The order of trait emergence in the evolution of cyanobacterial multicellularity

Katrin Hammerschmidt[1,2*], Giddy Landan[1,2], Fernando Domingues Kümmel Tria[1,3], Jaime Alcorta[4], Tal Dagan[1]

[1]Genomic Microbiology Group, Institute of Microbiology, Kiel University, Kiel, Germany.

[2]These authors contributed equally.

[3]Present address: Institute for Molecular Evolution, Heinrich Heine University Düsseldorf, Düsseldorf, Germany.

[4]Department of Molecular Genetics and Microbiology, Biological Sciences Faculty, Pontifical Catholic University of Chile, Santiago, Chile.

[*]Author for Correspondence: Katrin Hammerschmidt, Genomic Microbiology Group, Institute of General Microbiology, Kiel University, Kiel, Germany, Phone: + 49 431 880 5743, Fax: + 49 431 880 5747, khammerschmidt@ifam.uni-kiel.de

Abstract

The transition from unicellular to multicellular organisms is one of the most significant events in the history of life. Key to this process is the emergence of Darwinian individuality at the higher level: groups must become single entities capable of reproduction for selection to shape their evolution. Evolutionary transitions in individuality are characterized by cooperation between the lower level entities and by division of labor. Theory suggests that division of labor may drive the transition to multicellularity by eliminating the trade-off between two incompatible processes that cannot be performed simultaneously in one cell. Here we examine the evolution of the most ancient multicellular transition known today, that of cyanobacteria, where we reconstruct the sequence of ecological and phenotypic trait evolution. Our results show that the prime driver of multicellularity in cyanobacteria was the expansion in metabolic capacity offered by nitrogen fixation, which was accompanied by the emergence of the filamentous morphology and succeeded by a reproductive life cycle. This was followed by the progression of multicellularity into higher complexity in the form of differentiated cells and patterned multicellularity.

Key words: N_2 fixation, division of labor, filament, complexity, transition in individuality

Significance Statement

The emergence of multicellularity is a major evolutionary transition. The oldest transition, that of cyanobacteria, happened more than 3 to 3.5 billion years ago. We find N_2 fixation to be the prime driver of multicellularity in cyanobacteria. This innovation faced the challenge of incompatible metabolic processes since the N_2 fixing enzyme (nitrogenase) is sensitive to oxygen, which is abundantly found in cyanobacteria cells performing photosynthesis. At the same time, N_2-fixation conferred an adaptive benefit to the filamentous morphology as cells could divide their labour into performing either N_2-fixation or photosynthesis. This was

46 followed by the culmination of complex multicellularity in the form of differentiated cells and
47 patterned multicellularity.

48

49 **Introduction**

50 Multicellularity is considered a characteristic trait of eukaryotes, but has evolved
51 independently several times in diverse prokaryote taxa, including actinobacteria,
52 myxobacteria, and cyanobacteria (Bonner 1998). Bacterial multicellularity ranges from
53 transient associations, such as colonies, biofilms and cellular aggregations, to permanent
54 multicellular forms (Shapiro 1988)

55 Instances of multicellular bacterial species present the major traits of eukaryotic
56 multicellularity, including cell-to-cell adhesion, peri- or cytoplasmic continuity, intercellular
57 communication, patterning, programmed cell death (PCD), and division of labor (Claessen et
58 al. 2014). Aggregative forms of multicellularity are common among bacterial species, for
59 example, those that form a biofilm under specific external conditions (Tarnita et al. 2013).
60 *Bacillus subtilis*, for instance, forms biofilms upon nutrient deprivation in which cells
61 differentiate into motile, matrix producing, or spore cells depending on the environmental
62 cues (Claessen et al. 2014). Notably, cell differentiation in aggregates is adaptive at the level
63 of the individual cell as it directly confers a fitness benefit to that particular cell. In contrast,
64 under true division of labor, cells are interdependent upon each other and specialize in
65 performing complementary tasks. These tasks, e.g., somatic functions or PCD, are not
66 beneficial on the level of the individual cell, but are advantageous for the colony; thus, they
67 are emergent properties on a higher level of organization (van Gestel et al. 2015).

68 True division of labor in bacteria is best described in actinobacteria and cyanobacteria
69 (van Gestel et al. 2015). In cyanobacteria, the most complex of the filamentous species can
70 differentiate up to five different cell types: vegetative (photosynthetic) cells, akinetes (spore-
71 like cells), hormogonia (reproductive, motile filaments), necridia (dead cells resulting from

72 PCD/ apoptosis for hormogonia release), and heterocysts (Claessen et al. 2014; Herrero et
73 al. 2016). Heterocysts differentiate under nitrogen deprivation and are specialized in nitrogen
74 (N_2) fixation by the enzyme nitrogenase (Frías et al. 1994). As this enzyme is sensitive to
75 oxygen (O_2), these cells are characterized by the absence of oxygenic photosynthesis and by
76 a thick cell wall, which maintains an anaerobic environment. Heterocysts and vegetative cells
77 in the filament are metabolically interdependent with the heterocysts providing combined
78 nitrogen to the other cells within the filament and receiving fixed carbon compounds in return.
79 Heterocysts cannot reproduce hence they represent a prime example for emergent traits on
80 the level of a multicellular organism.

81 Cyanobacteria possess the hallmark traits reminiscent of complex eukaryotic
82 multicellularity, making the order of trait emergence essential for understanding the origin of
83 higher-level complexity in organismal evolution. Here we infer the evolutionary trajectory of
84 the emergence of traits in the evolution of multicellularity in cyanobacteria.

85

86 **Materials and Methods**

87 Data

88 The primary data underlying this study consists of the genomic sequences and phenotypic
89 traits of 199 representative cyanobacterial species. These were selected from the available
90 genomes so that the number of represented taxa will be as large as possible and genus-level
91 redundancy will be reduced (see supplementary table S1 for the complete list of species).

92

93 **Table 1**
94 Description of cyanobacterial cell types, morphological and physiological traits, their habitat
95 and life style.

