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 General 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, 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, has 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 followed by the culmination of complex multicellularity in the form of differentiated cells and patterned multicellularity.

Introduction

Multicellularity is considered a characteristic trait of eukaryotes, but has evolved independently several times in diverse prokaryote taxa, including actinobacteria, myxobacteria, and cyanobacteria (Bonner 1998). Bacterial multicellularity ranges from transient associations, such as colonies, biofilms and cellular aggregations, to permanent multicellular forms (Shapiro 1988). Instances of multicellular bacterial species present the major traits of eukaryotic multicellularity, including cell-to-cell adhesion, peri- or cytoplasmic continuity, intercellular communication, patterning, programmed cell death (PCD), and division of labor (Claessen et al. 2014). Aggregative forms of multicellularity are common among bacterial species, for example, those that form a biofilm under specific external conditions (Tarnita et al. 2013). *Bacillus subtilis*, for instance, forms biofilms upon nutrient deprivation in which cells differentiate into motile, matrix producing, or spore cells depending on the environmental cues (Claessen et al. 2014). Notably, cell differentiation in aggregates is adaptive at the level of the individual cell as it directly confers a fitness benefit to that particular cell. In contrast, under true division of labor, cells are interdependent upon each other and specialize in performing complementary tasks. These tasks, e.g., somatic functions or PCD, are not beneficial on the level of the individual cell, but are advantageous for the colony; thus, they 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 for hormogonia release), and heterocysts (Claessen et al. 2014; Herrero et al. 2016).
73 Heterocysts differentiate under nitrogen deprivation and are specialized in nitrogen (N_2)
74 fixation by the enzyme nitrogenase (Frías et al. 1994). As this enzyme is sensitive to oxygen
75 (O_2), these cells are characterized by the absence of oxygenic photosynthesis and by a thick
76 cell wall, which maintains an anaerobic environment. Heterocysts and vegetative cells in the
77 filament are metabolically interdependent with the heterocysts providing combined nitrogen
78 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 the
80 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).

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

Cell types	
Vegetative cells	Photosynthetic cells.
HORMOGONIA*	Motile reproductive cells 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.
Necridia*	Dead cells resulting from PCD for hormogonia release.
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 ammonium, in contrast to: Fixation of N_2 into ammonium.
SHEATH	Part of the cell envelope, located outside the cell wall.
MUCILAGE	Part of the envelope, located outside the cell wall, comprised of 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.

99 Phenotypic traits

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

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

110

111 Protein families and alignments

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

122 Gene families were then extended to include homologous sequences from non-
123 cyanobacteria species, serving as outgroups for rooting purposes. We identified outgroup

homologues by an rBBH analysis of the *Scytonema hofmanni* PCC 7110 genome (the most widely present species in the initial gene family dataset) against 26 high quality non-cyanobacteria genomes: Vampirovibrionia (12 genomes) and Sericytochromatia (2) (Soo et al. 2017; Carnevali et al. 2019), the closest phyla Margulisbacteria (6), Saganbacteria (2), Fusobacteria (1) and Firmicutes (1) (according to (Carnevali et al. 2019) and (Zhu et al. 2019); one reference anoxygenic photosynthetic genome from *Chloroflexus aurantiacus* J-10-fl and the *Escherichia coli* str. K-12 substr. MG1655 genome (see supplementary table S1). The number of gene families with homologs ranged between 204 and 451 for the 26 outgroup genomes. We selected six of these outgroups for further analyses: *Vampirovibrio chlorellavorus, Chloroflexales, Obscuri-PALSA-1081, Sericytochromatia-UBA7694, Bacillus subtilis,* and *Margulis-GWF2-35-9.* Protein sequences of these families were aligned using MAFFT version 7.027b employing the L-INS-i strategy (Katoh & Standley 2013). The alignments are available in supplementary material online.

Species tree reconstruction

The sequence data for the reconstruction of the cyanobacterial species tree consisted of 14 single-copy gene families that are present in all 199 cyanobacteria genomes and any of the six outgroup genomes. The species tree was inferred using IQ-TREE (Nguyen et al. 2015) in a partitioned analysis over the concatenated alignment of the 14 gene alignments (iqtree version 1.6.6.b with parameters -t BIONJ -keep-ident -mset LG -madd LG4X –spp). The unrooted species tree was rooted on the branch leading to the outgroup. The species tree is available in supplementary material online.

