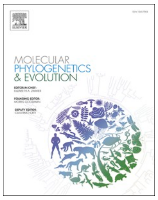
Molecular Phylogenetics and Evolution 212 (2025) 108419

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev



Phylogenomic analysis resolves the evolutionary pattern of diversification of cyanobacteria and reveals the evolutionary affinity of novel Antarctic lineages

Vadim Goremykin [a,*], Claudia Coleine [b], Caterina Ripa [b], Laura Selbmann [b,c], Claudio Donati [a]

[a] Unit of Computational Biology, Research and Innovation Centre, Fondazione Edmund Mach. Via Mach 1, 38098 San Michele all'Adige, TN, Italy
[b] Department of Ecological and Biological Sciences (DEB), University of Tuscia, L.go dell'Università snc, 01100 Viterbo, VT, Italy
[c] Italian Antarctic National Museum (MNA), Mycological Section, Genoa, Italy

ABSTRACT

Establishing phylogenetic relationships among the major cyanobacterial lineages remains elusive. Here we report a phylogenetic analysis of a large and well curated dataset, which allowed us to resolve with maximum statistical support the general pattern of cyanobacterial diversification and to address the limited knowledge regarding the diversity of Antarctic cyanobacteria. The analysis included novel Metagenome-Assembled Genomes (MAGs) sampled from diverse lithic habitats across Antarctica at elevations up to 3.4 km. The results presented here revealed previously undetected lineages and provided a robust phylogenetic framework for a natural cyanobacterial classification.

1. Introduction

Cyanobacteria are among the most abundant photosynthetic organisms on Earth and, in a variety of environments, they are the most important primary producers. Due to their resilience to harsh environmental conditions, including desiccation, extreme cold and heat, these organisms thrive in diverse environments such as hot and cold deserts, thermal springs, marine and freshwater, soil and rocks (Gaysina et al. 2019). In Antarctica, where environmental conditions are more extreme than anywhere else in the world, cyanobacteria have been identified in lakes, melt ponds and ice (Taton et al. 2003; Evans et al. 2022), soil (Rego et al. 2019; Pearce et al. 2012) and lithic environments (Velichko et al. 2021) playing dominant role as primary producers (Taton et al. 2003; Evans et al. 2022; Rego et al. 2019; Van Goethem and Cowan, 2019). In the ice-free areas of Continental Antarctica, cryptic modes of colonization (e.g. hypolithic, endolithic) are a widespread stress avoidance strategy and rocks act as refugia for microorganisms, providing a more buffered habitat that protects them from desiccation, extreme temperature fluctuations, and UV irradiation (De los Rios et al., 2014; Zakhia et al. 2008).

Given their ubiquity in many different and challenging environments and their importance as primary producers, developing a robust phylogeny of cyanobacteria is of fundamental importance. It should provide the evolutionary framework indispensable to understanding how adaptations and traits of interest emerged. However, the current state of cyanobacteria systematics was fittingly described as a “jungle” (Kaštovský, 2023). Considering that morphological similarity of even distantly related taxa is a common phenomenon in the cyanobacterial phylum (Strunecký et al. 2023), a need for phylogeny-based taxonomy (Strunecký et al. 2023) and for classification of novel strains has been recognized (Pessi et al. 2023).

The use of 16S rRNA gene for classification — a common practice in the past — was found to be unsatisfying. This gene is limited in its number of phylogenetically informative sites. A poor resolution of shallow and deeper nodes in the cyanobacterial trees was routinely noticed in the previous studies based on 16S rRNA analyses (Sánchez-Baracaldo et al., 2005; Zhaxybayeva et al. 2006; Swingley et al. 2008; Strunecký et al. 2013), which put in question the reliability of the inferred results (Strunecký et al. 2023).

The use of a broader sampling of sites for phylogeny reconstruction is justified based on theoretical considerations, since Maximum Likelihood and Bayesian tree inference has been shown to be statistically consistent (Chang 1996; Steel 2013; RoyChoudhury et al. 2015; Truszkowski and Goldman 2016). In other words, if the correct model is used for the phylogenetic inference, the chance that the correct tree will have the highest likelihood/posterior probability compared to all possible trees

---

* Corresponding author at: Unit of Computational Biology, Research and Innovation Centre, Fondazione Edmund Mach. Via Mach 1, 38098 San Michele all'Adige, TN, Italy
E-mail address: vadim.goremykin@fmach.it (V. Goremykin).

https://doi.org/10.1016/j.ympev.2025.108419
Received 26 June 2024; Received in revised form 6 March 2025; Accepted 18 July 2025
Available online 19 July 2025
1055-7903/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

increases with the increase in alignment length. Consequently, phylogenomic analyses are associated with lower probability of error in inference, provided that the model chosen for analysis is realistic.

The above rationale is not easy to apply. The most realistic, complex models require long computational times, which limits their use in the large data sets (Lozano-Fernandez 2022). This explains, perhaps, an almost uniform use of simple site-homogeneous models in recent phylogenomic studies of cyanobacterial phylogeny (Komárek et al., 2014; Mareš, 2018; Chen et al. 2021; Pessi et al. 2023; Strunecký et al. 2023). These models, with negligible exceptions, were shown to provide a much worse fit to amino-acid alignments of cyanobacteria compared to site-heterogeneous models in an extensive series of experiments (Pardo-De la Hoz et al. 2023). The choice of such simplified models does not fit well the theoretical reasoning for amassing large datasets.

The cases of reticulate evolution, common in cyanobacteria (Zhaxybayeva et al. 2006; Dagan et al. 2013), pose another problem, since phylogenetic methods assume that the data under analysis have evolved on the same tree. A widely accepted way to avoid phylogenetic errors under such circumstances is to use single copy orthologs for phylogenomic reconstructions (Aguileta et al. 2008). This limits the extent to which alignment length can be expanded. Recent phylogenomic studies of cyanobacterial phylogeny (Komárek et al., 2014; Mareš, 2018; Chen et al. 2021; Pessi et al. 2023; Mantas et al. 2021; Strunecký et al. 2023) utilized relatively small protein data sets ranging in length from 5,038 to 7,711 aligned positions. A broader sampling strategy was also tried by Chen et al. (2021), who built 263,124 pos. long alignment of the genes from a BUSCO database previously identified as near single copy orthologs in 112 cyanobacterial genomes (Dvořák et al., 2020). The assumption that these genes are single copies in a six times larger sampling of genomes used in Chen et al. was not checked. Considering that the number of single-copy orthologs can be expected to decrease with increase in the number of taxa (Emms and Kelly, 2018; Steenwyk et al. 2022), the above assumption it is unlikely.

Considering the above, the general strategy of data preparation chosen for this study was based on de-novo, uniform annotation of all the genomes under analysis, followed by de-novo sampling of single copy orthologs. This strategy arguably allows to identify the maximum number of markers which are suitable for phylogeny reconstruction given a fixed taxon sampling. The decision to use a carefully manually curated phylogenomic dataset was made here to ensure that the assumption of homology of character states at each alignment site — which is crucial for the correct reconstruction of phylogeny — was likely to hold in our analysis. The choice of a heterogeneous substitution model in our study was made in order to avoid errors in phylogenetic placement of deeper tree branches. As previously demonstrated (Talavera and Vila. 2011; Wang et al. 2019; Lartillot 2020; Szánthó et al. 2023 and citations therein), accounting for across-site heterogeneities is key to an accurate phylogenetic inference at deep taxonomic levels. However, since inference under heterogeneous models is computationally expensive, we had to somewhat limit taxon sample size in our experiments.

