A chronology of multicellularity evolution in cyanobacteria

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

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: kammerschmidt@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
the enzyme nitrogenase\textsuperscript{20}. As this enzyme is sensitive to oxygen (O\textsubscript{2}), these cells are characterized by the absence of photosynthesis and by a thick cell wall, which maintains an anaerobic environment. Heterocyst and vegetative cells in the filament are metabolically interdependent with the heterocysts providing combined nitrogen to the other cells within the filament and receiving fixed carbon compounds in return. With cyanobacteria possessing the hallmark traits reminiscent of complex eukaryotic multicellularity, except for tissues and organs, the trajectory of the evolution of multicellularity in cyanobacteria is of general interest and has implications beyond prokaryotic multicellularity.

\textbf{Results & Discussion}

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

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

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

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

To elaborate, N\textsubscript{2} fixation – the reduction of molecular dinitrogen (N\textsubscript{2}) to ammonium (NH\textsubscript{3}) – is catalysed by the enzyme nitrogenase. Whereas today’s cyanobacteria, other microorganisms, and most plants are able to take up nitrogen in various combined forms, such as nitrate, ammonium, organic nitrogen, or urea, these
combined forms of nitrogen are scarce in most environments (e.g., open oceans or terrestrial habitats\textsuperscript{21}). Combined nitrogen, which is critical for the biosynthesis of amino and nucleic acids, was likely a limiting resource in the early Earth environment\textsuperscript{22}. Hence, the capability of \textit{N} \textit{fixation} was key for cyanobacterial radiation into new habitats and subsequent diversification (phases ii and iii).

The realization of the full metabolic potential of N\textsubscript{2} fixation, however, faced the challenge of the incompatibility of nitrogenase with intracellular oxygen\textsuperscript{23}. When the cyanobacterial ancestor first acquired the capacity of N\textsubscript{2} fixation, it must have imposed a strong selection pressure on the individual cells. The trade-off between photosynthesis and nitrogen fixation led to the evolution of multiple solutions, which are still present in today's cyanobacteria: the circadian rhythm of N\textsubscript{2} fixation in unicellular cyanobacteria\textsuperscript{24} and the differentiation of the highly specialized heterocyst\textsuperscript{25}.

Theory predicts that within a population of genetically identical unicellular nitrogen fixing cyanobacteria, cell differentiation and phenotypic heterogeneity would have been adaptive if this increased the fitness of the organisms in groups\textsuperscript{10}. In the case of unicellular cyanobacteria this means that cells evolved adhesion and exchanged fixed nitrogen and carbon products within early cell groups such as filaments. Indeed, our results indicate that the \textit{filamentous} morphology emerged simultaneously with \textit{N} \textit{fixation} (phase ii). In filamentous cyanobacteria, where dividing cells remain linked in a chain, surface-associated growth, \textit{sheath}, and \textit{mucilage} lead to a localization of cells in close spatial proximity, facilitating metabolite exchange between the individual cells. When compared to the more transient associations in spatially structured communities, such as in extracellular polymeric substance (EPS) imbedded biofilms, the development of filaments opens possibilities for a more direct exchange of molecules with high specificity. Metabolic exchange could have evolved as described for the evolution of
metabolic cross-feeding\textsuperscript{26}, as the exchange of carbon and nitrogen against other products is generally common in photosynthetic or nitrogen-fixing organisms\textsuperscript{27}.

The trait \textsc{hormogonia} is inferred to occur simultaneously with the appearance of the \textsc{filamentous} morphology. Its co-occurrence with other traits, such as \textsc{not free-living benthic/sessile, epilithic/endothytic, sheath}, and \textsc{mucilage} underline the transition from a planktonic to a benthic lifestyle (phase ii). The differentiation of \textsc{hormogonia} can be induced by environmental stimuli, such as nitrogen deprivation\textsuperscript{28}. After breaking off from the mother filament at the necridia, \textsc{hormogonia} disperse via gliding motility or float thanks to \textsc{gas vesicles}, ensuring the reproduction of benthic species\textsuperscript{29,30}. The differentiation into \textsc{hormogonia} is reversible – they change back to the sessile lifestyle, where they grow into a new vegetative filament\textsuperscript{28}. Here we observe the emergence of a two-phases life cycle, which is important for the transition to multicellularity\textsuperscript{31,32}.

