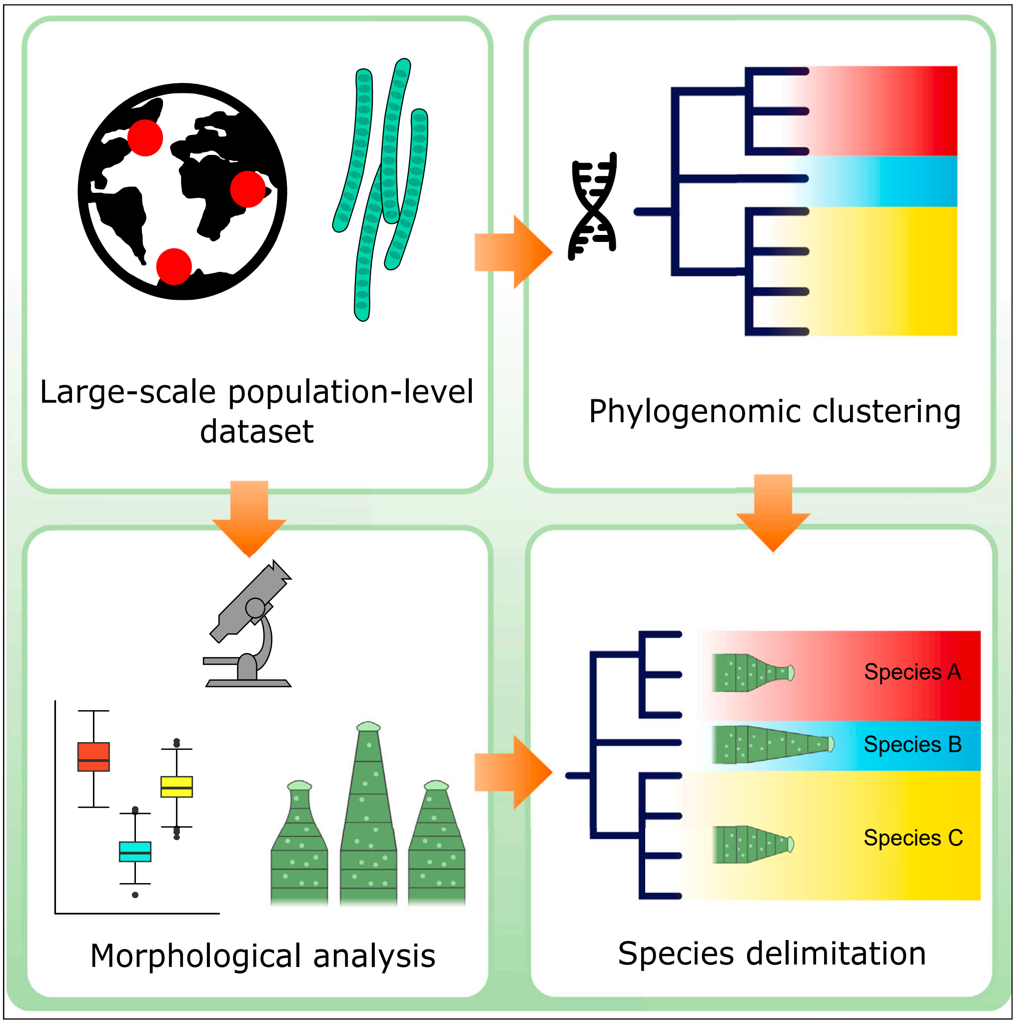
iScience

CellPress
OPEN ACCESS

Article

Population genomics and morphological data bridge the centuries of cyanobacterial taxonomy along the continuum of *Microcoleus* species



Svatopluk Skoupý, Aleksandar Stanojković, Dale A. Casamatta, ..., Anne D. Jungblut, Hester van Schalkwyk, Petr Dvořák

p.dvorak@upol.cz

Highlights

Soil cyanobacterium *Microcoleus* is among the most important primary producers

Microcoleus represents a speciation continuum of at least 12 distinct lineages

Genomic diversification in *Microcoleus* is reflected by morphological changes

Filament apices shape may have adaptive function to environmental conditions

Skoupý et al., iScience *27*, 109444
April 19, 2024
https://doi.org/10.1016/j.isci.2024.109444

iScience

CellPress
OPEN ACCESS

Article

Population genomics and morphological data bridge the centuries of cyanobacterial taxonomy along the continuum of *Microcoleus* species

Svatopluk Skoupý,[1] Aleksandar Stanojković,[1] Dale A. Casamatta,[3] Callahan McGovern,[3] Ana Martinović,[1] Jiří Jaskowiec,[1] Miriam Konderlová,[1] Viktória Dodoková,[1] Pavla Mikesková,[1] Eva Jahodářová,[2] Anne D. Jungblut,[4] Hester van Schalkwyk,[4] and Petr Dvořák[1,5,*]

SUMMARY

The filamentous cyanobacterium *Microcoleus* is among the most important global primary producers, especially in hot and cold desert ecosystems. This taxon represents a continuum consisting of a minimum of 12 distinct species with varying levels of gene flow and divergence. The notion of a species continuum is poorly understood in most lineages but is especially challenging in cyanobacteria. Here we show that genomic diversification of the *Microcoleus* continuum is reflected by morphological adaptation. We compiled a dataset of morphological data from 180 cultured strains and 300 whole genome sequences, including eight herbarium specimens and the type specimen of *Microcoleus*. We employed a combination of phylogenomic, population genomic, and population-level morphological data analyses to delimit species boundaries. Finally, we suggest that the shape of the filament apices may have an adaptive function to environmental conditions in the soil.

INTRODUCTION

Microcoleus is among the most abundant and significant terrestrial organisms on Earth. This cosmopolitan, filamentous cyanobacterium, known for its characteristic bundle-forming structure, plays a prominent role as a pioneering organism.[1] Predominantly inhabiting benthic and subaerophytic habitats, it is likely the most crucial component of biological soil crusts around the world.[2] Soil crusts represent intricate ecosystems consisting of a variety of cyanobacteria, bacteria, algae, and fungi, which cover the top layers of soil, especially in arid or disturbed environments. According to some estimates, biological soil crusts cover up to 24% of global terrestrial surfaces,[3] and in some regions, such as tropical or polar deserts, biological soil crusts can cover up to 90% of the ground surface.[4] Cyanobacteria belonging to the genus *Microcoleus* form a key component in such soil crusts, where their filaments and mucilaginous sheaths help stabilize the soil particles, retain water, and create a habitat for other organisms to gain a foothold.[5] The flagship species of this genus, *Microcoleus vaginatus,* has been identified from a wide range of climates, from polar to tropical regions. Widespread in temperate climates,[6] it is particularly vital as a primary producer in hot and cold deserts where, due to extreme climate and the absence of many vascular plants, soil crusts form the majority of primary production. A hypothesis proposed by Garcia-Pichel et al.[7] suggests that the calyptrate apices of the filaments, characteristic of *M. vaginatus*, aid in the burrowing process. The shape of the calyptra and filament apex could thus have an adaptive function, dependent on soil properties and environmental factors such as bulk soil density, average temperature, level of precipitation etc. However, the adaptive function of the calyptra has not yet been rigorously studied to date.

Despite several studies demonstrating that the genetic diversity within the species is large and it can be misidentified with similar species,[8] *Microcoleus vaginatus* is still mostly considered as a single species in diversity and ecological studies. Most recently, Stanojković et al.[6] explored the divergence and speciation patterns of 201 *Microcoleus* genomes, showing that *Microcoleus* represents a global continuum of species. This continuum consists of at least 12 species with varying degrees of divergence and gene flow. The diversification of these species began ca. 29.6 million years ago and coincided with aridification and adaptations to stress factors.[6]

In cyanobacteria, and prokaryotes in general, it is still unclear whether distinct species truly exist, or if vast population sizes, coupled with high mutation rates and high levels of dispersion, lead to the existence of more loosely definable fuzzy species and speciation continuums (e.g.,[9–11]). The concept of a speciation continuum has been previously employed in plant and animal models (e.g.,[12–14]). Stankowski and Ravinet[15] articulated this as a continuum of species separated by a continual level of reproductive isolation in accordance with the biological

---

[1] Department of Botany, Faculty of Science, Palacký University Olomouc, Šlechtitelů 27, 783 71 Olomouc, Czech Republic
[2] Department of Biology, Faculty of Education, Palacký University Olomouc, Purkrabská 2, 771 40 Olomouc, Czech Republic
[3] Department of Biology, College of Arts and Sciences, University of North Florida 1 UNF Drive, Jacksonville, FL 32224, USA
[4] Department of Science, Natural History Museum, Cromwell Road, London SW7 5BD, Great Britain
[5] Lead contact
*Correspondence: p.dvorak@upol.cz

https://doi.org/10.1016/j.isci.2024.109444

species concept (BSC[16,17]). Cyanobacteria and other prokaryotes are asexual but are subjected to episodic homologous recombination (HR) similar to sexual organisms, which is present in ca. 80% of prokaryotic species.[18,19] Prokaryotic HR may lead to similar patterns of evolution as seen in sexual organisms if the rates of HR are higher than the mutation rate. Thus, cyanobacteria and other prokaryotes may be considered quasi-sexual.[20] Consequently, a revised interpretation of the BSC can be employed, wherein the reproductive barriers between species are defined more flexibly as impediments to gene flow.[19] Similarly, the speciation continuum in prokaryotes can be understood as a continuum in gene-flow barriers.[6]

In our view, it is also important to use large datasets and population-genomic data to address questions of speciation. Such datasets are still uncommon in cyanobacteria and have, so far, only been explored in a few cosmopolitan, coccoid cyanobacteria such as *Prochlorococcus*, *Synechococcus*, and *Microcystis* (reviewed by Dvořák P. et al.[21]). Cyanobacterial species emerge by complex processes,[22–24] and population genomic data may help better articulate the processes that drive speciation and population structures in nature (e.g., Kashtan N. et al. and Stanojković A. et al.[25,26]). Large datasets may also help in assessing apomorphies, as some morphological traits important for species delimitation may only rarely be present or environmentally inducible.[27–30] It is also important to include herbarium specimens, especially those representing types, if possible. Herbarium collections represent an untapped source of data on the historical diversity of cyanobacteria (e.g., specimens collected in the 1700s) and sequencing herbarium specimens can help bridge the historical and current diversity. This approach has been employed with higher plants (e.g., Nevill P.G. et al.[31]), but has not been implemented to a larger extent in prokaryotes. While some prokaryotic, ancient genomes have been sequenced (e.g., *Yersinia pestis* from human remains[32]), scant genetic data exists from historical sources for cyanobacteria and other free-living prokaryotes.[33] In fact, the only historical herbarium-derived genomes of cyanobacteria were sequenced by Dvořák et al.[33] and Jungblut et al.,[34] who obtained two genomes of *Nostoc*, and Stanojković et al.,[6] who sequenced 8 herbarium specimens of *Microcoleus* up to two hundred years old (included in this paper). *Microcoleus* is an ideal cyanobacterium for obtaining herbarium genomes, due to its cosmopolitan distribution and ability to grow copious biomass.

