



## OPEN ACCESS

### EDITED BY

Srijak Bhatnagar,  
Athabasca University, Canada

### REVIEWED BY

Fei Tao,  
Shanghai Jiao Tong University, China  
Chinthala Paramageetham,  
Sri Venkateswara University, India

### \*CORRESPONDENCE

Luc Cornet  
✉ Luc.Cornet@uliege.be  
Jimmy H. Saw  
✉ jsaw@gwu.edu

RECEIVED 05 March 2026

REVISED 08 April 2026

ACCEPTED 09 April 2026

PUBLISHED 28 April 2026

### CITATION

Sudianto E, Shlafstein MD, Durieu B,  
Harmel M, Cornet L and Saw JH (2026)  
Taxonomic description of cyanobacteria  
from extreme habitats through  
genome-based classification.  
*Front. Microbiol.* 17:1824103.  
doi: 10.3389/fmicb.2026.1824103

### COPYRIGHT

© 2026 Sudianto, Shlafstein, Durieu,  
Harmel, Cornet and Saw. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which does  
not comply with these terms.

# Taxonomic description of cyanobacteria from extreme habitats through genome-based classification

Edi Sudianto<sup>1</sup>, Maximillian D. Shlafstein<sup>2</sup>, Benoit Durieu<sup>1</sup>,  
Marie Harmel<sup>1</sup>, Luc Cornet<sup>1\*</sup> and Jimmy H. Saw<sup>2\*</sup>

<sup>1</sup>InBioS–PhytoSYSTEMS, Eukaryotic Phylogenomics, University of Liège, Liège, Belgium, <sup>2</sup>Department of Biological Sciences, The George Washington University, Washington, DC, United States

**Introduction:** Cyanobacteria form a morphologically and phylogenetically diverse group of oxygenic phototrophic bacteria inhabiting a wide range of environments, including extreme habitats such as hot springs and volcanic steam vents. Many lineages, particularly those from these extreme environments, remain uncultured and are known only from metagenome-assembled genomes (MAGs), limiting their integration into formal taxonomy.

**Methods:** Analysis of 46 steam vent associated samples from Hawai'i using 16S rRNA amplicon sequencing revealed that cyanobacteria dominate these communities. *Gloeobacter kilaeensis* dominated pit-like environments with low-light conditions, while *Leptolyngbyaceae* and other families are more dominant in structured soil and wall communities. We further reconstructed 38 high-quality cyanobacterial MAGs and incorporated them into a phylogenomic analysis comprising 343 cyanobacterial genomes, followed by genome-based comparisons against 9,026 reference genomes.

**Results:** This revealed eight novel species and one novel genus spanning five orders: Chroococciopsidales, Leptolyngbyales, Nostocales, Oculatellales, and Oscillatoriales. Following SeqCode guidelines, we provide the first formal taxonomic descriptions of cyanobacterial MAGs and propose guidelines for integrating genome-based and cultivated material.

**Conclusion:** These findings highlight Hawaiian steam vents as hotspots of previously uncharacterized cyanobacterial diversity and underscore the importance of genome-based nomenclature.

### KEYWORDS

cyanobacteria, fumaroles, Hawai'i, SeqCode, steam vents, unculturable bacteria

## Introduction

Cyanobacteria represent a morphologically and phylogenetically diverse group of oxygenic phototrophic prokaryotes that inhabit a wide range of environments, from freshwater and marine ecosystems to terrestrial and extreme habitats such as hot springs, deserts, and polar regions (Whitton and Potts, 2012). Unlike most other prokaryotes, and due to their phototrophic nature, cyanobacteria are typically isolated and maintained as living cultures to distinguish them from potential contaminants. Accordingly, they are typically grown in a minimal medium containing only salts, minerals, and vitamin B12 (Rippka et al., 1979; Castenholz, 1981; Rippka et al., 1981; Waterbury, 2006). Weissman et al. (2021) demonstrated that

culture-based approaches are heavily biased toward copiotrophic organisms (i.e., adapted to nutrient-rich conditions), notably among cyanobacteria, while failing to capture many oligotrophs (i.e., adapted to low-nutrient environments). Recent comparisons of barcoding and genomic datasets indeed indicate that several evolutionarily important cyanobacterial lineages are absent from both cultured and genomic databases (Cornet et al., 2018, 2023). Consequently, many cyanobacteria present in natural environments remain uncultured and are known only through metagenome-assembled genomes (MAGs).

These MAGs mostly remain undescribed, with no validly published taxonomic names, which limits their integration into formal microbial taxonomy and ecological analyses. Indeed, validated naming ensures clear communication among microbiologists and prevents the instability and redundancy inherent to provisional *Candidatus* names that lack official standing and nomenclatural priority (Whitman et al., 2024). The Code of Nomenclature of Prokaryotes Described from Sequence Data (SeqCode) enables the formal taxonomic description of uncultured prokaryotes based solely on genome sequences, provided that specific quality and metadata criteria are met (Hedlund et al., 2022), thereby offering an improved integration of MAGs into prokaryotic systematics. Nevertheless, the duality of cyanobacterial nomenclature—historically governed by both the International Code of Nomenclature for algae, fungi, and plants (ICN) and the International Code of Nomenclature of Prokaryotes (ICNP) (Oren, 2020; Pinevich and Averina, 2024)—complicates the taxonomic treatment of uncultured cyanobacteria. While the ICN traditionally relies on morphological descriptions and/or 16S rRNA sequences, the ICNP is primarily genome-based; consequently, integrating taxa known only from MAGs, which often lack 16S rRNA sequences (Cornet et al., 2018), across these two nomenclatural codes remains particularly challenging.

Extreme environments host microorganisms capable of surviving extreme temperatures, high salinity, intense radiation, or deep-sea pressure (Taufiq et al., 2026). In these extreme ecosystems, cyanobacteria often constitute the main primary producers (Stal, 1995; Obbels et al., 2016), yet they are notoriously difficult to cultivate. Steam vents (fumaroles) are a striking example: driven by volcanic activity, they exhibit drastic fluctuations in temperature, pH, pressure, and moisture; all of which shape their microbial communities (Hadland et al., 2024). Despite these harsh conditions, the microbial communities of Hawaiian steam vents are unexpectedly diverse (Prescott et al., 2022). Wall et al. (2015) proposed that such fumaroles act as “extremophile collectors,” serving as biodiversity hotspots for novel extremophiles. Few cyanobacterial taxa have been isolated from Hawai‘i’s steam vents, including *Gloeobacter kilaueensis* (Saw et al., 2013) and *Fischerella* sp. JS2 and *Kovacicikia* sp. JS3 (Prescott et al., 2023). Additionally, 16S rRNA sequence-based studies reveal a substantial diversity of uncultured cyanobacteria within these environments (Wall et al., 2015; Prescott et al., 2022), underscoring Hawai‘i’s steam vents as promising reservoirs of novel cyanobacterial lineages.