Traits that are indicative of higher complexity emerged late in the evolution of cyanobacteria and dominate (phase iv) of the polarity matrix. We observe the occurrence of \textsc{akinetes} and the irreversibly differentiated \textsc{heterocysts}. \textsc{Heterocysts} represent not only a morphological adaptation to the obstacle of N\textsubscript{2} fixation under oxic conditions but also an elaborate and highly specialized communication and metabolite exchange system. In \textit{Anabaena} sp., for example, where several hundred cells communicate within a filament, a regular heterocyst formation pattern along the filament must be achieved to guarantee that every cell is adequately supplied with fixed nitrogen compounds\textsuperscript{19}. For this, the inhibitory signalling peptide PatS needs to be distributed along the filament with heterocyst formation occurring only in cells with low PatS concentration\textsuperscript{33}. Whether the exchange of metabolites and regulators happens via the continuous periplasm\textsuperscript{34,35} or through septal junctions\textsuperscript{36,37} is still not fully resolved.
Another central innovation that occurs in this phase of the polarity matrix is the ability to FISSION IN MULTIPLE PLANES. This trait co-occurred with the ability to produce BAEOCYTES, differentiated cells, which are the reproductive stages in the order Pleurocapsales\textsuperscript{38}. Notably, baeocyte-forming cyanobacteria, that have been traditionally grouped together with unicellular cyanobacteria\textsuperscript{29}, appear to immediately predate the evolution of AKINETES and HETEROCYSTS and thus emerge much later than filamentous forms. The ability to FISSION IN MULTIPLE PLANES is known to underlie the TRUE-BRANCHING morphology, where cells in a filament divide in more than one plane, and which is the last trait to emerge in our analysis. Members of the Haphalosiphon/Stigonematales clade having true-branching and multiseriate filament morphology are considered as the latest evolutionary innovations\textsuperscript{29} and this is further observed in cyanobacterial phylogenies\textsuperscript{39,40}.

Common features of evolutionary transitions in individuality comprise cooperation between the lower level units\textsuperscript{5-7} and the division of labour\textsuperscript{6,9}. The latter might be of particular advantage, and serve as the driver of the transition to multicellularity when there is a strong trade-off between processes that cannot be performed in a single cell at one time\textsuperscript{9,10}. Our current findings support this theory and point to nitrogen fixation, and its incompatibility with photosynthesis, as the trigger for the evolution of multicellularity in cyanobacteria. One open question concerns how the underlying genetics of novel traits, such as the division of labour, arise within a newly emerging multicellular individual. In the case of cyanobacteria multicellularity, we suggest that no new genes were required and that higher complexity was achieved by regulatory changes in gene expression patterns. Basic communication and metabolite exchange was pre-existing as single-celled bacteria frequently engage in cell-cell communication and cross-feeding of metabolites via the external environment\textsuperscript{26}. 
Division of labour between photosynthesis and nitrogen fixation was likely first established by the regulatory mechanism of temporal switching. Once simple forms of division of labour and metabolic exchange existed, the transition into spatial separation in differentiated cells could have evolved mainly by regulatory modifications.

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

**METHODS**

**Data**

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

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

Protein families and gene trees

The cyanobacteria protein families were constructed from completely sequenced genomes available in RefSeq database\(^1\) (ver. May 2016). For the construction of protein families, at the first stage, all protein sequences annotated in the genomes were blasted all-against-all using stand-alone BLAST\(^2\) ver. 2.2.26. Protein sequence pairs that were found as reciprocal best BLAST hits (rBBHs)\(^3\) with a threshold of E-value ≤ 1x10\(^{-5}\) were further compared by global alignment using needle\(^4\). Sequence pairs having ≥30% identical amino acids were clustered into protein families using the Markov clustering algorithm (MCL)\(^5\) ver. 12-135 with the default parameters. By requiring a gene to be present in the genomes of the SynProCya clade and in at least one other major cyanobacterial clade, we identified 1671 gene families that are present on both sides of the cyanobacterial root, i.e., ancient proteins families. Protein sequences of these families were aligned using MAFFT version 7.027b employing the L-INS-i strategy\(^6\).
Maximum likelihood trees were reconstructed with PhyML version 20120412 with parameters -b -4 -v e -m LG -c 4 -s SPR.