In our previous study, we obtained whole-genome sequences of 201 cultivated strains of *Microcoleus* spp. (we will use *Microcoleus* for short) from various geographical locations. We strove to infer the overall population structure and tested several clustering algorithms, but only one revealed clusters coinciding with the monophyletic clades in the tree (see details by Stanojković A. et al.[6]). Whole genome phylogenies and phylogenomic clustering based on optimized fastBAPS resulted in 13 distinct and well-supported clades. Notably, one clade was composed of a single genome, therefore disregarded for the population genomic analyses and morphological analyses. In this paper, we seek to articulate barriers between the remaining 12 putative species found along the speciation continuum that is currently *Microcoleus*, using population genomic, ecological, and morphological data. We will bridge hundreds of years of taxonomic knowledge by connecting the genome sequences of the type materials sampled centuries ago with the current species for the first time. We will demonstrate that articulating species boundaries in a speciation continuum can be laborious, but it is possible with extensive population genomics sampling and exhaustive phenotypic characterization.

RESULTS

We used two datasets labeled Dataset I and II. Dataset I is comprised of 201 genomes of cultivated strains of *Microcoleus* from various geographical locations. This dataset was used for morphological analysis and identification of putative species within the *Microcoleus* speciation continuum. Dataset II included 201 strains, 102 genome and metagenome sequences, and 8 herbarium genome sequences of *Microcoleus*. This dataset was used for expanded phylogenomic analysis.

Identification of species within *Microcoleus* speciation continuum

In our previous paper,[6] we employed Dataset I to test several methods of obtaining a consistent tree topology: single copy orthologues of protein and nucleotide sequences, species tree based on the genes from single copy protein orthologues using ASTRAL III,[35] single nucleotide polymorphisms over the whole genome. Finally, the recombinant sites were filtered from the alignment using Gubbins.[36] All methods revealed consistent and well-supported tree topologies. Thus, we included here the whole-genome phylogeny based on 2020 single-copy protein orthologues (Figure 1).

We also sought to infer the overall population structure, employing phylogenomic clustering which resulted in 13 distinct clades (M1–M13, Figure 1). Some clades were comprised of only a few strains from the same geographical region (e.g., clade M1, consisting of four strains; Table S1). Other clades, such as clade M6 and the most numerous clade M5 (81 strains), consisted of strains from disparate geographical locations and habitats. One clade (M13) was composed of a single genome, therefore disregarded for the population genomic analyses and morphological analyses.

Analysis of quantitative traits reveals morphological continuum

We documented traits such as cell dimensions, morphology of apical cells, color, presence and thickness of sheaths, cell content, etc. (Figure S1). Overall, the morphological variability of the 180 strains was higher than suspected (Table S2). The morphology was highly variable between the clades, but relatively stable within clades, with a notable exception of clades M5 and M10, which had high intra-clade morphological variability. Cell width varied highly, from ca. 3 μm up to 10 μm (Figure 2A). Cell length was also quite varied, from 1 μm up to 10 μm (Figure 2B). Cells were generally wider than long, with the exception of clades M10 and M2, where cells were often nearly isodiametric and rarely even longer than they were wide. In general, the difference in cell dimensions was significant ($p < 0.01$) between most clades, with some

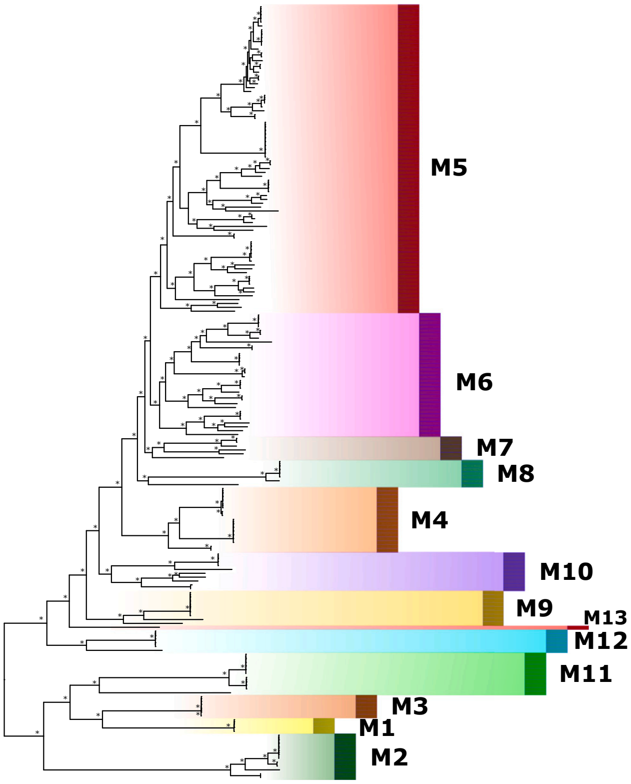


Figure 1. Phylogenomic inference of 201 *Microcoleus* genomes (Dataset I) and clustering analyses FastBAPS

Genomes clustered into 13 distinct clades labeled M1 to M13. Ultra-fast bootstrap support values of >98 are denoted with an asterisk.
See also Figures S1 and S3 and Tables S1 and S2.

notable exceptions (Table S3). The difference between clade M5 and clades M6, M4, and M11 was statistically insignificant (M5/M6 – p = 0.3517 cell width, 0.9987 cell length; M4/M11 – p = 0.9998 cell width, 0.8287 cell length) in both cell width and length, as well as the difference between clade M11 and clades M4 and M9. Between some clades, only the cell width was significant (e.g., clades M5 and M1, p = 1.76E-05), while in between other clades, only the difference in cell length was significant (e.g., clades M7 and M2, p = 0.000306).

The shape of calyptra and apical cells appear to be highly variable between the studied clades (Figures 3 and 4), but relatively stable within clades. A pointiness value was developed (see STAR Methods for details) to approximate the shape of the filament apex as a single value. The pointiness varied significantly, from 0.07 (the pointiest) to ca. 1 (the most round) (Figure 5A). The overall pointiest clades were clades M2 and M12, with the average pointiness of 0.2 and 0.21, while the "bluntest" was clade M1, with the average pointiness of 0.59. In general, the difference in pointiness was significant between most clades (p = 3.234E-73), with some notable exceptions, similar to the cell dimensions (Table S4). Notably, the difference between clades M5, M6, and M7 was not statistically significant pertaining to neither pointiness nor number of cells, while clades M12, M11, M1, and M2 were statistically different in pointiness value from nearly all other clades. Clade M2 was also significantly different from all other clades in the number of cells per attenuating segment (Figure 4N). While other clades average at 3.5–5.3 cells per attenuating segment, clade M2 averages at 7.5 cells per segment.

Adaptive function of calyptra and filament apices

To explore the hypothesis that the morphology of filament apices is dependent on the soil properties and that species are adapted to different environmental conditions, we performed a correlation study. In general, there was a statistically significant correlation between the pointiness value of filament apices and certain environmental variables (Table 1). The most significant correlation was linked to cation exchange capacity (at pH 7, mmol/kg) and bulk soil density (cg/m^3) (Table 1). Soil organic carbon (dg/kg), solar radiation and UV (annual mean UV-B and sum of monthly mean UV-B during highest quarter), and temperature (average temperature, maximum temperature, minimum temperature, annual temperature, minimal temperature of the coldest month, temperature warmest quarter, and temperature coldest quarter) were also strongly correlated (Table 1).

Analysis of qualitative traits further supports the morphological continuum

Strain color varied from light green to dark green, but also brown-green, bright blue, and rarely even purple. Color of the trichomes was relatively stable within clades. Sheaths varied from unobservable, to thin and diffluent, to distinct and thick (in some cases even thicker than the

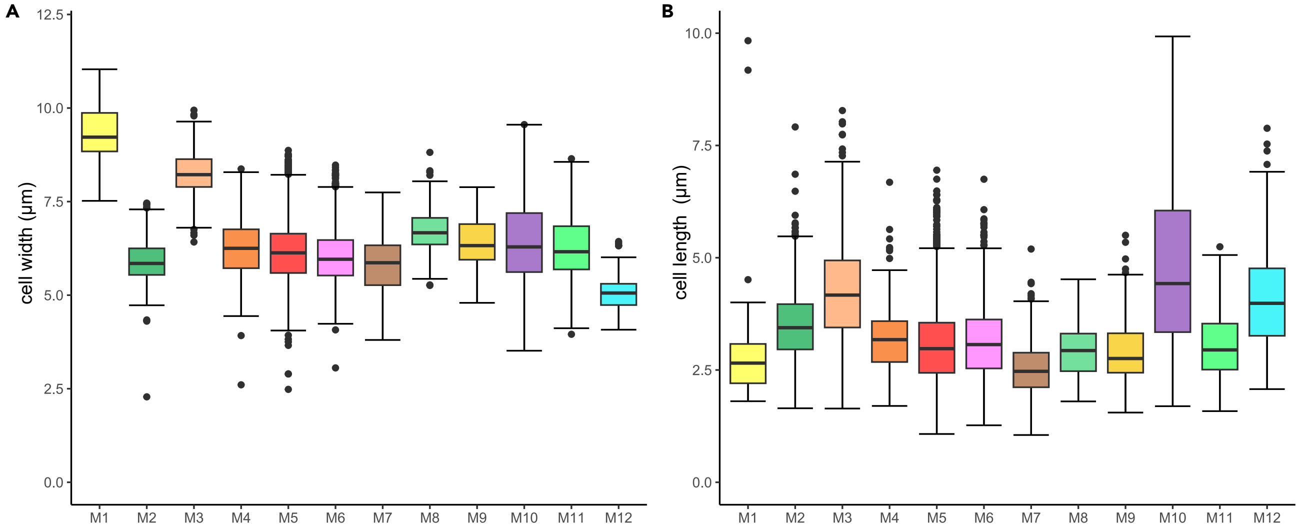


Figure 2. Cell dimension variability

Box-plots of cell widths (A) and lengths (B) within the investigated clades M1-M12.
See also Table S3.

trichome itself and layered) (Figure S1). In strains with layered sheaths, a spiral splitting of the sheath was usually observed when the sheath exceeded the filament. Filaments were mostly solitary within the sheath, but sometimes two or more filaments were present in a shared sheath. In strains with distinct sheaths, nodules or "knots" were sometimes present (Figure S1). Cell content was mostly homogeneous with small granules often concentrated toward crosswalls. Larger granules were occasionally present, and cell content was sometimes concentrated toward the cell periphery. Some traits, such as thick and layered sheaths or the presence of nodules, seem to appear in the phylogenetic tree in distinct nodes, which might imply a derived trait (Figure S2). Some traits were only observed rarely, meaning only in one or more of the strains within a clade, or only in some filaments in the sample. Such rare traits included long, thin crystals forming around the sheath (present in most clades, Figure S1). An overview of the morphological traits can be found in Table S2.

