that expanding our knowledge to a variety of diverse
254 environments may yield new insights about the pervasive role of gene flow. The *Synechococcus*
255 genomes also showcase another interesting variation on the dynamic acquisition of traits.
256 *Synechococcus* OS-A has acquired a complete urease operon, which is flanked by transposons,
257 suggestive of a HGT event. The urease operon would allow cells to import urea and use it as source of N
258 and C. The genome also has a second urease operon but appears to be rendered non-functional by
259 transposons and frameshift mutations. This second urease was the only urease identified in the
260 *Synechococcus* sp. OS-B' genome (10) (**Figure 2**). These examples highlight that transposons can be
261 powerful mediators of adaptive evolution (93), allowing microbes to acquire new traits, increase fitness
262 and occupy new nutrient niches (145) (H.Subramaniam et. al., unpublished). We anticipate that in the
263 future many other such examples will be uncovered leading to a broader understanding of the 'rules'
264 that govern the frequency, mechanisms and stability of HGT events.

265
266 Recently, Cyanobacteriota have been shown to have a rich reservoir of diverse CRISPR-associated
267 transposons which are Tn7-like elements associated with Type I CRISPR-Cas systems. It has been
268 suggested that these elements may play a role in targeting transposition, although the molecular details
269 are yet to be worked out in detail in Cyanobacteriota (62, 63, 107). In summary, MGEs play an
270 important role in cyanobacterial communities that grow in extreme environments or in rapidly growing
271 toxic algal blooms, but further studies are required to build a predictive understanding of their role in
272 the evolution, particularly as microbial communities adapt to climate change.

273

274 3. FACTORS SHAPING GENE FLOW

275 Several features of the host and the environment can influence gene flow and its outcomes. Some
276 examples include (i) the requirement to overcome host cellular barriers (the cell envelope and a variety
277 of defense systems), or (ii) the need for ecological or physiological opportunities for productive transfer
278 of genetic material, or (iii) the role of natural selection in stabilizing acquired genes or traits conferred by
279 auxiliary metabolic genes. However, it is worth noting that even after a successful HGT event, DNA can
280 be lost from the population for a variety of reasons, e.g., (i) if it is not advantageous to the host, (ii) if it
281 cannot be stably maintained, (iii) is not or poorly expressed or (iv) is too costly to maintain. The dynamic
282 features of HGT shape the rate of acquisition and maintenance of novel traits but so far they have been
283 difficult to capture or quantify. Deep sampling of microbial populations, in conjunction with time-
284 resolved or spatial metagenomics can help distinguish the contribution of drift, natural selection, and
285 genetic instabilities in the success of HGT events. Some of these features are likely to be common across
286 most bacteria. Interestingly, in Cyanobacteriota, the circadian clock also appears to play a role in gene

287 flow. This suggests that in phototrophs that must withstand harsh conditions such as excess absorption
288 of excitation energy and UV radiation, additional layers of control and regulation may be important and
289 unique.

290

291 *3.1 Barriers to gene flow*

292 The environment of the host can have a significant impact on the success or failure of HGT. For example,
293 transformation requires stable, free DNA to be present in the environment as well as close proximity to
294 the host cell which might be more likely in a dense biofilm than in a planktonic setting. In addition,
295 transformable species require specific DNA uptake sequences to favor intra-specific exchanges, whereas
296 others regulate competence for natural transformation by quorum-sensing (4). Conjugation can take
297 place between very widely separated bacterial phyla but requires direct cell-to-cell contact.
298 Transduction is often restricted to closely related bacterial hosts encoding specific cell receptors. On the
299 other hand, it can occur across large geographic distances. Once DNA enters a recipient cell, it must face
300 the arsenal of host immune defenses. In the last few years, over a hundred novel defense systems have
301 been uncovered and the list is growing and we refer the reader to recent reviews (9, 48, 138). Defense
302 systems are often categorized as innate or adaptive systems. Innate systems are the most common,
303 especially restriction-modification systems that provide a generic barrier to infections by MGEs. In
304 contrast, CRISPR-Cas systems adapt their response to infective MGEs, through a 'memory' of past
305 infections.

306

307 A recent assessment of 182 complete genomes of cyanobacteria identified 553 defense systems of 32
308 different types (138). Hence, the average genome of a cyanobacterium could encode three distinct
309 defense systems but these estimates are likely to rise in the future with the discovery of novel systems.
310 The most frequently identified defense systems are the restriction-modification (146 systems) and
311 CRISPR-Cas (103) systems, followed by less well-known systems like Septu (39) and PARIS (34). The
312 impact of defense systems in cyanobacteria-phage interactions and the effect on HGT has been analyzed
313 in a few cases and we anticipate more examples in the future. For example, matching CRISPR spacer
314 arrays to genomes of extant phage populations has indicated active infection of phages of thermophilic
315 cyanobacterial genera (25, 51, 58). Additionally, CRISPR-Cas defense modules have been identified in
316 *Microcystis* sp. and opportunities to observe how cyanophages impact gene flow in diverse
317 environments, where cyanobacteria are dominant is likely to reveal new features of host-phage
318 interactions (see next section on Auxiliary metabolic genes).

319

320 *3.2 Natural selection for auxiliary metabolic genes in MGEs*

321 Natural selection plays an important role in the fixation of genes acquired by HGT and has important
322 implications in ecological studies. This has been most extensively documented in marine Cyanobacterial
323 populations, particularly in *Prochlorococcus* and *Synechococcus*. Auxiliary metabolic genes (AMGs)
324 carried by phages infecting *Prochlorococcus* were first identified in 2004 (81), and include host-derived
325 genes that play a critical role on photosynthesis. These include photosystem II core reaction center
326 proteins D1 (*psbA*) and D2 (*psbD*), electron transport proteins, plastocyanin (*petE*) and ferredoxin (*petF*)
327 and a high-light-inducible protein (*hli*). These findings led to the hypothesis that AMGs permit the host

328 to survive longer, post-infection, thereby favoring phage reproduction. Several follow-up studies were
329 conducted and we refer the reader to excellent reviews (11, 24, 55, 65, 110).

330
331 The repertoire of potentially important AMGs in cyanophages has been expanded by the improvement
332 of metagenomic and viromic analyses (148). Genome sequences of phage isolates representing major
333 cyanophage clades, infecting different cyanobacterial genera, and isolated from different environments
334 (e.g. freshwater, lakes with toxic blooms, marine) are now available (86, 110, 139, 165). One bottleneck
335 in the field is the lack of experimental systems to verify the roles of AMGs. It is also possible that AMGs
336 benefit the phage as well as the host. A recent survey of 153 cyanophage genomes used homology-
337 derived functional assignments to identify AMGs involved in photosynthesis, carbon metabolism,
338 methylation, phosphorus metabolism and gene expression regulation (148). Of these, the forty-five
339 commonly observed AMGs likely to play a role in host metabolism have homologs in cyanobacterial host
340 genomes, suggesting that they were recruited from host genomes by a process of HGT.

341
342 Most AMGs identified in cyanophages encode proteins with a role in photosynthesis or energy
343 metabolism, with the potential to augment or control the infected host's photosynthetic activity.
344 Several cyanophages carry tRNAs that might facilitate protein synthesis and there are reports of
345 cyanophages carrying up to 15 tRNAs (86). Phages belonging to the Myoviridae, which usually have
346 larger genomes, tend to carry multiple AMGs (115). Several cyanophages infecting filamentous
347 *Pseudanabaena* sp., a species associated with cyanobacterial blooms in freshwater lakes were shown to
348 have several known AMGs, including a complete queuosine modification pathway (queuosine
349 modification of tRNAs is believed to modulate translation) (166). One of these phages shows significant
350 similarity to an alpha-proteobacterial phage, suggesting an extremely unusual phage-mediated
351 interphylum horizontal gene transfer between cyanobacteria and alphaproteobacteria.

352
353 An analysis of phages across bacterial hosts revealed that lifestyles (virulent versus temperate) are likely
354 to be the most important driver of AMG distribution, but habitat and host identity also play a role. For
355 example, it has been suggested that there are fewer AMGs in cyanophages from eutrophic waters (68).
356 Virulent phages exhibit greater AMG diversity and carry genes for photosynthesis, chaperone
357 biosynthesis, signaling proteins, and lipid metabolism, which could boost host burst size. On the other
358 hand, temperate phages tend to carry AMGs that favor host survival. A similar analysis carried out
359 specifically for Cyanobacteriota phages from various environments could reveal other patterns of AMG
360 acquisition.

361 362 *3.3 The circadian clock in Cyanobacteriota controls gene flow*

363 The Cyanobacteriota circadian molecular clock controls many endogenous rhythms including gene
364 expression and cell cycle (38). This circadian rhythm impacts gene flow via natural transformation and
365 phage transduction. Although some cyanobacterial genera are constitutively competent for natural
366 transformation, in most cases, entry into competence is controlled by regulatory signals triggered by
367 specific environmental or physiological conditions. For example, *Synechococcus* sp. PCC 7002 and
368 *Synechocystis* are most competent during exponential growth (49, 160). In contrast, competence in *S.*
369 *elongatus* appears to be regulated by the circadian clock. Pilus biogenesis occurs in the morning, while

370 natural transformation is maximal at the onset of darkness when cell division stops and DNA compaction
371 is maximal (135). Furthermore, some cyanobacteria are polyploid (152) which may help alleviate DNA
372 damage and also have mechanisms to correct damage caused by high light or UV radiation. Thus the
373 role of the circadian clock on natural transformation in Cyanobacteria deserves further study. It is
374 known that in bacteria transformation can be affected by several features, including availability of DNA
375 in the environment, the physiological state of the cell or the activity of cell receptors. In some bacteria,
376 transformation has been implicated to play a role in the DNA repair but there is no clear consensus of its
377 role.

