A chronology of multicellularity evolution in cyanobacteria

Katrin Hammerschmidt[1*], Giddy Landan[1*], Fernando Domingues Kümmel Tria[1,†], Tal Dagan[1]

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

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

[*] These authors contributed equally to this work.

Corresponding author: 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[1-4]. Evolutionary transitions in individuality are characterized by cooperation between the lower level entities[5-7] and by division of labour[8,9]. Theory suggests that division of labour may drive the transition to multicellularity by eliminating the trade-off between two incompatible processes that cannot be performed simultaneously in one cell[1,9,10]. Here we examine the evolution of the most ancient multicellular transition known today, that of cyanobacteria[11,12]. We developed a novel approach for the precedence polarization of phenotypic traits that employs gene phylogenies and does not require a species tree. Applying our procedure to cyanobacterial genomes we reconstruct the chronology of ecological and phenotypic trait evolution in cyanobacteria. 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 a reproductive life cycle. This was followed by a range of niche expansions and interactions with other species, and the progression of multicellularity into higher complexity 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[5]. Bacterial multicellularity ranges from transient associations, such as colonies, biofilms and cellular aggregations, to permanent multicellular forms[13]. 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 labour, often visible as cell differentiation[14]. Aggregative forms of multicellularity are common among bacterial species, for example, those that form a biofilm under specific external conditions[15]. *Bacillus subtilis*, for example, forms biofilms upon nutrient deprivation[16] in which cells differentiate into motile, matrix producing, or spore cells depending on the environmental cues[17]. 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 labour, cells are interdependent upon each other and specialize in performing complementary tasks. These tasks (e.g., somatic functions or PCD) cannot be regarded beneficial on the level of the individual cell, but are considered advantageous for the colony; thus, they are emergent properties on a higher level of organization[18].

Such true division of labour in bacteria is best described in actinobacteria and cyanobacteria[18]. In cyanobacteria, the most complex of the filamentous species can differentiate up to five different cell types: vegetative (photosynthetic) cells, akinetes (spore-like cells), hormogonia (reproductive, motile filaments), necridia (dead cells resulting from PCD for hormogonia release), and heterocysts[14,19]. Heterocysts differentiate under nitrogen deprivation and are specialized in nitrogen (N_2) fixation by

69 the enzyme nitrogenase[20]. As this enzyme is sensitive to oxygen (O_2), these cells are
70 characterized by the absence of photosynthesis and by a thick cell wall, which maintains
71 an anaerobic environment. Heterocyst and vegetative cells in the filament are
72 metabolically interdependent with the heterocysts providing combined nitrogen to the
73 other cells within the filament and receiving fixed carbon compounds in return. With
74 cyanobacteria possessing the hallmark traits reminiscent of complex eukaryotic
75 multicellularity, except for tissues and organs, the trajectory of the evolution of
76 multicellularity in cyanobacteria is of general interest and has implications beyond
77 prokaryotic multicellularity.

78 **Results & Discussion**

79 To reconstruct the order of trait emergence in the evolution of cyanobacterial
80 multicellularity, we evaluated 25 phenotypic traits variably present in 199
81 cyanobacterial species and determined the order of their first appearance (Table 1;
82 Supplementary Tables 1 and 2). The ability to perform photosynthesis is not included in
83 the analysis as it is common to all cyanobacteria. The chronology of trait emergence is
84 derived from the ancestor-decedent relations of pairs of traits, hereby termed *trait-pair*
85 *polarity.*

86 The problem of trait-pair polarity poses a general challenge in microbial
87 phylogenetic reconstruction due to two confounding aspects: the absence of reliable
88 species phylogenies and the frequent absence of clear outgroups for the resolution of
89 ancestor-descendent relations. We applied a novel phylogenomics approach for trait-
90 pair polarity inference (see Fig. 1 and legend; Methods). The phylogenomic procedure
91 does not rely on a particular species tree or an outgroup, but extracts the total evidence
92 for trait precedence from a genomic sample of gene trees for 1671 protein coding gene

93 families. The statistical inference from large samples is effective in overcoming
94 conflicting signals originating from individual gene phylogenies.

95 The reconstructed trait-pairs polarity matrix (Fig. 2a) reveals several groups of
96 traits that are inferred to appear simultaneously, that we term temporal phases. These
97 include: (phase i) the cyanobacteria ancestor (traits 1-7), (phase ii) the transition to
98 multicellular individuality (traits 8-16), (phase iii) the expansion of niche occupation
99 and species interactions (traits 17-20), and (phase iv) the evolution of higher complexity
100 (traits 21-25). Apart from providing an overall ordering of traits during the evolution of
101 multicellularity, the pairwise tests are instrumental in resolving several debates about
102 trait precedencies in cyanobacterial evolution (Table 2).

103 Considering the set of traits in its entirety, the polarity matrix leads to the
104 following scenario for the evolution of multicellularity in cyanobacteria (Fig. 2b). The
105 prime driver of multicellularity in cyanobacteria was the expansion in metabolic
106 capacity offered by N FIXATION (phase ii). This emergence was accompanied by two other
107 cardinal innovations: the emergence of the FILAMENTOUS morphology and the emergence
108 of HORMOGONIA, and thus a reproductive life cycle. Together, these traits form the
109 essential elements of true multicellularity. Secondary traits that played a role in
110 stabilizing the nascent multicellular organism emerged in close succession, and include
111 GAS VESICLES, SHEATH, and MUCILAGE. There followed a cascade of niche expansions and
112 species interactions, and the culmination of complex multicellularity in the form of
113 differentiated cells and patterned multicellularity.