Expanded phylogeny suggests seven more species within the *Microcoleus* continuum

To further assess the diversity of *Microcoleus*, we searched the NCBI database for available whole-genome assemblies. In total, 102 annotated cyanobacterial genomes were obtained from the NCBI database and included in the Dataset II (Table S5). We also included whole-genome sequences obtained from 8 herbarium specimens (Table S6). The phylogenomic analysis and FastBAPS clustering were performed again. After the dataset expansion, clades M5, M6, and M7 merged together, and 6 new clades emerged (labeled M14 to M21; Figure 6, uncollapsed tree can be found in Figure S2).

Clades M5, 6, and 7 form distinct branches while employing FastBAPS clustering on only our dataset (Dataset I), but upon expanding the dataset with genomes from the NCBI and herbarium specimens, clades M5, 6, and 7 collapsed into single clade. Clade M5 also includes a genome from an herbarium specimen of *M. vaginatus* collected and identified by Drouet in 1938. In our opinion, this clade can be assigned as *Microcoleus vaginatus sensu stricto*, and the herbarium specimen can be set as type. Clades M6 and M7 can either be delimited as diverging species, as subspecies of *M. vaginatus*, or taken together as a species complex of *M. vaginatus*.

Clades M4, M8, M9, M10, and M12 all form distinct and stable clusters both in Dataset I and II, and while there is still some gene flow between these clades and clades M5, 6, and 7, all these clades can be considered distinct species. Notable among these "species" is Clade M10. This clade contains an herbarium-derived genome of *Microcoleus terrestris* (Desmaziéres, approximately 1820–1825), likely the oldest available *Microcoleus* specimen, and thus the official type.

The majority of the metagenomes included in Dataset II clustered into clades M14 and M15 (Clade M14 comprised 55 and clade M15 11 genomes). Clade M21 consisted of one metagenome of *Microcoleus* and one genome of herbarium specimen denoted as *Phormidium subfuscum*. Clade M16 contained four herbarium specimens, two labeled as *Phormidium subfuscum*, one as *Phormidium uncinatum,* and one as *Microcoleus vaginatus* (Table S6). Clade M17 comprised 10 genomes of "uncultured oscillatorian cyanobacteria" and 3 genomes of *Microcoleus anatoxicus*. Clade M18 consisted of two genomes of uncultured *Microcoleus* and one genome denoted as *Tychonema bourrelyi*. Clade M19 comprised two uncultured *Microcoleus* and one strain from Mauritius. The last basal clade M20 comprised two uncultured *Microcoleus* genomes.

We also searched morphological literature for previously described species of *Microcoleus* and closely related genera (such as *Phormidium* or *Oscillatoria*) matching the morphology of our putative species, in an effort to prune the database of unrevised species. Clades, where we were unable to find matching unrevised species, will be described as new species. The formal species descriptions of new species and taxonomic revision of *Microcoleus* will be the subject of future work.

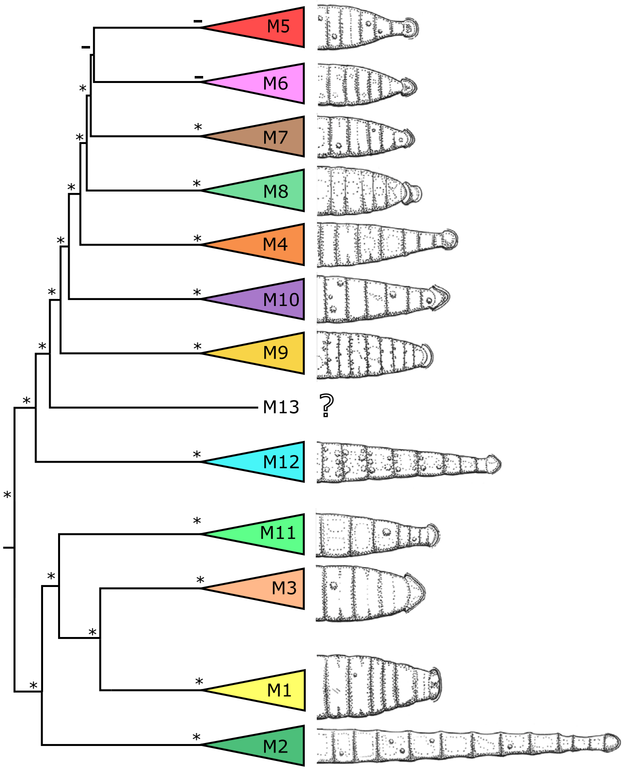


Figure 3. Phylogenomic tree of *Microcoleus* species with corresponding filament apex lineworks

The phylogenomic tree was constructed using maximum likelihood. Nodes with a support >98 are marked with an asterisk. Nodes with lower support are marked by (–). Clades with unknown morphology are denoted by a question mark.

16S rRNA phylogeny

We constructed a 16S rRNA sequence dataset containing all the genera from the order Oscillatoriales and representatives from all cyanobacterial lineages to test which sequences were monophyletic with the *Microcoleus vaginatus sensu stricto.* In total, we obtained 953 sequences of *Microcoleus* and performed phylogenetic analysis. All strains from our dataset (Dataset I) cluster to a monophyletic branch, forming a distinct monophyletic clade among other Oscillatoriales, but the topology differs from the phylogenomic tree (Dataset I, Figure 1). Only the deepest branching lineages coincide (the M4, 5, 6, 7, 8, 9, 10, 12, 13 branch and M1, 2, 3, 11 branch separation), but the topology within these branches differs drastically, as most strains are intermixed within the phylogenetic tree seemingly at random (Figure S3).

16S–23S rRNA ITS secondary structures

To test whether the ITS (internal transcribed spacer) secondary structures can be used as an apomorphy in *Microcoleus,* we reconstructed the ITS secondary structures of the 12 putative species (disregarding clade M13 as it is formed by a single strain). In total, there were 18 distinct secondary structures. Clades M1, 2, 3, 8, 10, and 12 had a single distinct ITS structure, which could be used as an apomorphy. Clades M4 and M9 had each 2 differing ITS structures, which complicates the use of ITS as an apomorphy in these clades. The M5-6-7 superclade had 8 different ITS structures, some of which were also polyphyletic. Thus, the ITS secondary structures cannot be used in the delineation of the clades M5, 6, and 7. In clade M11, we were unable to procure any ITS secondary structures (Figure S2).

DISCUSSION

Studies of population genomics have shown a continuum of species in cyanobacteria, ranging from well-defined (e.g., diverged) to fuzzy, and we propose a solution to this conundrum by delimiting species boundaries within the speciation continuum. We analyzed 201 strains of *Microcoleus* from diverse geographical locations and habitats in order to estimate phylogenetic history and hypothesize putative species along the speciation continuum. Next, we provided phenotypic evidence for putative species and we observed both genomic and phenotypic continua.

Species delineation along a continuum is a rarely explored topic.[6,37] Many researchers consider prokaryotic species as indistinct or "fuzzy".[18,38] In the most restrictive view, prokaryotes do not readily resolve into species but represent a series of transitional forms with

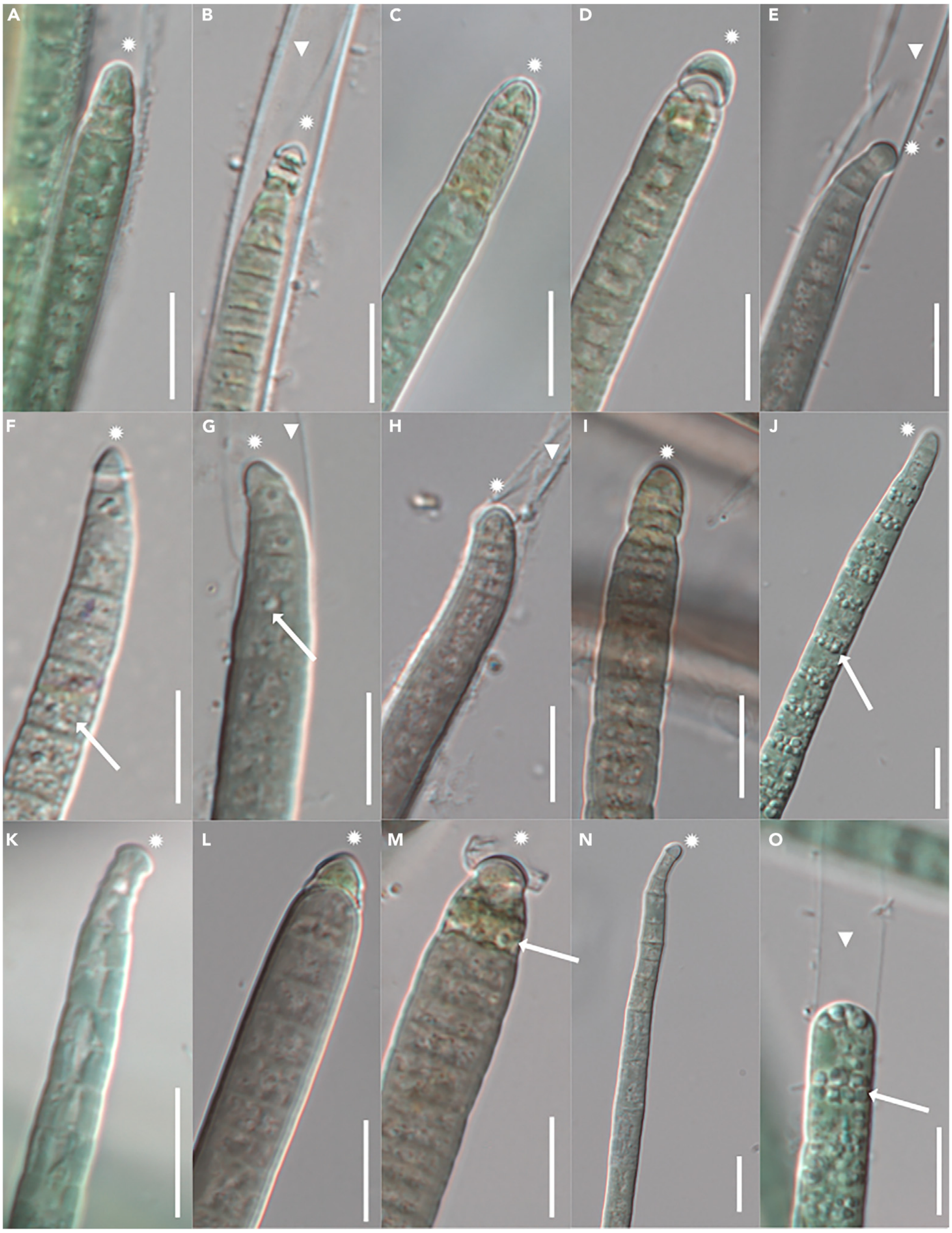


Figure 4. Morphological variability of filament apices and calyptras

A: clade M5, B: clade M6, C: clade M7, D: clade M8, E: clade M4, F and G: clade M10, H and I: clade M9, J: clade M12, K: clade M11, L: clade M3, M: clade M1, N: clade M2, O: filament without calyptra. Scale, 10 μm, asterisk = calyptra, arrowhead = granules, triangle = empty sheath.

different levels of divergence from one another, reflecting their past history.[9,10] However, we will show that "fuzzy" does not necessarily mean undefinable.