147 Gene trees reconstruction

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

156

157 Inference of trait order

158 The traits presence/absence pattern was mapped onto the rooted species tree. The LCA of
159 each trait was assigned to the most recent node where the trait is present in any of the node's
160 descendants. The species tree was traversed from root to tips to determine the order of trait
161 emergence. For each pair of traits we tested whether the order observed in the species tree was
162 reproduced in gene trees with at least two species displaying each of the two traits. For that
163 purpose, we repeated the trait-order analysis with the set of single-copy rooted gene trees,
164 including gene families that do not span the full taxa set. The support of the gene trees in the
165 conclusion based on the species trees is calculated as the proportion of gene trees where the
166 trait order is the same as in the species tree.

167

168 **Results and Discussion**

169 To reconstruct the order of trait emergence in the evolution of cyanobacterial multicellularity,
170 we evaluated 21 phenotypic traits variably present in 199 cyanobacterial species (table 1,
171 supplementary table S1). The ability to perform photosynthesis is not included in our study as

172 it is universal to all cyanobacteria hence it is considered an ancestral trait (Garcia Pichel et al.
173 2020). We inferred a species tree from a partitioned analysis of 14 single copy core genes that
174 are present in all 199 species. The species tree was rooted by inclusion of outgroup sequences
175 from six bacterial species (see methods), with the root separating the genus *Gloeobacter* from
176 all other cyanobacteria.

177 The rooting of cyanobacteria is a thorny issue, with two competing clades put forward
178 as basal lineages. Traditionally, *Gloeobacter* is assumed to be a basal lineage within
179 cyanobacteria since it is lacking well-defined thylakoids (Rippka et al. 1974; Mareš et al.
180 2019) that are considered a derived structure in the phylum. Several studies in the literature
181 thus used *Gloeobacter* for rooting the cyanobacterial species tree (e.g., Shih et al. 2012; Shi &
182 Falkowski 2008; Dagan et al. 2013; Sánchez-Baracaldo et al. 2014). Other studies, which
183 used midpoint rooting or MAD to root the cyanobacteria species tree, position the root on the
184 branch separating between the pico-cyanobacteria (*Synechococcus* & *Prochlorococcus,* and
185 hereafter SynPro clade) and the remaining species (Szöllősi et al. 2012; Tria et al. 2017). This
186 branch, however, is typically long in gene trees, as well as the species trees, hence it may
187 reflect a Long Branch Attraction (LBA) artifact (defined in Felsenstein 1978).

188 To evaluate whether the rooting placement is robust, we conducted gene-tree support
189 analyses by reconstructing trees for 553 cyanobacterial single-copy gene families along with
190 homologs from six different outgroup species (fig. 1, and methods). We then extracted the
191 rooted cyanobacterial subtree, while discarding 138 trees where the outgroup species formed
192 more than one group. We next characterize the trees by the pattern of the three cyanobacterial
193 subgroups: *Gloeobacter* (2 spp.), SynPro clade (32 spp.), and Other Cyanobacteria (165 spp.),
194 discarding 142 genes that are present in less than two genomes for each group. First, we
195 considered gene trees in which no group is present on both sides of the root. The group
196 appearing on its own as a lineage originating at the root is considered a basal group and the

197 tree is supporting a root located on the branch separating it from the others (left column in fig.
198 1). Next, if only one group appears on both sides of the root, this group is labeled as ancestral,
199 and supports a root position within it (right column in fig. 1). We found no gene trees where
200 more than one group appears on both sides of the root. We note that the gene trees may be
201 discordant with the species tree, and that any of the groups may seem to be paraphyletic,
202 either due to methodological artifacts (e.g. LBA involving the SynPro clade) or due to
203 biological processes such as lateral gene transfer (LGT).