114 To elaborate, N_2 fixation – the reduction of molecular dinitrogen (N_2) to
115 ammonium (NH_3) – is catalysed by the enzyme nitrogenase. Whereas today's
116 cyanobacteria, other microorganisms, and most plants are able to take up nitrogen in
117 various combined forms, such as nitrate, ammonium, organic nitrogen, or urea, these

118 combined forms of nitrogen are scarce in most environments (e.g., open oceans or
119 terrestrial habitats[21]). Combined nitrogen, which is critical for the biosynthesis of amino
120 and nucleic acids, was likely a limiting resource in the early Earth environment[22]. Hence,
121 the capability of N FIXATION was key for cyanobacterial radiation into new habitats and
122 subsequent diversification (phases ii and iii).

123 The realization of the full metabolic potential of N_2 fixation, however, faced the
124 challenge of the incompatibility of nitrogenase with intracellular oxygen[23]. When the
125 cyanobacterial ancestor first acquired the capacity of N_2 fixation, it must have imposed a
126 strong selection pressure on the individual cells. The trade-off between photosynthesis
127 and nitrogen fixation led to the evolution of multiple solutions, which are still present in
128 today's cyanobacteria: the circadian rhythm of N_2 fixation in unicellular cyanobacteria[24]
129 and the differentiation of the highly specialized heterocyst[25].

130 Theory predicts that within a population of genetically identical unicellular
131 nitrogen fixing cyanobacteria, cell differentiation and phenotypic heterogeneity would
132 have been adaptive if this increased the fitness of the organisms in groups[10]. In the case
133 of unicellular cyanobacteria this means that cells evolved adhesion and exchanged fixed
134 nitrogen and carbon products within early cell groups such as filaments. Indeed, our
135 results indicate that the FILAMENTOUS morphology emerged simultaneously with N
136 FIXATION (phase ii). In filamentous cyanobacteria, where dividing cells remain linked in a
137 chain, surface-associated growth, SHEATH, and MUCILAGE lead to a localization of cells in
138 close spatial proximity, facilitating metabolite exchange between the individual cells.
139 When compared to the more transient associations in spatially structured communities,
140 such as in extracellular polymeric substance (EPS) imbedded biofilms, the development
141 of filaments opens possibilities for a more direct exchange of molecules with high
142 specificity. Metabolic exchange could have evolved as described for the evolution of

143 metabolic cross-feeding[26], as the exchange of carbon and nitrogen against other
144 products is generally common in photosynthetic or nitrogen-fixing organisms[27].

145 The trait HORMOGONIA is inferred to occur simultaneously with the appearance of
146 the FILAMENTOUS morphology. Its co-occurrence with other traits, such as NOT FREE-LIVING
147 BENTHIC/SESSILE, EPILITHIC/ ENDOLYTHIC, SHEATH, and MUCILAGE underline the transition from
148 a planktonic to a benthic lifestyle (phase ii). The differentiation of HORMOGONIA can be
149 induced by environmental stimuli, such as nitrogen deprivation[28]. After breaking off
150 from the mother filament at the necridia, HORMOGONIA disperse via gliding motility or
151 float thanks to GAS VESICLES, ensuring the reproduction of benthic species[29,30]. The
152 differentiation into HORMOGONIA is reversible – they change back to the sessile lifestyle,
153 where they grow into a new vegetative filament[28]. Here we observe the emergence of a
154 two-phases life cycle, which is important for the transition to multicellularity[31,32].

155 Traits that are indicative of higher complexity emerged late in the evolution of
156 cyanobacteria and dominate (phase iv) of the polarity matrix. We observe the
157 occurrence of AKINETES and the irreversibly differentiated HETEROCYSTS. HETEROCYSTS
158 represent not only a morphological adaptation to the obstacle of N_2 fixation under oxic
159 conditions but also an elaborate and highly specialized communication and metabolite
160 exchange system. In *Anabaena* sp., for example, where several hundred cells
161 communicate within a filament, a regular heterocyst formation pattern along the
162 filament must be achieved to guarantee that every cell is adequately supplied with fixed
163 nitrogen compounds[19]. For this, the inhibitory signalling peptide PatS needs to be
164 distributed along the filament with heterocyst formation occurring only in cells with low
165 PatS concentration[33]. Whether the exchange of metabolites and regulators happens via
166 the continuous periplasm[34,35] or through septal junctions[36,37] is still not fully resolved.

167 Another central innovation that occurs in this phase of the polarity matrix is the
168 ability to FISSION IN MULTIPLE PLANES. This trait co-occurred with the ability to produce
169 BAEOCYTES, differentiated cells, which are the reproductive stages in the order
170 Pleurocapsales[38]. Notably, baeocyte-forming cyanobacteria, that have been traditionally
171 grouped together with unicellular cyanobacteria[29], appear to immediately predate the
172 evolution of AKINETES and HETEROCYSTS and thus emerge much later than filamentous
173 forms. The ability to FISSION IN MULTIPLE PLANES is known to underlie the TRUE-BRANCHING
174 morphology, where cells in a filament divide in more than one plane, and which is the
175 last trait to emerge in our analysis. Members of the Haphalosiphon/Stigonematales
176 clade having true-branching and multiseriate filament morphology are considered as the
177 latest evolutionary innovations[29] and this is further observed in cyanobacterial
178 phylogenies[39,40].