A commonly used criterion for species delimitation is the similarity of the 16S rRNA gene.[21,39] However, while the deeper branches in the 16S rRNA phylogenetic tree match those in the phylogenomic tree (i.e., the M4, 5, 6, 7, 8, 9, 10, 12, 13 and M1, 2, 3, 11 branches), the topology of the tree differs drastically (Figure S3). The 16S rRNA marker simply is not sensitive enough to distinguish between closely related species within the speciation continuum.[21,40,41] The commonly used genome-similarity ANI value also proved ineffective for the *Microcoleus* species.[6] The ANI value, like other similarity-based methods, does not take into account the dynamic processes of speciation, and the set boundaries between species are often arbitrary.[6] Employing the ANI analysis, the *Microcoleus* isolates (Dataset I) shared 86.94–99.9% sequence identity

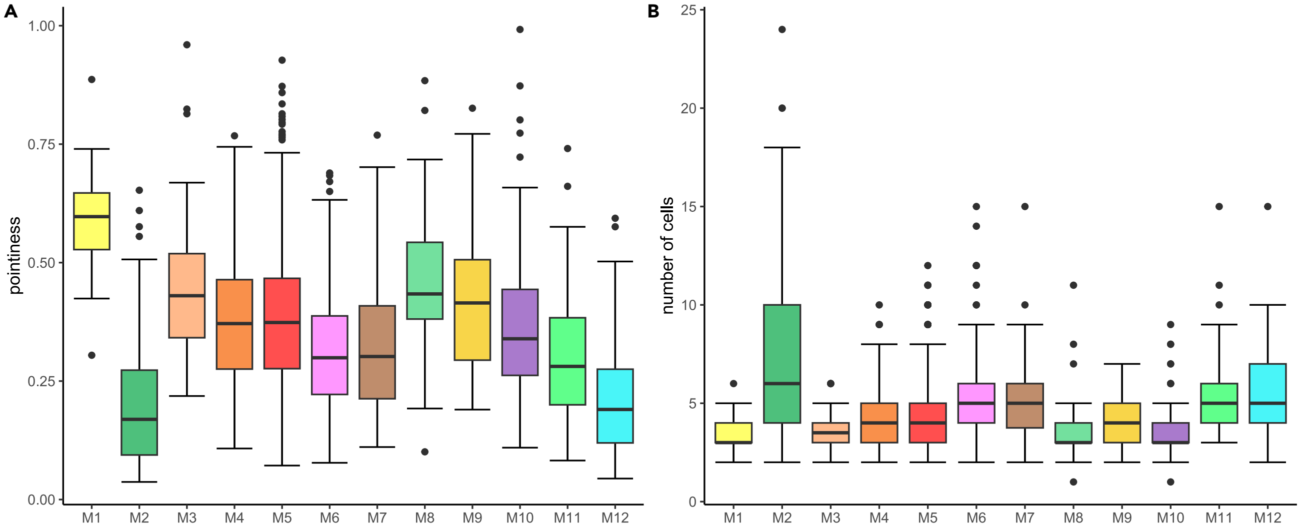


Figure 5. Filament apex shape variability

Box-plots of pointiness values (A) and numbers of cells per attenuating segment of filament apex (B) within the investigated clades. Pointiness value lies on a 0–1 scale, with 1 being the most round.

See also Table S4.

across their genomes, which generated 37 putative "species". The optimized Bayesian analysis we employed produced more conservative 13 (Dataset I) and 19 (Dataset II) putative species, which also followed monophyly in the species trees.[6] The question is: how can these putative species be articulated?

One potential solution could involve applying the biological species concept (BSC).[16–18] As prokaryotes reproduce asexually, the biological species concept has typically been eschewed. The BSC predicts reproductive barriers between species and free sexual reproduction within species. Prokaryotic homologous recombination[42] may produce similar patterns of evolution, making cyanobacteria quasi-sexual organisms. In this view, the gene flow within the prokaryotic species should be higher than between species. As explored in Stanojković et al.,[6] there is a substantial level of gene flow between the putative species of *Microcoleus*. While the mean genome fraction subjected to gene flow between strains of the same species varied from 15 to 53.2% (within the species), the genome fraction affected by gene flow from outside the species varied from 1.5 to 29.04%, meaning the species could be delineated with a biological species concept. The M5, M6, and M7 clades, however, pose an issue, as the genome fraction subjected to recombination is as similar between these species as within.

Perhaps a more fitting concept would be the monophyletic species concept (MSC) *sensu* Johansen and Casamatta[43]—a species is a monophyletic cluster of strains or natural populations that is diagnosable by some unique combination of traits, those traits being any combination of morphological, biochemical, molecular, or other characteristics. Most putative species in our Dataset I could be delineated based on this concept, as they form stable monophyletic clusters in both the un-expanded (Dataset I) and expanded dataset (Dataset II), and possess a distinct apomorphy or unique combination of traits that differentiates them from other clades. Some clades can be differentiated by a single phenotypic trait, such as the putative species M2, which is well supported, monophyletic, and all strains in this clade share a distinct apomorphy in the form of long, very gradually attenuating filament apex (Figures 3 and 4N). Furthermore, all strains in this clade originate in temperate climates, from either soils or puddles, which also suggests that they inhabit the same ecological niche. Clades M1, M3, M4, M8, M9, M11, and M12 could all also be delineated based on the MSC.

The remaining clades, M5, 6, 7, and 10 cannot be easily delineated based on MSC, albeit for differing reasons. Clade M10 forms a stable monophyletic clade in both Dataset I and II but has two rather distinct morphotypes. The sole unifying feature between these morphotypes is the shape of the filament apex. While this feature could be used as an apomorphy, this approach does not take into account the distinct difference in morphology between the two groups in this clade. The M5-6-7 group poses a different issue. These three clades form distinct, separate branches in Dataset I, but upon the inclusion of metagenomes and herbarium genomes (Dataset II), they merge into one "superclade". Furthermore, in the 16S rRNA phylogenetic tree, strains from these clusters were dispersed in the tree without any noticeable pattern (Figure S3). The M5-6-7 clade is composed of strains obtained from varied geographical locations, mainly from Europe, but also from Africa (Mauritius), the USA (Montana and Florida), and India (Ladakh). Strains in this clade also come from varied habitats: from cold, dry deserts to temperate and wet soils. As noted before, there is also a substantial level of gene flow between these 3 clades.[6] This clade also has larger inter-clade morphological variability than other studied clades in Dataset I. In the version of MSC modified for cyanobacteria,[43] secondary structures of ITS can be used as non-morphological apomorphies. In our dataset, however, the ITS secondary structures do not offer any help in resolving the M5-6-7 clade, as they are polyphyletic and incongruent with the species tree (Figure S2). All these factors complicate the delineation of these 3 clades as species in accordance with MSC, as they are likely incipient, insufficiently diverged species.

Table 1. Climatic variables with significant correlation to apex pointiness

Variable	cor value	p value
Average temperature (°C)	0.3644303	4.92E-07
Maximum temperature (average)	0.3776291	1.73E-07
Minimum temperature (average)	0.3425169	2.52E-06
Precipitation (average)	0.2335591	0.001602
Annual temperature	0.364445	4.92E-07
Minimal temperature coldest month	0.3384066	3.38E-06
Temperature warmest quarter	0.3429667	2.44E-06
Temperature coldest quarter	0.3552869	9.88E-07
Annual precipitation	0.2335591	0.001602
Precipitation wettest month	0.2766397	0.00017
Fertilizer global	−0.2174458	0.0126
Land-use-induced reductions in NPP as a percentage of NPP0	−0.3421253	4.88E-05
Total HANPP as a & of NPP0	−0.2470626	0.003485
Human appropriation of NPP in gc/m2/yr	−0.2192003	0.009792
NPP of the actual vegetation in gc/m2/yt	0.2088432	0.01396
NPP remaining in ecosystems after harvest in gC/m^2/yr	0.241554	0.004315
Organic carbon density (g/dm^3)	−0.3364964	0.000705
Soil organic carbon stock (t/ha)	−0.3529854	0.000339
Soil organic carbon (dg/kg)	−0.4379904	1.26E-05
Nitrogen (cg/kg)	−0.2520794	0.01183
Cation exchange capacity at pH 7 (mmol/kg)	−0.4088106	2.66E-05
Bulk density (cg/cm^3)	0.4166859	3.60E-05
Solar radiation (average)	0.3597114	7.07E-07
Vapor (average)	0.2788727	0.00015
Annual mean UV-B	0.2864779	9.67E-05
Sum of monthly mean UV-B during highest quarter	0.2907775	7.50E-05
Sum of monthly mean UV-B during lowest quarter	0.2647439	0.000329
Olsen phosphorus concentration (mg/kg)	−0.1752957	0.02614

The most significant (p) values and the highest correlation values are underlined.

Historically, cyanobacterial species were delineated based mainly on a phenetic species concept: individuals within the species are more similar to each other than outside of the species, which is exceedingly subjective. This concept is perhaps the most intuitive one for the human mind, but with the advent of molecular methods, it has been relegated to the background due to some key issues.[44] Cyanobacteria, in general, have high levels of cryptic diversity, and scant morphological diversity may conceal vast genetic diversity.[21,29,45,46] Further, some features (e.g., sheath characteristics) are likely environmentally plastic. That being said, the importance of well-constructed morphological analysis should not be underestimated. The morphology of many cyanobacterial groups is usually only studied in a handful of strains.[44] Traits that might aid in species identification and delimitation, however, might only be present relatively rarely and might escape capture in smaller datasets.[47–49] Population-level morphological studies might uncover unsuspected morphological diversity and, in turn, aid in understanding the diversity and ecology of species within the speciation continuum.[6] This study, analyzing the morphology of 180 strains of *Microcoleus* may be the most comprehensive one in cyanobacteria to date (perhaps among all prokaryotes).