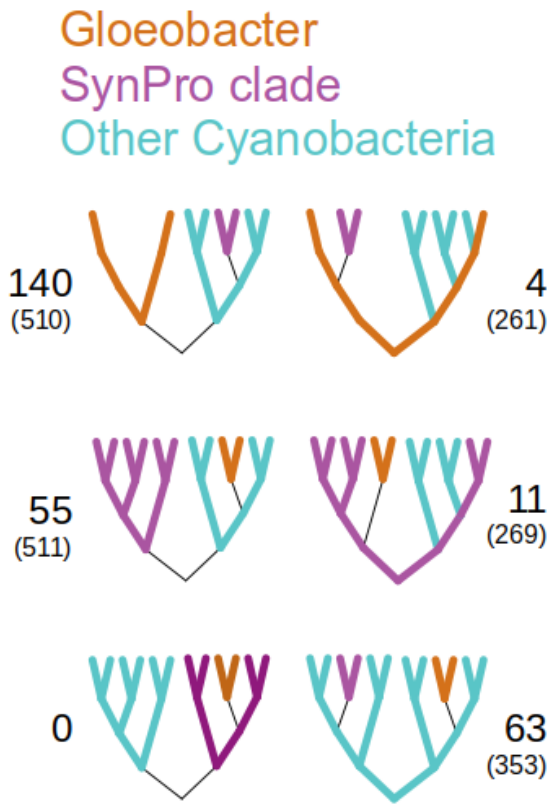

Figure 1. Support for three possible basal/ancestral cyanobacterial groups in 273 rooted gene trees. Numbers are number of gene families, median alignment length is in parenthesis.

204

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

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

209 We infer the order of trait emergence in cyanobacterial evolution by mapping the traits onto
210 the rooted species tree. The *Last Common Ancestor* (LCA) of each trait was assigned to the
211 most recent node where the trait is present in any of the node's descendants. This is a
212 conservative approach in that for each trait it allows for a single origin with possible
213 subsequent losses (Dagan & Martin 2007). When the LCA of two traits coincides, we label
214 the two traits as 'simultaneous' at the resolution of the current taxa sample. When an LCA
215 node of a trait is a descendent of a second trait's LCA node, we conclude that the first trait
216 emerged earlier. A third possibility is that the LCAs of two traits are not nested, but this
217 relationship was not observed in any of the 210 trait pairs. The order of trait emergence is
218 depicted in fig. 2.

219 Next, we evaluated whether the trait order derived from a single species tree is also
220 supported by individual gene trees. We considered the set of single-copy gene families where
221 the gene is present in at least one species from both sides of the species tree root and at least
222 one outgroup species. The gene trees were rooted by the outgroup, and the order of pairs of
223 traits determined. In fig. 2 we report the percentage of gene trees that reproduces the species
224 tree ordering. The vast majority of trait pair orderings are supported by more than 50% of the
225 individual gene trees. Excluding the sequences from the SynPro clade from the gene trees led
226 to slightly better support of the gene trees in the order observed in the species tree
227 (supplementary fig. S1), suggesting that LBA artifacts associated with the SynPro clade may
228 lead to disagreement between the gene and species trees. We note that gene trees that are
229 discordant with the species tree generally had shorter alignment length and lower bootstrap
230 support (supplementary table S2 and S3). Our results thus show a high level of agreement
231 between the species tree and the gene trees, as is expected under an overall homogeneous yet
232 low frequency of LGT during bacterial evolution (Dagan & Martin 2007).