In this study, we analyzed 46 environmental samples collected from Hawai‘i’s steam vents and found a high relative abundance of cyanobacteria. We reconstructed 38 high-quality MAGs, among which 8 represent novel cyanobacterial species and 1 novel cyanobacterial genus, as assessed by genome-based classification methods. Specifically, we conducted phylogenomic analyses on the 38 MAGs together with a curated set of 343 reference cyanobacterial and closely related non-photosynthetic genomes and performed average nucleotide identity (ANI) and average amino-acid identity (AAI)

comparisons against 9,026 cyanobacterial and melainobacterial genomes available from NCBI. We further propose the first taxonomic description of cyanobacterial MAGs in SeqCode, following the genomic criterion of  $\geq 95\%$  ANI for species delineation (Jain et al., 2018). Finally, we provide practical guidelines for taxonomic description of cyanobacterial MAGs, emphasizing the integration of cultivated material.

## Materials and methods

### Samples collection and DNA extraction

Detailed information regarding sampling locations and DNA extractions was described in a previous study (Prescott et al., 2022). Briefly, the samples were collected during an expedition in 2019. They came from three volcanic features on the island of Hawai‘i (also known as the Big Island) – a cave located inside the Kilauea Caldera within Hawaii Volcanoes National Park and two steam vent features located within the East Rift Zone (Prescott et al., 2022). Detailed methods describing the sequencing of the samples were reported in Saw et al. (2026).

### Amplicon sequencing

Amplicon sequencing data targeting the V3-4 region of the 16S rRNA gene, using primers (CCTACGGGNGGCWGCAG and GACTACHVGGGTATCTAATCC), were processed using the RASPAM pipeline (Harmel et al., 2025),<sup>1</sup> in ASV mode. Analyses were performed with DADA2 v1.22.0 (Callahan et al., 2016), using the core functions *plotQualityProfile*, *filterAndTrim*, *learnErrors*, *derepFastq*, *dada*, *mergePairs*, and *assignTaxonomy*. Quality profiles were inspected to determine optimal trimming and filtering parameters, resulting in the use of *truncLenF* = 290, *truncLenR* = 240, *trimLeftF* = 5, *trimLeftR* = 5, *maxEE* = 2 (forward) and 5 (reverse), a minimum overlap of 12 bp, and zero mismatches allowed during merging. Chimera removal was performed using the consensus method implemented in DADA2 (Callahan et al., 2016). ASV tables were further filtered according to abundance thresholds, applying both min occurrence per ASV (across samples) = 100 and min occurrence of ASV in the sample = 200, to retain high-confidence sequences. Taxonomic assignment of the resulting ASVs was performed using four independent reference databases: RDP trainset19 (Wang and Cole, 2024), CyanoSeq v1.3 (Lefler et al., 2023), CyanoSeq v1.3 with added species-level annotations, and CABO-16S (Eitel et al., 2025). No lineage-based filtering was applied to preserve the full taxonomic diversity. Diversity analyses and plots were generated in R v4.1.2 using the packages *phyloseq* v1.38.0 (McMurdie and Holmes, 2013), *Biostrings* v2.62.0 (Lifshitz et al., 2022), *dplyr* v1.1.4 (Wickham et al., 2023), *vegan* v2.7.1 (Dixon, 2003), and *ggplot2* v4.0.0 (Wickham, 2016). Sequence identity comparison was performed using BLAST+ (Camacho et al., 2009), against a Cyanobacteria 16S reference dataset retrieved with *edirect* v23.9 (Kans, 2025): (*esearch -db nucleotide -query “Cyanobacteria[Organism] AND 16S [Title]” | efetch -format fasta*). All analyses were run using Nextflow v21.08.0 using the following

<sup>1</sup> <https://bitbucket.org/phylogeno/raspam/src/main/>

commands: nextflow run main.nf --first --cpu = 40 --data = './data/2019\_amplicons' --outdir = "JIMMY\_SAW" --asv; nextflow run main.nf --second --cpu = 80 --data = './data/2019\_amplicons' --outdir = "JIMMY\_SAW" --asv --truncLenF = 290 --truncLenR = 240 --trimLeftF = 5 --trimLeftR = 5 --max\_error\_F = 2 --max\_error\_R = 5 --top\_taxa = 200 --feature\_sample\_min\_reads = 200 --feature\_min\_occurrence = 100 --filter\_taxo = "all."

## Metagenome assembly

Raw metagenomic sequences were processed using BBDuk v38.87<sup>2</sup> with input parameters: ktrim = r, minlen = 50, mink = 11, tbo = t, rcomp = f, k = 21, ow = t, zl = 4, qtrim = rl, trimq = 20, to remove Illumina adapter sequences and low quality regions. Quality filtering excluded regions with a score less than 20 and reads under 50 bp in length. FastQC v0.11.9 was then used to assess the quality of the trimmed sequences before downstream processing. Sequences from the same sample sites were combined before assembly to improve sequencing depth. MetaSPAdes v3.15.4 (Nurk et al., 2017) was used to assemble the combined trimmed sequences using default parameters but with k-mers of 21, 33, 55, and 77. Seqtk v1.3<sup>3</sup> was used to first create a list of the assembled contigs over 1 kb, which were then mapped back to the trimmed reads with BBWrap v38.87 (see footnote 2) using default parameters. For combined samples, contigs were individually mapped to each of the trimmed reads before combination. These mapped reads were sorted and indexed using Samtools v1.10 (Danecek et al., 2021) with default parameters. The JGI summarize script that was packaged with BBmap tool was then used to determine the depths of contigs over 1,500 bp, allowing for binning with MetaBAT2 v2.15 (Kang et al., 2019) using default parameters.