Microcoleus is predominantly a soil-surface cyanobacterium and as such the shape of the calyptra and filament apexes could aid in burrowing between soil particles.[2] The evolution of the shape of the filament apex could thus have an adaptive function, influenced by the soil composition, size of soil particles, precipitation, etc. In plants, root-cap shape has been tied to environmental factors such as bulk soil density.[50,51] The pointiness of the filament correlated to several environmental factors (Table 1), but the highest correlation was linked to cation exchange capacity (negatively) and bulk soil density (positively). In our case, the positive correlation means that the blunt phenotypes were generally associated with higher soil density, contrary to the findings of Colombi et al.[50] in higher plants. The "bluntness" of the root cap in wheat was negatively correlated with the root elongation rate under high to moderate bulk soil density (1.6–1.45 g/cm^3), meaning the roots with the "pointier" phenotypes were generally growing faster in more densely packed soils than the roots with blunter phenotype. This could possibly be explained by different scales. The root tips of wheat (*Triticum* sp.) are ca. 1 mm in diameter, while the filaments of *Microcoleus* are

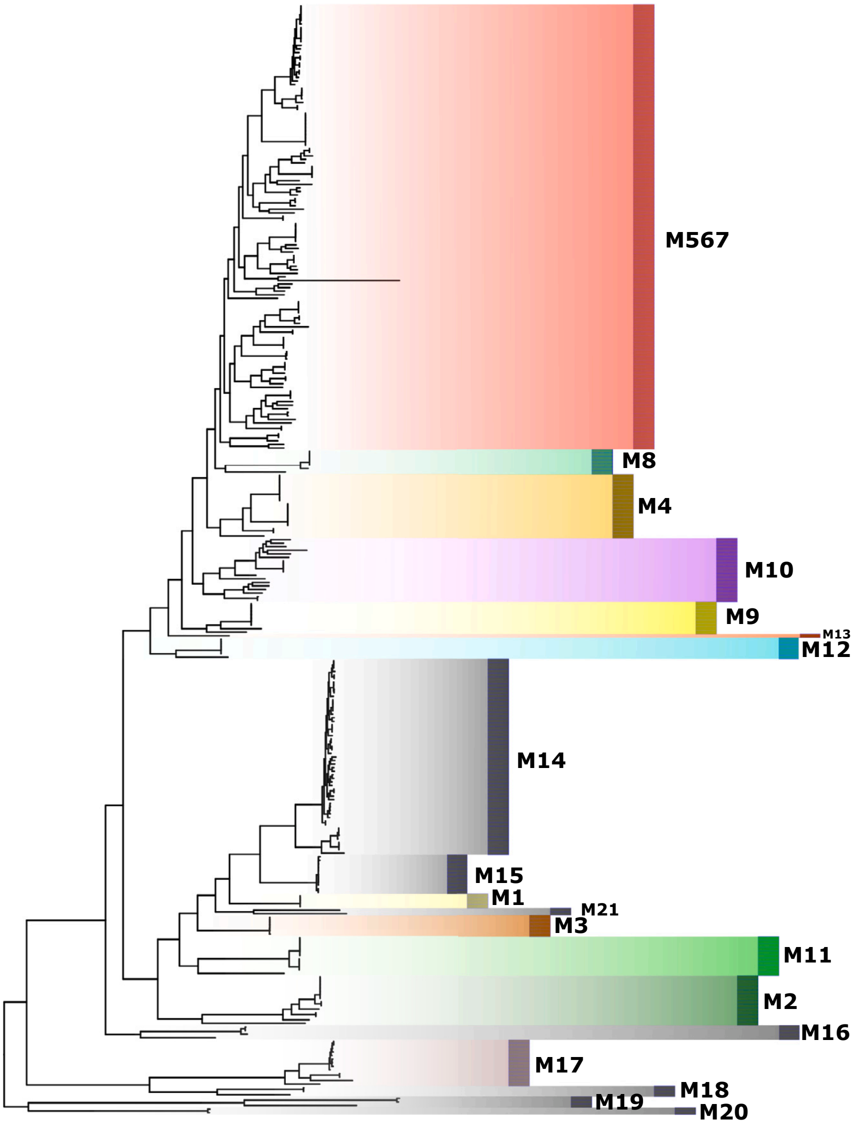


Figure 6. Phylogenomic clustering of the expanded dataset (Dataset II) of *Microcoleus* genomes based on FastBAPS

Genomes cluster into 19 distinct clades. Clades M5, M6, and M7 merged into one clade. New clades (M14 to M21) are marked in gray color. Bootstrap supports can be found in the expanded tree (Figure S2).

See also Tables S5 and S6.

ca. 5 μm in diameter. The forces and stresses exerted on the root/filament apex by soil particles are thus much different. In higher plants, the change of the root-cap shape also seems to be triggered by drought stress and other factors[51] while in cyanobacteria the shape of filament apex appears to be a relatively stable trait, regardless of cultivation conditions. The other strongly correlating variable is the cation exchange capacity at pH 7 (mmol/kg) (CEC). The CEC is a soil property that describes the capacity of soil to supply nutrient cations (Ca^{2+}, Mg^{2+}, K^{+}, etc.) for uptake. Organic soil carbon (dg/kg), and reduction in net primary productivity were both negatively correlated with the inverted pointiness value. This suggests that phenotypes with blunter filaments are generally correlated with nutrient-poorer soils, while filaments with pointier apices are generally associated with more nutrient-rich soils, and soils affected by land use.

The shape of filament apex also correlates with temperature and solar radiation. This suggests that phenotypes with "blunter" filament apices may be better adapted to higher temperatures and stronger solar radiation. The shape of the filament apex might aid in burrowing into deeper layers of soil to escape high temperatures and solar radiation. In our dataset, the "bluntest" phenotypes were in clades M1 and M3 which originated in Australia and Arizona, while the "pointiest" phenotypes came from clades M2 and M12, originating from Sweden, the Czech Republic, and Svalbard. The M5-6-7 superclade had the largest geographical and ecological territory, as well as the largest intra-species variance in pointiness value, suggesting these putative species are adapted to a broad range of conditions. The morphological variability of filament apices and significant correlation to soil conditions support the hypothesis that the apex shape in *Microcoleus* is dependent on soil properties and the putative species of *Microcoleus* are adapted to different conditions. It is likely that the blunt-ended phenotypes are better adapted to compact, nutrient poorer soils in warmer, drier regions, where the wider blunt apices of the filaments make them better able to take hold in the soil. The pointier phenotypes favor less hardy conditions in wetter and nutrient-rich soils, where the filament apices might provide some competitive advantage. It is important to note that we are only discussing a correlation and further research will be needed to confirm this hypothesis.

In terms of overall morphological features, most of the putative species within Dataset I possess unique morphology (or a combination of traits), meaning they could be delineated using the phenetic species concept. There were, however, some notable exceptions. As with the shape of the filament apex, the clades comprising the M5-6-7 "superclade" exhibited more intra-clade variability in comparison to the others. The dimensions of cells varied highly within the clades, as did the color of the filaments. Clades M5 and M6 cannot be morphologically distinguished in either cell dimensions, color, sheath thickness, etc., and clade M7 differs from clades M5 and M6 only in cell dimensions (Figure 5, and Table S3). This difference, however, is statistically significant. The unifying features of this superclade are the filament apex shape, the presence of distinct and often thick sheaths (and sometimes two filaments sharing a single sheath), and, relatively rarely, the presence of "nodules" or knots (Figure S1E). These nodules are morphologically similar to those found in *Nodosilinea*[52] but no such trait has been previously described in *Microcoleus*. It is unclear whether these knots serve some metabolic purpose or whether it is just by-product of the growth of filament in the thick sheath. Despite the morphological variability, the general morphology of the M5-6-7 superclade fits the description of *M. vaginatus sensu* Komárek and Anagnostidis (2005).[53]

As to where in this clade lies the true *M. vaginatus*, herbarium-derived genomes may assist with this systematic quandary. The inclusion of herbarium genomes may help alleviate the disparity in type materials and perhaps set new types that correspond better to the actual populations in nature.[54] In our dataset, sub-clade M5 includes the genome of the herbarium specimen of *M. vaginatus* obtained by Drouet.[55] While not the earliest available specimen of *M. vaginatus*, this specimen could be set as a new type specimen for *M. vaginatus sensu stricto*. In our opinion, subclade M5 can be confidently set as *M. vaginatus sensu stricto*, and clades M6 and M7 can be set as new, morphologically cryptic, incipient species that are not yet fully diverged.

A different issue is posed by the aforementioned clade M10. This lineage forms a stable monophyletic cluster in both Dataset I and II but consists of two distinct branches with notable differences in morphology (Figures 4F and 4G). In this case, the inclusion of herbarium genomes and metagenomes in Dataset II further complicates this issue. One genome is of the original herbarium specimen of *M. terrestris* by Desmaziéres (Table S6). *Microcoleus terrestris* was the first described species in *Microcoleus*, erected by the French researcher Jean Baptiste Henri Joseph Desmaziéres in 1823. However, Gomont[56] and Drouet[55] consider *M. terrestris* a synonym for the subsequently described *M. vaginatus*, today considered the type for *Microcoleus*. The *M. terrestris* herbarium genome clusters into clade M10 with cultivated strains originating from the Czech Republic (C2 strains). Furthermore, upon the inclusion of metagenomes, the metagenome of *M. asticus*[57] also clusters into the M10 clade, but into the second morphogroup. This issue warrants further study and is beyond the scope of this manuscript.