233

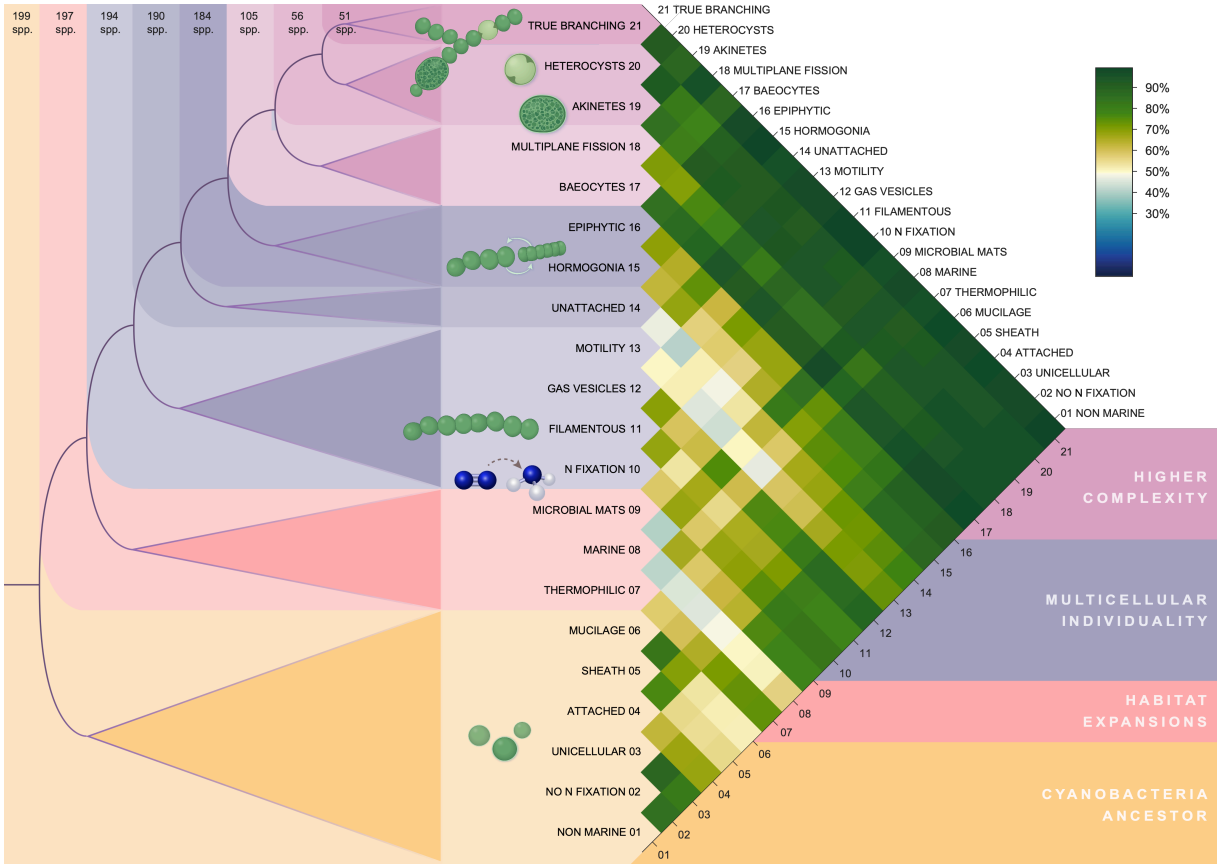

Figure 2. Order of trait emergence. Left: traits and their LCAs inferred from the rooted species tree. Colors mark traits with a common LCA. Colored boxes are nested, i.e., earlier traits are present also in the nested colors Right: Gene tree support for the relative order of pairs of traits. Cells in 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).

234

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

239

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

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

251 We can further deduce from our data that the cyanobacterial ancestor lived ATTACHED
252 and possessed a SHEATH and MUCILAGE. Whether the cyanobacteria ancestor lived ATTACHED
253 is a matter of debate and opposing views on the topic have been published (Sánchez-
254 Baracaldo et al. 2005; Uyeda et al. 2016; Schopf 1993; Garcia-Pichel 1998; Sánchez-
255 Baracaldo 2015). SHEATH and MUCILAGE are forms of extracellular polysaccharides (EPS),
256 located outside the cell wall, which in today's cyanobacteria are mainly involved in protecting
257 the cell from various stresses, such as UV and desiccation (Ehling-Schulz & Scherer 1999;
258 Potts 1994). Furthermore, we find the cyanobacteria ancestor to have most likely inhabited a
259 NON MARINE environment, agreeing with studies that suggest that early cyanobacteria lived in
260 freshwater or terrestrial habitats and subsequently diverged into marine environments (Dagan
261 et al. 2013; Uyeda et al. 2016).

262 The second phase in cyanobacterial evolution is expansion of the cyanobacterial
263 habitat, indicated by the traits MARINE and THERMOPHILIC, which are both inferred to
264 simultaneously occur with the ability to form MICROBIAL MATS. MICROBIAL MATS are dense

265 communities (Stal 1995) that typically present a laminated segregation of functional types.
266 They are often formed by cyanobacteria and are found in extreme habitats, such as deserts or
267 hot springs, characterized by temperatures between 30°C to 73°C (e.g., (Cox et al. 2011)).