Cell types	
Vegetative cells	Photosynthetic cells.
HORMOGONIA*	Motile reproductive filaments that result from repeated rounds of fission without intermittent growth phases. They break off the mother filament, ensuring the reproduction and dispersal of benthic species.
HETEROCYSTS*	Thick-walled cells that are specialized in fixing N_2.
AKINETES*	Thick-walled, spore-like cells that provide reproduction, dormancy, and resilience.
BAEOCYTES	Reproductive cells that result from repeated rounds of fission without intermittent growth phases.
Morphological and physiological traits	
UNICELLULAR	Single-celled morphology. After cell division cells separate.
FILAMENTOUS*	Multi-celled morphology. Cells remain attached after cell division.
NO N FIXATION N FIXATION	No Fixation of N_2 into ammonia, in contrast to: Fixation of N_2 into ammonia.
SHEATH	Part of the cell envelope, located outside the cell wall.
MUCILAGE	Part of the envelope, located outside the cell wall, comprised of extracellular polymeric substances (EPS), without a defined structure.
GAS VESICLES*	Intracellular gas-filled chambers for regulating buoyancy in the water column.
MOTILITY	Movement across surfaces or through a liquid medium.
MULTIPLANE FISSION	Cell division in two or three perpendicular planes.
TRUE BRANCHING*	Fission in multiple planes leads to branching filaments that remain attached to the main filament.
Habitat and life style	
NON MARINE	Aquatic environments with salinity between 0-0.5ppt, and terrestrial habitats.
MARINE	Environments with salinity between 30–50ppt.
THERMOPHILIC	Optimal growth temperature above 45°.
MICROBIAL MATS	Growth inside thick, laminated, microbial structures.
UNATTACHED	Organism that lives unattached to a substrate (planktonic), in contrast to:
ATTACHED	Organism that lives attached to a substrate (sessile/benthic).
EPIPHYTIC	Growth on plants.

***** Multicellularity markers: traits that are adaptations on the level of the filament.
Small caps indicate the traits that have been used in the analysis.

96

97 Phenotypic traits

98 Phenotypic traits were chosen for their potential relevance to the evolution of multicellularity
99 in cyanobacteria, such as environmental factors that might facilitate multicellularity and
100 markers that are indicative for the transition to multicellularity (table 1).

101 Information on presence and absence of traits was obtained from the published
102 literature and from the Pasteur Culture Collection of cyanobacteria, extending the work by
103 Uyeda et al. 2016, and coded as binary trait states. Traits included morphology (unicellular,
104 filamentous), nitrogen fixation (no N_2 fixation, N_2 fixation), habitat (marine/ non marine),
105 baeocytes, hormogonia, thermophilic, akinetes, heterocysts, true branching, epiphytic,
106 microbial mats, attached/ unattached, sheath, mucilage, gas vesicles, motility, and multiplane
107 fission (table 1, supplementary table S1).

108

109 Protein families and alignments

110 The cyanobacteria protein families were constructed from completely sequenced genomes
111 available in RefSeq database (O'Leary et al. 2016; ver. May 2016). For the construction of
112 protein families, at the first stage, all protein sequences annotated in the genomes were
113 blasted all-against-all using stand-alone BLAST (Altschul et al. 1990) ver. 2.2.26. Protein
114 sequence pairs that were found as reciprocal best BLAST hits (rBBHs) (Tatusov et al. 1997)
115 with a threshold of E-value ≤ $1x10^{-5}$ were further compared by global alignment using needle
116 (Rice et al. 2000). Sequence pairs having ≥30% identical amino acids were clustered into
117 protein families using the Markov clustering algorithm (MCL) (Enright et al. 2002) ver. 12-135
118 with the default parameters. Multiple-copy gene families were discarded, resulting in an initial
119 dataset of 18,873 single-copy gene families.

120 Gene families were then extended to include homologous sequences from non-
121 cyanobacteria species, serving as outgroups for rooting purposes. We identified outgroup
122 homologues by an rBBH analysis of the *Scytonema hofmanni* PCC 7110 genome (the most
123 widely present species in the initial gene family dataset) against 26 high quality non-
124 cyanobacteria genomes: Vampirovibrionia (12 genomes) and Sericytochromatia (2) (Soo et

125 al. 2017; Carnevali et al. 2019), the closest phyla Margulisbacteria (6), Saganbacteria (2),
126 Fusobacteria (1) and Firmicutes (1) (according to (Carnevali et al. 2019) and (Zhu et al.
127 2019); one reference anoxygenic photosynthetic genome from *Chloroflexus aurantiacus* J-
128 10-fl and the *Escherichia coli* str. K-12 substr. MG1655 genome (see supplementary table
129 S1). The number of gene families with homologs ranged between 204 and 451 for the 26
130 outgroup genomes. We selected six of these outgroups for further analyses: *Vampirovibrio*
131 *chlorellavorus, Chloroflexales, Obscuri-PALSA-1081, Sericytochromatia-UBA7694, Bacillus*
132 *subtilis,* and *Margulis-GWF2-35-9.* Protein sequences of these families were aligned using
133 MAFFT version 7.027b employing the L-INS-i strategy (Katoh & Standley 2013). The
134 alignments are available in supplementary material online.
135
136 Species tree reconstruction
137 The sequence data for the reconstruction of the cyanobacterial species tree consisted of 14
138 single-copy gene families that are present in all 199 cyanobacteria genomes and any of the
139 six outgroup genomes. The species tree was inferred using IQ-TREE (Nguyen et al. 2015) in
140 a partitioned analysis over the concatenated alignment of the 14 gene alignments (iqtree
141 version 1.6.6.b with parameters -t BIONJ -keep-ident -mset LG -madd LG4X –spp). The
142 unrooted species tree was rooted on the branch leading to the outgroup. The species tree is
143 available in supplementary material online.
144

145 Gene trees reconstruction

146 To evaluate the robustness of inferences drawn from the species tree, we also reconstructed
147 gene trees to provide a large sample of comparisons to the species tree. The gene trees
148 dataset consisted of 553 single-copy gene families present in at least one genome from both
149 sides of the root of the species tree, and at least one of the six outgroup species. Gene trees
150 were inferred using IQ-TREE (Nguyen et al. 2015) version 1.6.6.b with parameters -t BIONJ -
151 keep-ident -mset LG -madd LG4X). Trees were rooted on the branch separating the
152 outgroup from the ingroup. A total of 138 trees where the outgroup sequences did not form a
153 single partition were discarded. The gene trees are available in supplementary material
154 online.

155

156 Inference of trait order

157 The traits presence/absence pattern was mapped onto the rooted species tree. Most of the
158 traits in our study are rather complex (i.e., they involved multiple genes), hence their
159 emergence in the evolution of cyanobacteria is expected to be a rare event. Accordingly, we
160 used a parsimonious reconstruction approach and assigned the origin of a trait to the most
161 recent species tree node where the trait is present in any of the node's descendants.
162 Consistency index (CI) and retention index (RI) for each trait were calculated using the
163 PHYLIP program pars (Felsenstein 2005) as described in Farris (1989) The species tree was
164 traversed from root to tips to determine the order of trait emergence. For each pair of traits
165 we tested whether the order observed in the species tree was reproduced in gene trees with
166 at least two species displaying each of the two traits. For that purpose, we repeated the trait-
167 order analysis with the set of single-copy rooted gene trees, including gene families that do
168 not span the full taxa set. The agreement of the gene trees with the conclusion based on the
169 species trees is calculated as the proportion of gene trees where the trait order is the same
170 as in the species tree.