The site-heterogeneous phylogenetic analyses of the resulting data resolved the deeper nodes in the cyanobacterial tree that were poorly supported in the previous analyses. The analyses identified evolutionary affinity of the Antarctic MAGs generated in our study and revealed a number of large, well-supported monophyletic groups within the cyanobacterial tree backbone that were previously undetected. We discuss the obtained results using the framework of the higher order (families and above) classification of these organisms suggested by Strunecký et al. (2023). Relying on the phylogenetic inference based on the 5,038 positions long protein sequence alignment, Strunecký et al. (2023) provided the basis for a natural systematics of cyanobacteria. The observations presented here can be used to further enhance our understanding of the phylogeny of these organisms and, together with the increased availability of genomic resources, pose the basis for a systematic study of the evolutionary processes that led to their adaptation to a wide range of environments.

2. Materials and methods

2.1. Sampling of rocks for metagenomic analyses

Rocks colonized by endolithic microbial communities were collected in thirty-eight sites in Antarctica including Antarctic Peninsula (n = 3), McMurdo Dry Valleys, (n = 80), and Northern Victoria Land (n = 104) during more than 20 years of Italian Antarctic Expeditions. Different rock typologies (sandstone n = 137, granite n = 43, quartz n = 5, and basalt/dolerite n = 2) were collected along a latitudinal transect ranging from -62.10008 -58.51664 to -77.874 160.739 at different environmental conditions namely sun exposure (northern sun exposed and southern shady rocks) and an altitudinal transect from sea level to 3,400 m above sea level (a.s.l.) to provide a comprehensive overview of Antarctic endolithic cyanobacterial diversity. The presence of endolithic colonization was assessed by direct observation *in situ*. Rocks were excised using a geologic hammer and sterile chisel, and rock samples, preserved in sterile plastic bags, transported and stored at −20 °C in the Culture Collection of Antarctic fungi of the Mycological Section of the Italian Antarctic National Museum (MNA-CCFEE) until downstream analysis.

2.2. DNA extraction, library preparation, and sequencing

Metagenomic DNA was extracted from 1 g of crushed rocks using DNeasy PowerSoil Pro Kit (Qiagen, German), quality checked by electrophoresis using a 1.5% agarose gel and Nanodrop spectrophotometer (Thermofisher, USA) and quantified using the Qubit dsDNA HS Assay Kit (Life Technologies, USA). Shotgun metagenomic sequencing paired-end libraries were constructed and sequenced as 2×150 bp using the Illumina NovaSeq platform (Illumina Inc, San Diego, CA) at the Edmund Mach Foundation (San Michele all'Adige, Italy) and at the DOE Joint Genome Institute (JGI).

2.3. Assembly of novel metagenome-assembled genomes from Antarctic endolithic communities

In total, the dataset included 187 environmental DNA samples (their description, including Genbank accession details, is provided in Supplementary Table 1), of which 18 (JGI dataset) were sequenced and assembled at the DOE Joint Genome Institute (JGI) and binned into MAGs following the protocol from (Albanese et al., 2021), and 91 metagenomes (FEM dataset) were sequenced, assembled and binned into MAGs at the Fondazione Edmund Mach (FEM) as described in Coleine et al. (2024). For the remaining 78 samples (JGI-FICUS dataset) sequence data was generated at JGI and binned into MAGs as described in (Larsen et al, 2024).

The resulting 14503 bins were analysed using the metashot/prok-quality v1.2.3 (Albanese & Donati, 2021) (parameters –gunc_filter –gunc_db gunc_db_2.0.4.dmnd) workflow. Completeness, redundant and non-redundant contamination estimates were obtained by CheckM (Parks et al., 2015) v1.1.2 and GUNC (Orakov et al., 2021). Bins with completeness estimates of <50%, more than 10% contamination and that did not pass the GUNC filter were discarded, resulting in a total of 5769 filtered prokaryotic MAGs. MAGs were classified into a) high-quality draft (HQ) with >90% completeness and <5% contamination and b) medium-quality draft (MQ) with completeness estimates of ≥50% and less than 10% contamination. Species-level assignment of these metagenomic data was conducted by clustering HQ and MQ MAGs at 95% average nucleotide identity (ANI) using dRep v2.6.2 (Olm et al., 2017), resulting in a total of 2402 species-level metagenomic assemblies.

Species-level assemblies were subjected to a preliminary taxonomic binning using the metashot/prok-classify v1.3.1 workflow (https://github.com/metashot/prok-classify, parameters: –gtdbtk_db release207). Thirty-six high-quality draft genomes were assigned to cyanobacterial phylum and were selected for phylogenetic analysis. The sampling details for these genomes, such as latitude, longitude, sampling environment and elevation above the sea level are provided in Supplementary Table 2. The presence of each Antarctic MAG in individual samples was estimated using the command *mash screen* (Ondov et al. 2019) from the metashot/containment v1.1.0 workflow (https://github.com/metashot/containment) using the winner-takes-all strategy to avoid redundancy of the identified genomes.

2.4. *Gene prediction, functional annotation and preparation of alignments*

Complete and draft genome sequences and metagenome-assembled genomes (MAGs) of cyanobacteria were chosen for this study based on the preliminary average nucleotide identity-based taxonomic binning (ANI) performed emloying the GTDB-Tk v2.1.1 toolkit with the default parameters (Chaumeil et al. 2019). Considering the limitations of the ANI-based taxonomic assignment (Dvořák et al. 2023), which is based on arbitrarily selected thresholds of sequence similarity, we chose to perform a model-based phylogenomic analysis to provide an overall view of cyanobacterial phylogeny and to elucidate the placement of novel antarctic MAGs within the cyanobacterial phylum. Taking into account time constraints associated with phylogenetic inference under more realistic heterogeneous models, we selected a limited number of published cyanobacterial genomes (209 in total) for phylogeny reconstruction, to represent i) the lineages potentially related to our MAGs and ii) the overall diversity of the cyanobacteria. The accession details for these sequences and closely related outgroup melainabacterial MAGs used in our analyses are listed in Supplementary Table 3. In order to ensure homogeneity of gene predictions, these sequences (215 in total) and the aforementioned 36 MAGs of Antarctic cyanobacteria were subject to de novo gene prediction using the metashot/prok-annotate workflow v3.01 (https://github.com/metashot/prok-annotate) with the default settings. Coding DNA sequences (CDSs) were predicted using Prokka (Seemann 2014) v1.14.5 which in turn wraps the gene predictor Prodigal (Hyatt et al. 2010). The resulting coding gene sequences were translated into protein sequences. General functional overview of each of the resulting genome-specific protein sets was performed by assigning its members to pre-defined functionally characterized bacterial protein families from eggNOG 5.0 database (Huerta-Cepas et al. 2019) with the help of eggNOG-mapper v.2 (Cantalapiedra et al. 2021). For the purpose of identification of near-single copy orthologs specific to our taxon set we used a distinct procedure based on dedicated software program for phylogenetic orthology inference (OrthoFinder v. 2.5.4, Emms and Kelly, 2019). The two-step analysis by OrthoFinder employed here involved i) identification of the Hierarchical orthogroups (HOGs) for each node in the default species tree inferred in the first OrthoFinder run and ii) re-identification of HOGs under a different species tree built under LG+G model with IqTree v.1.6.12 (Nguyen et al. 2015) based on the concatenated alignment of protein sequences produced in the first OrthoFinder run.