## Metagenomic analyses

The 66 assembled MAGs were first checked for genome completeness and contamination using CheckM v1.2.2 (Parks et al., 2015) and GUNC v1.0.5 (Orakov et al., 2021) functions in the GENERA toolbox v3.0.0 (Cornet et al., 2023). Next, we selected 40 MAGs with >90% completeness and <5% contamination, meeting the data quality requirements for type genome assembly quality in the SeqCode Registry (Hedlund et al., 2022). The genomes were further classified using GTDB-Tk v2.0.0 (Parks et al., 2022) in the GENERA toolbox (Cornet et al., 2023). We run fastANI v1.33 (Jain et al., 2018) and FastAAI (Gerhardt et al., 2025) against the 9,026 Cyanobacteria/Melainabacteria genomes (7,913 cyanobacteria and 1,113 Vampirovibrionophyceae, last download date, July 10, 2025) available in the NCBI Genome Datasets (O'Leary et al., 2024) for MAG species identification based on ANI and AAI. The Hidden Markov Model profiles (pHMMs) for 115 genes associated with phycobilisomes, photosynthesis, and respiratory machinery (Lumian et al., 2021) were obtained from the KEGG database (Kanehisa et al., 2025) and queried against potential novel cyanobacterial genomes using the hmmsearch function in HMMER v3.4.<sup>4</sup>

To verify the phylogenetic placements of the MAGs, we combined the 38 identified cyanobacterial MAGs with 343 curated

cyanobacterial and closely related non-photosynthetic genomes from Sudianto et al. (2025). A supermatrix was generated for the 381 taxa using GToTree v1.8.8 (Lee, 2019) with 251 individual alignments produced using the cyanobacterial pHMMs available in the tool. Sequence alignment and trimming were performed with MUSCLE v5.1 (Edgar, 2022) and TrimAl v1.4.rev15 (Capella-Gutiérrez et al., 2009), respectively. GToTree was run with minor parameter adjustments ("c 1 -G 0"). We used IQ-TREE v2.3.6 (Minh et al., 2020), with automated ModelFinder for model selection (-m MFP; Kalyaanamoorthy et al., 2017) and 1,000 ultrafast bootstrap replicates (Hoang et al., 2018), to infer the phylogenomic tree. The resulting tree and associated information (e.g., genome completeness, contamination, ANI%) were visualized using iTOL v7 (Letunic and Bork, 2024).

## Results

### Amplicon sequencing

From 10,879,048 raw reads, 4,424,992 passed filtering, 4,005,210 merged successfully, and 2,438,362 non-chimeric reads were retained, yielding 18,830 ASVs. Abundance-based filtering kept ASVs present in  $\geq 200$  reads per sample and occurring  $\geq 100$  times across the dataset, resulting in 2,542 high-confidence ASVs. The four sampling sites harbor distinct bacterial communities, largely dominated by Cyanobacteriota and Chloroflexota (Figure 1). In the open area of the East Rift Zone, near the pit-like feature, Cyanobacteriota represent 49.2% of the community and are dominated by Hapalosiphonaceae (64.5%), Scytonemataceae (14.9%), and Oculatellaceae (11%) (Supplementary Table S1). These communities contrast with those found inside the pit-like feature, where Cyanobacteriota reach 79% and are strongly dominated by the family Gloeobacteraceae (85.2%). Inside the Big Ell cave (Kilauea Caldera), community composition varies between the soil-dominated by Chloroflexota (60%)—and the wall, which harbors a more balanced assemblage of Cyanobacteriota (47.7%) and Chloroflexota (40%). We further identified five rare cyanobacterial ASVs with BLAST sequence similarity ranging from 86.23–93.15% (Table 1). Following CyanoSeq classification, the lineage of these five ASVs is either Prochlorococcaceae (3), Leptolyngbyaceae (1), or broadly Cyanophyceae (1). Interestingly, these rare cyanobacterial ASVs are found only in pit-like areas, suggesting they may be the source of atypical or novel cyanobacterial spots.

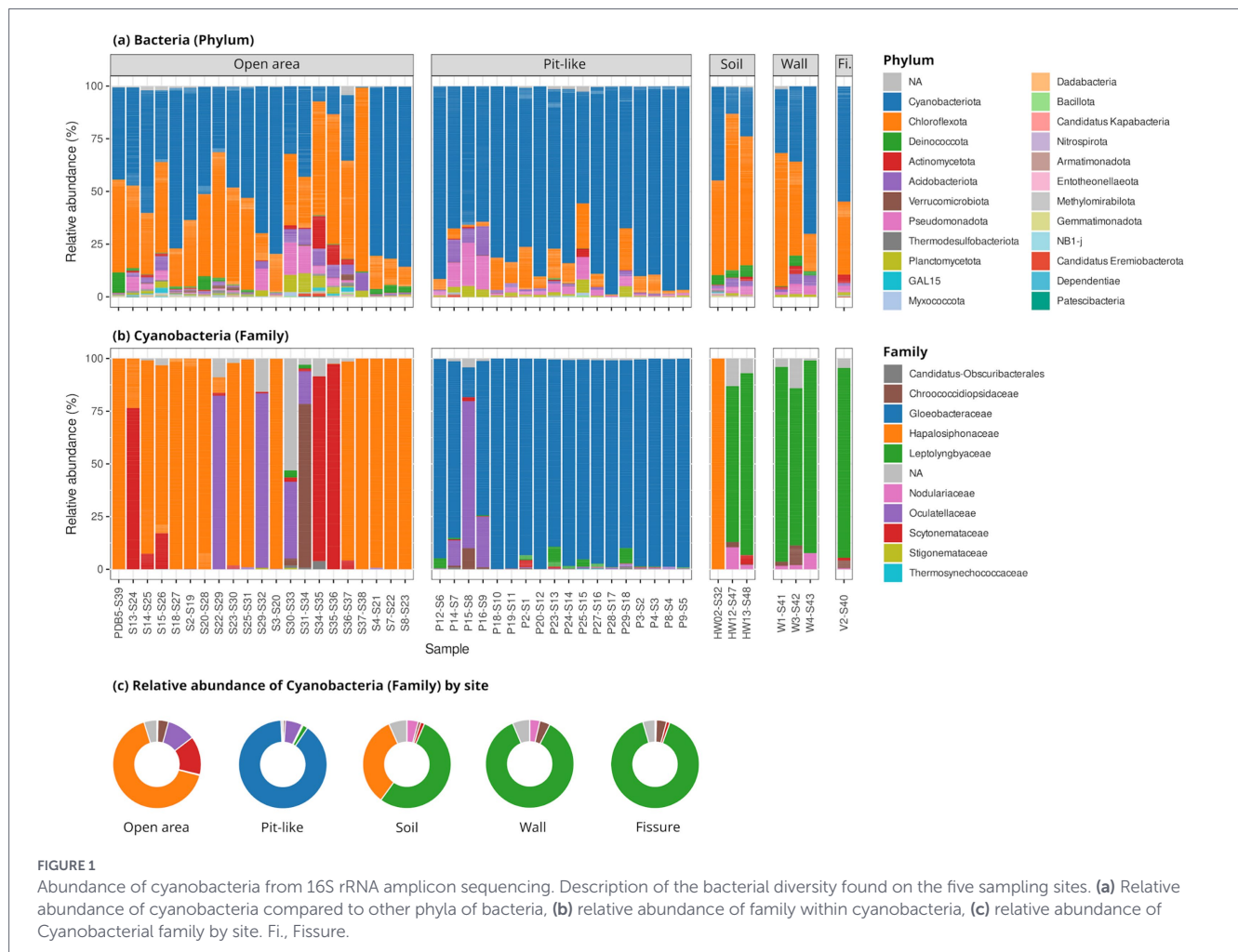
### Identification of novel cyanobacterial MAGs