171

Results and Discussion

To reconstruct the order of trait emergence in the evolution of cyanobacterial multicellularity, we evaluated 21 phenotypic traits variably present in 199 cyanobacterial species (table 1, supplementary table S1). The ability to perform photosynthesis is not included in our study as it is universal to all cyanobacteria hence it is considered an ancestral trait (Garcia Pichel et al. 2020). We inferred a species tree from a partitioned analysis of 14 single copy core genes that are present in all 199 species. The species tree was rooted by inclusion of outgroup sequences from six bacterial species (see methods), with the root separating the genus *Gloeobacter* from all other cyanobacteria.

The rooting of cyanobacteria is a thorny issue, with two competing clades put forward as basal lineages. Phylogenetic studies based on single gene phylogenies indicated *Gloeobacter* to be a basal lineage within cyanobacteria (e.g., Bhattacharya & Medlin 1995; Honda et al. 1999). The *Gloeobacter* basal position is in agreement with the cellular characteristics of species in this genus, which are lacking well-defined thylakoids (Rippka et al. 1974; Mareš et al. 2019) that are considered a derived structure in the phylum. Indeed, several studies in the literature thus used *Gloeobacter* for rooting the cyanobacterial species tree (e.g., Shih et al. 2012; Shi & Falkowski 2008; Dagan et al. 2013; Sánchez-Baracaldo et al. 2014). Nonetheless, other studies, which used midpoint rooting or minimal ancestor deviation (MAD) to root the cyanobacteria species tree, position the root on the branch separating between the pico-cyanobacteria (*Synechococcus* & *Prochlorococcus,* and hereafter SynPro clade) and the remaining species (Szöllősi et al. 2012; Tria et al. 2017). This branch, however, is typically long in gene trees, as well as the species trees, hence it may reflect a Long Branch Attraction (LBA) artifact (defined in Felsenstein 1978).

To evaluate whether the rooting placement is robust, we conducted gene-tree support analyses by reconstructing trees for 553 cyanobacterial single-copy gene families along with homologs from six different outgroup species (fig. 1, and methods). We then extracted the rooted cyanobacterial subtree, while discarding 138 trees where the outgroup species

199 formed more than one group. We next characterize the trees by the pattern of the three
200 cyanobacterial subgroups: *Gloeobacter* (2 spp.), SynPro clade (32 spp.), and Other
201 Cyanobacteria (165 spp.), discarding 142 genes that are present in less than two genomes
202 for each group. First, we considered gene trees in which no group is present on both sides of
203 the root. The group appearing on its own as a lineage originating at the root is considered a
204 basal group and the tree is supporting a root located on the branch separating it from the
205 others (left column in fig. 1). Next, if only one group appears on both sides of the root, this
206 group is labeled as ancestral, and supports a root position within it (right column in fig. 1). We
207 found no gene trees where more than one group appears on both sides of the root. We note
208 that the gene trees may be discordant with the species tree, and that any of the groups may
209 seem to be paraphyletic, either due to methodological artifacts (e.g. LBA involving the
210 SynPro clade) or due to biological processes such as lateral gene transfer (LGT).

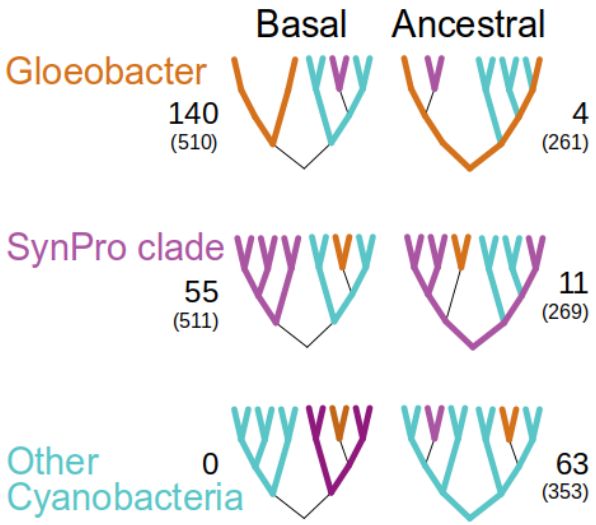

Figure 1. Support for three possible basal/ancestral cyanobacterial groups in 273 rooted gene trees. The number of gene families supporting each type of rooted topology is given (median alignment length is shown in parenthesis).

211

212 Our results reveal that the majority of gene trees identify the *Gloeobacter* lineage as a
213 distinct basal lineage stemming from the root, thus supporting the outgroup rooting of the
214 species tree on that branch.

215

216 **The order of trait emergence in cyanobacterial evolution**

217 We infer the order of trait emergence in cyanobacterial evolution by mapping the traits onto
218 the rooted species tree. The *origin node* corresponding to each trait was assigned as the
219 most recent node where the trait is present in any of the node's descendants. This is a
220 conservative approach in that for each trait it allows for a single origin with possible
221 subsequent losses (Dagan & Martin 2007). Our approach furthermore assumes that the traits
222 in our analysis are vertically inherited rather than acquired by lateral transfer. Indeed,
223 previous phylogenomic studies indicated that gene acquisition by lateral gene transfer is
224 frequent in the evolution cyanobacteria (Zhaxybayeva et al. 2009; Dagan et al. 2013) The
225 traits we included in our analysis, however, are complex phenotypic traits that are the
226 product of multiple genes whose expression is well coordinated with other physiological
227 processes in the cell. Such traits are unlikely to be acquired by lateral gene transfer, because
228 their retention requires that multiple coding and regulation elements will be functional on
229 arrival (Cohen et al. 2011). Nonetheless, since complex systems may have been sporadically
230 and rarely transferred (e.g., Brinkmann et al. 2018), we quantified the level of homoplasy and
231 synapomorphy in the evolution of the studied traits by the consistency and retention indices
232 (Farris 1989). Our results for most of the traits revealed high RI values (median 0.61) and low
233 CI values (median 0.048) (supplementary table S1). This pattern is as expected for complex
234 traits whose evolution is best characterized by a single origin and differential loses.

235 We next compare the origin nodes for pairs of traits. When the origin of two traits is
236 assigned to the same node, we label the two traits as 'simultaneous' at the resolution of the
237 current taxa sample. When an origin node of a trait is a descendent of a second trait's origin
238 node, we conclude that the first trait emerged earlier. A third possibility is that the origin

239 nodes of two traits are not nested, but this relationship was not observed in any of the 210
240 trait pairs. The order of trait emergence is depicted in fig. 2.