The resulting HOGs identified by OrthoFinder that contained Melainabacteria, Gloeobacteria and at least 190 species in total, with no more then 10% of species represented by multiple sequences were subject to multiple sequence alignment with MAFFT v7.397 (Katoh, 2005). The resulting 584 multiple sequence alignments (MSAs) were visually inspected to discard those with low similarity among melainabacterial and cyanobacterial gene sequences. This screening step was carried out in order to reduce potential orthology inference errors. Trees were built based on the resulting alignments under LG+G model with IqTree v. 1.6.12. The MSAs of the near-single copy genes and corresponding trees were further visually inspected to exclude i) short and/or poorly aligned sequences and ii) divergent sequences born on long tree branches. If, due to incomplete assembly and/or sequence errors, a gene belonging to a species was presented in an MSA by several non-overlapping partial sequences, the longest partial sequence was retained. If sequences from the same species shared a tree branch that contained no other species, the longest sequence was retained.

The above manual screening allowed us to produce 216 MSAs with no more than one protein sequence per species, which are unlikely to contain paralogs. These MSAs were re-aligned with MAFFT. The resulting data matrices are provided as Supplementary material. A set of vertical blocks in each MSA was sampled by the Gblocks program (Castresana 2000) embedded in the Seaview alignment editor (Gouy et al. 2010). The resulting blocks were inspected and manually edited in Seaview in order to ensure that only unambiguously aligned regions would be used for phylogenetic inference. Each set of blocks was saved to produce a curated MSA. These 216 curated single gene alignments were concatenated to produce a 52,867 positions long amino-acid alignment of 251 taxa (henceforth referred to as the observed alignment, available as Supplementary Material). The alignment contains 2.7 % of gaps.

2.5. *Phylogenetic Inference*

A relative fit of i) the homogeneous substitution models assuming empirical matrices of substitution rates (LG, JTT, WAG, etc) and ii) a GTR-based model that utilizes a substitution rate matrix derived from the observed alignment, to the observed alignment was estimated with the ModelFinder pipeline (Kalyaanamoorthy et al. 2017) implemented in IqTree. It was consistently observed that under a model composition that was fixed with the exception of the substitution matrix, the use of GTR-based substitution matrix resulted in a better model fit under the Bayesian Information criterion (BIC). Taking into account this preliminary observation, and the recent report of a much better fit of site-heterogeneous models in comparison to homogeneous models to amino-acid alignments of cyanobacteria observed in an extensive series of experiments (Pardo-De la Hoz et al. 2023), phylogenetic analyses here were performed under a site-heterogeneous model using a GTR-based substitution matrix and Dirichlet processes for modeling i) sites-specific frequency profiles and ii) the distribution of relative rates of substitution across sites (specified via “-gtr -cat -ratecat” options), henceforth referred as to “CAT+GTR+D” model, wherein +D model component stands for “Dirichlet”.

Two chains were run for approximately 6 months based on the observed alignment in PhyloBayes v.4.1 (Lartillot et al. 2009) to sample, for each chain, 4000 cycles under a CAT+GTR+D model. After checking the extent of the “burnin” zones for each chain, the consensus phylogenetic tree was built with the bpcomp program based on each of the last 2000 cycles sampled from each chain (4000 cycles in total). The largest (maxdiff) and mean (meandiff) discrepancies in clade support calculated with bpcomp were, respectively, 1 and 0.00801904.

3. Results

3.1. *Cyanobacterial diversification and phylogenetic placement of novel MAGs*

Almost each internal branch in the tree obtained based on the observed alignment (Fig. 1) was supported by the maximum posterior probability (PP) value. In agreement with the results of recent studies, the tree supports the earliest division within the cyanobacteria separating the thylacoid-less Gloeobacteria from Phycobacteria, the crown group cyanobacteria that have the chlorophyll contained in thylakoids. By contrast to Phycobacteria, Gloeobacteria encompass only a few species described so far. Recently, Gloeobacteria were found to contain two sister lineages (Grettenberger et al. 2020; Rahmatpour et al. 2021; Grettenberger, 2021; Pessi et al. 2023), each assigned a family rank in Strunecký et al. (2023): Gloeobacteraceae and Anthocerotibacteraceae. This division was confirmed in the present phylogenomic analysis that encompasses 12 species of Gloeobacteria (Fig. 1). An antarctic MAG sampled in this study from a sandstone rock at 3.4 km elevation in the