**Trait ordering**

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

**Polarization evidence from a single phylogenetic tree**

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

Next we conduct a MAD analysis\(^{48}\), yielding an Ancestor Deviation (AD) statistic for every branch of the tree. The AD measure quantifies the amount of lineage rate heterogeneity that is induced by postulating a branch as harbouring the root of tree. We have previously shown that the AD measure provides robust evidence for the inference of the root of the tree. In the current study we do not infer a single root, but use the AD measure to assess the relative strength of alternative rootings of the same tree.
Thus, for each possible root position we obtain a polarity state for \( A \) and \( B \) and a corresponding AD value. For each of the four polarities, we take the minimal AD value as the tree support of that polarity. Note that for certain trees and trait pairs, some polarities may not occur. We differentiate between two such possibilities. First, a certain tree is Uninformative regarding a specific trait-pair polarity if the species composition of the tree renders the observation of the polarity impossible in any possible topology. For example, the polarity \( 'A \) precedes \( B' \) is impossible to observe in trees where the \( A \) OTUs are a subset of the \( B \) OTUs, regardless of the specific topology or root position. In contrast, a certain polarity may be unobserved in any of the rootings of a specific unrooted tree, while still being a possible observation for a different tree topology. An example of such Informative absent-observation is when \( A \) and \( B \) label disjoint sets of OTUs yet the \( 'Not nested' \) polarity is not observed in the actual tree topology. In the latter case, the absence of an observation is evidence against that specific polarity, and the polarity is assigned the maximal observed AD as its score. When the tree is uninformative regarding a polarity, on the other hand, it is assigned a ‘missing’ value, and is excluded from subsequent analyses.

Statistical inference of pairwise polarity

Repeating the preceding procedure for trees derived from all different single-copy proteins families yields a phylogenomic sample of four variables - the AD scores of each of the four polarities for the pair of traits. The four distributions are paired, as each tree brings in a 4-tuple of values, and a significant difference in support values can be tested using the non-parametric Mann-Whitney U test\(^{58} \). In all, we conduct twelve one-sided tests of contrasting polarities, while employing an FDR\(^{59} \) correction for multiple comparisons. A trait-pair is considered polarized if there exists one polarity where all
three tests against the other polarities recover significantly lower AD values at the 1% FDR level. In exceptional situations we again encounter polarity contrasts that cannot be tested. For some polarity contrasts the sample size may not meet our threshold (10 paired observations), and the test is conservatively considered valid but not significant with a p-value of 1.0. A second exception occurs when one trait (A) is present only in a strict subset of the species that possess the other trait (B). In this case it is impossible to observe the polarities 'not nested' and 'A precedes B' in any tree, and we restrict the testing to the single contrast 'B precedes A' versus 'Simultaneous'.

The pairwise polarity inference is applied to each pair of traits and summarized in a trait-pair polarity matrix. We again apply the FDR correction, this time over all trait-pairs and polarity contrasts. In the present study, with 25 traits, we apply FDR over 3,480 tests. To derive an ordering of the traits, we apply 'Topological sort' to the significant polarities of type 'A precedes B', or vice versa. In the present study the significant polarities form a partial order, i.e., there are no self-contradicting precedence cycles, and the topological sorting order is used to order the polarities matrix and to reduce it to a feed-forward network (Fig. 2).

Data availability

The alignments and trees are available on our webpage at www.uni-kiel.de/genomik/ressourcen
References

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

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

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

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

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

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

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

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

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

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

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

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

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

14. Claessen, D., Rozen, D. E., Kuipers, O. P., Søgaard-Andersen, L. & van Wezel, G. P. Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. *Nat. Rev. Microbiol.* 12, 115–124 (2014).
15. Tarnita, C. E., Taubes, C. H. & Nowak, M. A. Evolutionary construction by staying
together and coming together. *J. Theor. Biol.* **320**, 10–22 (2013).

16. Karatan, E. & Watnick, P. Signals, regulatory networks, and materials that build and
break bacterial biofilms. *Microbiol. Mol. Biol. Rev.* **73**, 310–347 (2009).

17. Vlamakis, H., Aguilar, C., Losick, R. & Kolter, R. Control of cell fate by the formation of
an architecturally complex bacterial community. *Genes Dev.* **22**, 945–953 (2008).

18. van Gestel, J., Vlamakis, H. & Kolter, R. Division of Labor in Biofilms: the Ecology of
Cell Differentiation. *Microbiol. Spec.* **3**, 1–24 (2015).