241 The order of trait emergence is inferred from a species tree, yet, phylogenetic
242 inference of species phylogenies based on the concatenation of only few core genes may
243 suffer from a low resolution of the phylogenetic signal (e.g., Dagan & Martin 2006; Thiergart
244 et al. 2014). To evaluate the robustness of the trait order derived from the species tree, we
245 repeated the analysis by testing the trait order inference in individual gene trees. For that
246 purpose, we considered the set of single-copy gene families where the gene is present in at
247 least one species from both sides of the species tree root and at least one outgroup species.
248 The gene trees were rooted by the outgroup, and the order of pairs of traits determined. The
249 analysis of a large sample of gene trees provides a statistical view not possible in a single
250 species tree, but potentially introduces contradicting inferences due to both biological
251 reasons (LGT) and methodological uncertainties (alignment quality and phylogenetic artifacts
252 such as LBA). In fig. 2 we report the percentage of gene trees that reproduces the species
253 tree ordering. The vast majority of trait pair orderings are observed also in more than 50% of
254 the individual gene trees. Excluding the sequences from the SynPro clade from the gene
255 trees led to slightly better agreement of the gene trees with the order observed in the species
256 tree (supplementary fig. S1), suggesting that LBA artifacts associated with the SynPro clade
257 may lead to disagreement between the gene and species trees. We note that gene trees that
258 are discordant with the species tree generally had shorter alignment length and lower
259 bootstrap support (supplementary tables S2 and S3). Our results thus show a high level of
260 agreement between the species tree and the gene trees, as is expected under an overall
261 homogeneous yet low frequency of LGT during bacterial evolution (Dagan & Martin 2007).

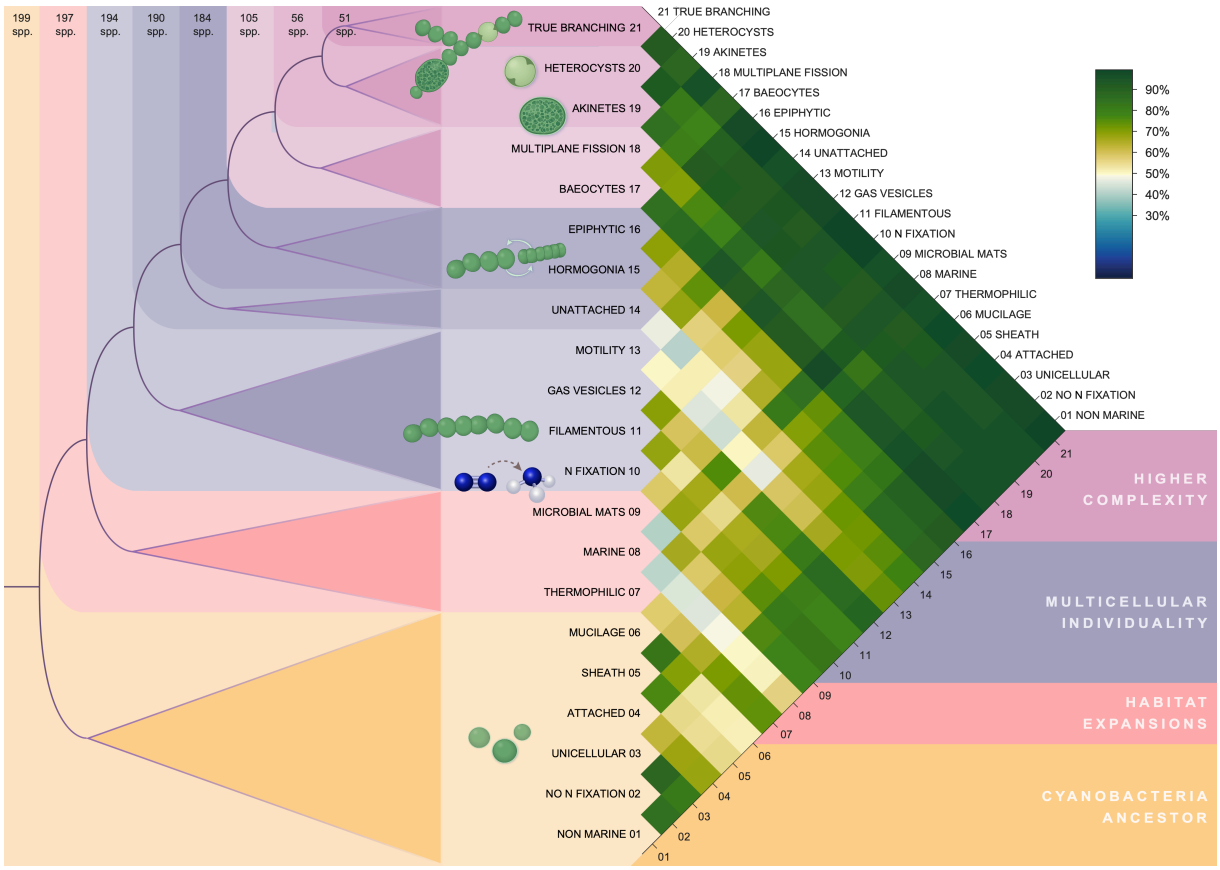

Figure 2. Order of trait emergence. Left: traits and their inferred origin node from the rooted species tree. Colors mark traits with a common origin node (note that the order of traits within the colored blocks is arbitrary). Colored boxes are nested, i.e., earlier traits are present also in the nested colors. Right: Frequency of gene trees in agreement with the relative order of pairs of traits. Cells in the matrix are shaded according to the color bar on the right.

In what follows we divide the inferred order of trait emergence into four temporal phases: (phase i) the cyanobacteria ancestor (traits 1-6); (phase ii) habitat expansions (traits 7-9); (phase iii) the transition to multicellular individuality (traits 10-16); and (phase iv) the evolution of higher complexity (traits 17-21).

268 **The cyanobacterial ancestor and subsequent habitat expansion**

269 The rooted tree topology supports the view that the cyanobacterial ancestor was
270 characterized by traits that include UNICELLULAR and NO N-FIXATION (fig. 2). The ancestral
271 state of both traits as preceding the emergence of filamentous forms is debated in the
272 literature - whereas one study suggested the ancestor to be unicellular and the filamentous
273 morphology to arise in independent lineages of the cyanobacterial tree (Sánchez-Baracaldo
274 et al. 2005), another view posed that the filamentous morphology evolved early during
275 cyanobacterial evolution and was subsequently lost and regained several times
276 (Schirrmeister et al. 2011). There are also claims that the last cyanobacterial common
277 ancestor already fixed N_2 (Tomitani et al. 2006), there are however others that concluded that
278 it could not fix N_2 and that cyanobacteria must have acquired this trait several times
279 independently (Sánchez-Baracaldo et al. 2005).