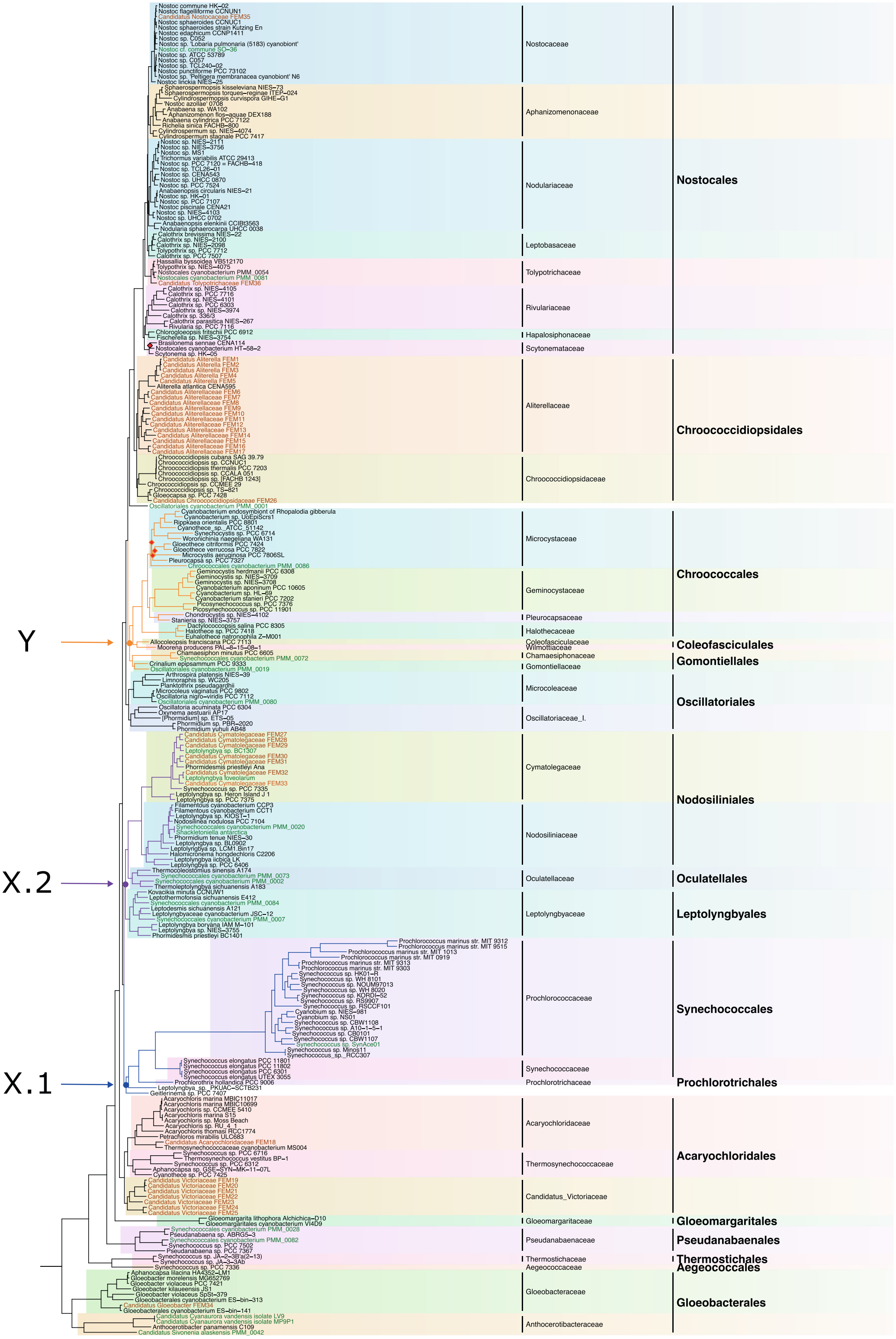


Fig. 1. Bayesian consensus phylogenetic tree showing the relationship between the major cyanobacterial clades as inferred under a CAT+GTR+D model based on the observed alignment. The four clades marked with diamonds received 0.5 posterior probability support (PP). All other clades, except the clade subtending five Acaryochloris strains (three A. marina, CCMEE 5410 and A. sp. Moss beach), which was supported by 0.99 PP, received maximum posterior probability support (1 PP). The taxonomic assignment of the clades, shown to the right of the phylogram, follows the higher order (families and above) classification of cyanobacteria proposed by Strunecký et al. (2023). The names of the MAGs generated in this study are highlighted in orange. The names of other antarctic species are highlighted in green. The clades X.1, X.2 and Y, mentioned in the main text, are shown to the left of the phylogram and highlighted in blue, pink and orange, respectively. The melainabacterial six-species outgroup is not shown.

Archambault ridge in the Southern Cross Mountains of Victoria Land (‘*Candidatus* Gloeobacter FEM34’) was recovered as a sister to the MAG ‘Gloeobacteriales cyanobacterium ES-bin-141’ sampled from exposed soil near the Little Firn Glacier in the northeast Greenland at near sea-level altitude (Grettenberger, 2021). The above two MAGs formed a sister group to the rest of the Gloeobacteraceae sensu Strunecký et al. The Anthocerotibacteraceae lineage in our recovered phylogeny were represented by *Sivonenia alaskensis* (Pessi et al. 2023), *Anthocerotibacter panamensis* (Rahmatpour et al. 2021), and two accessions of ‘*Candidatus* Cyanaurora vandensis’ (Grettenberger et al. 2020).

Within the Phycobacteria, the clade comprising members of Aegeococcales and Thermostichales, was resolved as the most basal lineage, followed next by Pseudanabenales, Gloeomargaritales and Acaryochloridales as successive sister lineages to the rest of the Phycobacteria. The order Acaryochloridales was recently defined based on phylogenetic inference (Strunecký et al. 2023) and was suggested to comprise two previously identified, although physiologically discrete sister phyla, each assigned to a rank of family — Acaryochloridaceae (Komárek et al. 2014) and Thermosynechococcaceae (Komárek et al. 2020). The analysis presented here confirmed the sister group relationship of Thermosynechococcaceae and Acaryochloridaceae lineages. One of the novel MAGs, ‘*Candidatus* Acaryochloridaceae FEM18’ recovered from granite rocks sampled at a sea level in Inexpressible Island in Terra Nova Bay, formed the most basal clade within the Acaryochloridaceae clade along with a MAG ‘Thermosynechococcaceae cyanobacterium MS004’, which was previously recovered from the Castaic Lake in Southern California (Fig. 1). Our analysis also revealed existence of a previously unknown sister lineage to the Thermosynechococcaceae plus Acaryochloridaceae clade. In our analysis, this novel lineage was represented by seven antarctic MAGs (*Candidatus* Victoriaceae FEM19.. 25), recovered from granite rocks in Northern Victoria Land at elevations ranging from 200 m to 2.2 km.

The next large clade to split off in the Phycobacteria in our reconstructed phylogeny, henceforth termed ‘clade X’, comprises two sister clades. In one sister clade (shown as ‘clade X.1’ in Fig. 1), we found: i) *Geitlerinema sp.* PCC 7407, ii) *Leptolyngbya sp.* PKUAC-SCTB231, iii) *Prochlorotrix hollandica* PCC 9006, and iv) the strains assigned to *Synechococcus elongatus* consecutively forming sister groups to the alpha-cyanobacterial branch, encompassing picoplanctonic cyanobacterial species mainly from *Prochlorococcus/Synechococcus* genera that possess a form IA RuBisCO and α-carboxysomes.

A sister clade to the clade X.1 (show as ‘clade X.2’ in Fig. 1) is formed by the (Leptolyngbyaceae (Oculatellaceae, (Nodosilineaceae, Cymatolegaceae)) lineage in our reconstructed phylogeny. A crown group of the Cymatolegaceae lineage (as shown in Strunecký et al. 2023) was found to encompass three previously sequenced species and seven MAGs (*Candidatus* Cymatolegaceae FEM27.. 33) recovered from rocks sampled from Antarctic Duffael, Inexpressible and Kay islands at the elevation up to 80 meters. In this 10-species lineage, only one species was sampled outside of Antarctica - a ‘*Phormidesmis priestleyi* Ana’ strain from a hypersaline Hot Lake, Washington, USA.

The next branch to split off in the Phycobacteria in our reconstructed phylogeny corresponds to the order Oscillatoriales. The next clade to separate from the main tree backbone (henceforth termed “clade Y”) is divided into Gomontiellales and Coleofasciculales plus Chroococcales subclades. The clade Y forms a sister group to a lineage encompassing ‘Oscillatoriales cyanobacterium PMM001’ Antarctic MAG generated by Pessi et al. (2023), being a sister to a clade comprising Nostocales plus Chroococcidiopsidales. Two of the newly generated MAGs were found to be affiliated with the order Nostocales. A MAG recovered from granite rock collected at 1.7 km above sea level in the Random Hills, Victoria Land (‘*Candidatus* Tolypotrichaceae FEM36’) was resolved in our analyses as basal within a lineage assigned in Strunecký et al. (2023) to Tolypotrichaceae, a family within the Nostocales. The group included an arctic ‘Nostocales cyanobacterium PMM_0054’ MAG and an antarctic ‘Nostocales cyanobacterium PMM_0081’ MAG generated by Pessi et al. (2023), *Tolypothrix_sp.* NIES-4075 from Japan and *Hassallia byssoidea* VB512170 from India. Another MAG recovered in this study from a granite rock sampled 1.4 km above sea level at Starr Nunatak, Victoria Land (‘*Candidatus* Nostocaceae FEM35’) formed a sister to the *Nostoc commune* HK-02 (Japan) plus *Nostoc flagelliforme* CCNUN1 (China) branch in the family Nostocaceae as defined in Strunecký et al.

Our phylogenetic survey revealed an abundance of cyanobacteria that belong to the Chroococcidiopsidales clade in lithic-associated microbial communities in Antarctica. The corresponding clade (Fig. 1) encompassed 18 out of 36 cyanobacterial MAGs generated in this study. Our reconstructed phylogeny is consistent with the division of the Chroococcidiopsidales onto *Aliterella* and *Chroococcidiopsis* lineages (Rigonato et al. 2016). The former lineage was represented in our study by 17 MAGs (*Candidatus* Aliterella FEM1.. 5 and *Candidatus* Aliterellaceae FEM6.. 17) that formed a cluster with a marine cyanobacterium *Aliterella atlantica* CENA595. The *Chroococcidiopsis* lineage was found to be far less abundant in lithic-associated microbial communities in Antarctica as it included a single Antarctic genome (‘*Candidatus* Chroococcidiopsidaceae FEM26’). This genome was resolved as a sister group to two hot spring isolates: *Chroococcidiopsis sp.* TS-821 (Thailand) and *Gloeocapsa sp.* PCC 7428 (Switzerland).