378
379 The infection patterns and gene expression of some cyanophages infecting *Prochlorococcus* also vary as
380 a function of light (83). Some phages do not adsorb in the absence of light, others adsorb but do not
381 replicate, while others are fully functional. Under light-dark cycles, all cyanophage types exhibited
382 rhythmic transcript abundance, which might result from interactions between host photosynthetic
383 activity and the diurnal gene expression of phage genes. This might have implications for phage-driven
384 gene transfer in the ocean. Size fractionation of particles from the Red Sea confirmed that light-
385 dependent interactions are prevalent in the natural populations of phages infecting Cyanobacteriota
386 (60). It remains to be understood how many of these dependencies are the result of light-associated
387 expression of phage receptors in the cell envelope, of the physiological state of the cell, or other
388 unknown factors.

389

390 **4. EFFECT OF HGT ON CYANOBACTERIAL EVOLUTION ACROSS DIFFERENT TIME SCALES**

391

392 The range of time scales over which evolution occurs is enormous. On long time scales, major transitions
393 such as the evolution of photosynthesis or the emergence of eukaryotes can be separated by billions of
394 years, while on short time scales, prokaryotes can evolve resistance to antibiotics over a few days.
395 Recombination and HGT events can potentially impact evolution across a wide range of time scales, but
396 in most bacteria the timing and frequency of these events can only be inferred indirectly from the
397 genomes of current populations. Cyanobacteriota afford a unique opportunity for assessment of the
398 role of HGT and recombination over long time scales because some major evolutionary transitions, such
399 as the evolution of photosynthesis, can be corroborated with the fossil and biogeochemical record. To
400 take advantage of this feature, we group results from phylogenomics and evolutionary studies into three
401 broad categories, based on the evolutionary time scales involved: long time scales (ranging from tens of
402 millions to billions of years), intermediate time scales (up to a few million years, where certain bacterial
403 species originated), and short time scales (from days to thousands of years, where micro-evolutionary
404 changes within populations occur). Our intention is to present the current state of information in
405 Cyanobacteriota as well as point out challenges and opportunities in this area.

406

407 Most research on Cyanobacterial evolution has focused on long time scales during which core
408 physiological traits, such as photosynthesis and N fixation, emerged. It was initially believed that HGT
409 did not play an important role in these processes, but this view is being increasingly questioned by
410 evidence from phylogenomics (19, 121, 125, 164). On intermediate time scales, the emergence of new
411 cyanobacterial species and their diversification can be studied using sequences derived from current

412 populations (70, 105, 116). On short time scales, within-species evolution can be studied experimentally
413 in the laboratory or in natural populations that have recently been established in new environments
414 (129). In this section, we draw from recent advances in genomic and population level approaches in four
415 communities where cyanobacteria are important members: hot spring microbial mats, cyanobacterial
416 blooms, lakes, and marine environments (**Figure 3**).

417

418 *4.1 The impact of HGT on the emergence of novel major functions over long time scales*

419 Early research on cyanobacterial macroevolution was driven by the discovery of microfossils with
420 morphologies similar to those of modern day Cyanobacteriota (121). This was taken as evidence that cell
421 morphology was conserved over macroevolutionary time scales and therefore, that it could be used to
422 infer the evolution of Cyanobacteriota. However, extensive genomic sequencing has revealed large
423 variation in gene content on macroevolutionary time scales as well as extensive HGT, including between
424 phyla that diverged hundreds of millions of years ago (6, 113, 123, 124). These results imply that most
425 genes have undergone many HGT events since the last common ancestor of Cyanobacteriota, making
426 inferences about their evolution from sequencing data alone particularly challenging. Nevertheless, a
427 combination of morphological, phylogenetic, and paleontological information has been used to infer the
428 evolution of core metabolic processes, and in particular the evolution of photosynthesis (19, 26, 39, 43,
429 61, 85, 101, 118, 141).

430

431 There are two main hypotheses for the emergence of oxygenic photosynthesis in Cyanobacteriota. The
432 first is that Photosystem I (PSI) and Photosystem II (PSII) evolved in an ancestral cyanobacterium via a
433 series of duplication and divergence events. In this scenario, current anaerobic phototrophs evolved
434 from these ancient Cyanobacteriota either via selective gene loss or HGT (89, 94). A second hypothesis is
435 that Cyanobacteriota acquired either PSI or PSII, or both, through HGT from anaerobic phototrophic
436 bacteria (158). Indirect support for this hypothesis comes from a recent study which found evidence of
437 HGT of reaction center genes within the phylum of anoxygenic phototrophs of Chloroflexi (151). This
438 second scenario is currently favored by most researchers (18, 39, 61, 101).

439

440 Recent evidence suggests that HGT may also have shaped other metabolic processes of
441 Cyanobacteriota, such as central carbon metabolism and N fixation. For instance, marine
442 *Prochlorococcus* and *Synechococcus* can use chitin as a supplementary C source via a chitin-degradation
443 pathway. Two genes in this pathway were acquired from an ancestor of the aquatic bacterium
444 Planctomycetes around 500 My ago (17). Similar HGT events have been observed in other phyla, such as
445 Chloroflexota, where two genes involved in the 3-hydroxypropionate bicycle carbon fixation pathway
446 were shown to be acquired horizontally from other Chloroflexi, possibly around 1 Gy ago (6, 151).
447 Another example involves a suggested role of HGT in the acquisition of the capacity for N fixation
448 capacity. Early studies proposed that the ability to fix N originated in methanogenic archaea and was
449 subsequently transferred to other phyla via HGT (13, 112). A recent analysis of 650 Cyanobacterial
450 genomes showed that N fixation is present across many phylogenetic clades with different morphologies
451 and habitats, suggesting horizontal transfer of this trait may be common. The same study also found a
452 group of Cyanobacterial genera across the phylogenetic tree that may have acquired their N fixation
453 genes from Deltaproteobacteria (20). Genes responsible for other metabolic processes, such as the

454 transport of sulfate, phosphonate, nitrate, and nitrite were enriched in Cyanobacteriota from freshwater
455 and terrestrial environments (21). Many of these genes appeared to be horizontally transferred from
456 other genera, as shown by their unusually high sequence similarity compared to the phylogenetic
457 distance between the strains.

458

459 *4.2 Impact of HGT in functional innovation at intermediate time scales*

460 The role of HGT and recombination on intermediate time scales has been studied by sequencing of
461 Cyanobacterial populations or isolates from different environments such as hot spring microbial mats
462 (10, 116, 117);(71), oceans (69, 70), freshwater lakes (29, 106), soil crusts (130), and mud (131). We will
463 focus on four different communities: hot spring microbial mats, cyanobacterial blooms in freshwater
464 lakes, marine environments and soil or desert crusts (**Figure 3**). We summarize recent advances that
465 have provided insights into HGT and recombination.

466

467 Cyanobacteria are keystone members in the microbial mats of hot springs and form dense biofilms with
468 a variety of anoxygenic phototrophs and heterotrophs (16, 74, 134, 150). Pioneering studies of the
469 *Synechococcus* populations from Yellowstone National Park found distinct population clusters at
470 different temperature regions of the hot spring. These clusters diverged by ~4% in the 16S rRNA region,
471 with other marker genes showing considerable fine-scale diversity (91, 149, 150). Comparative analysis
472 of two related *Synechococcus* isolate genomes (*Synechococcus* OS-A and OS-B') from these hot springs
473 indicated that their lack of synteny and extensive mosaicism was the result of extensive recombination,
474 active transposable elements and HGT events, that includes the acquisition of pathways for the
475 metabolism of urea and phosphonate (10) (see also section on MGEs). This high genomic diversity and
476 phenotypic plasticity contrasts with data from human commensal bacteria, in which single samples
477 often reveal clonal populations (1). It also highlights the possibility of enormous untapped genomic
478 diversity particularly in diverse environmental communities.

479

480 Several studies suggest that genetic diversity in hot spring microbial communities cluster strongly by
481 geographic location because dispersal between communities after initial colonization events may be
482 limited (66, 103, 150, 156). The limited dispersal in these communities was recently exploited to
483 reconstruct an evolutionary scenario for the *Synechococcus* populations of the Yellowstone National
484 Park. It is likely that current populations descended from ancestral colonizers (117) that arrived after the
485 Yellowstone caldera was formed in a massive volcanic explosion around 640 Ky (116). Consistent with
486 this, 16S sequences from strains of *Synechococcus* OS-A and OS-B' like populations are most similar to
487 each other, and are clearly separated from other lineages outside North America and are not found in
488 hot springs from other continents (66). Sequence analysis also revealed extensive recombination (116)
489 and this view was recently confirmed by an analysis of over 300 single-cell genomes of *Synechococcus*
490 sp. (12) which showed that homologous recombination between species clusters was responsible for 60-
491 95% of the genetic variation in the population. In addition, several hundred genes were transferred
492 between clusters and underwent selective sweeps through the population. For example, a 23 Kb region
493 encoding an entire N fixation pathway with several dozen genes was reported in both species and
494 shown to be active in N fixation (10, 132) (Shelton et al Bioarchives)). The genomic diversity of this
495 region suggested that the *Synechococcus* OS-A like population acquired the N fixation module at some

496 point after it colonized the hot springs, and therefore within the 640 Ky since the formation of the
497 Yellowstone caldera (Birzu 2023). Thus, the current scenario suggests that even core metabolic genes
498 may have been transferred several times across strains on these time scales, and highlights the difficulty
499 of categorically identifying evolutionary histories.