Out of the 66 cyanobacterial MAGs, 40 MAGs with high CheckM completeness (>95%) and low contamination scores (<5%) were selected. These MAG sizes range from 4.5–7.9 Mb and ANI of 78.7–99.9% to available cyanobacterial genomes (Supplementary Table S2). Two MAGs (S29-31\_bin.128 and P18-20\_bin.4) that are potentially Vampirovibrionophyceae were excluded from further analyses, resulting in 38 cyanobacterial MAGs. Phylogenomic tree indicates that these 38 MAGs are nested within the Gloeobacterales, Oculatellales, Leptolyngbyales, Oscillatoriales, Chroococidiopsiales, and Nostocales (Figure 2). Twenty two of these MAGs can be identified down to the species level (ANI  $\geq 95\%$ ; Figure 2; Supplementary Table S2), such as *Gloeobacter kilaueensis* (P23-25\_bin.50, P27-29\_bin.2, P18-20\_bin.17, P2-4\_bin.3, and P8-12\_bin.7), cyanobacteriota bacterium JS3

2 <https://sourceforge.net/projects/bbmap/>

3 <https://github.com/lh3/seqtk>

4 <http://hmmmer.org/>



(W1-4\_bin.70<sup>TS</sup>, HW-13\_bin.4, P23-25\_bin.53, and V2\_bin.1), *Fischerella* sp. FACHB-330 (S18-20\_bin.32, S34-36\_bin.84, S13-15\_bin.40, and S2-4\_bin.62), and cyanobacterium PCC 7702 (HW-02\_bin.18, S2-4\_bin.85, S34-36\_bin.68, PDB5\_bin.16, S13-15\_bin.112, S22-25\_bin.34, S7-8\_bin.7, and S18-20\_bin.8). The 16 remaining MAGs have ANI < 95%, indicating the genomes of potential new cyanobacteria. Here, we describe nine novel cyanobacterial genomes according to the SeqCode's classification, comprising eight novel species and one novel genus (Table 2). The AAI scores are >65% to the closest cyanobacteria genome available in NCBI for these nine MAGs (Supplementary Table S2). Nevertheless, the phylogenetic placement of W1-4\_bin.36<sup>TS</sup> between the Nostocaceae and the Scytonemataceae (Figure 2) indicates the presence of a novel genus.

## Photosynthetic potential of the novel cyanobacterial MAGs

As oxygenic photosynthesis is the synapomorphy of cyanobacteria, we used 115 genes associated with the phycobilisome, photosynthetic apparatus, and respiratory machinery to characterize the novel cyanobacterial species identified from our MAGs (Supplementary Table S3). W1-4\_bin.70<sup>TS</sup> encodes the fewest, largely because it lacks the cytochrome bd complex (*cydABX*). In contrast, the genomes of S29-31\_bin.237<sup>TS</sup> and W1-4\_bin.36<sup>TS</sup> each encode 100 of the 115 genes. All

MAGs encode a complete or nearly complete phycobilisome gene set. The core allophycocyanin genes (*apcABCDE*) and rod phycocyanin/phycoerythrocyanin genes (*cpcABCDEFG*) are present in all MAGs. The phycoerythrin gene set (*cpeABCDEFSTUYZ*) is broadly conserved; however, the *cpeR* gene is absent from five out of the nine MAGs, specifically P23-25\_bin.64<sup>TS</sup>, S29-31\_bin.125<sup>TS</sup>, S29-31\_bin.237<sup>TS</sup>, S29-31\_bin.60<sup>TS</sup>, and W1-4\_bin.70<sup>TS</sup>. Several photosystem I and II genes (e.g., *psaGHNO* and *psbQRW*) were absent from these nine cyanobacterial MAGs as they are eukaryote-specific subunits. All nine MAGs contain a full complement of 29 genes associated with F-type ATPase, NADH dehydrogenase, and photosynthetic electron transport chains. The only gene missing from any genome is *hoxE*, which is absent from S13-15\_bin.75<sup>TS</sup>, S29-31\_bin.125<sup>TS</sup>, and W1-4\_bin.70<sup>TS</sup> (Supplementary Table S3). Several genes responsible for the cytochrome b6f complex (*petGN*), cytochrome c oxidase (*ctaF*), and succinate dehydrogenase (*sdhCD*) are missing from most of the nine MAGs. Specifically, *petG* occurs in 2/9 MAGs and *petN* in 1/9, whereas *ctaF*, *sdhC*, and *sdhD* are not detected in any MAG (Supplementary Table S3).

## SeqCode description

### Description of *Kovacikia ellensis* sp. nov

*Kovacikia ellensis* (el'len.sis (N. L. masc./fem. Adj. *ellensis*), from Ell-, referring to the "Big Ell" cave; L. suffix *-ensis*, meaning

TABLE 1 Rare cyanobacterial ASV sequences from steam vents.