280 We can further deduce from our data that the cyanobacterial ancestor lived ATTACHED
281 and possessed a SHEATH and MUCILAGE. Whether the cyanobacteria ancestor lived
282 ATTACHED is a matter of debate and opposing views on the topic have been published
283 (Sánchez-Baracaldo et al. 2005; Uyeda et al. 2016; Schopf 1993; Garcia-Pichel 1998;
284 Sánchez-Baracaldo 2015)

285 SHEATH and MUCILAGE are forms of extracellular polymeric substances (EPS), located
286 outside the cell wall, which in today's cyanobacteria are mainly involved in protecting the cell
287 from various stresses, such as UV and desiccation (Ehling-Schulz & Scherer 1999; Potts
288 1994). Furthermore, we find the cyanobacteria ancestor to have most likely inhabited a NON
289 MARINE environment, agreeing with studies that suggest that early cyanobacteria lived in
290 freshwater or terrestrial habitats and subsequently diverged into marine environments
291 (Dagan et al. 2013; Uyeda et al. 2016).

292 The second phase in cyanobacterial evolution is the expansion of the cyanobacterial
293 habitat, indicated by the traits MARINE and THERMOPHILIC, which are both inferred to
294 simultaneously occur with the ability to form MICROBIAL MATS. MICROBIAL MATS are dense
295 communities (Stal 1995) that typically present a laminated segregation of functional types.

296 They are often formed by cyanobacteria and are frequently found in extreme habitats, such
297 as deserts or hot springs, characterized by temperatures between 30°C to 73°C (e.g., (Cox
298 et al. 2011)).

299

300 **The emergence of N_2 fixation is at the origin of cyanobacterial multicellularity**

301 Phase iii in our reconstruction comprises three sets of cyanobacterial traits (fig. 2). First, the
302 simultaneous emergence of the FILAMENTOUS morphology, N FIXATION, GAS VESICLES and
303 MOTILITY, followed by the trait UNATTACHED, and lastly by the co-occurrence of HORMOGONIA
304 and EPIPHYTIC. During cyanobacterial N_2 fixation, molecular dinitrogen (N_2) is reduced to
305 ammonia (NH_3), a process that is catalyzed by the enzyme nitrogenase. Whereas present
306 day cyanobacteria, other microorganisms, and most plants are able to take up nitrogen in
307 various combined forms, such as nitrate, ammonium, organic nitrogen, or urea, these
308 combined forms of nitrogen are scarce in most environments (e.g., open oceans or terrestrial
309 habitats (Zehr 2011)). Combined nitrogen, which is critical for the biosynthesis of amino and
310 nucleic acids, was likely a limiting resource in the early Earth environment (Kasting & Siefert
311 2001).

312 The realization of the full metabolic potential of N_2 fixation, however, faced the
313 challenge of the incompatibility of nitrogenase with intracellular oxygen (Gallon 1981). When
314 the cyanobacterial ancestor first acquired the capacity of N_2 fixation, it must have imposed a
315 strong selection pressure on the individual cells. The trade-off between photosynthesis and
316 nitrogen fixation led to the evolution of multiple solutions, which are still present in today's
317 cyanobacteria: the circadian rhythm of N_2 fixation in unicellular cyanobacteria (Mitsui et al.
318 1986), specific cells devoted to N_2 fixation in undifferentiated filaments (Berman-Frank et al.
319 2003; Bergman et al. 2013), and the differentiation of the highly specialized heterocyst in
320 filamentous cyanobacteria (Flores et al. 2018).

321 The need to compartmentalize the two incompatible functions, photosynthesis and N_2
322 fixation, has been proposed to drive the emergence of multicellular forms in cyanobacteria
323 (Ispolatov et al. 2012). More specifically, the result from a quantitative theoretical model
324 predicts that within a population of genetically identical unicellular nitrogen fixing
325 cyanobacteria, cell differentiation and phenotypic heterogeneity would have been adaptive if
326 this increased the fitness of the organisms in multicellular groups. In the case of unicellular
327 cyanobacteria this means that cells evolved adhesion and exchanged fixed nitrogen and
328 carbon products within early cell groups, such as filaments.

329 In filamentous cyanobacteria, dividing cells remain linked in a chain, resulting in a
330 localization of cells in close spatial proximity, facilitating metabolite exchange between the
331 individual cells. When compared to the more transient associations in spatially structured
332 communities, such as in EPS imbedded biofilms, the development of filaments opens
333 possibilities for a more direct and permanent exchange of molecules between neighboring
334 cells with high specificity. Metabolic exchange could have evolved as described for the
335 evolution of metabolic cross-feeding (D'Souza et al. 2018), as the exchange of carbon and
336 nitrogen against other products is generally common in photosynthetic or nitrogen-fixing
337 organisms (Kaiser et al. 2015).

338 The emergence of GAS VESICLES and MOTILITY traits signify the evolution from a
339 stationary to a more active lifestyle, enabling cells to regulate their buoyancy in the water
340 column. This result is further supported by the subsequent inference of UNATTACHED, which
341 indicates the transition from a benthic to a planktonic lifestyle.

342 Thereafter the traits HORMOGONIA and EPIPHYTIC are inferred to occur simultaneously.
343 The differentiation of HORMOGONIA can be induced by environmental stimuli, such as nitrogen
344 deprivation (Flores & Herrero 2010). HORMOGONIA are released from the mother filament
345 through the formation of necridia, dead cells resulting from PCD (Nürnberg et al. 2014). After
346 their release from the main trichome, HORMOGONIA disperse via gliding motility or float thanks
347 to GAS VESICLES, ensuring the reproduction of benthic species (Rippka et al. 1979).

348 HORMOGONIA with GAS VESICLES are thus important for distribution in aquatic environments as
349 known from modern cyanobacteria (*Fischerella, Hapalosiphon, Tolypothrix*) (Komárek 2013).
350 The close local association with plants, as indicated by EPIPHYTIC, however might have been
351 the first step towards the initiation of one of the many symbioses between higher organisms
352 and cyanobacteria, where HORMOGONIA serve as the infection units (Meeks & Elhai 2002).

353 Notably, the differentiation into HORMOGONIA is reversible, as they develop a sessile
354 lifestyle, where they grow into a new vegetative filament (Flores & Herrero 2010). Here we
355 observe the emergence of a life cycle with two distinct cell types, which is important for the
356 transition to multicellularity (Hammerschmidt et al. 2014; Rose 2020). Such a life cycle
357 results in selection operating at the higher, the filament level and thus ensures the
358 reproduction of the newly formed collective entity.

359

360 **The evolution of cell differentiation leads to higher cyanobacterial complexity**

361 A central innovation that is associated with this phase (iv) in the species tree is MULTIPLANE
362 FISSION. This trait co-occurred with the ability to produce BAEOCYTES, differentiated cells,
363 which are the reproductive stages in the order Pleurocapsales (Waterbury & Stanier 1978).
364 Notably, baeocyte-forming cyanobacteria, that have been traditionally grouped together with
365 unicellular cyanobacteria (Rippka et al. 1979), appear to immediately predate the evolution of
366 spore-like AKINETES and nitrogen-fixing HETEROCYSTS and thus emerge much later than
367 filamentous forms. Indeed, a recent study suggested that *Gloeocapsopsis* sp., a baeocytous
368 cyanobacterium, harbors several characteristics that are in common with filamentous
369 cyanobacteria, including mechanisms of cell-cell communication (Urrejola et al. 2020).