500
501 Dramatic effects of HGT and homologous recombination on evolution can also be found in freshwater
502 Cyanobacteriota such as *Microcystis*, which is responsible for toxic blooms (29, 67, 105). The average
503 nucleotide identity of *Microcystis* samples from around the world is >94% (29). Rates of mutation
504 accumulation for several strains have been measured under laboratory conditions and were found to be
505 ~1 SNP/y (106)—very similar to estimates from the human gut microbiome (163). While extrapolating
506 from microevolutionary rates to macroevolution is difficult in bacteria (57), these results suggest the
507 common ancestor of *Microcystis* is likely less than several million years old. Moreover, strains sampled
508 from the same location were found to differ by <300 SNPs, implying their divergence time is only a few
509 hundred years (106). Despite these relatively short time scales, studies have found widespread evidence
510 of both HGT and homologous recombination in these populations. For example, an analysis of 79
511 genomes revealed that almost 5% of the gene families had evidence of HGT within the local population
512 (105).

513
514 In marine ecosystems, analysis of single-cell data from a closely-related group of strains from the highly
515 abundant *Prochlorococcus* clade HLII revealed around half a dozen genomic islands in which frequent
516 HGT led to highly variable gene content (69, 70). The rapid turnover of these genomic islands results in a
517 large pan-genome that contains several tens of thousands of genes, even though an individual genome
518 typically has less than 2000 genes. Recent evidence suggests tycheposons might play an important role
519 in the rapid gene turnover within islands (11, 102). Within the core genome, the evidence for the impact
520 of homologous recombination on the genetic diversity within *Prochlorococcus* is mixed. Earlier studies
521 proposed a scenario in which the diversity in the core genome of HLII is explained by the presence of up
522 to several thousand subpopulations that diverged asexually around ~1 Mya and have been coexisting
523 since (69, 70). These subpopulations were assumed to be ecologically distinct and would explain the
524 presence of genomic clusters in the single-cell data. A more recent study suggested that genomic
525 clusters could be explained by rapid recombination within closely-related subpopulations and reduced
526 recombination between them (Arevalo 2019). In this scenario, each cluster would be more similar to a
527 quasisexual species rather than a collection of asexually reproducing strains (Shapiro 2012; Arevalo
528 2019). Reconciling these views presents exciting new opportunities for both theoretical and data
529 analysis work.

530 531 4.3 *The impact of HGT on population genetics of Cyanobacteriota*

532 In recent decades, experimental evolution has greatly increased our quantitative understanding of
533 microbial evolution. Most experimental evolution studies on cyanobacteria have focused on applications
534 to biotechnology and agriculture and less on fundamental research related to evolutionary dynamics
535 (80). In general, studies of cyanobacteria have been hampered by their slow doubling times and lack of
536 genetic tools (but see (146)). We believe there are many opportunities for future studies in this
537 direction. For example, experiments involving recombination in competent populations of

538 *Synechococcus* OS-A/B', like those previously done on *Bacillus* sp. could uncover the early stages of the
539 hybridization process that led to the diversification of the population (109).

540

541 Among natural populations, bloom-forming Cyanobacteriota are a valuable model system for studying
542 short-term evolution. For example, one study investigated the genetic diversity of populations of
543 *Planktothrix* from four Norwegian lakes (129). They found very low genetic diversity in the population
544 and estimated the time since their common ancestor at ~10-100 y. Despite the short timeframe,
545 individual genomes from the population were a mixture of alleles from different strains. The authors
546 also found alleles that were mosaics of ancestral types, suggesting recombination played an important
547 role in the diversification of the population. Another study found only 68% of genes were shared
548 between 4 MAGs from a bloom population, despite the high similarity in the core genes of these
549 genomes (>99% ANI) (56).

550

551 Sequencing multiple isolates and metagenomes from diverse locations that are either geographically
552 separated or across a spatial continuum is likely to be adopted by several groups in the future as
553 sequencing costs drop. Increases in high quality genomic data will empower stronger quantitative
554 analysis and interpretations of the cyanobacterial pan-genome, and the roles of geographic isolation and
555 selection, recombination and HGT in establishing and maintaining genetic diversity in populations. For
556 example, a recent study explored the diversity of *Microcoleus* sp., a filamentous cyanobacterium, by
557 sequencing about 200 isolates from various locations and herbarium specimens. They noted a high level
558 of genetic divergence possibly driven by a complex combination of factors such as the environment,
559 geography, and selection (34, 128, 130). Cyanobacteria such as *Microcoleus* play an important role in
560 biocrusts to help in stabilizing the soil of arid environments that are found the world over (46). In
561 summary, we anticipate that insights into the role of gene flow in Cyanobacteriota are only likely to
562 grow in the next few years.

563

564 **5. PERSPECTIVES**

565 We now have ample evidence from various approaches that gene flow had a profound effect on
566 Cyanobacteriota. HGT and recombination over short and long time scales result in the acquisition of
567 novel functions and facilitate dynamic genomic rearrangements. This impacts the ability of
568 Cyanobacteriota to adapt to their environment and to rise to the challenges associated with colonizing
569 new niches. This versatile phylum is vital to the health of terrestrial and marine ecosystems (72, 73, 108)
570 and has untapped potential for translational research (14, 84, 92, 120). In addition, studies of
571 Cyanobacteriota could shed light on the evolution of photosynthesis—a major evolutionary transition
572 that fundamentally reshaped the course of life on Earth. For these reasons there is an urgency in
573 understanding their adaptability and evolvability, but knowledge about phototrophic microbes lags
574 behind that of other phyla.

575 The mechanisms of genetic exchange are scarcely explored in Cyanobacteria. Powerful computational
576 approaches using extensive genomic and metagenomic databases have identified numerous loci for
577 conjugative systems, transducing agents, phages and satellites but the ways they operate in
578 Cyanobacteriota is still unknown. Elucidation of molecular mechanisms that accelerate or limit HGT in

579 the environment or in synthetic communities, would be a major step forward. We also have almost no
580 knowledge of cyanophage receptors even though it is clear that the role of phage on host evolution and
581 survival can have significant environmental repercussions. To cope with the effects of climate change on
582 ecosystems, we need to harness microbial (104) solutions which would require cross-disciplinary efforts.

583
584 The vast genetic and morphological diversity that already exists in Nature can provide unexpected
585 insights and provide bio-inspired solutions. The unparalleled ability of Cyanobacteria to form
586 differentiated structures and bio-compartments provides a toolkit for synthetic biology that we have
587 barely exploited (33, 45). New insights into the role of HGT and gene flow can be gained from stable
588 microbial communities where phototrophic microbes are keystone members and have co-evolved to
589 extreme environments and to phages (77). These communities can now be interrogated by powerful
590 and complementary techniques such as metagenomics, proximity ligation methods and the construction
591 of synthetic communities.

592
593 Finally, we propose that it may be the right time to build experimental evolution regimes capable of
594 testing hypotheses of ways in which core metabolic pathways evolve, although this may be considerably
595 more challenging than for very fast growing bacteria (but see (146)). Such research could potentially
596 resolve inconsistencies between our understanding of short and long-term evolution in Cyanobacteriota
597 and address fundamental questions about the role of HGT and HR in adaptive evolution.

598

599 **FUTURE ISSUES**

- 600 1. Mechanisms of genetic exchange are poorly characterized in Cyanobacteriota, even though
601 many conjugative systems, transducing agents, prophages, defense systems and satellites have
602 been identified using computational methods. Elucidating gene flow in Cyanobacteriota,
603 particularly as the Earth's environment changes and episodes like harmful algal blooms increase
604 worldwide.
- 605 2. Transduction, which is likely to play an important role in gene flow, depends on the ability of
606 phages to recognize cell receptors, but there is little information regarding phage receptors in
607 Cyanobacteriota and how they might evolve under phage pressure.
- 608 3. Building synthetic communities to be queried using high throughput technologies (e.g.
609 microfluidics) will allow testing hypotheses about gene flow. Experimental evolution of
610 community members can be used to study the fitness landscape and formation of hybrid genes
611 and genotypes.
- 612 4. Phototrophic biofilms harbor dense, relatively stable diverse communities and provide naturally
613 evolving model communities to track HGT and recombination.
- 614 5. The development of MGEs for genetic engineering will accelerate the use of Cyanobacteriota as
615 effective chassis for synthetic biology.

616

617

618 **Acknowledgments**

619 DB was supported by the Carnegie Institution and NSF grant #1921429 (BBSRC-NSF/BIO) and NSF grant
620 #2125965 (Emerging Frontiers). GB was partially supported by the Simons Foundation Postdoctoral
621 Fellowship Award 730295. ER was supported by the INCEPTION project (PIA/ANR-16-CONV-0005), the
622 Laboratoire d'Excellence IBEID Integrative Biology of Emerging Infectious Diseases (ANR-10-LABX-62-
623 IBEID), and regular funding from the Institut Pasteur and the CNRS. DB and ER thank Boris Shraiman for
624 hosting them at the Kavli Institute for Theoretical Physics, Santa Barbara, which was supported in part
625 by NSF grant # PHY-2309135. We thank Charles Coluzzi for the data used to draw Figure 1. GB thanks
626 Daniel Fisher for helpful discussions.