3.2. Genomic characteristics of cyanobacterial genomes

The analyzed cyanobacterial genomes showed a remarkable variability in GC content, genome size, and number of predicted genes (Supplementary Fig. 1 and Supplementary Table 3) that correlate with their placement in the phylogenomic tree. While species close to the root of the cyanobacterial tree (orders Gleobacterales, Aegeococcales, Thermostichales) have smaller genomes with lower number of predicted genes and higher GC content, species branching later from the tree (order Nostocales) show a lower GC content, larger genome and higher number of predicted genes, in some cases exceeding 10 Mbases and 10,000 predicted genes, respectively (family Rivulariaceae). The data reveal a high variability also amongst phylogenetically closely related species. Especially noteworthy is the large variations in GC content within the family Procholococcaceaee, (30.8% - 68.6%) that exceed those previously reported (30%-50.7%) (Tschoeke et al. 2020). The trend towards an increasing genomic complexity for taxa branching later from the phylogenetic tree is confirmed if the number of distinct core protein families (i.e. present in more than 90% of the members of the taxon) is considered, with the number of distinct core protein families exceeding 2,000 in some families of the Order Nostocales (Families Leptobasaceae and Nostocaceae), while it is much smaller for taxa closer to the root of the tree, such as Families Gloeobacteraceae and Pseudanabaenaceae, that include 1,225 and 1,133 distinct core protein families, respectively (Fig. S2).

3.3. Genomic and ecological characteristics of Antarctic MAGs

Although each of these 36 MAGs was assembled from a single sampling site (Supplementary Table 2), mapping of the raw reads (see Methods) from other locations on these reconstructed species-specific genomes allowed us to assess their geographical distribution (Supplementary Fig. 3). In the following we discuss the results of these analyses for each taxon.

Order Gleobacterales, Family Gloeobacteraceae. We assembled a single MAG of the order Gleobacterales. Beside the single sandstone sample collected in Archambault Ridge (-73,66942 N 162,59369 W) at an elevation of 3,400 masl, we could not find evidence of this organism in any other sample in the whole dataset.

Order Acharyochloridales, Families Acaryochloridaceae and *Candidatus* Victoriaceae. We assembled 8 genomes from the Order Acharyochloridales. All genomes were assembled from granite samples collected at altitudes between sea level (Inexpressible Island, -74,88931 N 163,74550 W) and 2240 masl (Mt. Nansen, -74,62861 N 162,59389 W).

However, these organisms were also detected in a small number of sandstone, quartz and mixed samples (Supplementary figure 3).

Order Nodosiliniales, Family Cymatolegaceae. We assembled 7 MAGs from the family Cymatolegaceae. All the members of this family were assembled from samples collected at low elevation (between 0 and 80 masl) in different geographical locations from Quartz, Dolerite and Granite samples. However, by analyzing the distribution of these organisms in all the samples we found that one of them ('*Candidatus* Cymatolegaceae FEM27') is also present in sandstone samples.

Order Nostocales, Families Nostocaceae and Tolypotrichaceae. We assembled 2 MAGs from the Order Nostocaceae. The two genomes were assembled from two granite samples from Random Hills (-74.01000 N, 164.36667 W, 1700 m asl) and Starr Nunatak (-75.89853 N, 162.59289 W, 1420 m asl), respectively. Analysis of their distribution in other samples revealed no evidence of their presence in other rock types.

Order Chroococcidiopsidales, Families Aliterellaceae and Chroococcidiopsidaceae. Half (18/36) of the newly assembled MAGs were classified as members of the Order Chroococcidiopsidaceae, 17 from Family Aliterellaceae and one from Family Chroococcidiopsidaceae. With the exception of '*Candidatus* Aliterellaceae FEM7' and '*Candidatus* Aliterellaceae FEM12' MAGs that were assembled from two samples collected at elevation of 385 and 700 meters above mean sea level (masl), respectively, all other MAGs from this group were assembled from an elevation comprised in a relatively narrow range between 1,420 and 1,840 masl. Within Family *Aliterellaceae,* we can distinguish 2 monophyletic groups, one including also the strain Aliterella atlantica CENA595, and the other composed only by Antarctic MAGs assembled in this study with distinct environmental distribution. Specifically, while members of the group including only Antarctic MAGs were prevalently found in sandstone samples, the other that included *Aliterella atlantica* CENA595 had a more diverse distribution with a preference for granite samples.

4. Discussion

4.1. Overall tree structure

The branching order among early-divergent cyanobacterial lineages inferred here is concordant with the results of the recent phylogenomic studies (Chen et al. 2021; Pessi et al. 2023; Strunecký et al. 2023). The primitive thylakoid-less Gloeobacteria represent the most basal lineage in Cyanobacteria. The clade comprising thermophilic *Synechococcus* isolates from Octopus Spring, Yellowstone National Park (*Synechococcus* PCC 7336, *Synechococcus* JA-2-3B'a(2–13), and *Synechococcus* JA-3-3Ab) branches off next, followed by Pseudanabenales, Gloeomargaritales and Acaryochloridales as successive sister lineages to the rest of the Cyanobacteria.

The next clade (termed 'clade X') to split off in our reconstructed phylogeny, was, according to our knowledge, never recovered with strong support in phylogenomic studies. The clade includes two sister lineages, one lineage ('clade X.1', as shown in the Fig. 1) with i) *Geitlerinema sp.* PCC 7407, ii) *Leptolyngbya sp.* PKUAC-SCTB231, iii) *Prochlorotrix hollandica* PCC 9006 and iv) *Synechococcus elongatus* strains as consecutive sisters to alpha-cyanobacteria and the other lineage ('clade X.2', as shown in the Fig. 1) divided into two sublcades, one subtending Leptolyngbyales and the other (Oculatellales, Nodosiliniales) cluster.

In Chen et al. (2021), Pessi et al. (2023) and Strunecký et al. (2023), the taxa included in clade X formed ladder-like phylogenetic structures. The common origin of some taxa included in the clade X was poorly supported in Komárek et al. (2014) that did not include Oculatellales (32% bootstrap support, (BP)), in the analysis by Mantas et al. (2021) that did not include Oculatellales and Cymatolegaceae (69% BP) and in the Bayesian analysis in Mareš (2018) (0.8 PP) that also did not include Oculatellales. In the Maximum Likelihood analyses also performed by Mareš (2018), the common origin of the clade X taxa was not supported. The internal branching order among the clade X taxa as inferred here is not compatible with tree topologies recovered by Komárek et al. (2014), Mareš (2018) and Mantas et al. (2021). For instance, these sudies do not support a common origin of the taxa subtended by the clade X.2. Also, according to our knowledge, the most basal division among the clade X taxa corresponding to X.1 and X.2 clades has been recovered in none of previous phylogenomic studies.

The next clade to split off from the main cyanobacterial lineage after branching off of all other taxa mentioned in this section (Fig. 1), corresponds to the order Oscillatoriales, which is consistent with the previously reported results (Komárek et al. 2014; Mareš, 2018; Chen et al. 2021; Pessi et al. 2023; Mantas et al. 2021). The clade Y containing Gomontiellales as sister to Coleofasciculales plus Chroococcales splits off next from the main backbone of the tree, which is consistent with the reconstructed phylogenies presented in Komárek et al. (2014) and Mareš (2018). Other recent phylogenomic studies (Chen et al. 2021; Mantas et al. 2021; Pessi et al. 2023; Strunecký et al. 2023) do not support the monophyly of the above mentioned three orders.