179 Common features of evolutionary transitions in individuality comprise
180 cooperation between the lower level units[5-7] and the division of labour[8,9]. The latter
181 might be of particular advantage, and serve as the driver of the transition to
182 multicellularity when there is a strong trade-off between processes that cannot be
183 performed in a single cell at one time[9,10]. Our current findings support this theory and
184 point to nitrogen fixation, and its incompatibility with photosynthesis, as the trigger for
185 the evolution of multicellularity in cyanobacteria. One open question concerns how the
186 underlying genetics of novel traits, such as the division of labour, arise within a newly
187 emerging multicellular individual. In the case of cyanobacteria multicellularity, we
188 suggest that no new genes were required and that higher complexity was achieved by
189 regulatory changes in gene expression patterns. Basic communication and metabolite
190 exchange was pre-existing as single-celled bacteria frequently engage in cell-cell
191 communication and cross-feeding of metabolites via the external environment[26].

192 Division of labour between photosynthesis and nitrogen fixation was likely first
193 established by the regulatory mechanism of temporal switching. Once simple forms of
194 division of labour and metabolic exchange existed, the transition into spatial separation
195 in differentiated cells could have evolved mainly by regulatory modifications.

196 Differentiated cells are one of the hallmarks of complex multicellularity. It is
197 therefore significant that we observe six distinct cell types in cyanobacteria:
198 photosynthetic, hormogonia, necridia, akinetes, baeocytes, and heterocysts. Such a
199 plurality indicates that the underlying regulatory mechanisms are well developed and
200 that their plasticity and adaptability are a matter of course. It is also significant that
201 three of the differentiated cell types, hormogonia, akinetes, and baeocytes, offer novel
202 reproductive potential and the establishment of a multicellular life cycle. Moreover,
203 signs of a nascent developmental plan can be observed in both the distribution of
204 heterocysts along filaments and in the patterning of true branching cyanobacteria. These
205 elements have no fitness value for the individual cell, but are selectable adaptations on
206 the higher level, the filament. The chronology of the evolution of multicellularity in
207 cyanobacteria shows that, once established, multicellular individuality opens new vistas
208 of opportunities.

209 **METHODS**

210 **Data**

211 The data underlying this study consists of the genomic sequences and phenotypic traits
212 of 199 cyanobacterial species. These were selected from the available genomes so that
213 the number of represented taxa will be as large as possible and genus-level redundancy
214 will be reduced (See Supplementary Table 1 for the complete list of species).

215 **Phenotypic traits**

216 Phenotypic traits were chosen for their potential relevance to the evolution of
217 multicellularity in cyanobacteria, such as environmental factors that might facilitate
218 multicellularity and markers that are indicative for the transition to multicellularity
219 (Table 1). Information on presence and absence of traits was obtained from the
220 published literature and from the Pasteur Culture Collection of cyanobacteria, extending
221 the work by Uyeda et al[44], and coded as binary trait states. Traits included MORPHOLOGY,
222 NITROGEN FIXATION, FRESHWATER, MARINE, BAEOCYTES, HORMOGONIA, THERMOPHILIC, AKINETES,
223 HETEROCYSTS, TRUE BRANCHING, EPI/ ENDOLITHIC, EPIPHYTIC, PERIPHYTIC, MATS, FREE-LIVING,
224 HABITAT, SHEATH, MUCILAGE, GAS VESICLES, MOTILITY, and FISSION IN MULTIPLE PLANES
225 (Supplementary Table 1).

226 **Protein families and gene trees**

227 The cyanobacteria protein families were constructed from completely sequenced
228 genomes available in RefSeq database[51] (ver. May 2016). For the construction of protein
229 families, at the first stage, all protein sequences annotated in the genomes were blasted
230 all-against-all using stand-alone BLAST[52] ver. 2.2.26. Protein sequence pairs that were
231 found as reciprocal best BLAST hits (rBBHs)[53] with a threshold of E-value ≤ $1x10^{-5}$ were
232 further compared by global alignment using needle[54]. Sequence pairs having ≥30%
233 identical amino acids were clustered into protein families using the Markov clustering
234 algorithm (MCL)[55] ver. 12-135 with the default parameters. By requiring a gene to be
235 present in the genomes of the SynProCya clade and in at least one other major
236 cyanobacterial clade, we identified 1671 gene families that are present on both sides of
237 the cyanobacterial root, i.e., ancient proteins families. Protein sequences of these
238 families were aligned using MAFFT version 7.027b employing the L-INS-i strategy[56].

239 Maximum likelihood trees were reconstructed with PhyML version 20120412[57] with
240 parameters -b -4 -v e -m LG -c 4 -s SPR.

241 **Trait ordering**

242 We consider four possible polarizations of a pair of traits ***A*** and ***B***: ***A*** originates before ***B***;
243 ***B*** originates before ***A***; ***A*** and ***B*** emerge simultaneously; and ***A*** and ***B*** are not nested but
244 originate in independent lineages. We infer the pairwise emergence order of the traits
245 by collecting evidence from a set of phylogenetic trees of independent protein families,
246 and conducting a formal statistical test for the best supported polarization. Finally, the
247 set of all pairwise polarities is used to derive the order of emergence of all traits (Fig. 1).