627

628 REFERENCES

- 629 1. Achtman M, Wagner M. 2008. Microbial diversity and the genetic nature of microbial species. *Nat.*
630 *Rev. Microbiol.* 6(6):431–40
- 631 2. Adams MM, Gómez-García MR, Grossman AR, Bhaya D. 2008. Phosphorus deprivation responses
632 and phosphonate utilization in a thermophilic *Synechococcus* sp. from microbial mats. *J. Bacteriol.*
633 190(24):8171–84
- 634 3. Agostinho DP, Fu Y, Menon VK, Metcalf GA, Treangen TJ, Sedlazeck FJ. 2024. Unveiling microbial
635 diversity: harnessing long-read sequencing technology. *Nat. Methods.* 21(6):954–66
- 636 4. Ambur OH, Engelstädter J, Johnsen PJ, Miller EL, Rozen DE. 2016. Steady at the wheel: conservative
637 sex and the benefits of bacterial transformation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371(1706):
- 638 5. Arnold BJ, Huang I-T, Hanage WP. 2022. Horizontal gene transfer and adaptive evolution in
639 bacteria. *Nat. Rev. Microbiol.* 20(4):206–18
- 640 6. Assié A, Leisch N, Meier DV, Gruber-Vodicka H, Tegetmeyer HE, et al. 2020. Horizontal acquisition
641 of a patchwork Calvin cycle by symbiotic and free-living *Campylobacterota* (formerly
642 *Epsilonproteobacteria*). *ISME J.* 14(1):104–22
- 643 7. Bar-On YM, Phillips R, Milo R. 2018. The biomass distribution on Earth. *Proc. Natl. Acad. Sci. U. S. A.*
644 115(25):6506–11
- 645 8. Bernheim A, Bikard D, Touchon M, Rocha EPC. 2019. A matter of background: DNA repair pathways
646 as a possible cause for the sparse distribution of CRISPR-Cas systems in bacteria. *Philos. Trans. R.*

- 647 *Soc. Lond. B Biol. Sci.* 374(1772):20180088
- 648 9. Bernheim A, Sorek R. 2020. The pan-immune system of bacteria: antiviral defence as a community
649 resource. *Nat. Rev. Microbiol.* 18(2):113–19
- 650 10. Bhaya D, Grossman AR, Steunou A-S, Khuri N, Cohan FM, et al. 2007. Population level functional
651 diversity in a microbial community revealed by comparative genomic and metagenomic analyses.
652 *ISME J.* 1(8):703–13
- 653 11. Biller SJ, Berube PM, Lindell D, Chisholm SW. 2015. Prochlorococcus: the structure and function of
654 collective diversity. *Nat. Rev. Microbiol.* 13(1):13–27
- 655 12. Birzu G, Muralidharan HS, Goudeau D, Malmstrom RR, Fisher DS, Bhaya D. 2023. Hybridization
656 breaks species barriers in long-term coevolution of a cyanobacterial population. *bioRxiv*
- 657 13. Bolhuis H, Cretoiu MS, Stal LJ. 2014. Molecular ecology of microbial mats. *FEMS Microbiol. Ecol.*
658 90(2):335–50
- 659 14. Bozan M, Berreth H, Lindberg P, Bühler K. 2024. Cyanobacterial biofilms: from natural systems to
660 applications. *Trends Biotechnol.*
- 661 15. Brito IL. 2021. Examining horizontal gene transfer in microbial communities. *Nat. Rev. Microbiol.*
662 19(7):442–53
- 663 16. Brock TD. 1995. The road to Yellowstone--and beyond. *Annu. Rev. Microbiol.* 49:1–28
- 664 17. Capovilla G, Braakman R, Fournier GP, Hackl T, Schwartzman J, et al. 2023. Chitin utilization by
665 marine picocyanobacteria and the evolution of a planktonic lifestyle. *Proc. Natl. Acad. Sci. U. S. A.*
666 120(20):e2213271120
- 667 18. Cardona T. 2019. Thinking twice about the evolution of photosynthesis. *Open Biol.* 9(3):180246
- 668 19. Cardona T, Rutherford AW. 2019. Evolution of photochemical reaction centres: More twists? *Trends*
669 *Plant Sci.* 24(11):1008–21
- 670 20. Chen M-Y, Teng W-K, Zhao L, Han B-P, Song L-R, Shu W-S. 2022. Phylogenomics uncovers

- 671 evolutionary trajectory of nitrogen fixation in Cyanobacteria. *Mol. Biol. Evol.* 39(9):
- 672 21. Chen M-Y, Teng W-K, Zhao L, Hu C-X, Zhou Y-K, et al. 2021. Comparative genomics reveals insights
673 into cyanobacterial evolution and habitat adaptation. *ISME J.* 15(1):211–27
- 674 22. Chevallereau A, Pons BJ, van Houte S, Westra ER. 2022. Interactions between bacterial and phage
675 communities in natural environments. *Nat. Rev. Microbiol.* 20(1):49–62
- 676 23. Conradi FD, Mullineaux CW, Wilde A. 2020. The Role of the Cyanobacterial Type IV Pilus Machinery
677 in Finding and Maintaining a Favourable Environment. *Life.* 10(11):
- 678 24. Crummett LT, Puxty RJ, Weihe C, Marston MF, Martiny JBH. 2016. The genomic content and context
679 of auxiliary metabolic genes in marine cyanomyoviruses. *Virology.* 499:219–29
- 680 25. Davison M, Treangen TJ, Koren S, Pop M, Bhaya D. 2016. Diversity in a Polymicrobial Community
681 Revealed by Analysis of Viromes, Endolysins and CRISPR Spacers. *PLoS One.* 11(9):e0160574
- 682 26. Demoulin CF, Lara YJ, Cornet L, François C, Baurain D, et al. 2019. Cyanobacteria evolution: Insight
683 from the fossil record. *Free Radic. Biol. Med.* 140:206–23
- 684 27. Denise R, Abby SS, Rocha EPC. 2019. Diversification of the type IV filament superfamily into
685 machines for adhesion, protein secretion, DNA uptake, and motility. *PLoS Biol.* 17(7):e3000390
- 686 28. de Vries S, de Vries J. 2022. Evolutionary genomic insights into cyanobacterial symbioses in plants.
687 *Quant. Plant Biol.* 3(e16):e16
- 688 29. Dick GJ, Duhaime MB, Evans JT, Errera RM, Godwin CM, et al. 2021. The genetic and
689 ecophysiological diversity of *Microcystis*. *Environ. Microbiol.* 23(12):7278–7313
- 690 30. Dittmann E, Gugger M, Sivonen K, Fewer DP. 2015. Natural Product Biosynthetic Diversity and
691 Comparative Genomics of the Cyanobacteria. *Trends Microbiol.* 23(10):642–52
- 692 31. Dreher TW, Matthews R, Davis EW 2nd, Mueller RS. 2023. *Woronichinia naegeliana*: A common
693 nontoxic component of temperate freshwater cyanobacterial blooms with 30% of its genome
694 in transposons. *Harmful Algae.* 125:102433

- 695 32. Dubnau D, Blokesch M. 2019. Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annu.*
696 *Rev. Genet.* 53:217–37
- 697 33. Dutka P, Metskas LA, Hurt RC, Salahshoor H, Wang T-Y, et al. 2023. Structure of *Anabaena flos-*
698 *aquae* gas vesicles revealed by cryo-ET. *Structure.* 31(5):518–28.e6
- 699 34. Dvořák P, Jahodářová E, Stanojković A, Skoupy S, Casamatta DA. 2023. Population genomics meets
700 the taxonomy of cyanobacteria. *Algal Res.* 72(103128):103128
- 701 35. Dyhrman ST, Chappell PD, Haley ST, Moffett JW, Orchard ED, et al. 2006. Phosphonate utilization by
702 the globally important marine diazotroph *Trichodesmium*. *Nature.* 439(7072):68–71
- 703 36. Encinas D, Garcillán-Barcia MP, Santos-Merino M, Delaye L, Moya A, de la Cruz F. 2014. Plasmid
704 conjugation from proteobacteria as evidence for the origin of xenologous genes in cyanobacteria. *J.*
705 *Bacteriol.* 196(8):1551–59
- 706 37. Eren AM, Kiehl E, Shaiber A, Veseli I, Miller SE, et al. 2021. Community-led, integrated, reproducible
707 multi-omics with anvi'o. *Nat. Microbiol.* 6(1):3–6
- 708 38. Fang M, LiWang A, Golden SS, Partch CL. 2024. The inner workings of an ancient biological clock.
709 *Trends Biochem. Sci.* 49(3):236–46
- 710 39. Fischer WW, Hemp J, Johnson JE. 2016. Evolution of Oxygenic Photosynthesis. *Annu. Rev. Earth*
711 *Planet. Sci.* 44(1):647–83
- 712 40. Flemming H-C, Wuertz S. 2019. Bacteria and archaea on Earth and their abundance in biofilms. *Nat.*
713 *Rev. Microbiol.* 17(4):247–60
- 714 41. Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, et al. 2013. Present and future global
715 distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad.*
716 *Sci. U. S. A.* 110(24):9824–29
- 717 42. Foster RA, Zehr JP. 2019. Diversity, Genomics, and Distribution of Phytoplankton-Cyanobacterium
718 Single-Cell Symbiotic Associations. *Annu. Rev. Microbiol.* 73:435–56

- 719 43. Fournier GP, Moore KR, Rangel LT, Payette JG, Momper L, Bosak T. 2021. The Archean origin of
720 oxygenic photosynthesis and extant cyanobacterial lineages. *Proc. Biol. Sci.* 288(1959):20210675
- 721 44. Frangeul L, Quillardet P, Castets A-M, Humbert J-F, Matthijs HCP, et al. 2008. Highly plastic genome
722 of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. *BMC Genomics.*
723 9(1):274
- 724 45. Gao Y-L, Cournoyer JE, De BC, Wallace CL, Ulanov AV, et al. 2024. Introducing carbon assimilation in
725 yeasts using photosynthetic directed endosymbiosis. *Nat. Commun.* 15(1):5947
- 726 46. Garcia-Pichel F. 2023. The microbiology of biological soil crusts. *Annu. Rev. Microbiol.* 77:149–71
- 727 47. Garcillán-Barcia MP, Francia MV, de la Cruz F. 2009. The diversity of conjugative relaxases and its
728 application in plasmid classification. *FEMS Microbiol. Rev.* 33(3):657–87
- 729 48. Georjon H, Bernheim A. 2023. The highly diverse antiphage defence systems of bacteria. *Nat. Rev.*
730 *Microbiol.* 21(10):686–700
- 731 49. Golden SS, Sherman LA. 1984. Optimal conditions for genetic transformation of the
732 cyanobacterium *Anacystis nidulans* R2. *J. Bacteriol.* 158(1):36–42
- 733 50. Grasso CR, Pokrzywinski KL, Waechter C, Rycroft T, Zhang Y, et al. 2022. A review of Cyanophage-
734 host relationships: Highlighting cyanophages as a potential Cyanobacteria control strategy. *Toxins*
735 *(Basel)*. 14(6):385
- 736 51. Guajardo-Leiva S, Pedrós-Alió C, Salgado O, Pinto F, Díez B. 2018. Active Crossfire Between
737 Cyanobacteria and Cyanophages in Phototrophic Mat Communities Within Hot Springs. *Front.*
738 *Microbiol.* 9:2039
- 739 52. Guglielmini J, de la Cruz F, Rocha EPC. 2013. Evolution of conjugation and type IV secretion systems.
740 *Mol. Biol. Evol.* 30(2):315–31
- 741 53. Guglielmini J, Néron B, Abby SS, Garcillán-Barcia MP, la Cruz F de, Rocha EPC. 2014. Key
742 components of the eight classes of type IV secretion systems involved in bacterial conjugation or