Since neither phenetic, ecological, monophyletic nor biological species concepts seem sufficient to delineate all of the putative species within our dataset, we seem to be running out of usable species concepts. Perhaps the efforts to identify precise boundaries and characteristic features for prokaryotic species along speciation continuums have been doomed to failure from the outset. As noted by the French impressionist painter Édouard Manet, "There are no lines in nature, only areas of color, one against another". Perhaps the best solution for the case of the "fuzzy" species within species continuums is not to search for a concrete border at all. Species in nature emerge by complex evolutionary processes and attempts to pigeonhole species into precise boundaries can obscure their true diversity and complexity. Some authors even advocate for the abolition of the species rank altogether.[58,59] However, the species-less approach does not take into account the practicality of the species rank.[60–62] A potential framework for delineating species in a continuum was recently formulated by Kollár et al.[63] The universal probabilistic concept of evolutionary lineages (UPCEL) allows for delineation at any stage of speciation using the probability of divergence. As noted previously, the morphological similarity (and inter-clade variability), coupled with high levels of gene flow, complicates the delimitation of species in the M5-6-7 superclade using phenetic, monophyletic, or biological species concepts. These three lineages could, however, be described as incipient, diverging species on the UPCEL scale, as there is a spectrum of probabilities of divergence based on the fraction of genome subjected to gene flow (1.5–29.4%, as noted in Stanojković et al.[6]). The advantage of this concept is that species can be defined on the basis of genomic divergence and there is no need to establish precise boundaries and apomorphies.

In this paper, we strove to articulate species boundaries along the speciation continuum of the ubiquitous, cosmopolitan cyanobacterium *Microcoleus*. We described a lineage composed of a whole spectrum of evolutionary relationships, from incipient to fully diverged species. Using large datasets can provide some information that would not be captured in smaller datasets, such as morphological variability (e.g., rare morphological features or morphological continuum) or gene flow. While employing population-level data, previously cryptic or fuzzy species can also become distinguished, whether based on a morphological apomorphy, genetic signature, or distinct genome divergence. Furthermore, the inclusion of herbarium genomes can help alleviate confusion in type material and can provide a bridge between historical and modern taxonomy, which can help with more grounded species delimitation.

Limitations of the study

While we employed perhaps the largest morphological dataset in cyanobacteria to-date, there are some limitations to this study. Due to the distribution of sampling efforts and the cost and difficulty of sampling expeditions, we do not have equivalent coverage of all continents. Australia, for example, is under-sampled in comparison to other geographical locations, and South America is missing from our dataset altogether. We would like to correct this limitation in the future. The advantage of this large dataset is that it can be updated with new genomes at any point in the future, hopefully filling in the blanks on the map and gaining an even deeper understanding of the diversity of the global speciation continuum of *Microcoleus*.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Phylogenomic analysis
  - Population clustering
  - 16S rRNA phylogenetic analysis
  - 16S–23S ITS secondary structures
  - Morphology assessment
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109444.

ACKNOWLEDGMENTS

The research was supported by the Grant Agency of the Czech Republic (grant no. 23-06507S), the Internal Grant Agency of Palacký University (grant no. IGA PrF-2024-001), Internal Fund of Faculty of Education UP (grant no. GFD_PdF_2023_06), and SYNTHESYS Project http://www.synthesys.info/ which is financed by European Community Research Infrastructure Action under the FP7 "Capacities" Program.

AUTHOR CONTRIBUTIONS

S.S., A.S., D.A.C., C.M., and P.D. analyzed data, wrote the manuscript, and prepared figures. S.S., A.S., A.M., M.K., V.D., P.M., and E.J. cultivated strains, isolated DNA for genome sequencing, and obtained microphotographs and measurements. J.J. conceived and calculated the filament pointiness value. D.A.C. provided language corrections. H.v.S., A.J., and P.D. obtained and analyzed herbarium-derived genomes. P.D. and E.J. coordinated work, designed experiments, and obtained funding. All authors reviewed and edited the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 14, 2023
Revised: January 30, 2024
Accepted: March 5, 2024
Published: March 6, 2024

REFERENCES

1. Nelson, C., Giraldo-Silva, A., Warsop Thomas, F., and Garcia-Pichel, F. (2022). Spatial self-segregation of pioneering cyanobacterial species drives microbiome organization in biocrusts. ISME Commun. *2*, 114. https://doi.org/10.1038/s43705-022-00199-0.
2. Garcia-Pichel, F., and Belnap, J. (1996). Microenvironments and microscale productivity of cyanobacterial desert crusts. J. Phycol. *32*, 774–782. https://doi.org/10.1111/j.0022-3646.1996.00774.x.
3. Rodriguez-Caballero, E., Belnap, J., Büdel, B., Crutzen, P.J., Andreae, M.O., Pöschl, U., and Weber, B. (2018). Dryland photoautotrophic soil surface communities endangered by global change. Nat. Geosci. *11*, 185–189. https://doi.org/10.1038/s41561-018-0072-1.
4. Williams, L., Borchhardt, N., Colesie, C., Baum, C., Komsic-Buchmann, K., Rippin, M., Becker, B., Karsten, U., and Büdel, B. (2017). Biological soil crusts of Arctic Svalbard and of Livingston Island, Antarctica. Polar Biol. *40*, 399–411. https://doi.org/10.1007/s00300-016-1967-1.
5. Garcia-Pichel, F., and Wojciechowski, M.F. (2009). The evolution of a capacity to build supra-cellular ropes enabled filamentous cyanobacteria to colonize highly erodible substrates. PLoS One *4*, e7801. https://doi.org/10.1371/journal.pone.0007801.
6. Stanojković, A., Skoupý, S., Johannesson, H., and Dvořák, P. (2024). The global speciation continuum of the cyanobacterium *Microcoleus*. Nat. Commun. *15*, 2122. https://doi.org/10.1038/s41467-024-46459-6.
7. Garcia-Pichel, F., Prufert-Bebout, L., and Muyzer, G. (1996). Phenotypic and phylogenetic analyses show *Microcoleus chthonoplastes* to be a cosmopolitan cyanobacterium. Appl. Environ. Microbiol. *62*, 3284–3291. https://doi.org/10.1128/aem.62.9.3284-3291.1996.