248 **Polarization evidence from a single phylogenetic tree**

249 Given an unrooted phylogenetic tree, we first consider all possible rooted versions of the
250 tree. In an *n*-OTU (*Operative Taxonomic Unit*) unrooted tree, the root can be placed on
251 any of the 2*n-3* branches of the tree. For each of the 2*n-3* rooted trees, we label the OTUs
252 by the presence of the two traits ***A*** and ***B***, and infer the most parsimonious *Last Common*
253 *Ancestor* (LCA) of each trait, assuming a single trait origin and possible trait losses. We
254 then record the polarity induced by the putative root position according to whether the
255 LCAs of ***A*** and ***B*** coincide, are descended one from the other, or are located on
256 independent lineages.

257 Next we conduct a MAD analysis[48], yielding an *Ancestor Deviation* (AD) statistic
258 for every branch of the tree. The AD measure quantifies the amount of lineage rate
259 heterogeneity that is induced by postulating a branch as harbouring the root of tree. We
260 have previously shown that the AD measure provides robust evidence for the inference
261 of the root of the tree. In the current study we do not infer a single root, but use the AD
262 measure to assess the relative strength of alternative rootings of the same tree.

263 Thus, for each possible root position we obtain a polarity state for ***A*** and ***B*** and a
264 corresponding AD value. For each of the four polarities, we take the minimal AD value as
265 the tree support of that polarity. Note that for certain trees and trait pairs, some
266 polarities may not occur. We differentiate between two such possibilities. First, a certain
267 tree is *Uninformative* regarding a specific trait-pair polarity if the species composition of
268 the tree renders the observation of the polarity impossible in any possible topology. For
269 example, the polarity ***'A precedes B'*** is impossible to observe in trees where the ***A*** OTUs
270 are a subset of the ***B*** OTUs, regardless of the specific topology or root position. In
271 contrast, a certain polarity may be unobserved in any of the rootings of a specific
272 unrooted tree, while still being a possible observation for a different tree topology. An
273 example of such *Informative absent-observation* is when ***A*** and ***B*** label disjoint sets of
274 OTUs yet the ***'Not nested'*** polarity is not observed in the actual tree topology. In the
275 latter case, the absence of an observation is evidence *against* that specific polarity, and
276 the polarity is assigned the maximal observed AD as its score. When the tree is
277 uninformative regarding a polarity, on the other hand, it is assigned a *'missing'* value,
278 and is excluded from subsequent analyses.

279 **Statistical inference of pairwise polarity**

280 Repeating the preceding procedure for trees derived from all different single-copy
281 proteins families yields a phylogenomic sample of four variables - the AD scores of each
282 of the four polarities for the pair of traits. The four distributions are paired, as each tree
283 brings in a 4-tuple of values, and a significant difference in support values can be tested
284 using the non-parametric Mann-Whitney U test[58]. In all, we conduct twelve one-sided
285 tests of contrasting polarities, while employing an FDR[59] correction for multiple
286 comparisons. A trait-pair is considered polarized if there exists one polarity where all

287 three tests against the other polarities recover significantly lower AD values at the 1%
288 FDR level. In exceptional situations we again encounter polarity contrasts that cannot be
289 tested. For some polarity contrasts the sample size may not meet our threshold (10
290 paired observations), and the test is conservatively considered valid but not significant
291 with a p-value of 1.0. A second exception occurs when one trait (***A***) is present only in a
292 strict subset of the species that possess the other trait (***B***). In this case it is impossible to
293 observe the polarities ***'not nested'*** and ***'A precedes B'*** in any tree, and we restrict the
294 testing to the single contrast ***'B precedes A'*** versus ***'Simultaneous'***.
295 The pairwise polarity inference is applied to each pair of traits and summarized
296 in a trait-pair polarity matrix. We again apply the FDR correction, this time over all trait-
297 pairs and polarity contrasts. In the present study, with 25 traits, we apply FDR over
298 3,480 tests. To derive an ordering of the traits, we apply 'Topological sort'[60] to the
299 significant polarities of type ***'A precedes B'***, or vice versa. In the present study the
300 significant polarities form a partial order, i.e., there are no self-contradicting precedence
301 cycles, and the topological sorting order is used to order the polarities matrix and to
302 reduce it to a feed-forward network (Fig. 2).
303
304 **Data availability**
305 The alignments and trees are available on our webpage at www.uni-
306 kiel.de/genomik/ressourcen
307
308

309 **References**

310 1. Buss, L. W. *The evolution of individuality*. (Princeton University Press, 1987).

311 2. Smith, J. M. & Szathmáry, E. *The major transitions in evolution*. (Oxford University
312 Press, 1995).

313 3. Okasha, S. *Evolution and the levels of selection*. (Oxford University Press, 2006).

314 4. Godfrey-Smith, P. *Darwinian populations and natural selection*. (OUP Oxford, 2009).

315 5. Bonner, J. T. The origins of multicellularity. *Integr. Biol.* **1,** 27–36 (1998).

316 6. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314,** 1560–1563
317 (2006).

318 7. Rainey, P. B. & De Monte, S. Resolving conflicts during the evolutionary transition to
319 multicellular life. *Annu. Rev. Ecol. Evol. Syst.* **45,** 599–620 (2014).

320 8. Kirk, D. L. A twelve-step program for evolving multicellularity and a division of labor.
321 *BioEssays* **27,** 299–310 (2005).

322 9. Michod, R. E. Evolution of individuality during the transition from unicellular to
323 multicellular life. *Proc. Natl. Acad. Sci. USA* **104 Suppl 1,** 8613–8618 (2007).

324 10. Ispolatov, I., Ackermann, M. & Doebeli, M. Division of labour and the evolution of
325 multicellularity. *Proc. Biol. Sci.* **279,** 1768–1776 (2012).