- 743 protein secretion. *Nucleic Acids Res.* 42(9):5715–27
- 744 54. Guglielmini J, Quintais L, Garcillán-Barcia MP, de la Cruz F, Rocha EPC. 2011. The repertoire of ICE in
745 prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet.*
746 7(8):e1002222
- 747 55. Hackl T, Laurenceau R, Ankenbrand MJ, Bliem C, Cariani Z, et al. 2023. Novel integrative elements
748 and genomic plasticity in ocean ecosystems. *Cell.* 186(1):47–62.e16
- 749 56. Halary S, Duperron S, Kim Tiam S, Duval C, Bernard C, Marie B. 2023. Intra-population genomic
750 diversity of the bloom-forming cyanobacterium, *Aphanizomenon gracile*, at low spatial scale. *ISME*
751 *Commun.* 3(1):57
- 752 57. Harmon LJ, Pennell MW, Henao-Diaz LF, Rolland J, Siple BN, Uyeda JC. 2021. Causes and
753 consequences of apparent timescaling across all estimated evolutionary rates. *Annu. Rev. Ecol.*
754 *Evol. Syst.* 52(1):587–609
- 755 58. Heidelberg JF, Nelson WC, Schoenfeld T, Bhaya D. 2009. Germ warfare in a microbial mat
756 community: CRISPRs provide insights into the co-evolution of host and viral genomes. *PLoS One.*
757 4(1):e4169
- 758 59. Hein S, Scholz I, Voß B, Hess WR. 2013. Adaptation and modification of three CRISPR loci in two
759 closely related cyanobacteria. *RNA Biol.* 10(5):852–64
- 760 60. Hevroni G, Flores-Urbe J, Béjà O, Philosofof A. 2020. Seasonal and diel patterns of abundance and
761 activity of viruses in the Red Sea. *Proc. Natl. Acad. Sci. U. S. A.* 117(47):29738–47
- 762 61. Hohmann-Marriott MF, Blankenship RE. 2011. Evolution of photosynthesis. *Annu. Rev. Plant Biol.*
763 62:515–48
- 764 62. Hsieh S-C, Peters JE. 2023. Discovery and characterization of novel type I-D CRISPR-guided
765 transposons identified among diverse Tn7-like elements in cyanobacteria. *Nucleic Acids Res.*
766 51(2):765–82

- 767 63. Hsieh S-C, Peters JE. 2024. Natural and engineered guide RNA-directed transposition with CRISPR-
768 associated Tn7-like transposons. *Annu. Rev. Biochem.* 93(1):139–61
- 769 64. Hüner NPA, Ivanov AG, Szyszka-Mroz B, Savitch LV, Smith DR, Kata V. 2024. Photostasis and
770 photosynthetic adaptation to polar life. *Photosynth. Res.* 161(1-2):51–64
- 771 65. Hurwitz BL, Brum JR, Sullivan MB. 2015. Depth-stratified functional and taxonomic niche
772 specialization in the “core” and “flexible” Pacific Ocean Virome. *ISME J.* 9(2):472–84
- 773 66. Ionescu D, Hindiyeh M, Malkawi H, Oren A. 2010. Biogeography of thermophilic cyanobacteria:
774 insights from the Zerka Ma'in hot springs (Jordan). *FEMS Microbiol. Ecol.* 72(1):103–13
- 775 67. Jackrel SL, White JD, Evans JT, Buffin K, Hayden K, et al. 2019. Genome evolution and host-
776 microbiome shifts correspond with intraspecific niche divergence within harmful algal bloom-
777 forming *Microcystis aeruginosa*. *Mol. Ecol.* 28(17):3994–4011
- 778 68. Jiang T, Guo C, Wang M, Wang M, Zhang X, et al. 2020. Genome analysis of two novel
779 *Synechococcus* phages that lack common auxiliary metabolic genes: Possible reasons and ecological
780 insights by comparative analysis of cyanomyoviruses. *Viruses.* 12(8):800
- 781 69. Kashtan N, Roggensack SE, Berta-Thompson JW, Grinberg M, Stepanauskas R, Chisholm SW. 2017.
782 Fundamental differences in diversity and genomic population structure between Atlantic and
783 Pacific *Prochlorococcus*. *ISME J.* 11(9):1997–2011
- 784 70. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, et al. 2014. Single-Cell Genomics
785 Reveals Hundreds of Coexisting Subpopulations in Wild *Prochlorococcus*. *Science.* 344(6182):416
- 786 71. Kees ED, Murugapiran SK, Bennett AC, Hamilton TL. 2022. Distribution and Genomic Variation of
787 Thermophilic Cyanobacteria in Diverse Microbial Mats at the Upper Temperature Limits of
788 Photosynthesis. *mSystems*, p. e0031722
- 789 72. Kim M, Jung J, Kim W, Park Y, Jeon CO, Park W. 2024. Extensive genomic rearrangement of
790 catalase-less cyanobloom-forming *Microcystis aeruginosa* in freshwater ecosystems. *J. Microbiol.*

- 791 73. Kim M-J, Kang D, Lee G, Kim K, Kim J, et al. 2023. Interplays between cyanobacterial blooms and
792 antibiotic resistance genes. *Environ. Int.* 181(108268):108268
- 793 74. Klatt CG, Inskeep WP, Herrgard MJ, Jay ZJ, Rusch DB, et al. 2013. Community structure and function
794 of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal
795 environments. *Front. Microbiol.* 4:106
- 796 75. Knoll AH. *Life on a Young Planet: The First Three Billion Years of Evolution on Earth* . Princeton, NJ:
797 Princeton University Press. 2003
- 798 76. Koksharova OA, Wolk CP. 2002. Genetic tools for cyanobacteria. *Appl. Microbiol. Biotechnol.*
799 58(2):123–37
- 800 77. Kolan D, Cattan-Tsaushu E, Enav H, Freiman Z, Malinsky-Rushansky N, et al. 2024. Tradeoffs
801 between phage resistance and nitrogen fixation drive the evolution of genes essential for
802 cyanobacterial heterocyst functionality. *ISME J.* 18(1):
- 803 78. Kumar HD, Ueda K. 1984. Conjugation in the cyanobacterium *Anacystis nidulans*. *Mol. Gen. Genet.*
804 195(1-2):356–57
- 805 79. Laurenceau R, Raho N, Forget M, Arellano AA, Chisholm SW. 2021. Frequency of mispackaging of
806 *Prochlorococcus* DNA by cyanophage. *ISME J.* 15(1):129–40
- 807 80. Leister D. 2018. Experimental evolution in photoautotrophic microorganisms as a means of
808 enhancing chloroplast functions. *Essays Biochem.* 62(1):77–84
- 809 81. Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. 2004. Transfer of
810 photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl. Acad. Sci. U. S. A.*
811 101(30):11013–18
- 812 82. Lin S, Haas S, Zemojtel T, Xiao P, Vingron M, Li R. 2011. Genome-wide comparison of cyanobacterial
813 transposable elements, potential genetic diversity indicators. *Gene.* 473(2):139–49
- 814 83. Liu R, Liu Y, Chen Y, Zhan Y, Zeng Q. 2019. Cyanobacterial viruses exhibit diurnal rhythms during

- 815 infection. *Proc. Natl. Acad. Sci. U. S. A.* 116(28):14077–82
- 816 84. Li Z, Li S, Chen L, Sun T, Zhang W. 2024. Fast-growing cyanobacterial chassis for synthetic biology
817 application. *Crit. Rev. Biotechnol.* 44(3):414–28
- 818 85. Louca S, Shih PM, Pennell MW, Fischer WW, Parfrey LW, Doebeli M. 2018. Bacterial Diversification
819 through Geological Time. *Nature Ecology & Evolution.* 2(9):1458–67
- 820 86. Luo X-Q, Wang P, Li J-L, Ahmad M, Duan L, et al. 2022. Viral community-wide auxiliary metabolic
821 genes differ by lifestyles, habitats, and hosts. *Microbiome.* 10(1):190
- 822 87. Lyons TW, Tino CJ, Fournier GP, Anderson RE, Leavitt WD, et al. 2024. Co-evolution of early Earth
823 environments and microbial life. *Nat. Rev. Microbiol.* 22(9):572–86
- 824 88. Magnabosco C, Husain F, Paoletti MM, Parsons C, Payette JG, et al. 2024. Toward a natural history
825 of microbial life. *Annu. Rev. Earth Planet. Sci.* 52(1):85–108
- 826 89. Martin WF, Bryant DA, Beatty JT. 2018. A physiological perspective on the origin and evolution of
827 photosynthesis. *FEMS Microbiol. Rev.* 42(2):205–31
- 828 90. McDaniel LD. 2011. Viruses of Cyanobacteria. In *Studies in Viral Ecology*, ed CJ Hurst. 155:169–87.
829 Hoboken, NJ, USA: John Wiley & Sons, Inc.
- 830 91. Melendrez MC, Lange RK, Cohan FM, Ward DM. 2011. Influence of molecular resolution on
831 sequence-based discovery of ecological diversity among *Synechococcus* populations in an alkaline
832 siliceous hot spring microbial mat. *Appl. Environ. Microbiol.* 77(4):1359–67
- 833 92. Melis A, Hidalgo Martinez DA, Betterle N. 2023. Perspectives of cyanobacterial cell factories.
834 *Photosynth. Res.*
- 835 93. Miller SR, Abresch HE, Ulrich NJ, Sano EB, Demaree AH, et al. 2021. Bacterial Adaptation by a
836 Transposition Burst of an Invading IS Element. *Genome Biol. Evol.* 13(11):
- 837 94. Mulkidjanian AY, Koonin EV, Makarova KS, Mekhedov SL, Sorokin A, et al. 2006. The cyanobacterial
838 genome core and the origin of photosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 103(35):13126–31