268

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

270 Phase iii in our reconstruction comprises three sets of cyanobacterial traits (fig. 2). First, the
271 simultaneous emergence of the FILAMENTOUS morphology, N FIXATION, GAS VESICLES and
272 MOTILITY, followed by the trait UNATTACHED, and lastly by the co-occurrence of
273 HORMOGONIA and EPIPHYTIC. During cyanobacterial N_2 fixation, molecular dinitrogen (N_2) is
274 reduced to ammonium (NH_3), a process that is catalyzed by the enzyme nitrogenase. Whereas
275 present day cyanobacteria, other microorganisms, and most plants are able to take up nitrogen
276 in various combined forms, such as nitrate, ammonium, organic nitrogen, or urea, these
277 combined forms of nitrogen are scarce in most environments (e.g., open oceans or terrestrial
278 habitats (Zehr 2011)). Combined nitrogen, which is critical for the biosynthesis of amino and
279 nucleic acids, was likely a limiting resource in the early Earth environment (Kasting & Siefert
280 2001).

281 The realization of the full metabolic potential of N_2 fixation, however, faced the
282 challenge of the incompatibility of nitrogenase with intracellular oxygen (Gallon 1981). When
283 the cyanobacterial ancestor first acquired the capacity of N_2 fixation, it must have imposed a
284 strong selection pressure on the individual cells. The trade-off between photosynthesis and
285 nitrogen fixation led to the evolution of multiple solutions, which are still present in today's
286 cyanobacteria: the circadian rhythm of N_2 fixation in unicellular cyanobacteria (Mitsui et al.
287 1986), specific cells devoted to N_2 fixation in undifferentiated filaments (Berman-Frank et al.

288 2003; Bergman et al. 2013), and the differentiation of the highly specialized heterocyst in
289 filamentous cyanobacteria (Flores et al. 2018).

290 Theory predicts that within a population of genetically identical unicellular nitrogen
291 fixing cyanobacteria, cell differentiation and phenotypic heterogeneity would have been
292 adaptive if this increased the fitness of the organisms in groups (Ispolatov et al. 2012). In the
293 case of unicellular cyanobacteria this means that cells evolved adhesion and exchanged fixed
294 nitrogen and carbon products within early cell groups such as filaments.

295 In filamentous cyanobacteria, dividing cells remain linked in a chain, resulting in a
296 localization of cells in close spatial proximity, facilitating metabolite exchange between the
297 individual cells. When compared to the more transient associations in spatially structured
298 communities, such as in extracellular polymeric substance (EPS) imbedded biofilms, the
299 development of filaments opens possibilities for a more direct exchange of molecules with
300 high specificity. Metabolic exchange could have evolved as described for the evolution of
301 metabolic cross-feeding (D'Souza et al. 2018), as the exchange of carbon and nitrogen against
302 other products is generally common in photosynthetic or nitrogen-fixing organisms (Kaiser et
303 al. 2015).

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

308 Thereafter the traits HORMOGONIA and EPIPHYTIC are inferred to occur simultaneously.
309 The differentiation of HORMOGONIA can be induced by environmental stimuli, such as
310 nitrogen deprivation (Flores & Herrero 2010). After breaking off from the mother filament at
311 the necridia, HORMOGONIA disperse via gliding motility or float thanks to GAS VESICLES,

312 ensuring the reproduction of benthic species (Rippka et al. 1979). HORMOGONIA with GAS
313 VESICLES are thus important for distribution in aquatic environments as known from modern
314 cyanobacteria (*Fischerella, Hapalosiphon, Tolypothrix*) (Komárek 2013). The close local
315 association with plants, as indicated by EPIPHYTIC, however might have been the first step
316 towards the initiation of one of the many symbioses between higher organisms and
317 cyanobacteria, where HORMOGONIA serve as the infection units (Meeks & Elhai 2002).

318 Notably, the differentiation into HORMOGONIA is reversible, as they develop a sessile
319 lifestyle, where they grow into a new vegetative filament (Flores & Herrero 2010). Here we
320 observe the emergence of a life cycle with two distinct cell types, which is important for the
321 transition to multicellularity (Hammerschmidt et al. 2014; Rose 2020).

322

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

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

333 The late timing and the co-occurrence of the two traits AKINETES and HETEROCYSTS,
334 indicative of higher complexity, are in line with the view that the evolution of the heterocyst
335 was relatively late in the history of filamentous cyanobacteria (Tomitani et al. 2006), and

336 where a common origin of akinetes and heterocysts has been proposed (Adams & Duggan
337 1999). HETEROCYSTS represent not only a morphological adaptation to the obstacle of N_2
338 fixation under oxic conditions but also an elaborate and highly specialized communication
339 and metabolite exchange system. In *Anabaena* sp., for example, where several hundred cells
340 communicate within a filament, a regular heterocyst formation pattern along the filament
341 must be achieved to guarantee that every cell is adequately supplied with fixed nitrogen
342 compounds (Herrero et al. 2016). For this, the inhibitory signaling peptide PatS needs to be
343 distributed along the filament with heterocyst formation occurring only in cells with low PatS
344 concentration (Yoon & Golden 1998). Whether the exchange of metabolites and regulators
345 happens via the continuous periplasm (Flores et al. 2006) or through septal junctions
346 (Mullineaux et al. 2008) is still not fully resolved.