326 11. Schopf, J. W. Microfossils of the Early Archean Apex chert: new evidence of the
327 antiquity of life. *Science* **260,** 640–646 (1993).

328 12. Grosberg, R. K. & Strathmann, R. R. The evolution of multicellularity: a minor major
329 transition? *Annu. Rev. Ecol. Evol. Syst.* **38,** 621–654 (2007).

330 13. Shapiro, J. Bacteria as multicellular organisms. *Sci. Am.* **258,** 82–89 (1988).

331 14. Claessen, D., Rozen, D. E., Kuipers, O. P., Søgaard-Andersen, L. & van Wezel, G. P.
332 Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting
333 bodies. *Nat. Rev. Microbiol.* **12,** 115–124 (2014).

334 15. Tarnita, C. E., Taubes, C. H. & Nowak, M. A. Evolutionary construction by staying
335 together and coming together. *J. Theor. Biol.* **320,** 10–22 (2013).
336 16. Karatan, E. & Watnick, P. Signals, regulatory networks, and materials that build and
337 break bacterial biofilms. *Microbiol. Mol. Biol. Rev.* **73,** 310–347 (2009).
338 17. Vlamakis, H., Aguilar, C., Losick, R. & Kolter, R. Control of cell fate by the formation of
339 an architecturally complex bacterial community. *Genes Dev.* **22,** 945–953 (2008).
340 18. van Gestel, J., Vlamakis, H. & Kolter, R. Division of Labor in Biofilms: the Ecology of
341 Cell Differentiation. *Microbiol. Spec.* **3,** 1–24 (2015).
342 19. Herrero, A., Stavans, J. & Flores, E. The multicellular nature of filamentous
343 heterocyst-forming cyanobacteria. *FEMS Microbiol. Rev.* **40,** 831–854 (2016).
344 20. Frías, J. E., Flores, E. & Herrero, A. Requirement of the regulatory protein NtcA for
345 the expression of nitrogen assimilation and heterocyst development genes in the
346 cyanobacterium *Anabaena* sp. PCC 7120. *Mol. Microbiol.* **14,** 823–832 (1994).
347 21. Zehr, J. P. Nitrogen fixation by marine cyanobacteria. *Trends Microbiol.* **19,** 162–173
348 (2011).
349 22. Kasting, J. F. & Siefert, J. L. Biogeochemistry. The nitrogen fix. *Nature* **412,** 26–27
350 (2001).
351 23. Gallon, J. R. The oxygen sensitivity of nitrogenase: a problem for biochemists and
352 micro-organisms. *Trends Biochem. Sci.* **6,** 19–23 (1981).
353 24. Mitsui, A. *et al.* Strategy by which nitrogen-fixing unicellular cyanobacteria grow
354 photoautotrophically. *Nature* **323,** 720–722 (1986).
355 25. Flores, E., Picossi, S., Valladares, A. & Herrero, A. Transcriptional regulation of
356 development in heterocyst-forming cyanobacteria. *BBA - Gene Regul. Mech.* (2018).
357 26. D'Souza, G. *et al.* Ecology and evolution of metabolic cross-feeding interactions in
358 bacteria. *Nat. Prod. Rep.* **35,** 455–488 (2018).

359 27. Kaiser, D., Kowalski, N., Böttcher, M. E., Yan, B. & Unger, D. Benthic nutrient fluxes
360 from mangrove sediments of an anthropogenically impacted estuary in Southern
361 China. *J. Mar. Sci. Eng.* **3,** 466–491 (2015).
362 28. Flores, E. & Herrero, A. Compartmentalized function through cell differentiation in
363 filamentous cyanobacteria. *Nat. Rev. Microbiol.* **8,** 39–50 (2010).
364 29. Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. & Stanier, R. Y. Generic
365 assignments, strain histories and properties of pure cultures of cyanobacteria.
366 *Microbiol.* **111,** 1–61 (1979).
367 30. Herdman, M. & Rippka, R. Cellular differentiation: hormogonia and baeocytes.
368 *Methods Enzymol.* **167,** 232–242 (1988).
369 31. Rainey, P. B. & Kerr, B. Cheats as first propagules: A new hypothesis for the
370 evolution of individuality during the transition from single cells to multicellularity.
371 *BioEssays* **32,** 872–880 (2010).
372 32. Hammerschmidt, K., Rose, C. J., Kerr, B. & Rainey, P. B. Life cycles, fitness decoupling
373 and the evolution of multicellularity. *Nature* **515,** 75–79 (2014).
374 33. Yoon, H.-S. & Golden, J. W. Heterocyst pattern formation controlled by a diffusible
375 peptide. *Science* **282,** 935–938 (1998).
376 34. Flores, E., Herrero, A., Wolk, C. P. & Maldener, I. Is the periplasm continuous in
377 filamentous multicellular cyanobacteria? *Trends Microbiol.* **14,** 439–443 (2006).
378 35. Mariscal, V., Herrero, A. & Flores, E. Continuous periplasm in a filamentous,
379 heterocyst-forming cyanobacterium. *Mol. Microbiol.* **65,** 1139–1145 (2007).
380 36. Mullineaux, C. W. *et al.* Mechanism of intercellular molecular exchange in
381 heterocyst-forming cyanobacteria. *EMBO J.* **27,** 1299–1308 (2008).
382 37. Bornikoel, J. *et al.* Role of two cell wall amidases in septal junction and nanopore
383 formation in the multicellular cyanobacterium *Anabaena* sp. PCC 7120. *Front. Cell.*