ASV name	BLAST Searches			CyanoSeqV1.3 lineage	No. of detected samples at sampling sites						
	Hit ID	Description	Sequence similarity (%)		P14	P16	P24	P25	P27	P28	P29
ASV_1789	OR727807.1	Unicellular cyanobacterium A47	93.151	Bacteria; Cyanobacteriota; Cyanophyceae; Synechococcales; Prochlorococcaceae; Synechococcus_B	0	0	0	0	0	0	135
ASV_2010	AM940917.1	Uncultured cyanobacterium partial 16S rRNA gene; clone B9_86	93.072	Bacteria; Cyanobacteriota; Cyanophyceae; Leptolyngbyales; Leptolyngbyaceae	61	0	53	0	0	6	0
ASV_2139	AM940917.1	Uncultured cyanobacterium partial 16S rRNA gene; clone B9_86	93.023	Bacteria; Cyanobacteriota; Cyanophyceae	0	86	0	20	0	8	0
ASV_2222	KC262697.1	Uncultured cyanobacterium clone OTUC33	86.404	Bacteria; Cyanobacteriota; Cyanophyceae; Synechococcales; Prochlorococcaceae; Synechococcus_B	0	0	0	0	111	0	0
ASV_2472	GU271464.1	Uncultured Trichodesmium sp. clone 1_241	86.23	Bacteria; Cyanobacteriota; Cyanophyceae; Synechococcales; Prochlorococcaceae; Synechococcus_B	0	0	0	0	0	0	102

originating from or pertaining to; N. L. masc./fem. Adj. *ellensis*, originating from the Ell cave). Four MAGs corresponding to this species were derived from samples collected from diverse sites (open area, wall, soil) of the Big Ell cave, Hawai'i. The type genome (W1-4\_bin.70<sup>TS</sup>) is 6.30 Mb, with CheckM completeness of 98.82% and contamination of 0.47%. This species has been previously cultured and has its genome deposited as "cyanobacteriota bacterium JS3" in the NCBI (GCA\_034541955.1, ANI: 99.6%; AAI: 82.8%) by Prescott et al. (2023), but has not been officially named using any nomenclature guidelines. Here, we proposed using the SeqCode nomenclature to name this species. The genome lacks a complete cytochrome bd complex (*cydABX*), but retains genes related to the cytochrome b6f complex (*pet-ABCDLM*) and cytochrome c oxidase (*ctaCDE*). In total, this MAG encodes 96 (out of 115) genes related to phycobilisome, photosynthesis, and respiratory machinery. The type genome has been deposited in ENA with the accession number ERZ28674871<sup>TS</sup>. Phylogenomic analysis placed this genome in the order Leptolyngbyales. *Kovackikia ellensis* belongs to the family Leptolyngbyaceae, order Leptolyngbyales, class Cyanophyceae, and phylum Cyanobacteriota. The name has been registered under SeqCode: seqco.de/r:s5njm2s9.

### Description of *Paranostoc* gen. nov.

*Paranostoc* (Pa.ra.nos'toc Gr. prep. *Para*, beside, near; N. L. neut. n. *Nostoc*, a bacterial genus; N. L. neut. n. *Paranostoc*, a genus related to or resembling *Nostoc*).

### Description of *Paranostoc speluncae* sp. nov.

*Paranostoc speluncae* (spe.lun'cae (N. L. gen. n./adj. *speluncae*), L. fem. n. *spelunca*, cave; L. gen. sg. ending -ae, indicating association or inhabitation; N. L. gen. n./adj. *Speluncae*, inhabiting or associated with a cave). This MAG (W1-4\_bin.36<sup>TS</sup>) has a size of 7.8 Mb, completeness of 99.04%, and contamination of 1.59%. This MAG was assembled from sequencing reads generated from samples taken from a cave wall located near the entrance of Big Ell. This is the same wall from which *Gloeobacter kilaueensis* JS1 was isolated in 2013 (Saw et al., 2013). The closest identified species is a Nostocales cyanobacterium (GCA\_021324415.1, ANI: 81.4%; AAI: 74.3%). In total, this MAG encodes 100 (out of 115) genes related to phycobilisome, photosynthesis, and respiratory machinery. This type genome has been deposited in ENA with the accession number ERZ28674878<sup>TS</sup>. Phylogenomic analysis placed