- 839 95. Mullineaux CW, Wilde A. 2021. The social life of cyanobacteria. *Elife*. 10:
- 840 96. Mutalipassi M, Riccio G, Mazzella V, Galasso C, Somma E, et al. 2021. Symbioses of Cyanobacteria in
841 marine environments: Ecological insights and biotechnological perspectives. *Mar. Drugs*. 19(4):227
- 842 97. Nelson WC, Wollerman L, Bhaya D, Heidelberg JF. 2011. Analysis of insertion sequences in
843 thermophilic cyanobacteria: exploring the mechanisms of establishing, maintaining, and
844 withstanding high insertion sequence abundance. *Appl. Environ. Microbiol.* 77(15):5458–66
- 845 98. Nies F, Mielke M, Pochert J, Lamparter T. 2020. Natural transformation of the filamentous
846 cyanobacterium *Phormidium lacuna*. *PLoS One*. 15(6):e0234440
- 847 99. Nies F, Springstein BL, Hanke DM, Dagan T. 2022. Natural Competence in the Filamentous,
848 Heterocystous Cyanobacterium *Chlorogloeopsis fritschii* PCC 6912. *mSphere*. 7(4):e0099721
- 849 100. Nies F, Wein T, Hanke DM, Springstein BL, Alcorta J, et al. 2023. Role of natural transformation in
850 the evolution of small cryptic plasmids in *Synechocystis* sp. PCC 6803. *Environ. Microbiol. Rep.*
851 15(6):656–68
- 852 101. Oliver T, Kim TD, Trinugroho JP, Cerdón-Preciado V, Wijayatilake N, et al. 2023. The evolution and
853 evolvability of photosystem II. *Annu. Rev. Plant Biol.* 74:225–57
- 854 102. Pachiadaki MG, Brown JM, Brown J, Bezuidt O, Berube PM, et al. 2019. Charting the Complexity of
855 the Marine Microbiome through Single-Cell Genomics. *Cell*. 179(7):1623–35.e11
- 856 103. Papke RT, Ramsing NB, Bateson MM, Ward DM. 2003. Geographical isolation in hot spring
857 cyanobacteria. *Environ. Microbiol.* 5(8):650–59
- 858 104. Peixoto R, Voolstra CR, Stein LY, Hugenholtz P, Salles JF, et al. 2024. Microbial solutions must be
859 deployed against climate catastrophe. *Nat. Microbiol.*, pp. 1–2
- 860 105. Pérez-Carrascal OM, Terrat Y, Giani A, Fortin N, Greer CW, et al. 2019. Coherence of *Microcystis*
861 species revealed through population genomics. *ISME J.* 13(12):2887–2900
- 862 106. Pérez-Carrascal OM, Tromas N, Terrat Y, Moreno E, Giani A, et al. 2021. Single-colony sequencing

- 863 reveals microbe-by-microbiome phyllosymbiosis between the cyanobacterium *Microcystis* and its
864 associated bacteria. *Microbiome*. 9(1):194
- 865 107. Peters JE, Makarova KS, Shmakov S, Koonin EV. 2017. Recruitment of CRISPR-Cas systems by Tn7-
866 like transposons. *Proc. Natl. Acad. Sci. U. S. A.* 114(35):E7358–66
- 867 108. Plaas HE, Paerl HW. 2021. Toxic Cyanobacteria: A Growing Threat to Water and Air Quality. *Environ.*
868 *Sci. Technol.* 55(1):44–64
- 869 109. Power JJ, Pinheiro F, Pompei S, Kovacova V, Yüksel M, et al. 2021. Adaptive evolution of hybrid
870 bacteria by horizontal gene transfer. *Proc. Natl. Acad. Sci. U. S. A.* 118(10):
- 871 110. Puxty RJ, Millard AD. 2023. Functional ecology of bacteriophages in the environment. *Curr. Opin.*
872 *Microbiol.* 71:102245
- 873 111. Quevarec L, Brasseur G, Aragnol D, Robaglia C. 2024. Tracking the early events of photosymbiosis
874 evolution. *Trends Plant Sci.* 29(4):406–12
- 875 112. Raymond J, Siefert JL, Staples CR, Blankenship RE. 2004. The natural history of nitrogen fixation.
876 *Mol. Biol. Evol.* 21(3):541–54
- 877 113. Raymond J, Zhaxybayeva O, Gogarten JP, Gerdes SY, Blankenship RE. 2002. Whole-genome analysis
878 of photosynthetic prokaryotes. *Science*. 298(5598):1616–20
- 879 114. Rebière M-C, Castets A-M, Houmard J, Tandeau de Marsac N. 1986. Plasmid distribution among
880 unicellular and filamentous cyanobacteria: Occurrence of large and mega-plasmids. *FEMS*
881 *Microbiol. Lett.* 37(3):269–75
- 882 115. Rong C, Zhou K, Li S, Xiao K, Xu Y, et al. 2022. Isolation and characterization of a novel Cyanophage
883 encoding multiple auxiliary metabolic genes. *Viruses*. 14(5):887
- 884 116. Rosen MJ, Davison M, Bhaya D, Fisher DS. 2015. Microbial diversity. Fine-scale diversity and
885 extensive recombination in a quasisexual bacterial population occupying a broad niche. *Science*.
886 348(6238):1019–23

- 887 117. Rosen MJ, Davison M, Fisher DS, Bhaya D. 2018. Probing the ecological and evolutionary history of
888 a thermophilic cyanobacterial population via statistical properties of its microdiversity. *PLoS One*.
889 13(11):e0205396
- 890 118. Sánchez-Baracaldo P, Bianchini G, Wilson JD, Knoll AH. 2022. Cyanobacteria and biogeochemical
891 cycles through Earth history. *Trends Microbiol.* 30(2):143–57
- 892 119. Schirmacher AM, Hanamghar SS, Zedler JAZ. 2020. Function and benefits of natural competence in
893 Cyanobacteria: From ecology to targeted manipulation. *Life (Basel)*. 10(11):249
- 894 120. Schmelling NM, Bross M. 2024. What is holding back cyanobacterial research and applications? A
895 survey of the cyanobacterial research community. *Nat. Commun.* 15(1):6758
- 896 121. Schopf JW. 1994. Disparate rates, differing fates: tempo and mode of evolution changed from the
897 Precambrian to the Phanerozoic. *Proc. Natl. Acad. Sci. U. S. A.* 91(15):6735–42
- 898 122. Sessions AL, Doughty DM, Welander PV, Summons RE, Newman DK. 2009. The continuing puzzle of
899 the great oxidation event. *Curr. Biol.* 19(14):R567–74
- 900 123. Shih PM, Hemp J, Ward LM, Matzke NJ, Fischer WW. 2017. Crown group Oxyphotobacteria
901 postdate the rise of oxygen. *Geobiology*. 15(1):19–29
- 902 124. Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, et al. 2013. Improving the coverage of the
903 cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.*
904 110(3):1053–58
- 905 125. Shi T, Falkowski PG. 2008. Genome evolution in cyanobacteria: the stable core and the variable
906 shell. *Proc. Natl. Acad. Sci. U. S. A.* 105(7):2510–15
- 907 126. Shu W-S, Huang L-N. 2021. Microbial diversity in extreme environments. *Nat. Rev. Microbiol.*, pp.
908 1–17
- 909 127. Sibbald SJ, Archibald JM. 2020. Genomic Insights into Plastid Evolution. *Genome Biol. Evol.*
910 12(7):978–90

- 911 128. Skoupý S, Stanojković A, Casamatta DA, McGovern C, Martinović A, et al. 2024. Population
912 genomics and morphological data bridge the centuries of cyanobacterial taxonomy along the
913 continuum of *Microcoleus* species. *iScience*. 27(4):109444
- 914 129. Sogge H, Rohrlack T, Rounge TB, Sønstebø JH, Tooming-Klunderud A, et al. 2013. Gene flow,
915 recombination, and selection in cyanobacteria: population structure of geographically related
916 *Planktothrix* freshwater strains. *Appl. Environ. Microbiol.* 79(2):508–15
- 917 130. Stanojković A, Skoupý S, Johannesson H, Dvořák P. 2024. The global speciation continuum of the
918 cyanobacterium *Microcoleus*. *Nat. Commun.* 15(1):1–15
- 919 131. Stanojković A, Skoupý S, Škaloud P, Dvořák P. 2022. High genomic differentiation and limited gene
920 flow indicate recent cryptic speciation within the genus *Laspinema* (cyanobacteria). *Front.*
921 *Microbiol.* 13:977454
- 922 132. Steunou A-S, Bhaya D, Bateson MM, Melendrez MC, Ward DM, et al. 2006. In situ analysis of
923 nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot
924 spring microbial mats. *Proc. Natl. Acad. Sci. U. S. A.* 103(7):2398–2403
- 925 133. Strunecký O, Ivanova AP, Mareš J. 2023. An updated classification of cyanobacterial orders and
926 families based on phylogenomic and polyphasic analysis. *J. Phycol.* 59(1):12–51
- 927 134. Tank M, Thiel V, Ward DM, Bryant DA. 2017. A Panoply of Phototrophs: An Overview of the
928 Thermophilic Chlorophototrophs of the Microbial Mats of Alkaline Siliceous Hot Springs in
929 Yellowstone National Park, WY, USA. In *Modern Topics in the Phototrophic Prokaryotes:*
930 *Environmental and Applied Aspects*, ed PC Hallenbeck, pp. 87–137. Cham: Springer International
931 Publishing
- 932 135. Taton A, Erikson C, Yang Y, Rubin BE, Rifkin SA, et al. 2020. The circadian clock and darkness control
933 natural competence in cyanobacteria. *Nat. Commun.* 11(1):1688
- 934 136. Taton A, Gilderman TS, Ernst DC, Omega CA, Cohen LA, et al. 2023. *Synechococcus elongatus*