The sister group to the clade Y as recovered in our analyses includes an antarctic MAG ('Oscillatoriales cyanobacterium PMM_001') from Pessi et al. (2023) as sister to Nostocales plus Chroococcidiopsidales, which is in agreement with the tree topology presented in Pessi et al. The clade subtending Nostocales plus Chroococcidiopsidales was also recovered in a number of recent analyses of cyanobacterial phylogeny based on multilocus phylogenetic inference (Komárek et al. 2014; Mareš 2018; Chen et al. 2021; Mantas et al. 2021; Strunecký et al. 2023).

4.2. Phylogenetic placement of novel MAGs and its implications

The MAGs recovered in this study were placed in the phylogenomic trees in the clades corresponding to Chroococcidiopsidales (n=18), Acaryochloridiales (n=8), Nodosileniales (Cymatolegaceae) (n=7), Nostocales (n=2), and Gloeobacterales (n=1) lineages. As such, prevalence of Chroococcidiopsidales in the Antarctic desert is not surprising, since this order has been frequently found in a variety of extreme environments (Rigonato et al. 2016; Jung et al. 2019, 2020, 2021; Schulze-Makuch et al. 2021). Outstanding tolerance to ionizing and UV radiation, and desiccation, which is common among Chroococcidiopsidales, have already prompted proposals of employment of these cyanobacteria in the bioregenerative life support systems for colonization of Mars (Billi et al., 2019a; Billi et al 2019b; Napoli et al. 2022). Rather conspicuous is that a great majority of the Chroococcidiopsidales MAGs recovered in Antarctica clustered not with the *Chroococcidiopsis* clade, but with *Aliterella atlantica* CENA595, a marine species inhabiting waters near the continental shelf of south-eastern Brazil (Rigonato et al., 2016). Based on phylogenetic analyses, *Aliterella* was recently erected as the second genus of the order Chroococcidiopsidales, belonging to a new family Aliterellaceae (Rigonato et al. 2016). Initially, *Aliterella* species were described from aquatic environments, however, recent assignment of a strain sampled from the Atacama Desert (Jung et al. 2020) to *Aliterella* indicated existence of a specialized species in this lineage that is adapted to desiccation. The results presented here show that extremophiles are common within the Aliterellaceae. Considering abundance of the *Aliterella* strains in the inhospitable Antarctic environment, that, arguably, is most similar to the Martian one, it might well be that the most suitable species for supporting human space exploration should be sought within this lineage.

Discovery of a previously undetected Antarctic lineage being sister to the Thermosynechococcaceae plus Acaryochloridaceae clade suggests revision of the order Acaryochloridales. Considering that the length of the branches subtending Thermosynechococcaceae and Acaryochloridaceae is comparable with the length of the branch subtending the above lineage, we suggest to assign to this lineage the rank of a family within Acaryochloridales for which we propose the name '*Candidatus* Victoriaceae' in reference to the geographic preference of the family.

Cymatolegaceae, a new family erected within Nodosilineales based on phylogenetic evidence was initially described to include solely marine or brackish species (Strunecký et al. 2023). Present study expands the known ecological distribution of species of Cymatolegaceae to include rocks in the Antarctic islands. Considering the results of phylogeny reconstruction obtained here, adaptation to terrestrial habitat is an ubiquitous feature in Cymatolegaceae that independently occurred several times during family diversification. Nostocaceae and Tolypotrichaceae (Nostocales) also show high ecological plasticity in the choice of habitats, being able to thrive in polar and near-equatorial zones, where they populate water, soil and lithic environments.

The discovery of a novel Antarctic lineage of Gloeobacterales at 3.4 km altitude in the Archambault Ridge, where the climatic conditions are among the harshest sampled here, is consistent with the hypothesis (Grettenberger, 2021) that slow-growing Gloeobacteria are k-selected species that avoid competition by requiring fewer resources and live in environments near their carrying capacity. A half of Gloeobacterales specimens in this study, including representatives of the most basal lineages in Gloeobacteraceae and Anthocerotibacteraceae, were sampled from polar regions. This encourages that a search for new Gloeobacterales taxa in these areas can provide new insights into the evolutionary history of oxygenic photosynthesis.

Taken together, the analyses presented this study provided new insights on the general pattern of cyanobacterial diversification and biodiversity of Antarctic cyanobacteria. These were obtained based on a large, well-curated dataset, and, arguably, the use of realistic substitution model. Under such expirimental setup, the complexity of analyses is limited by computation time (e.g. half a year in our case). Provided the availability of computational resources, the robustness of these results should be checked in the future on a more representative sample of taxa.

Data availability

The observed alignment, individual protein alignments and the annotated MAGs presented in this study are available from https://doi.org/10.5281/zenodo.14143379.

CRediT authorship contribution statement

Vadim Goremykin: Data curation, Investigation, Writing – review & editing, Conceptualization, Writing – original draft. **Claudia Coleine:** Writing – review & editing, Data curation. **Caterina Ripa:** Data curation. **Laura Selbmann:** Writing – review & editing, Data curation, Funding acquisition. **Claudio Donati:** Supervision, Writing – review & editing, Writing – original draft, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

C.C., C.R., and L.S. wish to thank the Italian National Antarctic Research Program. The Italian Antarctic National Museum (MNA) is kindly acknowledged for financial support to the Mycological Section of the MNA and for providing rock samples used in this study stored in the Culture Collection of Antarctic fungi (MNA-CCFEE), University of Tuscia, Italy. The MNA funded sampling campaigns and research activities in Italy in the frame of PNRA projects.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2025.108419.

References

Albanese, D., Coleine, C., Rota-Stabelli, O., Onofri, S., Tringe, S.G., Stajich, J.E., Selbmann, L., Donati, C., 2021. Pre-Cambrian roots of novel Antarctic cryptoendolithic bacterial lineages. Microbiome 9. https://doi.org/10.1186/s40168-021-01021-0.

Albanese, D., Donati, C., 2021. Genome recovery, functional profiling, and taxonomic classification from metagenomes. In: Methods in molecular biology, p. 153–172. doi: 10.1007/978-1-0716-1099-2_10.

Aguileta, G., Marthey, S., Chiapello, H., Lebrun, M.-h., Rodolphe, F., Fournier, E., Gendrault-Jacquemard, A., Giraud, T., 2008. Assessing the performance of single-copy genes for recovering robust phylogenies. Syst. Biol. 57, 613–627, doi: 10.1080/10635150802306527.

Billi, D., Staibano, C., Verseux, C., Fagliarone, C., Mosca, C., Baqué, M., Rabbow, E., Rettberg, P., 2019a. Dried biofilms of desert strains of chroococcidiopsis survived prolonged exposure to space and mars-like conditions in low earth orbit. Astrobiology 19, 1008–1017. https://doi.org/10.1089/ast.2018.1900.

Billi, D., Verseux, C., Fagliarone, C., Napoli, A., Baqué, M., De Vera, J.-P., 2019b. A desert cyanobacterium under simulated mars-like conditions in low earth orbit: implications for the habitability of mars. Astrobiology 19, 158–169. https://doi.org/10.1089/ast.2017.1807.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. 38, 5825–5829. https://doi.org/10.1093/molbev/msab293.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Chang, J.T., 1996. Full reconstruction of Markov models on evolutionary trees: Identifiability and consistency. Math. Biosci. 137, 51–73. https://doi.org/10.1016/s0025-5564(96)00075-2.

Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2019. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics 36, 1925–1927.

Chen, M.-Y., Teng, W.-K., Zhao, L., Hu, C.-X., Zhou, Y.-K., Han, B.-P., Song, L.-R., Shu, W.-S., 2021. Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. the ISME J. 15. https://doi.org/10.1038/s41396-020-00775-z.

Coleine, C., Albanese, D., Ray, A.E., Delgado-Baquerizo, M., Stajich, J.E., Williams, T.J., Larsen, S., Tringe, S., Pennacchio, C., Ferrari, B.C., et al., 2024. Metagenomics untangles potential adaptations of Antarctic endolithic bacteria at the fringe of habitability. Sci. Total Environ. 917, 170290. https://doi.org/10.1016/j.scitotenv.2024.170290.

Dagan, T., Roettger, M., Stucken, K., Landan, G., Koch, R., Major, P., Gould, S.B., Goremykin, V.V., Rippka, R., De Marsac, N.T., et al., 2013. Genomes of stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. Genome Biol. Evol. 5, 31–44. https://doi.org/10.1093/gbe/evs117.

De los Rios, A., Cary, C., Cowan, D., 2014. The spatial structures of hypolithic communities in the Dry Valleys of East Antarctica. Polar Biol. 37, 1823–1833. https://doi.org/10.1007/s00300-014-1564-0.

Dvořák, P., Hašler, P., Poulíčková, A., 2020. New insights into the genomic evolution of cyanobacteria using herbarium exsiccatae. Eur. J. Phycol. 55, 30–38. https://doi.org/10.1080/09670262.2019.1638523.

Dvořák, P., Jahodářová, E., Stanojković, A., Skoupý, S., Casamatta, D.A., 2023. Population genomics meets the taxonomy of cyanobacteria. Algal Res. 72, 103128. https://doi.org/10.1016/j.algal.2023.103128.

Emms, D.M., Kelly, S., 2018. STAG: species tree inference from all genes. BioRxiv. https://doi.org/10.1101/267914.

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 1–14. https://doi.org/10.1186/s13059-019-1832-y.

Evans, T.W., Kalambokidis, M.J., Jungblut, A.D., Millar, J.L., Bauersachs, T., Grotheer, H., Mackey, T.J., Hawes, I., Summons, R.E., 2022. Lipid biomarkers from microbial mats on the Mcmurdo ice shelf, Antarctica: signatures for life in the cryosphere. Front. Microbiol. 13, 903621. https://doi.org/10.3389/fmicb.2022.903621.

Gaysina, L.A., Saraf, A., Singh, P., 2019. Cyanobacteria in diverse habitats. Cyanobacteria. Academic Press 2019, 1–28. https://doi.org/10.1016/b978-0-12-814667-5.00001-5.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView Version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224. https://doi.org/10.1093/molbev/msp259.

Grettenberger, C.L., Sumner, D.Y., Wall, K., Brown, C.T., Eisen, J.A., Mackey, T.J., Hawes, I., Jospin, G., Jungblut, A.D., 2020. A phylogenetically novel cyanobacterium most closely related to Gloeobacter. the ISME J. 14. https://doi.org/10.1038/s41396-020-0668-5.

Grettenberger, C.L., 2021. Novel Gloeobacterales spp. from diverse environments across the globe. MSphere. 6. https://doi.org/10.1128/msphere.00061-21.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47, D309–D314. https://doi.org/10.1093/nar/gky1085.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 11, 1–11. https://doi.org/10.1186/1471-2105-11-119.

Jung, P., Schermer, M., Briegel-Williams, L., Baumann, K., Leinweber, P., Karsten, U., Lehnert, L., Achilles, S., Bendix, J., Büdel, B., 2019. Water availability shapes edaphic and lithic cyanobacterial communities in the Atacama Desert. J. Phycol. 55, 1306–1318. https://doi.org/10.1111/jpy.12908.

Jung, P., Mikhailyuk, T., Emrich, D., Baumann, K., Dultz, S., Büdel, B., 2020. Shifting boundaries: ecological and geographical range extension based on three new species in the cyanobacterial genera cyanocohniella, oculatella, and, aliterell. J. Phycol. 56, 1216–1231. https://doi.org/10.1111/jpy.13025.

Jung, P., Azua-Bustos, A., Gonzalez-Silva, C., Mikhailyuk, T., Zabicki, D., Holzinger, A., Lakatos, M., Büdel, B., 2021. Emendation of the Coccoid Cyanobacterial Genus Gloeocapsopsis and Description of the New Species Gloeocapsopsis diffluens sp. nov. and Gloeocapsopsis dulcis sp. nov. Isolated From the Coastal Range of the Atacama Desert (Chile). Front. Microbiol. 12:671742, DOI: 10.3389/fmicb.2021.671742.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.

Katoh, K., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–518. https://doi.org/10.1093/nar/gki198.

Kaštovský, J., 2023. Welcome to the jungle!: An overview of modern taxonomy of cyanobacteria. Hydrobiologia 851, 1063–1077. https://doi.org/10.1007/s10750-023-05356-7.

Komárek, J., Johansen, J.R., Šmarda, J., Strunecký, O., 2020. Phylogeny and taxonomy of Synechococcus-like cyanobacteria. Fottea. 20, 171–191. https://doi.org/10.5507/fot.2020.006.

Komárek, J., Kastovsky, J., Mares, J., Johansen, J.R., 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. Preslia 86, 295–335. https://cyanobacteria.myspecies.info/node/184.

Larsen, S., Coleine, C., Albanese, D., Stegen, J.C., Selbmann, L., Donati, C., 2024. Geology and elevation shape bacterial assembly in Antarctic endolithic communities. Sci. Total Environ. 907, 168050. https://doi.org/10.1016/j.scitotenv.2023.168050.

Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288. https://doi.org/10.1093/bioinformatics/btp368.

Lartillot, N., 2020. PhyloBayes: Bayesian phylogenetics using site-heterogeneous models. In: Scornavacca, C., Delsuc, F., Galtier, N., (eds), Phylogenetics in the genomic era. No commercial publisher | Authors open access book: 1.5:1-1.5:16. https://hal.science/hal-02535342.

Lozano-Fernandez, J., 2022. A practical guide to design and assess a phylogenomic study. Genome Biol. Evol. 14, evac129. https://doi.org/10.1093/gbe/evac129.

Mantas, M.J.Q., Nunn, P.B., Ke, Z., Codd, G.A., Barker, D., 2021. Genomic insights into the biosynthesis and physiology of the cyanobacterial neurotoxin 2,4-diaminobutanoic acid (2,4-DAB). Phytochem. 192, 112953. https://doi.org/10.1016/j.phytochem.2021.112953.

Mareš, J., 2018. Multilocus and SSU rRNA gene phylogenetic analyses of available cyanobacterial genomes, and their relation to the current taxonomic system. Hydrobiologia 811, 19–34. https://doi.org/10.1007/s10750-017-3373-2.

Napoli, A., Micheletti, D., Pindo, M., Larger, S., Cestaro, A., De Vera, J.-P., Billi, D., 2022. Absence of increased genomic variants in the cyanobacterium Chroococcidiopsis exposed to Mars-like conditions outside the space station. Sci. Rep. 12, 8437. https://doi.org/10.1038/s41598-022-12631-5.

Nguyen, L.-T., Schmidt, H.A., Von Haeseler, A., 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.

Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F., 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. the ISME J. 11. https://doi.org/10.1038/ismej.2017.126.