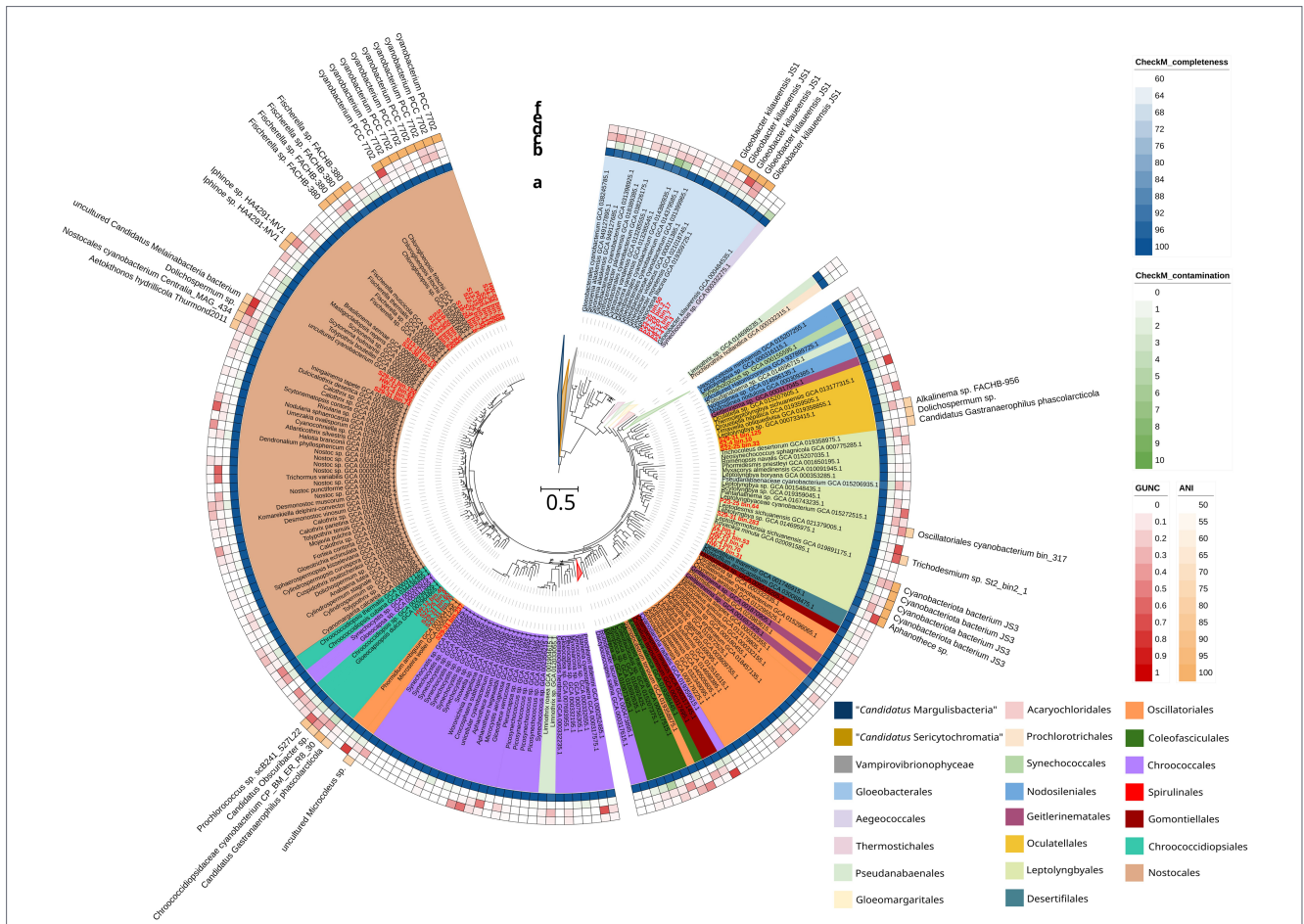


FIGURE 2

Maximum likelihood phylogenomic tree of 381 cyanobacterial genomes, including newly sequenced genomes labeled in red. The tree was inferred from 46,754 unambiguously aligned sites under the best-fit model (LG + F + I + R10) and 1,000 ultrafast bootstrap replicates. Only nodes with <90% bootstrap supports are labeled. Annotations include: (a) Cyanobacterial order assignments following Strunecký et al. (2023); (b) CheckM completeness; (c) CheckM contamination; (d) GUNC CSS; (e) GUNC contamination; (f) Percentage ANI and best-match taxa.

this genome in the order Nostocales. *Paranostoc* belongs to the family Stigonemataceae, order Nostocales, class Cyanophyceae, and phylum Cyanobacteriota. The name has been registered under SeqCode: seqco.de/r:s5njm2s9.

**Description of *Aetokthonos rifticola* sp. nov.**

*Aetokthonos rifticola* (rif.ti'co.la (N. L. masc. n. *rifticola*), N. L. neut. n. *riftum*, “rift”; derived from English rift; L. n. *-cola*, an inhabitant or dweller; N. L. masc. n. *rifticola*, one that dwells in a rift). This MAG (S29-31\_bin.74<sup>TS</sup>) has a size of 7.6 Mb, completeness of 96.75%, and contamination of 1.81%. This MAG was assembled from sequencing reads generated from samples taken from an open area located within a publicly accessible steam vent features located within the East Rift Zone. The closest identified species is *Aetokthonos hydrillicola* (GCA\_017591595.2, ANI: 83.9%; AAI: 79.5%). In total, this MAG encodes 99 (out of 115) genes related to phycobilisome, photosynthesis, and respiratory machinery. This type genome has been deposited in ENA with the accession number ERZ28674877<sup>TS</sup>. Phylogenomic analysis placed this genome in the order Nostocales. *Aetokthonos rifticola* belongs to the family Scytonemataceae, order Nostocales, class Cyanophyceae, and phylum

Cyanobacteriota. The name has been registered under SeqCode: seqco.de/r:s5njm2s9.

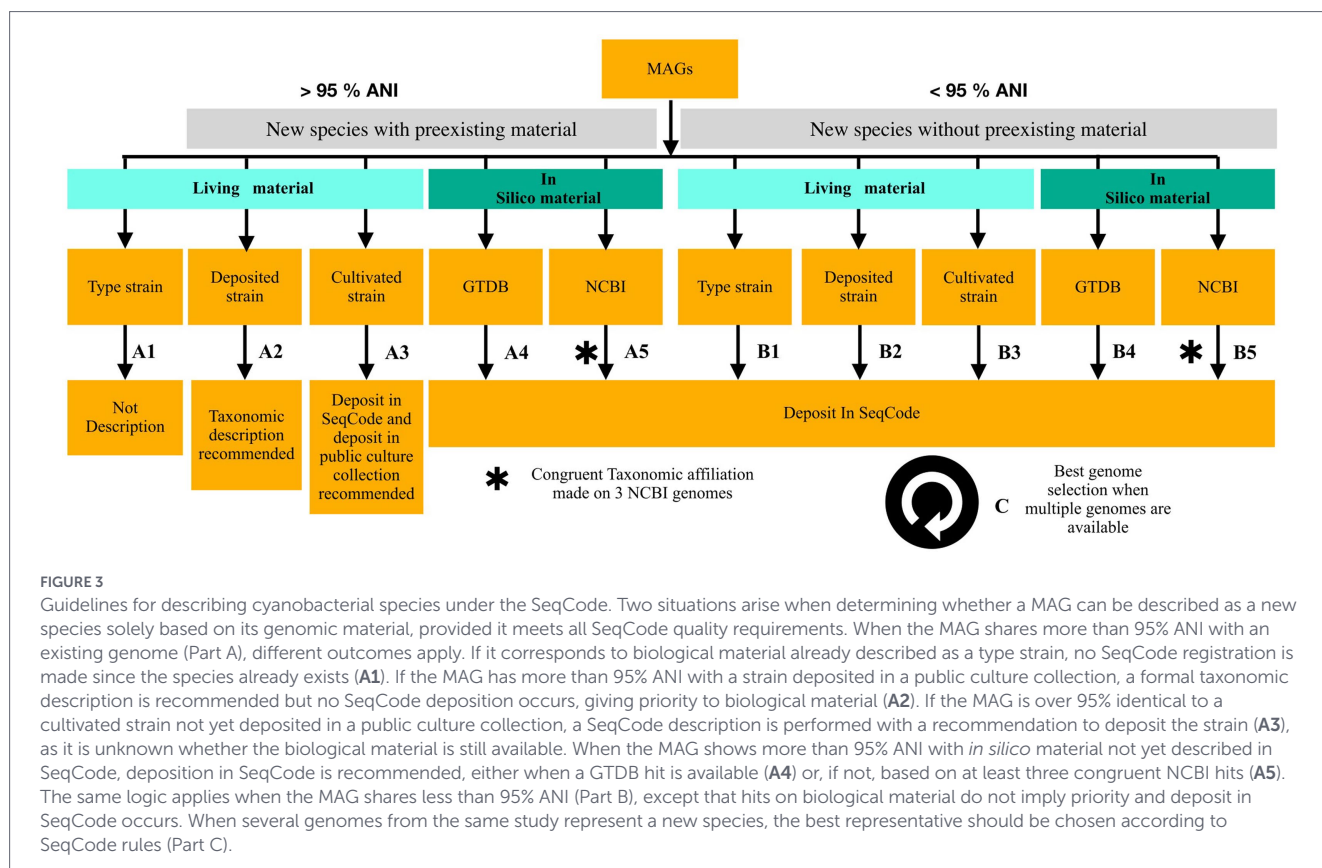
**Description of *Chroococcidiopsis rifticola* sp. nov.**