384 *Infect. Microbiol.* **7,** 92–15 (2017).
385 38. Waterbury, J. B. & Stanier, R. Y. Patterns of growth and development in
386 pleurocapsalean cyanobacteria. *Microbiol. Rev.* **42,** 2–44 (1978).
387 39. Dagan, T. *et al.* Genomes of stigonematalean cyanobacteria (subsection V) and the
388 evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biol.*
389 *Evol.* **5,** 31–44 (2013).
390 40. Shih, P. M. *et al.* Improving the coverage of the cyanobacterial phylum using
391 diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. USA* **110,** 1053–1058
392 (2013).
393 41. Giovannoni, S. J. *et al.* Evolutionary relationships among cyanobacteria and green
394 chloroplasts. *J. Bacteriol.* **170,** 3584–3592 (1988).
395 42. Baracaldo, P. S., Hayes, P. K. & Blank, C. E. Morphological and habitat evolution in the
396 cyanobacteria using a compartmentalization approach. *Geobiol.* **3,** 145–165 (2005).
397 43. Schirrmeister, B. E., Antonelli, A. & Bagheri, H. C. The origin of multicellularity in
398 cyanobacteria. *BMC Evol. Biol.* **11,** 45 (2011).
399 44. Uyeda, J. C., Harmon, L. J. & Blank, C. E. A comprehensive study of cyanobacterial
400 morphological and ecological evolutionary dynamics through deep geologic time.
401 *PLoS ONE* **11,** e0162539–32 (2016).
402 45. Tomitani, A., Knoll, A. H., Cavanaugh, C. M. & Ohno, T. The evolutionary
403 diversification of cyanobacteria: Molecular–phylogenetic and paleontological
404 perspectives. *Proc. Natl. Acad. Sci. USA* **103,** 5442–5447 (2006).
405 46. Shi, T. & Falkowski, P. G. Genome evolution in cyanobacteria: The stable core and the
406 variable shell. *Proc. Natl. Acad. Sci. USA* **105,** 2510–2515 (2008).
407 47. David, L. A. & Alm, E. J. Rapid evolutionary innovation during an Archaean genetic
408 expansion. *Nature* **469,** 93–96 (2010).

409 48. Tria, F. D. K., Landan, G. & Dagan, T. Phylogenetic rooting using minimal ancestor
410 deviation. *Nat. Ecol. Evol.* **1,** 0193 (2017).
411 49. Tomitani, A., Knoll, A. H., Cavanaugh, C. M., & Ohno, T. The evolutionary
412 diversification of cyanobacteria: molecular–phylogenetic and paleontological
413 perspectives. *Proc. Natl. Acad. Sci. USA* **103**, 5442-5447 (2006).
414 50. Wolk, C. P., Ernst, A. & Elhai, J. in *The molecular biology of cyanobacteria* 769–823
415 (Springer, Dordrecht, 1994).
416 51. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status,
417 taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44,** D733–45
418 (2016).
419 52. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
420 search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
421 53. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families.
422 *Science* **278,** 631–637 (1997).
423 54. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open
424 Software Suite. *Trends Gen.* **16,** 276–277 (2000).
425 55. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale
426 detection of protein families. *Nucleic Acids Res.* **30,** 1575–1584 (2002).
427 56. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software version 7:
428 improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
429 57. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood
430 phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321
431 (2010).
432 58. Zahr, J. H. *Biostatistical analysis*. (2014).

59. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29, 1165–1188 (2001).

60. Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. in *Section 22.4: Topological sort* 549–552 (2001).

Acknowledgments

We thank Tom Williams, Anne Kupczok, Tanita Wein, and Caroline Rose for comments on the manuscript. KH received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014–2020) under the Marie Skłodowska-Curie Grant Agreement No. 657096 and from the German Academic Exchange Service (DAAD). F.D.K.T. was supported by CAPES (Coordination for the Improvement of Higher Education Personnel–Brazil).

Author contributions

K.H. and T.D. conceived the study. K.H. collected the traits data. G.L., T.D. and F.D.K.T. developed and implemented the method. G.L. performed all analyses. K.H., G.L., and T.D. wrote the manuscript with contributions from F.D.K.T..

Author information

The authors declare no competing financial interests and no conflict of interest.

Correspondence and requests for materials should be addressed to K.H.

455 **Table 1. Description of cyanobacterial cell types, morphological and physiological**
456 **traits, and 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 of 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.
NITROGEN FIXATION	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.
FISSION IN MULTIPLE PLANES	Cell division in two or three perpendicular planes.
TRUE BRANCHING*	Fission in multiple planes leads to branches that remain attached to the main filament.
Habitat and life style	
FRESHWATER	Aquatic environments with salinity between 0-0.5ppt.
MARINE	Aquatic environments with salinity between 30–50ppt.
THERMOPHILIC	Optimal growth temperature above 45°.
PLANKTONIC	Organism that lives in the plankton (not attached).
SESSILE/ BENTHIC	Attachment to a substrate.
MATS	Growth inside thick, laminated, microbial structures.
FREE-LIVING	Organism that lives autonomously, in contrast to:
NOT FREE-LIVING	Organism that lives in a symbiotic relationship.
EPILITHIC/ ENDOLITHIC	Growth on or inside rocky substrates.
EPIPHYTIC	Growth on plants.
PERIPHYTIC	Attachment to underwater substrates.

457 * Multicellularity markers: traits that are adaptations on the level of the filament.