- 935 Argonaute reduces natural transformation efficiency and provides immunity against exogenous
936 plasmids. *MBio*, p. e0184323
- 937 137. Tempel S, Bedo J, Talla E. 2022. From a large-scale genomic analysis of insertion sequences to
938 insights into their regulatory roles in prokaryotes. *BMC Genomics*. 23(1):451
- 939 138. Tesson F, Hervé A, Mordret E, Touchon M, d’Humières C, et al. 2022. Systematic and quantitative
940 view of the antiviral arsenal of prokaryotes. *Nat. Commun.* 13(1):2561
- 941 139. Tian F, Wainaina JM, Howard-Varona C, Domínguez-Huerta G, Bolduc B, et al. 2024. Prokaryotic-
942 virus-encoded auxiliary metabolic genes throughout the global oceans. *Microbiome*. 12(1):159
- 943 140. Timms VJ, Hassan KA, Pearson LA, Neilan BA. 2023. Cyanobacteria as a critical reservoir of the
944 environmental antimicrobial resistome. *Environ. Microbiol.*
- 945 141. Tomitani A, Knoll AH, Cavanaugh CM, Ohno T. 2006. The evolutionary diversification of
946 cyanobacteria: molecular-phylogenetic and paleontological perspectives. *Proc. Natl. Acad. Sci. U. S.*
947 *A.* 103(14):5442–47
- 948 142. Touchon M, Moura de Sousa JA, Rocha EP. 2017. Embracing the enemy: the diversification of
949 microbial gene repertoires by phage-mediated horizontal gene transfer. *Curr. Opin. Microbiol.*
950 38:66–73
- 951 143. Treangen TJ, Rocha EPC. 2011. Horizontal Transfer, Not Duplication, Drives the Expansion of Protein
952 Families in Prokaryotes. *PLoS Genet.* 7(1):e1001284
- 953 144. Tria FDK, Martin WF. 2021. Gene Duplications Are At Least 50 Times Less Frequent than Gene
954 Transfers in Prokaryotic Genomes. *Genome Biol. Evol.* 13(10):
- 955 145. Ulrich NJ, Miller SR. 2024. Integration of horizontally acquired light-harvesting genes into an
956 ancestral regulatory network in the cyanobacterium *Acaryochloris marina* MBIC11017. *MBio*, p.
957 e0242324
- 958 146. Ungerer J, Wendt KE, Hendry JI, Maranas CD, Pakrasi HB. 2018. Comparative genomics reveals the

- 959 molecular determinants of rapid growth of the cyanobacterium *Synechococcus elongatus* UTEX
960 2973. *Proc. Natl. Acad. Sci. U. S. A.* 115(50):E11761–70
- 961 147. Veseli I, Cooper ZS, DeMers MA, Schechter MS, Miller S, et al. 2024. *Digital Microbe: A Genome-*
962 *Informed Data Integration Framework for Collaborative Research on Emerging Model Organisms.*
963 Work. Pap.
- 964 148. Wang Y, Ferrinho S, Connaris H, Goss RJM. 2023. The impact of viral infection on the chemistries of
965 the earth's most abundant photosynthesizers: Metabolically talented aquatic Cyanobacteria.
966 *Biomolecules.* 13(8):
- 967 149. Ward DM, Bateson MM, Ferris MJ, Kühl M, Wieland A, et al. 2006. Cyanobacterial ecotypes in the
968 microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-
969 like units linking microbial community composition, structure and function. *Philos. Trans. R. Soc.*
970 *Lond. B Biol. Sci.* 361(1475):1997–2008
- 971 150. Ward DM, Ferris MJ, Nold SC, Bateson MM. 1998. A natural view of microbial biodiversity within
972 hot spring cyanobacterial mat communities. *Microbiol. Mol. Biol. Rev.* 62(4):1353–70
- 973 151. Ward LM, Hemp J, Shih PM, McGlynn SE, Fischer WW. 2018. Evolution of Phototrophy in the
974 Chloroflexi Phylum Driven by Horizontal Gene Transfer. *Front. Microbiol.* 9:260
- 975 152. Watanabe S. 2020. Cyanobacterial multi-copy chromosomes and their replication. *Biosci.*
976 *Biotechnol. Biochem.* 84(7):1309–21
- 977 153. Weisberg AJ, Chang JH. 2023. Mobile genetic element flexibility as an underlying principle to
978 bacterial evolution. *Annu. Rev. Microbiol.* 77:603–24
- 979 154. Wendt KE, Pakrasi HB. 2019. Genomics Approaches to Deciphering Natural Transformation in
980 Cyanobacteria. *Front. Microbiol.* 10:1259
- 981 155. Wendt KE, Walker P, Sengupta A, Ungerer J, Pakrasi HB. 2021. Engineering natural competence into
982 the fast-growing cyanobacterium *Synechococcus elongatus* UTEX 2973. *Appl. Environ. Microbiol.*, p.

- 983 AEM0188221
- 984 156. Whitaker RJ, Grogan DW, Taylor JW. 2003. Geographic barriers isolate endemic populations of
985 hyperthermophilic archaea. *Science*. 301(5635):976–78
- 986 157. Wu X, Jia W, Fang Z, Sun H, Wang G, et al. 2024. Cyanobacteria mediate the dissemination of
987 bacterial antibiotic resistance through conjugal transfer. *Environ. Pollut.* 359(124592):124592
- 988 158. Xiong J, Fischer WM, Inoue K, Nakahara M, Bauer CE. 2000. Molecular evidence for the early
989 evolution of photosynthesis. *Science*. 289(5485):1724–30
- 990 159. Yancey CE, Kiledal EA, Chaganti SR, Deneff VJ, Errera RM, et al. 2023. The Western Lake Erie culture
991 collection: A promising resource for evaluating the physiological and genetic diversity of *Microcystis*
992 and its associated microbiome. *Harmful Algae*. 126:102440
- 993 160. Zang X, Liu B, Liu S, Arunakumara KKIU, Zhang X. 2007. Optimum conditions for transformation of
994 *Synechocystis* sp. PCC 6803. *J. Microbiol.* 45(3):241–45
- 995 161. Zborowsky S, Lindell D. 2019. Resistance in marine cyanobacteria differs against specialist and
996 generalist cyanophages. *Proc. Natl. Acad. Sci. U. S. A.* 116(34):16899–908
- 997 162. Zehr JP, Capone DG. 2020. Changing perspectives in marine nitrogen fixation. *Science*. 368(6492):
- 998 163. Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, et al. 2019. Adaptive Evolution within
999 Gut Microbiomes of Healthy People. *Cell Host Microbe*. 25(5):656–67.e8
- 1000 164. Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT. 2006. Phylogenetic analyses of
1001 cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res*.
1002 16(9):1099–1108
- 1003 165. Zhou K, Wong TY, Long L, Anantharaman K, Zhang W, et al. 2023. Genomic and transcriptomic
1004 insights into complex virus-prokaryote interactions in marine biofilms. *ISME J.* 17(12):2303–12
- 1005 166. Zhu J, Yang F, Du K, Wei Z-L, Wu Q-F, et al. 2023. Phylogenomics of five *Pseudanabaena*
1006 cyanophages and evolutionary traces of horizontal gene transfer. *Environ Microbiome*. 18(1):3