*Chroococcidiopsis rifticola* (rif.ti'co.la (N. L. masc. n. *rifticola*), N. L. neut. n. *riftum*, “rift”; derived from English rift; L. n. *-cola*, an inhabitant or dweller; N. L. masc. n. *rifticola*, one that dwells in a rift). This MAG (S29-31\_bin.60<sup>TS</sup>) was assembled from sequencing reads generated from samples taken from an open area within publicly accessible steam vent features located within the East Rift Zone. S29-31\_bin.60<sup>TS</sup> has a genome size of 6.4 Mb, completeness of 97.6%, and contamination of 3.2%. Both ANI and AAI identified it as a Chroococcidiopsidaceae cyanobacterium (GCA\_019244235.1, ANI: 94.3%; GCA\_019242465.1, AAI: 76.9%). In total, this MAG encodes 99 (out of 115) genes related to phycobilisome, photosynthesis, and respiratory machinery. This type genome has been deposited in ENA with the accession number ERZ28674876<sup>TS</sup>. Phylogenomic analysis placed this genome in the order Chroococcidiopsidales. *Chroococcidiopsis rifticola* belongs to the family Chroococcidiopsidaceae, order Chroococcidiopsidales, class Cyanophyceae, and phylum Cyanobacteriota. The name has been registered under SeqCode: seqco.de/r:s5njm2s9.

TABLE 2 Statistics and applied guidelines of selected cyanobacterial MAGs deposited in SeqCode.

MAG	Accession No.	Genome size (Mb)	No. of contigs	Completeness (%)	Contamination (%)	GTDB classification	ANI (%)	Proposed names	Guidelines applied*
W1-4_bin.70 <sup>TS</sup>	ERZ28674871 <sup>TS</sup>	6.30	227	98.82	0.47	s__Kovacikia sp034541955	99.6772	<i>Kovacikia ellensis</i>	A3: Description and recommend deposit
W1-4_bin.36 <sup>TS</sup>	ERZ28674878 <sup>TS</sup>	7.88	195	99.04	1.59	s__JAJPJF01 sp021324415	81.405	<i>Paranostoc speluncae</i>	B4: < 95% ANI
S29-31_ bin.74 <sup>TS</sup>	ERZ28674877 <sup>TS</sup>	7.66	587	96.75	1.81	s__Aetokthonos hydrillicola	85.3098	<i>Aetokthonos rifticola</i>	B4: < 95% ANI
S29-31_ bin.60 <sup>TS</sup>	ERZ28674876 <sup>TS</sup>	6.44	173	97.56	3.22	s__JAFVL01 sp019244235	94.3411	<i>Chroococciopsis rifticola</i>	B4: < 95% ANI
HW-12_ bin.40 <sup>TS</sup>	ERZ28674869 <sup>TS</sup>	5.72	367	96.89	2.23	s__Prochlorococcus_A sp001989455	86.9307	<i>Chroococciopsis edaphicus</i>	B4: < 95% ANI
S29-31_ bin.237 <sup>TS</sup>	ERZ28674875 <sup>TS</sup>	7.83	105	98.67	1.56	s__JANRIL01 sp033764955	79.1589	<i>Microcoleus pahoensis</i>	B4: < 95% ANI
S29-31_ bin.125 <sup>TS</sup>	ERZ28674874 <sup>TS</sup>	6.86	283	98.82	2.2	s__MUGG01 sp014697025	79.9028	<i>Elainella pahoensis</i>	B4: < 95% ANI
S13-15_ bin.75 <sup>TS</sup>	ERZ28674873 <sup>TS</sup>	6.81	326	98.92	0.44	s__Brasilonema sp019359415	89.9154	<i>Brasilonema rifticola</i>	B4: < 95% ANI
P23-25_ bin.64 <sup>TS</sup>	ERZ28674872 <sup>TS</sup>	7.54	213	97.41	0.71	s__Leptodesmis sp037442585	79.7857	<i>Leptodesmis pahoensis</i>	B4: < 95% ANI

\*The applied guidelines followed the classification in Figure 3.



## Description of *Chroococidiopsis edaphicus* sp. nov.

*Chroococidiopsis edaphicus* (ed.ə'phi.cus (N. L. masc. Adj. *edaphicus*), Gr. neut. n. ἔδαφος *edaphos*, “soil, ground”; N. L. adj. Suffix *-icus* (from Gr. -ικός, “pertaining to”); N. L. masc. Adj. *edaphicus*, pertaining to soil). This MAG (HW-12\_bin.40<sup>TS</sup>) was assembled from sequencing reads generated from soil samples taken from Big Ell, a cave located within Kilauea Caldera inside Hawaii Volcanoes National Park. HW-12\_bin.40<sup>TS</sup> has a genome size of 5.7 Mb, completeness of 96.9%, and contamination of 2.2%. The closest species identification was *Prochlorococcus* (GCA\_000635355.1, ANI: 86.9%) and *Chroococidiopsis* sp. CCME29 (GCA\_023558375.1, AAI: 65.9%). In total, this MAG encodes 98 (out of 115) genes related to phycobilisome, photosynthesis, and respiratory machinery. This type genome has been deposited in ENA with the accession number ERZ28674869<sup>TS</sup>. Phylogenomic analysis placed this genome in the order Chroococidiopsidales. *Chroococidiopsis edaphicus* belongs to the family Chroococidiopsidaceae, order Chroococidiopsidales, class Cyanophyceae, and phylum Cyanobacteriota. The name has been registered under SeqCode: seqco.de/r:s5njm2s9.