458 SMALL CAPS indicate the traits that have been used in the analysis.

459

460 **Table 2. Resolution of on-going debates regarding trait precedencies in**
461 **cyanobacterial evolution.**

Debate and resolution	Description of debate
Unicellular – Filamentous Trait-pair polarity tests show that the cyanobacterial ancestor was UNICELLULAR and that the FILAMENTOUS morphology arose later (FDR adjusted U-test p-value 2.3×10^{-110}).	Whereas early work[41] suggested that both morphologies had multiple origins, a subsequent analysis found the ancestor to be unicellular and the filamentous morphology to arise in independent lineages of the cyanobacterial tree[42]. Another view is that the filamentous morphology evolved early during cyanobacterial evolution and was subsequently lost and regained several times[43].
Planktonic - Benthic Our results show that the cyanobacteria ancestor traits include PLANKTONIC, MOTILITY and FREE LIVING (all FDR adjusted U-test p-values $<3.2\times10^{-69}$, see Supplementary table 2).	Whether the cyanobacteria ancestor was planktonic is a matter of debate and opposing views on the topic have been published[11,42,44].
N_2 fixation – No N_2 fixation Trait-pair polarity tests show that N FIXATION is a derived trait and that the ancestor of cyanobacteria was lacking the ability to fix N_2 (FDR adjusted U-test p-value 2.6×10^{-125}).	Whereas some studies claimed the last cyanobacterial common ancestor to fix N_2[45], there are others that concluded that it could not fix N_2 and that cyanobacteria must have acquired this trait several times independently[42,46,47].
Freshwater – Marine Our results show that there is no evidence for either MARINE or FRESHWATER environments as ancestral or derived habitat (Simultaneous polarity with FDR adjusted U-test p-value 5.4×10^{-148}).	Some studies suggest that early cyanobacteria lived in freshwater and subsequently diverged into marine environments[39,42,44], whereas others provide evidence in support of a marine origin[48]. The cyanobacteria ancestor most likely inhabited an aquatic environment and colonized both environments early.
Akinetes – Heterocysts Our results show that AKINETES and HETEROCYSTS emerged simultaneously (FDR adjusted U-test p-value 1.8×10^{-124}).	There is a common agreement that these cell types appeared late in cyanobacterial evolution[49], but there is a controversy about whether they shared a common ancestor and appeared simultaneously[50] or successively[42,49].