19. Herrero, A., Stavans, J. & Flores, E. The multicellular nature of filamentous
heterocyst-forming cyanobacteria. *FEMS Microbiol. Rev.* **40**, 831–854 (2016).

20. Frías, J. E., Flores, E. & Herrero, A. Requirement of the regulatory protein NtcA for
the expression of nitrogen assimilation and heterocyst development genes in the
cyanobacterium *Anabaena* sp. PCC 7120. *Mol. Microbiol.* **14**, 823–832 (1994).

21. Zehr, J. P. Nitrogen fixation by marine cyanobacteria. *Trends Microbiol.* **19**, 162–173
(2011).

22. Kasting, J. F. & Siefert, J. L. Biogeochemistry. The nitrogen fix. *Nature* **412**, 26–27
(2001).

23. Gallon, J. R. The oxygen sensitivity of nitrogenase: a problem for biochemists and
micro-organisms. *Trends Biochem. Sci.* **6**, 19–23 (1981).

24. Mitsui, A. *et al.* Strategy by which nitrogen-fixing unicellular cyanobacteria grow
photoautotrophically. *Nature* **323**, 720–722 (1986).

25. Flores, E., Picossi, S., Valladares, A. & Herrero, A. Transcriptional regulation of
development in heterocyst-forming cyanobacteria. *BBA - Gene Regul. Mech.* (2018).

26. D’Souza, G. *et al.* Ecology and evolution of metabolic cross-feeding interactions in
bacteria. *Nat. Prod. Rep.* **35**, 455–488 (2018).
27. Kaiser, D., Kowalski, N., Böttcher, M. E., Yan, B. & Unger, D. Benthic nutrient fluxes from mangrove sediments of an anthropogenically impacted estuary in Southern China. *J. Mar. Sci. Eng.* **3**, 466–491 (2015).

28. Flores, E. & Herrero, A. Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nat. Rev. Microbiol.* **8**, 39–50 (2010).

29. Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. & Stanier, R. Y. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Microbiol.* **111**, 1–61 (1979).

30. Herdman, M. & Rippka, R. Cellular differentiation: hormogonia and baeocytes. *Methods Enzymol.* **167**, 232–242 (1988).

31. Rainey, P. B. & Kerr, B. Cheats as first propagules: A new hypothesis for the evolution of individuality during the transition from single cells to multicellularity. *BioEssays* **32**, 872–880 (2010).

32. Hammerschmidt, K., Rose, C. J., Kerr, B. & Rainey, P. B. Life cycles, fitness decoupling and the evolution of multicellularity. *Nature* **515**, 75–79 (2014).

33. Yoon, H.-S. & Golden, J. W. Heterocyst pattern formation controlled by a diffusible peptide. *Science* **282**, 935–938 (1998).

34. Flores, E., Herrero, A., Wolk, C. P. & Maldener, I. Is the periplasm continuous in filamentous multicellular cyanobacteria? *Trends Microbiol.* **14**, 439–443 (2006).

35. Mariscal, V., Herrero, A. & Flores, E. Continuous periplasm in a filamentous, heterocyst-forming cyanobacterium. *Mol. Microbiol.* **65**, 1139–1145 (2007).

36. Mullineaux, C. W. *et al.* Mechanism of intercellular molecular exchange in heterocyst-forming cyanobacteria. *EMBO J.* **27**, 1299–1308 (2008).

37. Bornikoel, J. *et al.* Role of two cell wall amidases in septal junction and nanopore formation in the multicellular cyanobacterium *Anabaena* sp. PCC 7120. *Front. Cell.*
Infect. Microbiol. 7, 92–15 (2017).

38. Waterbury, J. B. & Stanier, R. Y. Patterns of growth and development in pleurocapsal cyanobacteria. Microbiol. Rev. 42, 2–44 (1978).

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

40. Shih, P. M. et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc. Natl. Acad. Sci. USA 110, 1053–1058 (2013).

41. Giovannoni, S. J. et al. Evolutionary relationships among cyanobacteria and green chloroplasts. J. Bacteriol. 170, 3584–3592 (1988).

42. Baracaldo, P. S., Hayes, P. K. & Blank, C. E. Morphological and habitat evolution in the cyanobacteria using a compartmentalization approach. Geobiol. 3, 145–165 (2005).