370 The late timing and the co-occurrence of the two traits AKINETES and HETEROCYSTS,
371 indicative of higher complexity, are in line with the view that the evolution of the heterocyst
372 was relatively late in the history of filamentous cyanobacteria (Tomitani et al. 2006), and
373 where a common origin of akinetes and heterocysts has been proposed (Adams & Duggan

374 1999). HETEROCYSTS represent not only a morphological adaptation to the obstacle of N_2
375 fixation under oxic conditions but also an elaborate and highly specialized communication
376 and metabolite exchange system. In *Anabaena* sp., for example, where several hundred
377 cells communicate within a filament, a regular heterocyst formation pattern along the filament
378 must be achieved to guarantee that every cell is adequately supplied with fixed nitrogen
379 compounds (Herrero et al. 2016). For this, the inhibitory signaling peptide PatS needs to be
380 distributed along the filament with heterocyst formation occurring only in cells with low PatS
381 concentration (Yoon & Golden 1998). Whether the exchange of metabolites and regulators
382 happens via the continuous periplasm (Flores et al. 2006) or through septal junctions
383 (Mullineaux et al. 2008) is still not fully resolved.

384 The trait that evolved last, based on the analysis, is TRUE-BRANCHING, where cells in a
385 filament perform MULTIPLANE FISSION. TRUE-BRANCHING is characteristic for the members of
386 the Haphalosiphon/Stigonematales clade. Our results confirm previous morphological and
387 phylogenetic studies that found TRUE-BRANCHING to be the latest evolutionary innovation in
388 cyanobacteria (Rippka et al. 1979; Dagan et al. 2013; Koch et al. 2017).

389

390 **Trade offs between incompatible processes lead to division of labor and stable**
391 **multicellularity**

392 Common features of evolutionary transitions in individuality comprise cooperation between
393 the lower level units (Bonner 1998) and the division of labor (Michod 2007). The latter might
394 be of particular advantage, and serve as the driver of the transition to multicellularity when
395 there is a strong trade-off between processes that cannot be performed in a single cell at one
396 time (Michod 2007; Ispolatov et al. 2012). Our current findings support this theory and point
397 to nitrogen fixation, and its incompatibility with photosynthesis, as the trigger for the evolution
398 of multicellularity in cyanobacteria. One open question concerns how the underlying genetics
399 of novel traits, such as the division of labor, arise within a newly emerging multicellular
400 individual. In the case of cyanobacteria multicellularity, as also suggested for animal

401 multicellularity (Brunet & King 2017), we conjecture that no new genes were required and
402 that higher complexity was achieved by regulatory changes in gene expression patterns.
403 Basic communication and metabolite exchange was pre-existing as single-celled bacteria
404 frequently engage in cell-cell communication and cross-feeding of metabolites via the
405 external environment (D'Souza et al. 2018). Division of labor between photosynthesis and
406 nitrogen fixation was likely first established by the regulatory mechanism of temporal
407 switching. Once simple forms of division of labor and metabolic exchange existed, the
408 transition into spatial separation in differentiated cells could have evolved mainly by
409 regulatory modifications.

410 Differentiated cells are one of the hallmarks of complex multicellularity. It is therefore
411 significant that we observe six distinct cell types in cyanobacteria: photosynthetic,
412 hormogonia, necridia, akinetes, baeocytes, and heterocysts. Such a plurality indicates that
413 the underlying regulatory mechanisms are well developed and that their plasticity and
414 adaptability are a matter of course. It is also significant that three of the differentiated cell
415 types, hormogonia, akinetes, and baeocytes, offer novel reproductive potential and the
416 establishment of a multicellular life cycle. Moreover, signs of a nascent developmental plan
417 can be observed in both the distribution of heterocysts along filaments and in the patterning
418 of true branching cyanobacteria. These elements have no fitness value for the individual cell,
419 but are selectable adaptations on the higher level, the filament. The chronology of the
420 evolution of multicellularity in cyanobacteria shows that, once established, multicellular
421 individuality opens new vistas of opportunities.

Acknowledgements

We thank Tom Williams, Anne Kupczok, Tanita Wein, Claudia Taubenheim, Peter Deines, and Caroline Rose for comments on previous versions of the manuscript. We thank Fenna Stücker for illustrations. This work was supported by European Union's Framework Programme for Research and Innovation Horizon 2020 (2014–2020) under the Marie Skłodowska-Curie Grant Agreement No. 657096, by the German Academic Exchange Service (DAAD) to KH, by CAPES (Coordination for the Improvement of Higher Education Personnel–Brazil) to F.D.K.T., and by the Chilean National Agency for Research and Development (ANID) Doctoral scholarship N° 21191763 to J.A..

Author Contributions

K.H. and T.D. conceived the study. K.H. collected the traits data. G.L., F.D.K.T and J.A. performed the analyses. K.H., G.L., and T.D. wrote the manuscript with contributions from F.D.K.T. and J.A..

Declaration of Interests

The authors declare no competing interests.

Literature Cited

Adams DG, Duggan PS. 1999. Heterocyst and akinete differentiation in cyanobacteria. New Phytol. 144:3–33.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Bhattacharya D, Medlin L. 1995. The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. J Phycol. 31: 489–498.

Bergman B, Sandh G, Lin S, Carpenter, EJ. 2013. *Trichodesmium* - a widespread marine cyanobacterium with unusual nitrogen fixation properties. FEMS Microbiol Rev. 37:285–302.

Berman-Frank I, Lundgren P, Falkowski P. 2003. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. Res Microbiol. 154:157–164.

Bonner JT. 1998. The origins of multicellularity. Integr Biol. 1:27–36.

Brinkmann H, Göker M, Koblížek M, Wagner-Döbler I, Petersen J. 2018. Horizontal operon transfer, plasmids, and the evolution of photosynthesis in *Rhodobacteraceae*. ISME J. 8:a018127.

Brunet T, King N. 2017. The origin of animal multicellularity and cell differentiation. Dev Cell. 43:124–140.

Carnevali PBM et al. 2019. Hydrogen-based metabolism as an ancestral trait in lineages sibling to the cyanobacteria. Nat Commun. 10:1–15.

Claessen D, Rozen DE, Kuipers OP, Søgaard-Andersen L, van Wezel GP. 2014. Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. Nat Rev Microbiol. 12:115–124.

Cohen O, Gophna U, Pupko T. 2011. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. Mol Biol Evol. 28:1481–1489.

Cox A, Shock EL, Havig JR. 2011. The transition to microbial photosynthesis in hot spring ecosystems. Chem Geol. 280:344–351.

D'Souza G et al. 2018. Ecology and evolution of metabolic cross-feeding interactions in bacteria. Nat Prod Rep. 35:455–488.