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

352

353 **Trade offs between incompatible processes lead to division of labor and stable**
354 **multicellularity**

355 Common features of evolutionary transitions in individuality comprise cooperation between
356 the lower level units (Bonner 1998) and the division of labor (Michod 2007). The latter might
357 be of particular advantage, and serve as the driver of the transition to multicellularity when
358 there is a strong trade-off between processes that cannot be performed in a single cell at one
359 time (Michod 2007; Ispolatov et al. 2012). Our current findings support this theory and point

360 to nitrogen fixation, and its incompatibility with photosynthesis, as the trigger for the
361 evolution of multicellularity in cyanobacteria. One open question concerns how the
362 underlying genetics of novel traits, such as the division of labor, arise within a newly
363 emerging multicellular individual. In the case of cyanobacteria multicellularity, as also
364 suggested for animal multicellularity (Brunet & King 2017), we propose that no new genes
365 were required and that higher complexity was achieved by regulatory changes in gene
366 expression patterns. Basic communication and metabolite exchange was pre-existing as
367 single-celled bacteria frequently engage in cell-cell communication and cross-feeding of
368 metabolites via the external environment (D'Souza et al. 2018). Division of labor between
369 photosynthesis and nitrogen fixation was likely first established by the regulatory mechanism
370 of temporal switching. Once simple forms of division of labor and metabolic exchange
371 existed, the transition into spatial separation in differentiated cells could have evolved mainly
372 by regulatory modifications.

373 Differentiated cells are one of the hallmarks of complex multicellularity. It is therefore
374 significant that we observe six distinct cell types in cyanobacteria: photosynthetic,
375 hormogonia, necridia, akinetes, baeocytes, and heterocysts. Such a plurality indicates that the
376 underlying regulatory mechanisms are well developed and that their plasticity and
377 adaptability are a matter of course. It is also significant that three of the differentiated cell
378 types, hormogonia, akinetes, and baeocytes, offer novel reproductive potential and the
379 establishment of a multicellular life cycle. Moreover, signs of a nascent developmental plan
380 can be observed in both the distribution of heterocysts along filaments and in the patterning of
381 true branching cyanobacteria. These elements have no fitness value for the individual cell, but
382 are selectable adaptations on the higher level, the filament. The chronology of the evolution of
383 multicellularity in cyanobacteria shows that, once established, multicellular individuality
384 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.

403 **Literature Cited**

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

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

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

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

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

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

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

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

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

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

424 Dagan T et al. 2013. Genomes of Stigonematalean cyanobacteria (Subsection V) and the
425 evolution of oxygenic photosynthesis from prokaryotes to plastids. Genome Biol Evol. 5:31–
426 44.

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

429 Ehling-Schulz M, Scherer S. 1999. UV protection in cyanobacteria. Eur J Phycol. 34:329–
430 338.

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

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

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

437 Flores E, Herrero A, Wolk CP, Maldener I. 2006. Is the periplasm continuous in filamentous

438 multicellular cyanobacteria? Trends Microbiol. 14:439–443.

439 Flores E, Picossi S, Valladares A, Herrero A. 2018. Transcriptional regulation of development
440 in heterocyst-forming cyanobacteria. Biochim Biophys Acta Gene Regul Mech. 1862: 673–
441 684.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

477 Nguyen L-T, Schmidt HA, Haeseler von A, Minh BQ. 2015. IQ-TREE: a fast and effective
478 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32:268–
479 274.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

507 Stal LJ. 1995. Physiological ecology of cyanobacteria in microbial mats and other

508 communities. New Phytol. 131:1–32.

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

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

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

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

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

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

524 Uyeda JC, Harmon LJ, Blank CE. 2016. A comprehensive study of cyanobacterial
525 morphological and ecological evolutionary dynamics through deep geologic time. PLoS ONE.
526 11:e0162539–32.

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

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

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

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

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

536