43. Schirrmeister, B. E., Antonelli, A. & Bagheri, H. C. The origin of multicellularity in cyanobacteria. BMC Evol. Biol. 11, 45 (2011).

44. Uyeda, J. C., Harmon, L. J. & Blank, C. E. A comprehensive study of cyanobacterial morphological and ecological evolutionary dynamics through deep geologic time. PLoS ONE 11, e0162539–32 (2016).

45. Tomitani, A., Knoll, A. H., Cavanaugh, C. M. & Ohno, T. The evolutionary diversification of cyanobacteria: Molecular–phylogenetic and paleontological perspectives. Proc. Natl. Acad. Sci. USA 103, 5442–5447 (2006).

46. Shi, T. & Falkowski, P. G. Genome evolution in cyanobacteria: The stable core and the variable shell. Proc. Natl. Acad. Sci. USA 105, 2510–2515 (2008).

47. David, L. A. & Alm, E. J. Rapid evolutionary innovation during an Archaean genetic expansion. Nature 469, 93–96 (2010).
48. Tria, F. D. K., Landan, G. & Dagan, T. Phylogenetic rooting using minimal ancestor deviation. *Nat. Ecol. Evol.* **1**, 0193 (2017).

49. Tomitani, A., Knoll, A. H., Cavanaugh, C. M., & Ohno, T. The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. *Proc. Natl. Acad. Sci. USA* **103**, 5442-5447 (2006).

50. Wolk, C. P., Ernst, A. & Elhai, J. in *The molecular biology of cyanobacteria* 769–823 (Springer, Dordrecht, 1994).

51. O’Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–45 (2016).

52. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

53. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278**, 631–637 (1997).

54. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Gen.* **16**, 276–277 (2000).

55. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).

56. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

57. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

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

**Cell types**

| Cell types | Description |
|------------|-------------|
| 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 N₂. |
| 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**

| Trait | Description |
|-------|-------------|
| UNICELLULAR | Single-celled morphology. After cell division cells separate. |
| FILAMENTOUS* | Multi-celled morphology. Cells remain attached after cell division. |
| NITROGEN FIXATION | Fixation of N₂ 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**

| Environment | Description |
|-------------|-------------|
| 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. |

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

**SMALL CAPS** indicate the traits that have been used in the analysis.
| Debate and resolution | Description of debate |
|-----------------------|-----------------------|
| **Unicellular – Filamentous** | Whereas early work 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. Another view is that the filamentous morphology evolved early during cyanobacterial evolution and was subsequently lost and regained several times. |
| **Planktonic - Benthic** | Whether the cyanobacteria ancestor was planktonic is a matter of debate and opposing views on the topic have been published. |
| **N\textsubscript{2} fixation – No N\textsubscript{2} fixation** | Whereas some studies claimed the last cyanobacterial common ancestor to fix N\textsubscript{2}, there are others that concluded that it could not fix N\textsubscript{2} and that cyanobacteria must have acquired this trait several times independently. |
| **Freshwater – Marine** | Some studies suggest that early cyanobacteria lived in freshwater and subsequently diverged into marine environments, whereas others provide evidence in support of a marine origin. The cyanobacteria ancestor most likely inhabited an aquatic environment and colonized both environments early. |
| **Akinetes – Heterocysts** | There is a common agreement that these cell types appeared late in cyanobacterial evolution, but there is a controversy about whether they shared a common ancestor and appeared simultaneously or successively. |

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

| 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\textsuperscript{-110}). |
| Our results show that the cyanobacteria ancestor included PLANKTONIC, MOTILITY and FREE LIVING (all FDR adjusted U-test p-values <3.2×10\textsuperscript{-69}, see Supplementary table 2). |
| Trait-pair polarity tests show that N\textsubscript{2} fixation is a derived trait and that the ancestor of cyanobacteria was lacking the ability to fix N\textsubscript{2} (FDR adjusted U-test p-value 2.6×10\textsuperscript{-125}). |
| 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\textsuperscript{-140}). |
| Our results show that AKINETES and HETEROCYSTS emerged simultaneously (FDR adjusted U-test p-value 1.8×10\textsuperscript{-124}). |
**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\(^4\)). 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
four phases of trait emergence. Traits that can be regarded as multicellularity markers in cyanobacteria are labelled in red (Table 1).