Ondov, B.D., Starrett, G.J., Sappington, A., Kostic, A., Koren, S., Buck, C.B., Phillippy, A. M., 2019. Mash Screen: high-throughput sequence containment estimation for genome discovery. Genome Biol. 20, 1–13. https://doi.org/10.1186/s13059-019-1841-x.

Orakov, A., Fullam, A., Coelho, L.P., Khedkar, S., Szklarczyk, D., Mende, D.R., Schmidt, T.S.B., Bork, P., 2021. GUNC: detection of chimerism and contamination in prokaryotic genomes. Genome Biol. 22, 1–19. https://doi.org/10.1186/s13059-021-02393-0.

Pardo-De la Hoz, C.J., Magain, N., Piatkowski, B., Cornet, L., Forno, M.D., Carbone, I., Miadlikowska, J., Lutzoni, F., 2023. Ancient rapid radiation explains most conflicts among gene trees and well-supported phylogenomic trees of nostocalean cyanobacteria. Syst. Biol. 72, 694–712. https://doi.org/10.1093/sysbio/syad008.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055. https://doi.org/10.1101/gr.186072.114.

Pearce, D.A., Newsham, K.K., Thorne, M.a.S., Calvo-Bado, L.A., Krsek, M., Laskaris, P., Hodson, A., Wellington, E.M.H., 2012. Metagenomic analysis of a southern maritime antarctic soil. Front. Microbiol. 3, 403. doi: 10.3389/fmicb.2012.00403.

Pessi, I.S., Popin, R.V., Durieu, B., Lara, Y., Tytgat, B., Savaglia, V., Roncero-Ramos, B., Hultman, J., Verleyen, E., Vyverman, W., et al., 2023. Novel diversity of polar Cyanobacteria revealed by genome-resolved metagenomics. Microb. Genom. 9, 001056. https://doi.org/10.1099/mgen.0.001056.

Rahmatpour, N., Hauser, D.A., Nelson, J.M., Chen, P.Y., A, J.C.V., Ho, M.-Y., Li, F.-W., 2021. A novel thylakoid-less isolate fills a billion-year gap in the evolution of Cyanobacteria. Curr. Biol. 31:2857-2867.e4. doi: 10.1016/j.cub.2021.04.042.

Rego, A., Raio, F., Martins, T.P., Ribeiro, H., Sousa, A.G.G., Séneca, J., Baptista, M.S., Lee, C.K., Cary, S.C., Ramos, V., et al., 2019. Actinobacteria and cyanobacteria diversity in terrestrial antarctic microenvironments evaluated by culture-dependent and independent methods. Front. Microbiol. 10, 1018. https://doi.org/10.3389/fmicb.2019.01018.

Rigonato, J., Gama, W.A., Alvarenga, D.O., Branco, L.H.Z., Brandini, F.P., Genuário, D.B., Fiore, M.F., 2016. Aliterella atlantica gen. nov., sp. nov., and Aliterella antarctica sp. nov., novel members of coccoid Cyanobacteria. Int. J. Syst. Evol. Microbiol. 66, 2853–2861. https://doi.org/10.1099/ijsem.0.001066.

RoyChoudhury, A., Willis, A., Bunge, J., 2015. Consistency of a phylogenetic tree maximum likelihood estimator. J. Stat. Plan. Inference. 161, 73–80. https://doi.org/10.1016/j.jspi.2015.01.001.

Sánchez-Baracaldo, P., Hayes, P.K., Blank, C.E., 2005. Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach. Geobiology 3, 145–165. https://doi.org/10.1111/j.1472-4669.2005.00050.x.

Schulze-Makuch, D., Lipus, D., Arens, F.L., Baqué, M., Bornemann, T.L.V., De Vera, J.-P., Flury, M., Frösler, J., Heinz, J., Hwang, Y., et al., 2021. Microbial hotspots in lithic microhabitats inferred from DNA fractionation and metagenomics in the Atacama Desert. Microorganisms. 9, 1038. https://doi.org/10.3390/microorganisms9051038.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Steel, M., 2013. Consistency of Bayesian inference of resolved phylogenetic trees. J. Theor. Biol. 336, 246–249. https://doi.org/10.1016/j.jtbi.2013.08.012.

Steenwyk, J.L., Goltz, D.C., Buida III, T.J., Li, Y., Shen, X.X., Rokas, A., 2022. OrthoSNAP: A tree splitting and pruning algorithm for retrieving single-copy orthologs from gene family trees. PLoS Biol. 20, e3001827. https://doi.org/10.1371/journal.pbio.3001827.

Strunecký, O., Komárek, J., Johansen, J., Lukešová, A., 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.

Strunecký, O., Ivanova, A.P., Mareš, J., 2023. An updated classification of cyanobacterial orders and families based on phylogenomic and polyphasic analysis. J. Phycol. 59, 12–51. https://doi.org/10.1111/jpy.13304.

Swingley, W.D., Blankenship, R.E., Raymond, J., 2008. Integrating Markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. Mol. Biol. Evol. 25, 643–654. https://doi.org/10.1093/molbev/msn034.

Szánthó, L.L., Lartillot, N., Szöllősi, G.J., Schrempf, D., 2023. Compositionally constrained sites drive long-branch attraction. Syst. Biol. 72, 767–780. https://doi.org/10.1093/sysbio/syad013.

Talavera, G., Vila, R., 2011. What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. BMC Evol. Biol. 11, 1–15. https://doi.org/10.1186/1471-2148-11-315.

Taton, A., Grubisic, S., Brambilla, E., De Wit, R., Wilmotte, A., 2003. Cyanobacterial diversity in natural and artificial microbial mats of Lake Fryxell (McMurdo Dry Valleys, Antarctica): a morphological and molecular approach. Appl. Environ. Microbiol. 69, 5157–5169.

Tschoeke, D., Salazar, V.W., Vidal, L., Campeão, M., Swings, J., Thompson, F., Thompson, C., 2020. Unlocking the genomic taxonomy of the prochlorococcus collective. Microb. Ecol. 80, 546–558. https://doi.org/10.1007/s00248-020-01526-5.

Truszkowski, J., Goldman, N., 2016. Maximum likelihood phylogenetic inference is consistent on multiple sequence alignments, with or without gaps. Syst. Biol. 65, 328–333. https://doi.org/10.1093/sysbio/syv089.

Van Goethem, M.W., Cowan, D.A., 2019. Role of Cyanobacteria in the ecology of polar environments, in: Springer polar sciences. p. 3–23. DOi: 10.1007/978-3-030-02786-5_1.

Velichko, N., Smirnova, S., Averina, S., Pinevich, A., 2021. A survey of Antarctic cyanobacteria. Hydrobiologia 848, 2627–2652. https://doi.org/10.1007/s10750-021-04588-9.

Wang, H.C., Susko, E., Roger, A.J., 2019. The relative importance of modeling site pattern heterogeneity versus partition-wise heterotachy in phylogenomic inference. Syst. Biol. 68, 1003–1019. https://doi.org/10.1093/sysbio/syz021.

Zakhia, F., Jungblut, A.-D., Taton, A., Vincent, W.F., Wilmotte, A., 2008. Cyanobacteria in Cold Ecosystems. In: Psychrophiles: from Biodiversity to Biotechnology, Springer eBooks. p. 121–135. DOi: 10.1007/978-3-540-74335-4_8.

Zhaxybayeva, O., Gogarten, J.P., Charlebois, R.L., Doolittle, W.F., Papke, R.T., 2006. Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. Genome Res. 16, 1099–1108. https://doi.org/10.1101/gr.5322306.