8. Strunecký, O., Komárek, J., Johansen, J., Lukešová, A., and Elster, J. (2013). Molecular and morphological criteria for revision of the genus *Microcoleus* (Oscillatoriales, Cyanobacteria). J. Phycol. *49*, 1167–1180. https://doi.org/10.1111/jpy.12128.
9. Doolittle, W.F., and Zhaxybayeva, O. (2009). On the origin of prokaryotic species. Genome Res. *19*, 744–756. https://doi.org/10.1101/gr.086645.108.
10. Baptiste, E., Ómalley, M.A., Beiko, R.G., Ereshefsky, M., Gogarten, J.P., Franklin-Hall, L., Lapointe, F.J., Dupre, J., Dagan, T., Boucher, Y., and Martin, W. (2009). Prokaryotic evolution and tree of life are two different things. Biol. Direct *4*, 34. https://doi.org/10.1186/1745-6150-4-34.
11. Gonzalez, J.M., Puerta-Fernández, E., Santana, M.M., and Rekadwad, B. (2020). On a Non-Discrete Concept of Prokaryotic Species. Microorganisms *8*, 1723. https://doi.org/10.3390/microorganisms8111723.
12. Johnson, E.A., and Miyanishi, K. (2008). Testing the assumptions of chronosequences in succession. Ecol. Lett. *11*, 419–431. https://doi.org/10.1111/j.1461-0248.2008.01173.x.
13. White, N.J., Snook, R.R., and Eyres, I. (2020). The past and future of experimental speciation. Trends Ecol. Evol. *35*, 10–21. https://doi.org/10.1016/j.tree.2019.08.009.
14. Shaw, K.L., and Mullen, S.P. (2014). Speciation Continuum. J. Hered. *105*, 741–742. https://doi.org/10.1093/jhered/esu060.
15. Stanowski, S., and Ravinet, M. (2020). Defining the speciation continuum. Evolution *75–6*, 1256–1276. https://doi.org/10.1111/evo.14215.
16. Mayr, E. (1942). Systematics and the Origin of Species (Columbia Univ. Press).
17. Coyne, J.A., and Orr, H.A. (2004). Speciation (Sinauer Associates).
18. Bobay, L.M. (2020). The Prokaryotic Species Concept and Challenges. In The Pangenome: Diversity, Dynamics and Evolution of Genomes, H. Tettelin and D. Medini, eds. (Springer). https://doi.org/10.1007/978-3-030-38281-0_2.
19. Bobay, L.M., and Ochman, H. (2017). Biological species are universal across life's domains. Genome Biol. Evol. *9*, 491–501. https://doi.org/10.1093/gbe/evx026.
20. Fraser, C., Hanage, W.P., and Spratt, B.G. (2007). Recombination and the nature of bacterial speciation. Science *315*, 476–480. https://doi.org/10.1126/science.1127573.
21. Dvořák, P., Jahodářová, E., Stanojković, A., Skoupý, S., and Casamatta, D.A. (2023). Population genomics meets the taxonomy of cyanobacteria. Algal Res. *72*, 103128. https://doi.org/10.1016/j.algal.2023.103128.
22. Hohenlohe, P.A., Hand, B.K., Andrews, K.R., and Luikart, G. (2018). Population Genomics Provides Key Insights in Ecology and Evolution. In Population Genomics. Population Genomics, O. Rajora, ed. (Springer). https://doi.org/10.1007/13836_2018_20.
23. Black, W.C., 4th, Baer, C.F., Antolin, M.F., and Du Teau, N.M. (2001). Population genomics: genome-wide sampling of insect populations. Annu. Rev. Entomol. *46*, 441–469. https://doi.org/10.1146/annurev.ento.46.1.441.
24. Luikart, G., England, P.R., Tallmon, D., Jordan, S., and Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. Nat. Rev. Genet. *4*, 981–994. https://doi.org/10.1038/nrg1226.
25. Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., Ding, H., Marttinen, P., Malmstrom, R.R., Stocker, R., et al. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. Science *344*, 416–420. https://doi.org/10.1126/science.1248575.
26. Stanojković, A., Skoupý, S., Hašler, P., Poulíčková, A., and Dvořák, P. (2022). Geography and climate drive the distribution and diversification of the cosmopolitan cyanobacterium *Microcoleus* (Oscillatoriales, Cyanobacteria). Eur. J. Phycol. *57*, 396–405. https://doi.org/10.1080/09670262.2021.2007420.
27. Skoupý, S., Stanojković, A., Pavlíková, M., Poulíčková, A., and Dvořák, P. (2022). New cyanobacterial genus *Argonema* is hiding in soil crusts around the world. Sci. Rep. *12*, 7203. https://doi.org/10.1038/s41598-022-11288-4.
28. Komárek, J. (2010). Recent Changes in Cyanobacteria Taxonomy Based on a Combination of Molecular Background with Phenotype and Ecological Consequences (Genus and Species Concept). Hydrobiologia *639*, 245–259. https://doi.org/10.1007/s10750-009-0031-3.
29. Casamatta, D.A., Vis, M.L., and Sheath, R.G. (2003). Cryptic species in cyanobacterial systematics: a case study of *Phormidium retzii* (Oscillatoriales) using 16S rDNA and RAPD analyses. Aquat. Bot. *77*, 295–309. https://doi.org/10.1016/j.aquabot.2003.08.005.
30. Jahodářová, E., Dvořák, P., Hašler, P., Holušová, K., and Poulíčková, A. (2018). *Elainella* gen. nov.: a new tropical cyanobacterium characterized using a complex genomic approach. Eur. J. Phycol. *53*, 39–51. https://doi.org/10.1080/09670262.2017.1362591.
31. Nevill, P.G., Zhong, X., Tonti-Filippini, J., Byrne, M., Hislop, M., Thiele, K., van Leeuwen, S., Boykin, L.M., and Small, I. (2020). Large scale genome skimming for herbarium material for accurate plant identification and phylogenomics. Plant Methods *16*, 1. https://doi.org/10.1186/s13007-019-0534-5.
32. Swali, P., Schulting, R., Gilardet, A., Kelly, M., Anastasiadou, K., GLocke, I., McCabe, J., Williams, M., Audsley, T., Loe, L., et al. (2023). Yersinia pestis genomes reveal plague in Britain 4000 years ago. Nat. Commun. *14*, 2930. https://doi.org/10.1038/s41467-023-38393-w.
33. Dvořák, P., Hašler, P., and Poulíčková, A. (2020). New insights into the genomic evolution of cyanobacteria using herbarium exsiccate. Eur. J. Phycol. *55*, 30–38. https://doi.org/10.1080/09670262.2019.1638523.
34. Jungblut, A.D., Raymond, F., Dion, M.B., Moineau, S., Mohit, V., Nguyen, G.Q., Déraspe, M., Francovic-Fontaine, É., Lovejoy, C., Culley, A.I., et al. (2021). Genomic diversity and CRISPR-Cas systems in the cyanobacterium *Nostoc* in the High Arctic. Environ. Microbiol. *23*, 2955–2968. https://doi.org/10.1111/1462-2920.15481.
35. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. *19*, 153. https://doi.org/10.1186/s12859-018-2129-y.
36. Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., Parkhill, J., and Harris, S.R. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. *43*, e15. https://doi.org/10.1093/nar/gku1196.
37. Kollár, J., Poulíčková, A., and Dvořák, P. (2022). On the relativity of species, or the probabilistic solution to the species problem. Mol. Ecol. *31*, 411–418. https://doi.org/10.1111/mec.16218.
38. Hanage, W.P. (2013). Fuzzy species revisited. BMC Biol. *11*, 41. https://doi.org/10.1186/1741-7007-11-41.
39. Johansen, J.R., Mareš, J., Pietrasiak, N., Bohunická, M., Zima, J., Štenclová, L., and Hauer, T. (2017). Highly divergent 16S rRNA sequences in ribosomal operons of *Scytonema hyalinum* (Cyanobacteria). PLoS One *12*, e0186393. https://doi.org/10.1371/journal.pone.0186393E.
40. Casamatta, D.A., Villanueva, C.D., Garvey, A.D., Stocks, H.S., Vaccarino, M., Dvořák, P., Hašler, P., and Johansen, J.R. (2020). *Reptodigitus Chapmanii* (Nostocales, Hapalosiphonaceae) Gen. Nov.: A Unique Nostocalean (Cyanobacteria) Genus Based on a Polyphasic Approach. J. Phycol. *56*, 425–436. https://doi.org/10.1111/jpy.12954.
41. Dvořák, P., Poulíčková, A., Hašler, P., Belli, M., Casamatta, D.A., and Papini, A. (2015). Species concepts and speciation factors in cyanobacteria, with connection to the problems of diversity and classification. Biodivers. Conserv. *24*, 739–757. https://doi.org/10.1007/s10531-015-0888-6.
42. Smith, G.R. (1989). Homologous recombination in prokaryotes: enzymes and controlling sites. Genome *31*, 520–527. https://doi.org/10.1139/g89-100.
43. Johansen, J.R., and Casamatta, D.A. (2005). Recognizing cyanobacterial diversity through adoption of a new species paradigm. Algol. Stud. *117*, 71–93. https://doi.org/10.1127/1864-1318/2005/0117-0071.
44. Komárek, J., Kaštovský, J., Mareš, J., and Johansen, J.R. (2014). Taxonomic classification of cyanoprokaryotes (cyanobacterial genera), using a polyphasic approach. Preslia *86*, 295–335.
45. Osorio-Santos, K., Pietrasiak, N., Bohunická, M., Miscoe, L.H., Kováčik, L., Martin, M.P., and Bohunická, M. (2014). Seven new species of *Oculatella* (Pseudanabaenales, Cyanobacteria): taxonomically recognizing cryptic diversification. Eur. J. Phycol. *49*, 450–470. https://doi.org/10.1080/09670262.2014.976843.
46. Komárek, J. (2020). Quo vadis, taxonomy of cyanobacteria. Fottea *20*, 104–110. https://doi.org/10.5507/fot.2019.020.
47. Komárek, J. (2006). Cyanobacterial taxonomy: current problems and prospects for the integration of traditional and molecular approaches. ALGAE *21*, 349–375. https://doi.org/10.4490/algae.2006.21.4.349.
48. Koch, R., Kupczok, A., Stucken, K., Ilhan, J., Hammerschmidt, K., and Dagan, T. (2017). Plasticity first: Molecular signatures of a complex morphological trait in filamentous cyanobacteria. BMC Evol. Biol. *17*, 209. https://doi.org/10.1186/s12862-017-1053-5.
49. Pietrasiak, N., Reeve, S., Osorio-Santos, K., Lipson, D.A., and Johansen, J.R. (2021). *Trichotorquatus* gen. nov. - a new genus of soil cyanobacteria discovered from American drylands. J. Phycol. *57*, 886–902. https://doi.org/10.1111/jpy.13147.
50. Colombi, T., Kirchgessner, N., Walter, A., and Keller, T. (2017). Root tip shape governs root elongation rate under increased soil strength.

Plant Physiol. *174*, 2289–2301. https://doi.org/10.1104/pp.17.00357.