Dagan T et al. 2013. Genomes of Stigonematalean cyanobacteria (Subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. Genome Biol Evol. 5:31–44.

Dagan T, Martin W. 2006. The tree of one percent. Genome Biol. 7:118.

Dagan T, Martin W. 2007. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. Proc Natl Acad Sci USA. 104:870–875.

Ehling-Schulz M, Scherer S. 1999. UV protection in cyanobacteria. Eur J Phycol. 34:329–338.

Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30:1575–1584.

477 Farris JS. 1989. The retention index and the rescaled consistency index. Cladistics. 5: 417-
478 419.

479 Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively
480 misleading. Syst Zool. 27:401–410.

481 Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) version 3.697. Distributed by
482 the author. Department of Genome Sciences. University of Washington, Seattle.

483 Flores E, Herrero A. 2010. Compartmentalized function through cell differentiation in
484 filamentous cyanobacteria. Nat Rev Microbiol. 8:39–50.

485 Flores E, Herrero A, Wolk CP, Maldener I. 2006. Is the periplasm continuous in filamentous
486 multicellular cyanobacteria? Trends Microbiol. 14:439–443.

487 Flores E, Picossi S, Valladares A, Herrero A. 2018. Transcriptional regulation of development
488 in heterocyst-forming cyanobacteria. Biochim Biophys Acta Gene Regul Mech. 1862: 673–
489 684.

490 Frías JE, Flores E, Herrero A. 1994. Requirement of the regulatory protein NtcA for the
491 expression of nitrogen assimilation and heterocyst development genes in the
492 cyanobacterium *Anabaena* sp. PCC7120. Mol Microbiol. 14:823–832.

493 Gallon JR. 1981. The oxygen sensitivity of nitrogenase: a problem for biochemists and micro-
494 organisms. Trends Biochem Sci. 6:19–23.

495 Garcia Pichel F, Zehr JP, Bhattacharya D, Pakrasi HB. 2020. What's in a name? The case of
496 cyanobacteria. J Phycol. 56:1–5.

497 Garcia-Pichel F. 1998. Solar ultraviolet and the evolutionary history of cyanobacteria. Origins
498 Life Evol B. 28:321–347.

499 Hammerschmidt K, Rose CJ, Kerr B, Rainey PB. 2014. Life cycles, fitness decoupling and
500 the evolution of multicellularity. Nature. 515:75–79.

501 Herrero A, Stavans J, Flores E. 2016. The multicellular nature of filamentous heterocyst-
502 forming cyanobacteria. FEMS Microbiol Rev. 40:831–854.

503 Honda D, Yokota A, Sugiyama J. 1999. Detection of seven major evolutionary lineages in
504 cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five
505 marine *Synechococcus* strains. J Mol Evol 48:723–739.

506 Ispolatov I, Ackermann M, Doebeli M. 2012. Division of labour and the evolution of
507 multicellularity. Proc Biol Sci. 279:1768–1776.

508 Kaiser D, Kowalski N, Böttcher ME, Yan B, Unger D. 2015. Benthic nutrient fluxes from
509 mangrove sediments of an anthropogenically impacted estuary in Southern China. J Mar Sci
510 Engineer. 3:466–491.

511 Kasting JF, Siefert JL. 2001. Biogeochemistry. The nitrogen fix. Nature. 412:26–27.

512 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
513 improvements in performance and usability. Mol Biol Evol. 30:772–780.

514 Koch R et al. 2017. Plasticity first: molecular signatures of a complex morphological trait in
515 filamentous cyanobacteria. BMC Evol Biol. 17:209.

516 Komárek J. 2013. *Süßwasserflora von Mitteleuropa, Bd. 19/3: Cyanoprokaryota; 3rd part:*
517 *Heterocystous Genera*. Springer Spektrum: Berlin.

518 Mareš J, Strunecký O, Bučinská L, Wiedermannová J. 2019. Evolutionary patterns of
519 thylakoid architecture in cyanobacteria. Front Microbiol. 10:277.

520 Meeks JC, Elhai J. 2002. Regulation of cellular differentiation in filamentous cyanobacteria in
521 free-living and plant-associated symbiotic growth states. Microbiol Mol Biol R. 66:94–121.

522 Michod RE. 2007. Evolution of individuality during the transition from unicellular to
523 multicellular life. Proc Natl Acad Sci USA. 104:8613–8618.

524 Mitsui A et al. 1986. Strategy by which nitrogen-fixing unicellular cyanobacteria grow
525 photoautotrophically. Nature. 323:720–722.

526 Mullineaux CW et al. 2008. Mechanism of intercellular molecular exchange in heterocyst-
527 forming cyanobacteria. EMBO J. 27:1299–1308.

528 Nguyen L-T, Schmidt HA, Haeseler von A, Minh BQ. 2015. IQ-TREE: a fast and effective
529 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32:268–
530 274.

531 Nürnberg DJ, Mariscal V, Parker J, Mastroianni G, Flores E, Mullineaux CW. 2014.
532 Branching and intercellular communication in the section V cyanobacterium *Mastigocladus*
533 *laminosus*, a complex multicellular prokaryote. Mol Microbiol. 91:935-949.

534 O'Leary NA et al. 2016. Reference sequence (RefSeq) database at NCBI: current status,
535 taxonomic expansion, and functional annotation. Nucleic Acids Res. 44:D733–745.

536 Potts M. 1994. Desiccation tolerance of prokaryotes. Microbiol Rev. 58:755–805.

537 Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open
538 Software Suite. Trends Genet. 16:276–277.

539 Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. 1979. Generic assignments,
540 strain histories and properties of pure cultures of cyanobacteria. Microbiol. 111:1–61.

541 Rippka R, Waterbury J, Cohen-Bazire G. 1974. A cyanobacterium which lacks thylakoids.
542 Arch Microbiol. 100:419–436.

543 Rose CJ. 2020. Germ lines and extended selection during the evolutionary transition to
544 multicellularity. J Exp Zool. (Mol Dev Evol). 1-7. https://doi.org/10.1002/jez.b.22985

545 Sánchez-Baracaldo P. 2015. Origin of marine planktonic cyanobacteria. Sci. Rep. 5, 17418.

546 Sánchez-Baracaldo P, Hayes PK, Blank CE. 2005. Morphological and habitat evolution in the
547 Cyanobacteria using a compartmentalization approach. Geobiol. 3:145–165.

548 Sánchez-Baracaldo P, Ridgwell A, Raven JA. 2014. A neoproterozoic transition in the marine
549 nitrogen cycle. Curr Biol. 24:652–657.

550 Schirrmeister BE, Antonelli A, Bagheri HC. 2011. The origin of multicellularity in
551 cyanobacteria. BMC Evol Biol. 11:45–4.

552 Schopf JW. 1993. Microfossils of the early Archean Apex chert: new evidence of the antiquity
553 of life. Science. 260:640–646.