1007

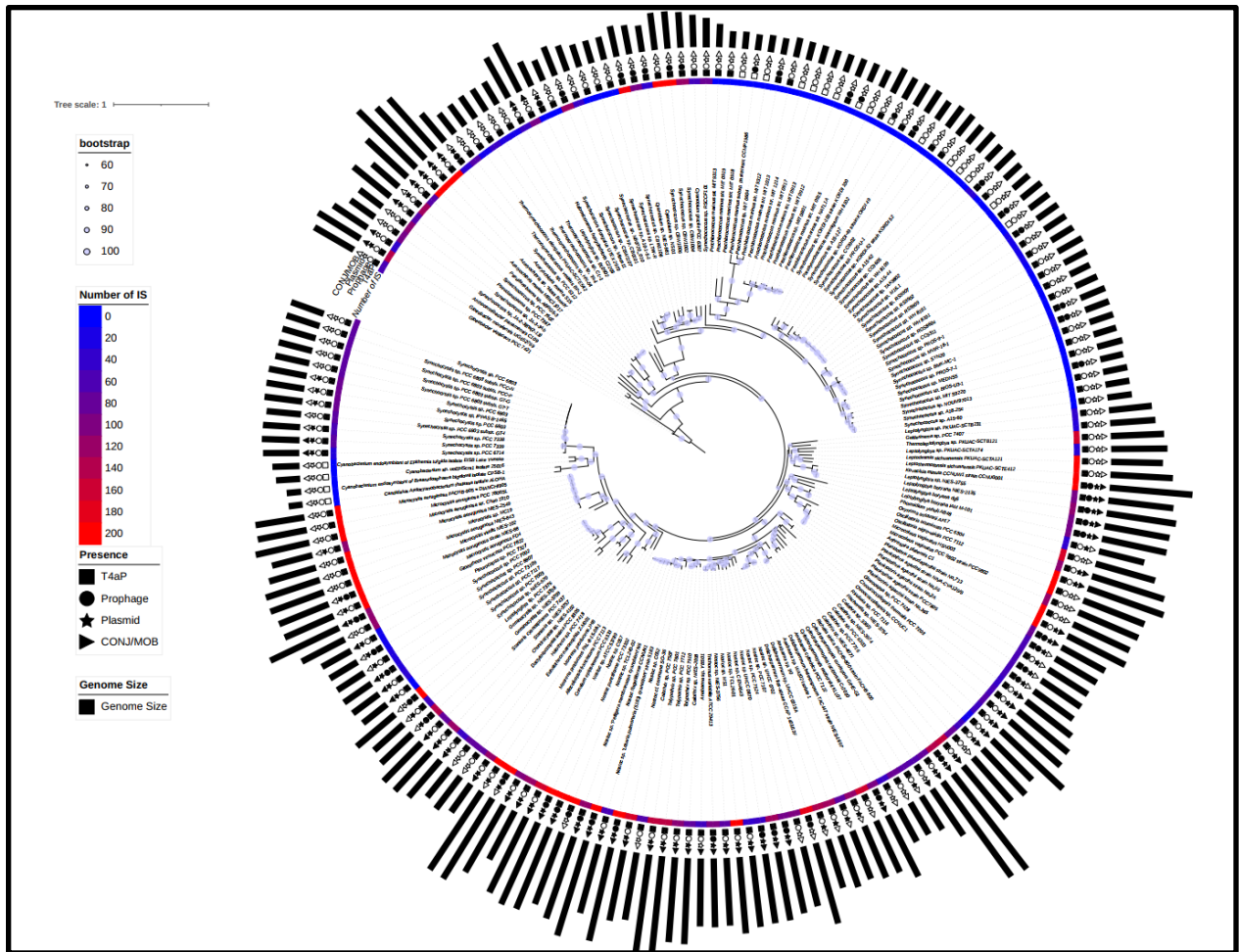


FIGURE 1. Features associated with genome size and gene flow identified in 193 complete genomes of Cyanobacteriota.

Distribution of insertion sequences (ISs, identified with ISEscan (12), T4P for natural transformation, identified with the module TFFscan of MacSyFinder v2 (10), prophages (identified with Genomad) (3), plasmids (identified as such in GenBank files), and proteins involved in conjugation (conjugative system and/or relaxase, identified with the module CONJscan of MacSyFinder v2) (4) in 193 genomes of Cyanobacteriota retrieved from the NCBI RefSeq in March 2023. All software was used with default parameters. The outer circle represents the genome size. In the inner part of the figure is represented the phylogenetic tree of the phylum (with bootstrap values for assessment of robustness of the tree). The tree was built using 126 almost universal protein profiles from PFAM (9), a subset of those given in (11), to identify the best hit in each genome per profile using hmmer v3 .3.2 with the -cut_ga option (5)(as in (2)). The protein families were aligned with MAFFT v7.407 (6)and concatenated to infer a phylogenetic tree with IQ-tree 1.6.12 (8), using the best model according to BIC (LG+I+G4) and 1000 replicates to obtain the ultra-fast bootstraps. Visualization was done with (7).

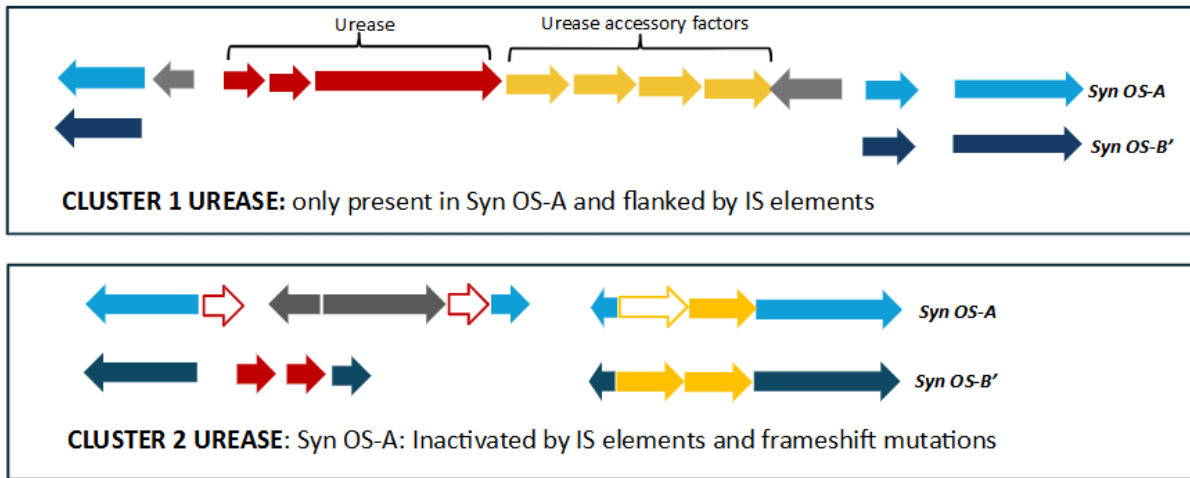


FIGURE 2. Comparative analysis of urease operons in two thermophiles, *Synechococcus* OS-A and OS-B' illustrating transposition events.

UPPER PANEL: *Top row* shows Cluster 1 urease in *Syn OS-A* (red and yellow arrows: urease and accessory factor encoding genes). Urease module is flanked by IS elements (gray arrows), possibly indicative of an HGT event. *Bottom row* homologous region in *Syn OS-B'*, to identify position where the urease cluster entered *Syn OS-A* by transposition.

LOWER PANEL: *Top row* shows Cluster 2 urease in *Syn OS-A* (red and yellow arrows: urease and accessory factor encoding genes; outlined arrows indicate frameshift mutations). *Bottom row* homologous region in *Syn OS-B'*, all genes appear intact.

Light blue and dark blue arrows represent homologs in *Syn OS-A* and *Syn OS-B'* respectively., in both panels. Modified and redone from Figure in Bhaya et al (1) which has additional details about gene designations and functions.





	 THERMAL SPRINGS	 FRESHWATER LAKES	 OCEANS	 DESERT CRUSTS
KEY FEATURES	Thermal alkaline springs with phototrophic biofilms; worldwide occurrence; limit of photosynthesis is ~70C.	Large or small water freshwater; periodic toxic cyanobacterial blooms are a worldwide and growing problem	Saltwater covers ~70% of the Earth's surface; host to complex ecosystems; ~25% net ocean productivity from cyanobacteria.	Biological soil crusts cover and stabilize deserts; dominated by cyanobacteria, also contain lichens, algae, fungi and bacteria.
CYANOBACTERIA	<i>Synechococcus</i> sp. <i>Pseudoanabaena</i> sp. <i>Gloeomargarita</i> sp. <i>Oscillatoria</i> sp.	<i>Microcystis</i> sp; <i>Dolichospermum</i> sp. <i>Nodularia</i> sp. and several others	<i>Prochlorococcus</i> sp. <i>Synechococcus</i> sp. <i>Trichodesmium</i> sp. <i>A. thalassa</i> (symbiont)	<i>Microcoleus</i> sp.
DATA TYPES	Many isolate genomes from thermal springs , metagenomes, MAGs, SAGs, viromes	Isolate genomes from many locations, MAGs, Cyanophages	Several isolate genomes from various locations, herbaria, SAGs, MAGs, cyanophages	Several <i>Microcoleus</i> sp isolate genome from any locations
WEBSITES	https://www.nps.gov/yell/learn/nature/thermophilic-bacteria.htm	https://www.epa.gov/habs	https://www.usgs.gov/programs/cmhrp/news/top-10-things-you-didnt-know-about-ocean	https://www.nps.gov/articles/seug-soil-crust.htm

FIGURE 3. Four diverse ecosystems where cyanobacteria are important community members:

Photos show examples of Thermal springs, Freshwater lakes, Oceans and Desert crusts located in the U.S. Table shows key features, major cyanobacterial genera, data types available and popular educational websites. *Photographs*: Thermal spring in Yellowstone National Park, WY USA; Ocean shore, Santa Barbara, USA; Desert crusts New Mexico USA, courtesy Devaki Bhaya; Satellite image of Pyramid lake with cyanobacterial bloom from NASA website (2024).

REFERENCES

1. Bhaya D, Grossman AR, Steunou A-S, Khuri N, Cohan FM, et al. 2007. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J.* 1(8):703–13
2. Bongiovanni TR, Latario CJ, Le Cras Y, Trus E, Robitaille S, et al. 2024. Assembly of a unique membrane complex in type VI secretion systems of Bacteroidota. *Nat. Commun.* 15(1):429
3. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, et al. 2024. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* 42(8):1303–12
4. Cury J, Abby SS, Doppelt-Azeroual O, Néron B, Rocha EPC. 2020. Identifying conjugative plasmids and integrative conjugative elements with CONJscan. *Methods Mol. Biol.* 2075:265–83
5. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput. Biol.* 7(10):e1002195
6. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30(4):772–80
7. Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 52(W1):W78–82
8. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37(5):1530–34
9. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, et al. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49(D1):D412–19
10. Néron B, Denise R, Coluzzi C, Touchon M, Rocha EPC, Abby SS. 2023. MacSyFinder v2: Improved modelling and search engine to identify molecular systems in genomes. *Peer Community J.* 3(e28):
11. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, et al. 2013. Insights into the

phylogeny and coding potential of microbial dark matter. *Nature*. 499(7459):431–37

12. Xie Z, Tang H. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics*. 33(21):3340–47