51. Abdirad, S., Ghaffari, M.R., Majd, A., Irian, S., Soleymaniniya, A., Daryani, P., Koobaz, P., Shobbar, Z.S., Farsad, L.K., Yazdanpanah, P., et al. (2022). Genome-Wide Expression Analysis of Root Tips in Contrasting Rice Genotypes Revealed Novel Candidate Genes for Water Stress Adaptation. Front. Plant Sci. *13*, 792079. https://doi.org/10.3389/fpls.2022.792079.
52. Perkerson Iii, R.B., Johansen, J.R., Kovácik, L., Brand, J., Kaštovský, J., and Casamatta, D.A. (2011). A unique Pseudanabaenalean (Cyanobacteria) genus *Nodosilinea* gen. nov. based on morphological and molecular data. J. Phycol. *47*, 1397–1412. https://doi.org/10.1111/j.1529-8817.2011.01077.x.
53. Komárek, J., and Anagnostidis, K. (2005). Cyanoprokaryota 2nd Part: Oscillatoriales. In Süβwasserflora von MItteleuropa, B. Büdel, L. Krienitz, G. Gärtner, and M. Schagerl, eds. (Elsevier GmbH).
54. Beck, J.B., and Semple, J.C. (2015). Next-generation sampling: Pairing genomics with herbarium specimens provides species-level signal in Solidago (Asteraceae). Appl. Plant Sci. *3*, 1500014. https://doi.org/10.3732/apps.1500014[.
55. Drouet, F. (1968). Revision of the Classification of the Oscillatoriaceae (The academy of Natural Sciences of Philadelphia, Fulton Press Inc).
56. Gomont, M. (1892). Monographie des Oscillariées (Nostocacées homocystées), *15* (Annales des Sciences Naturelles, Botanique), pp. 263–368.
57. Churro, C., Semedo-Aguiar, A.P., Silva, A.D., Pereira-Leal, J.B., and Leite, R.B. (2020). A novel cyanobacterial geosmin producer, revising GeoA distribution and dispersion patterns in Bacteria. Sci. Rep. *10*, 8679. https://doi.org/10.1038/s41598-020-64774-y.
58. Mishler, B.D. (2022). Ecology, Evolution and Systematics in a Post-Species world. In Species Problem and Beyond, Contemporary Issue in Philosophy and Practice, J.S. Wilkins, F.E. Zachos, and I.Y. Pavlinov, eds. (CRC Press).
59. Wilkins, J.S. (2003). How to be a chaste species pluralist-realist: The origins of species modes and the synapomorphies species concept. Biol. Philos. *18*, 621–638. https://doi.org/10.1023/A:1026390327482.
60. Haber, M.H. (2019). Species in the age of discordance. Philos. Theor. Pract. Biol. *11*, 21.
61. Quinn, A. (2022). Species in the Time of Big Data: The multi-Species Coalescent, the General Lineage Concept, and Species Delimitation. In Species Problems and Beyond, Contemporary Issues in Philosophy and Practice, J.S. Wilkins, F.E. Zachos, and I.Y. Pavlinov, eds. (CRC Press).
62. Smith, M.L., and Carstens, B.C. (2022). Species Delimitation Using Molecular Data. In Species Problems and Beyond, Contemporary Issues in Philosophy and Practice, J.S. Wilkins, F.E. Zachos, and I.Y. Pavlinov, eds. (CRC Press).
63. Kollár, J., Poulíčková, A., and Dvořák, P. (2022). On the relativity of species, or the probabilistic solution to the species problem. Mol. Ecol. *31*, 411–418. https://doi.org/10.1111/mec.16218.
64. Staub, R. (1961). Ernährungsphysiologisch-autökologische Untersuchungen an der planktischen Blaualge *Oscillatoria rubescens* DC. ETH Zürich. https://doi.org/10.3929/ethz-a-000092335.
65. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.
66. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274. https://doi.org/10.1093/molbev/msu300.
67. Tonkin-Hill, G., Lees, J.A., Bentley, S.D., Frost, S.D.W., and Corander, J. (2019). Fast hierarchical Bayesian analysis of population structure. Nucleic Acids Res. *47*, 5539–5549. https://doi.org/10.1093/nar/gkz361.
68. R Core Team (2022). R: A Language and Environment for Statistical Compu-ting (R Foundation for Statistical Com-puting). https://www.R-project.org/.
69. Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large data sets. Bioinformatics *30*, 3276–3278. https://doi.org/10.1093/bioinformatics/btu531.
70. Labrada, N.A., McGovern, C.A., Thomas, A.L., Hurley, A.C., Mooney, M.R., and Casamatta, D.A. (2023). The CIMS (Cyanobacterial ITS motif slicer) for molecular systematics. Fottea *24*, 23–26. https://doi.org/10.5507/fot.2023.008.
71. Singh, J., Hanson, J., Paliwal, K., and Zhou, Y. (2019). RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. Nat. Commun. *10*, e5407. https://doi.org/10.1038/s41467-019-13395-9.
72. Hammer, Ø., Harper, D.A., and Ryan, P.D. (2001). Past: Paleontological Statistics Software Package For Education And Data Analysis. Paleontologia Electronica *4*, 1–9. http://palaeo-electronica.org/2001_1/past/issue1_01.htm.
73. QGIS.org, %Y (2021). QGIS Geographic Information System (QGIS Association). http://www.qgis.org.
74. Hašler, P., Dvořák, P., Johansen, J.R., Kitner, M., Ondřej, V., and Poulíčková, A. (2012). Morphological and molecular study of epipelic filamentous genera Phormidium, *Microcoleus* and *Geitlerinema* (Oscillatoriales, Cyanophyta/Cyanobacteria). Fottea *12*, 341–356. https://doi.org/10.5507/fot.2012.024.
75. Stanojković, A., Skoupý, S., Škaloud, P., and Dvořák, P. (2022). High genomic differentiation and limited gene flow indicate recent cryptic speciation within the genus *Laspinema* (cyanobacteria). Front. Microbiol. *13*, 977454. https://doi.org/10.3389/fmicb.2022.977454.
76. Dvořák, P., Jahodářová, E., Casamatta, D.A., Hašler, P., and Poulíčková, A. (2018). Difference without distinction? Gaps in cyanobacterial systematics; when more is just too much. Fottea *18*, 130–136. https://doi.org/10.5507/fot.2017.023.
77. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340.
78. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. Nat. Methods *14*, 587–589. https://doi.org/10.1038/nmeth.4285.
79. Fick, S.E., and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. *37*, 4302–4315. https://doi.org/10.1002/joc.5086.
80. McDowell, R.W., Noble, A., Pletnyakov, P., and Haygarth, P.M. (2023). A Global Database of Soil Plant Available Phosphorus. Sci. Data *10*, e125. https://doi.org/10.1038/s41597-023-02022-4.
81. Beckmann, M., Václavík, T., Manceur, A.M., Šprtová, L., Von Wehrden, H., Welk, E., and Cord, A.F. (2014). glUV: a global UV-B radiation data set for macroecological studies. Methods Ecol. Evol. *5*, 372–383. https://doi.org/10.1111/2041-210X.12168.
82. Haberl, H., Erb, K.H., Krausmann, F., Gaube, V., Bondeau, A., Plutzar, C., Gingrich, S., Lucht, W., and Fischer-Kowalski, M. (2007). Quantifying and mapping the human appropriation of net primary production in earth's terrestrial ecosystems. Proc. Natl. Acad. Sci. USA *104*, 12942–12947. https://doi.org/10.1073/pnas.0704243104.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Cultivated strains	Stanojković et al.[6]	Table S1
Herbarium specimens	National History Museum – London, UK	Table S6
Chemicals, peptides, and recombinant proteins		
Z – medium	Staub[64]	N/A
Deposited data		
Accesion numbers of cultivated strains and herbarium specimens	Stanojković et al.[6]	Tables S1 and S6, NCBI BioProject: PRJNA985139
All cell dimension measurements	This study	https://doi.org/10.17632/vbnmztmgpy.2
Correlation values for all strains against all climatic variables	This study	https://doi.org/10.17632/vbnmztmgpy.2
R script: pointiness and variables	This study	https://doi.org/10.17632/vbnmztmgpy.2
Software and algorithms		
Orthofinder v2.3.1	Emms, Kelly[65]	https://github.com/davidemms/OrthoFinder/releases
IQ-TREE v1.6.1	Nguyen et al.[66]	http://www.iqtree.org/
fastBAPS	Tonkin-Hill et al.[67]	https://github.com/gtonkinhill/fastbaps
R v4.1.3	R Core Team[68]	https://www.R-project.org
AliView	Larsson[69]	https://doi.org/10.1093/bioinformatics/btu531
CIMS	Labrada et al.[70]	https://doi.org/10.5507/fot.2023.008
SPOT-RNA	Singh et al.[71]	https://doi.org/10.1038/s41467-019-13395-9
PAST	Hammer et al.[72]	http://palaeo-electronica.org/2001_1/past/issue1_01.htm.
QGis v3.22.8	QGIS.org [73]	www.qgis.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and code should be directed to and will be fulfilled by the lead contact, Petr Dvořák (p.dvorak@upol.cz).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The genetic data analyzed in this paper have been deposited in the National Center for Biotechnology Information (NCBI). Accession numbers can be found in Tables S1, S5, and S6.
- Custom scripts for the correlation analyses were stored at Mendeley Data: https://doi.org/10.17632/vbnmztmgpy.2.
- All data will be publicly available upon publication of this study. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cyanobacterial samples were collected from diverse climates and habitats, such as soil crusts, freshwater benthos, and moss and rock surfaces. Strains were isolated according to,[74] and maintained in liquid Zehnder medium (Z-medium, Zehnder in Staub[64]) under the following conditions: temperature 24+/-1°C, illumination 20 μmol photons $m^{-2}s^{-1}$ and light regime of 12h light and 12h dark. Strains were isolated over the period of 4 years from 2019 to 2022. Sampling and strain isolation are more in-depth described in Stanojković et al.[75] The entire culture collection is currently maintained at the Department of Botany, Palacký University in Olomouc, Czech Republic. Furthermore, 8 herbarium specimens were included in this study (Table S6). The specimens were obtained from the herbarium collections of the Nature History Museum (London, UK).

METHOD DETAILS

Phylogenomic analysis

In this study, we analyzed two phylogenomic datasets. Dataset I consisted of 201 whole-genome *Microcoleus* sequences of our cultivated strains sequenced in Stanojković et al.[6] Dataset II is the expanded dataset including metagenome assembled genomes and herbarium genomes. Due to the unavailability of *M. anatoxicus* genome assemblies in the GenBank, we downloaded the raw reads from the SRA archive (SRR10997084, SRR10997083, and SRR10997082). We assembled the genomes and annotated them as described in Stanojković et al.[6] and Stanojković et al.[75] Altogether, Dataset II included 102 whole-genome sequences. The species trees were inferred based on the amino acid sequences (Dataset I) and nucleotide sequences (Dataset II) of single-copy orthologues. Orthofinder v2.3.1[68] with default parameters was used to identify single-copy orthologues, and acquire multiple sequence alignment (MSA). A maximum likelihood (ML) tree was inferred in IQ-TREE v1.6.1[69]. The best model (Dataset I: JTT+F+I+G4; Dataset II: SYM+I+G) was selected using modeltest implemented in IQ-TREE and the tree topology was tested using 2000 ultrafast bootstrap replicates. Phylogenomic trees were modified in Inkscape (https://inkscape.org/).

Population clustering

To elucidate the population structure of *Microcoleus* in Dataset II and assign them to genetically distinct clusters, we used a hierarchical Bayesian clustering algorithm implemented in package fastBAPS[67] in R v4.1.3.[68] Fastbaps clusters were determined using functions optimised.symmetric prior (optimized), and baps prior (unoptimized). The code was adopted from Stanojković et al.[6]

16S rRNA phylogenetic analysis

We performed a database search for all *Microcoleus* 16S rRNA sequences. We searched NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/) with the following terms: 16S rRNA AND *Phormidium*, *Oscillatoria*, or *Microcoleus.* We could not simply download only *Microcoleus* sequences because the database contained erroneous identifications, and the species were not renamed after the revisions (reviewed in[76]). All the sequences were downloaded and added to the multiple sequence alignment from Skoupý et al.,[27] which contains all the genera from the order Oscillatoriales and representatives from all cyanobacterial lineages to test which sequences were monophyletic with the *Microcoleus vaginatus sensu stricto,* i.e., with our thoroughly defined strains. The MSA was performed using the Muscle algorithm[77] in AliView.[69] The phylogenetic was reconstructed using the maximum likelihood optimality criterion in IQ-TREE[65] and the best model was determined based on the BIC using the modeltest[78] within IQ-TREE. The topology was tested by 2000 ultrafast bootstrap replicates.

16S–23S ITS secondary structures

For the reconstruction of ITS secondary structures, CIMS[70] was used to identify Box B and D1D1′ helices in the sequences and SPOT-RNA software was used for the folding.[71]

Morphology assessment

We assessed the morphology of 180 isolated cyanobacterial strains. The strain genomes were sequenced in our previous study Stanojković et al.,[6] corresponding to Dataset I above. The strains were studied using a light microscope Zeiss AxioImager (objectives EC Plan-Neofluar 40×/1.3 N.A., oil immersion, DIC; Plan-Apochromat 100×/1.4 N.A., oil immersion, DIC) and documented via high-resolution camera (Axio Cam D512 12MPx). Not all strains were viable enough for morphological assessment, but in each clade from our dataset there were at least 3 strains available. The following traits were observed for each strain: cell shape and dimensions, colour, sheath presence and thickness, calyptra and filament apex shape, presence of granules and presence of unique traits (e.g. nodules).

QUANTIFICATION AND STATISTICAL ANALYSIS

For each of the selected strains, 80 measurements of cell dimensions and 10 measurements of calyptra/filament apex were taken. To identify whether the difference in cell dimensions between the strains was statistically significant, a One-Way ANOVA test was performed (Table S3) using PAST software.[72] We assessed the attenuation of the filament apex, 3 measurements were taken for each apex. The width before attenuation, the width at calyptra, and the length of the attenuating segment (Figure 6A). The number of cells per attenuating segment was also counted. To assess the shape of the filament apex statistically, a "pointiness" value was developed as an approximation of the pointiness into the shape of ellipse (Figure S4). To determine statistically significant differences in pointiness values and the number of cells per attenuating segment of the filament apex, a One-way ANOVA test was performed (Table S4).

To test the hypothesis that the shape of filament apex is dependent on the environmental properties and that species are adapted to the different conditions, we performed a correlation analysis. The correlation was done in R v4.1.3[71] using the function cor.test. The bioclimatic variables were downloaded from WorldClim v2.1 database[79] at 2.5 arc-minutes resolution and extracted in R. Soil variables were obtained from ISRIC SoilGrids (www.isric.org) and Global Database of Soil Plant Available Phosphorus,[80] global UV-B radiation parameters from the glUV database[81] and the global human appropriation of net primary production (HANPP) parameters from the data published by Haberl et al.[82] and then extracted with QGis v3.22.8 software.[73] All environmental data were downloaded as of June 2023, except for phosphorous concentrations, obtained from the Global Database of Soil Plant Available Phosphorous[80] in January 2024.