462

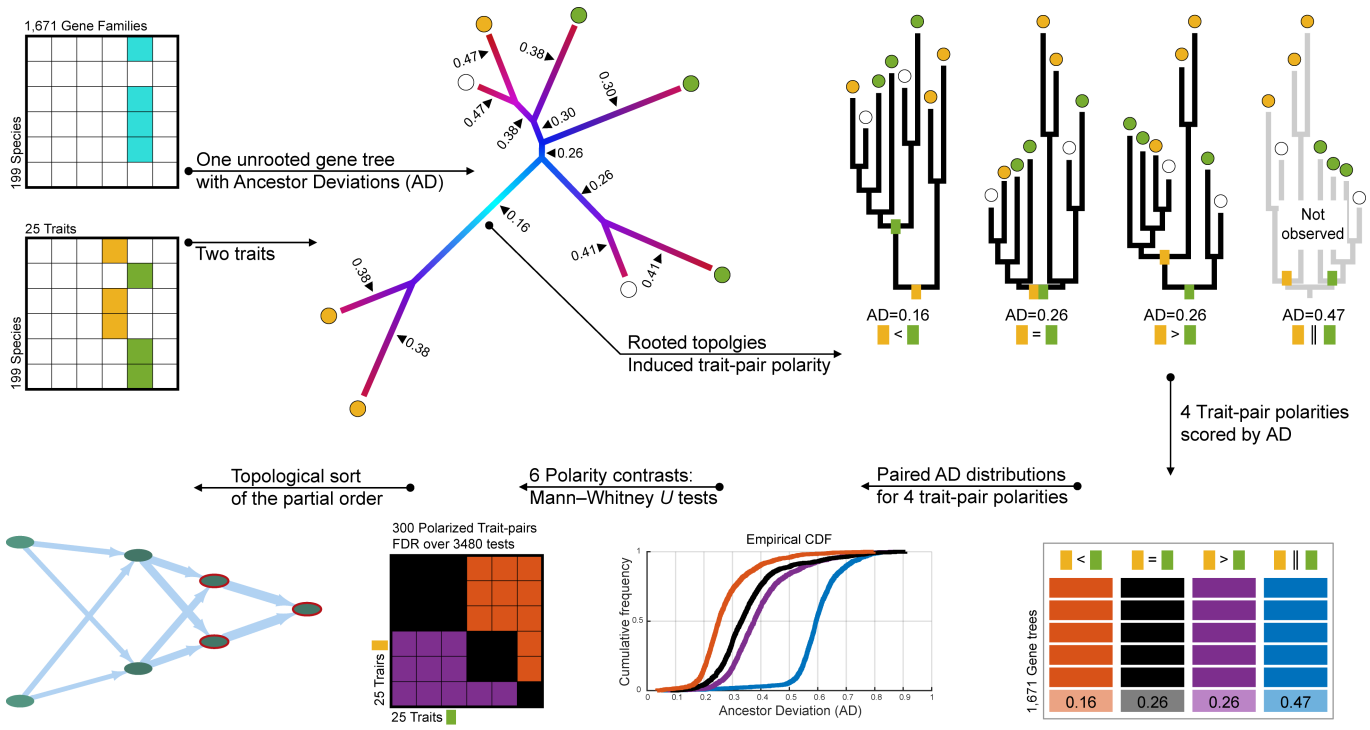

Fig. 1. Phylogenomic reconstruction of trait emergence chronology. Stages in the procedure are depicted clockwise from top-left. A genome-wide sample of single-copy protein coding genes provides measures of ancestor-descendant relations (Ancestor Deviation, AD)[48]. The AD support is then coupled to traits presence-absence patterns for each gene tree. Considering a specific pair of traits, their emergence polarity is deduced from formal statistical tests that contrast alternative hypothesized ancestor-descendant relations, utilizing paired information from all gene trees. All pairwise comparisons of traits are then combined into an overall trait-pairs polarity matrix, and if partial ordering of the traits (i.e., no polarity cycles) is possible, a feed-forward network can be inferred by topological sorting.

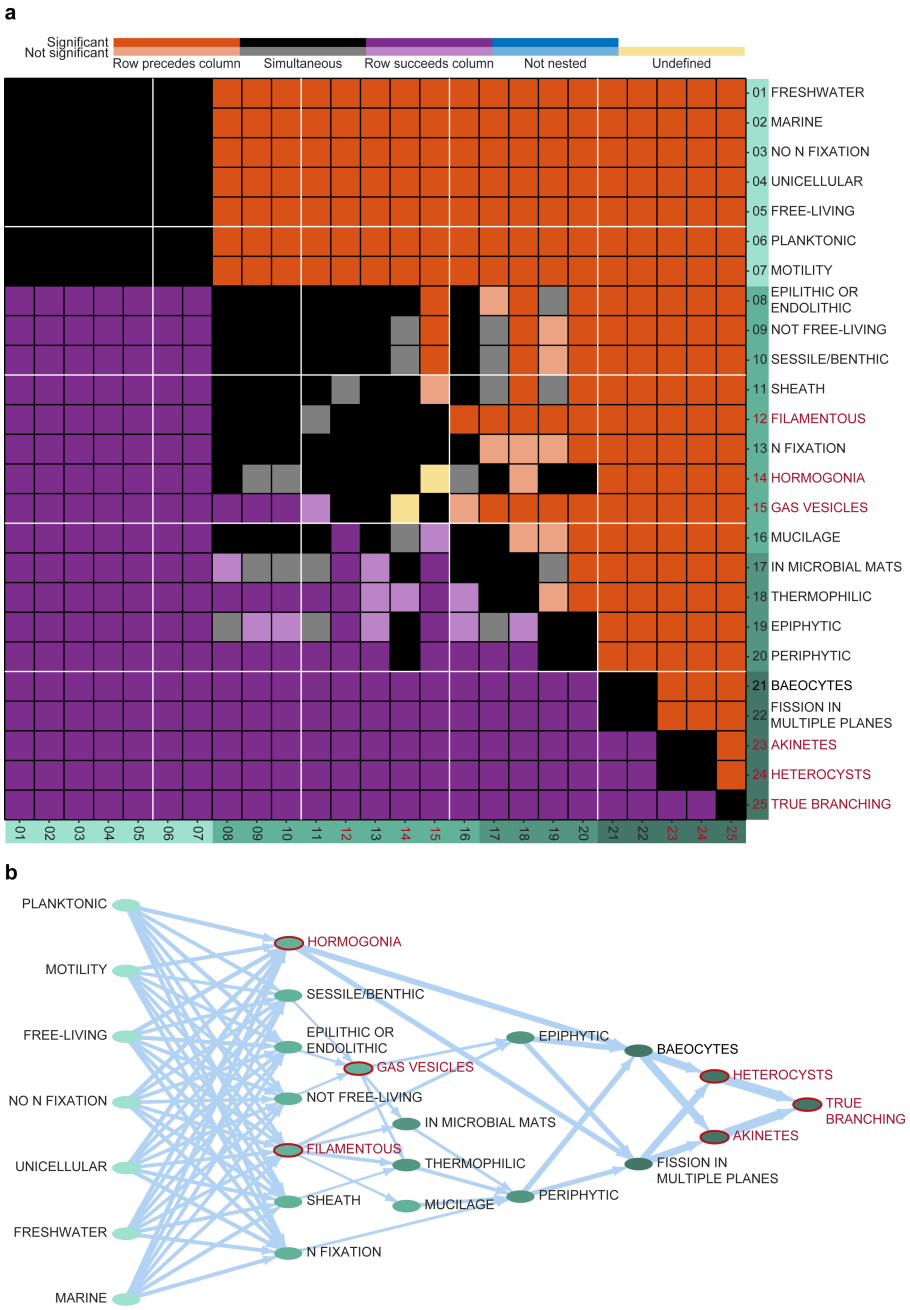

Fig. 2. Trait precedence in the evolution of cyanobacteria. (**a**) Trait-pairs polarity matrix. (**b**) Trait emergence order network. Cells in the matrix are shaded according to the polarity of the row and column traits, and the FDR-corrected significance of the *U* tests (for trait definition and test results see Supplementary Tables 1 and 2). In this dataset there are no cycles of conflicting ordering among the 300 inferred pairwise polarities. Thus, the traits form a *partial order* that is used to determine the trait order in the matrix (**a**) and is visualized as a feed-forward network (**b**). Shades of green mark the

482 four phases of trait emergence. Traits that can be regarded as multicellularity markers

483 in cyanobacteria are labelled in red (Table 1).