554 Shapiro J. 1988. Bacteria as multicellular organisms. Sci Am. 258:82–89.

555 Shi T, Falkowski PG. 2008. Genome evolution in cyanobacteria: the stable core and the
556 variable shell. Proc Natl Acad Sci USA. 105:2510–2515.

557 Shih PM et al. 2012. Improving the coverage of the cyanobacterial phylum using diversity-
558 driven genome sequencing. Proc Natl Acad Sci USA. 110:201217107–1058.

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

561 Stal LJ. 1995. Physiological ecology of cyanobacteria in microbial mats and other
562 communities. New Phytol. 131:1–32.

563 Szöllősi GJ, Boussau B, Abby SS, Tannier E, Daubin V. 2012. Phylogenetic modeling of
564 lateral gene transfer reconstructs the pattern and relative timing of speciations. Proc Natl
565 Acad Sci USA. 109:17513–17518.

566 Tarnita CE, Taubes CH, Nowak MA. 2013. Evolutionary construction by staying together and
567 coming together. J Theor Biol. 320:10–22.

568 Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families.
569 Science. 278:631–637.

570 Thiergart T, Landan G, Martin WF. 2014. Concatenated alignments and the case of the
571 disappearing tree. BMC Evolutionary Biology 14:266.

572 Tomitani A, Knoll AH, Cavanaugh CM, Ohno T. 2006. The evolutionary diversification of
573 cyanobacteria: molecular–phylogenetic and paleontological perspectives. Proc Natl Acad Sci
574 USA. 103:5442–5447.

575 Tria, FDK, Landan, G, Dagan, T. 2017 Phylogenetic rooting using minimal ancestor
576 deviation. Nat Ecol Evol. 1:0193–0197.

577 Urrejola C et al. 2020. Loss of filamentous multicellularity in cyanobacteria: the extremophile
578 *Gloeocapsopsis* sp. Strain UTEX B3054 retained multicellular features at the genomic and
579 behavioral levels. J Bacteriol. 202:e00514–19.

580 Uyeda JC, Harmon LJ, Blank CE. 2016. A comprehensive study of cyanobacterial
581 morphological and ecological evolutionary dynamics through deep geologic time. PLoS ONE.
582 11:e0162539–32.

583 van Gestel J, Vlamakis H, Kolter R. 2015. Division of labor in biofilms: the ecology of cell
584 differentiation. Microbiol Spec. 3:1–24.

585 Waterbury JB, Stanier RY. 1978. Patterns of growth and development in pleurocapsalean
586 cyanobacteria. Microbiol Rev. 42:2–44.

587 Yoon H-S, Golden JW. 1998. Heterocyst pattern formation controlled by a diffusible peptide.
588 Science. 282:935–938.

589 Zehr JP. 2011. Nitrogen fixation by marine cyanobacteria. Trends Microbiol. 19:162–173.

590 Zhaxybayeva O, Doolittle WF, Papke RT, Gogarten JP. 2009. Intertwined evolutionary
591 histories of marine *Synechococcus* and *Prochlorococcus marinus*. Genome Biol Evol. 1:325–
592 339.

Zhu Q et al. 2019. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains bacteria and archaea. Nat Commun. 10:1–14.

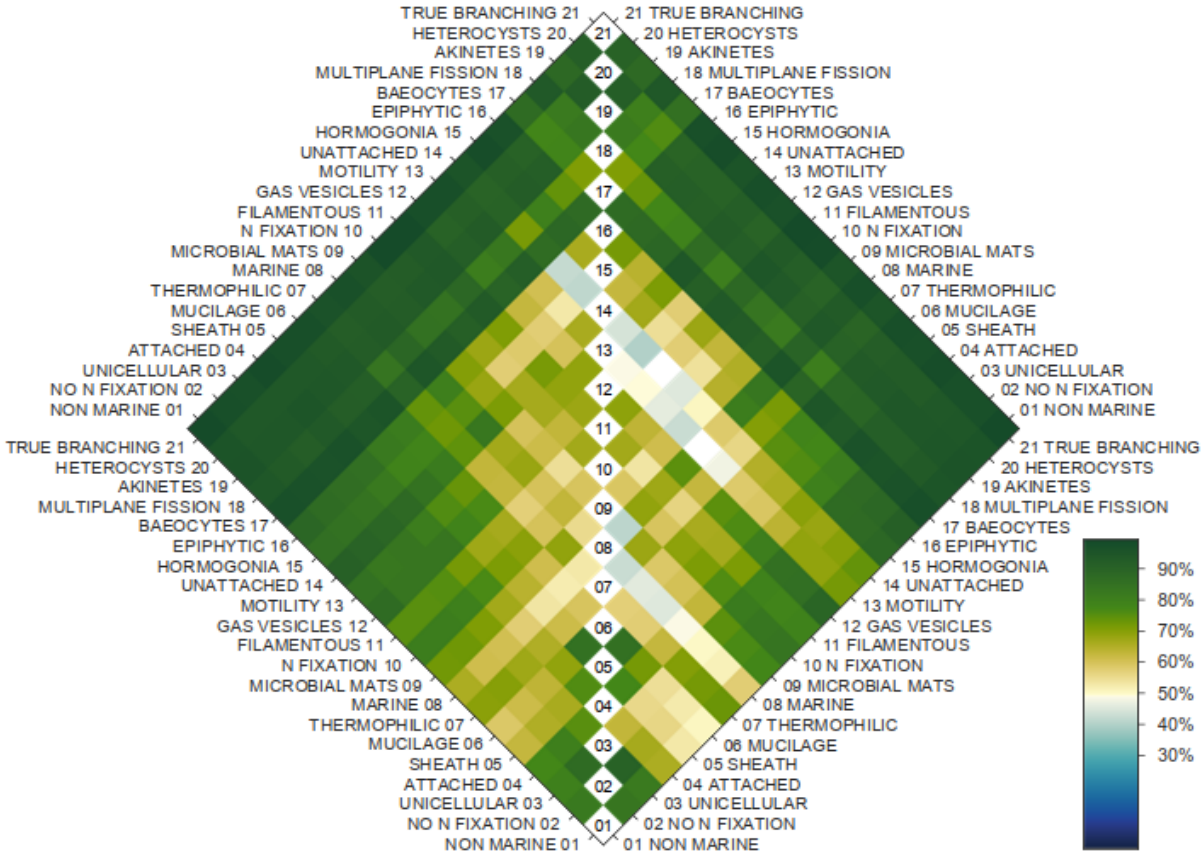

Figure S1. Gene tree support for the relative order of pairs of traits. Left: support from gene trees with SynPro clade excluded (167 spp.). Right: original gene trees as in fig. 1 (199 spp.). Cells in the matrix are shaded according to the proportion of gene trees that support the conclusion based on the species tree (according to the color bar on the right).