## Description of *Microcoleus pahoensis* sp. nov.

*Microcoleus pahoensis* (pa.ho.a.ensis (N. L. adj. *pahoensis*), from Pāhoa, a toponym in Hawai'i; L. suffix *-ensis*, pertaining to or originating from; N. L. adj. *Pahoensis*, originating from Pāhoa). This MAG (S29-31\_bin.237<sup>TS</sup>) has a size of 7.8 Mb, completeness of 98.67%, and contamination of 1.56%. This MAG was assembled from sequencing reads generated from samples taken from an open area within publicly accessible steam vent features located within the East Rift Zone. The closest identified species is a *Thermostichaceae* (GCA\_033764955.1, ANI: 79.1%) or an uncultured *Microcoleus* (GCA\_964657555.1, AAI:

71.1%). In total, this MAG encodes 100 (out of 115) genes related to phycobilisome, photosynthesis, and respiratory machinery. This type genome has been deposited in ENA with the accession number ERZ28674875<sup>TS</sup>. Phylogenomic analysis placed this genome in the order Oscillatoriales. *Microcoleus pahoensis* belongs to the family Microcoleaceae, order Oscillatoriales, class Cyanophyceae, and phylum Cyanobacteriota. The name has been registered under SeqCode: seqco.de/r:s5njm2s9.

## Description of *Elainella pahoensis* sp. nov.

*Elainella pahoensis* (pa.ho.a.ensis (N. L. adj. *pahoensis*), from Pāhoa, a toponym in Hawai'i; L. suffix *-ensis*, pertaining to or originating from; N. L. adj. *Pahoensis*, originating from Pāhoa). This MAG (S29-31\_bin.125<sup>TS</sup>) has a size of 6.8 Mb, completeness of 98.82%, and contamination of 2.2%. This MAG was assembled from sequencing reads generated from samples taken from an open area within publicly accessible steam vent features located within the East Rift Zone. The closest identified species is an *Alkalinema* (GCA\_014697025.1, ANI: 79.9%) or *Elainella* (GCA\_000733415.1, AAI: 74.3%). In total, this MAG encodes 97 (out of 115) genes related to phycobilisome, photosynthesis, and respiratory machinery. This type genome has been deposited in ENA with the accession number ERZ28674874<sup>TS</sup>. Phylogenomic analysis placed this genome in the order Oculatellales. *Elainella pahoensis* belongs to the family Oculatellaceae, order Oculatellales, class Cyanophyceae, and phylum Cyanobacteriota. The name has been registered under SeqCode: seqco.de/r:s5njm2s9.

## Description of *Brasilonema rifticola* sp. nov.

*Brasilonema rifticola* (rif.ti'co.la (N. L. masc. n. *rifticola*), N. L. neut. n. *riftum*, “rift”; derived from English rift; L. n. *-cola*, an

inhabitant or dweller; N. L. masc. n. *rifticola*, one that dwells in a rift). This MAG (S13-15\_bin.75<sup>TS</sup>) was assembled from sequencing reads generated from samples taken from an open area within publicly accessible steam vent features located within the East Rift Zone. S13-15\_bin.75<sup>TS</sup> has a genome size of 6.8 Mb, completeness of 98.9%, and contamination of 0.4%. Both ANI and AAI identified it as a *Brasilonema* (GCA\_019359415.1, ANI: 89.9%; AAI: 78.1%). In total, this MAG encodes 99 (out of 115) genes related to phycobilisome, photosynthesis, and respiratory machinery. This type genome has been deposited in ENA with the accession number ERZ28674873<sup>TS</sup>. Phylogenomic analysis placed this genome in the order Nostocales. *Brasilonema rifticola* belongs to the family Scytonemataceae, order Nostocales, class Cyanophyceae, and phylum Cyanobacteriota. The name has been registered under SeqCode: seqco.de/r:s5njm2s9.

### Description of *Leptodesmis pahoensis* sp. nov.

*Leptodesmis pahoensis* (pa.ho.a.en'sis (N. L. adj. *pahoensis*), from Pāhoa, a toponym in Hawai'i; L. suffix *-ensis*, pertaining to or originating from; N. L. adj. *Pahoensis*, originating from Pāhoa). This MAG (P23-25\_bin.64<sup>TS</sup>) has a size of 6.8 Mb, completeness of 98.82%, and contamination of 2.2%. This MAG was assembled from sequencing reads generated from samples taken from a publicly accessible pit-like steam vent feature located within the East Rift Zone. The closest identified species is a *Leptodesmis* (GCA\_037442585.1, ANI: 79.8%; GCA\_021379005.1, AAI: 77.2%). In total, this MAG encodes 98 (out of 115) genes related to phycobilisome, photosynthesis, and respiratory machinery. This type genome has been deposited in ENA with the accession number ERZ28674872<sup>TS</sup>. Phylogenomic analysis placed this genome in the order Leptolyngbyales. *Leptodesmis pahoensis* belongs to the family Leptolyngbyaceae, order Leptolyngbyales, class Cyanophyceae, and phylum Cyanobacteriota. The name has been registered under SeqCode: seqco.de/r:s5njm2s9.

## Discussion

### Hawaiian steam vents contain diverse, novel, and largely uncultured cyanobacteria

Bacterial community composition in steam vents can vary markedly even within closely located areas, largely due to steep gradients in temperature and moisture. Soils adjacent to active vents are typically dominated by photoautotrophic taxa (notably cyanobacteria), whereas peripheral, drier soils support drought-resistant heterotrophs, such as actinomycetes (Costello et al., 2009; Hadland et al., 2024). Previous studies report substantial diversity in the microbial assemblages of steam vent systems worldwide. In particular, the steam vents of Hawai'i (Wall et al., 2015; Prescott et al., 2022), Paricutín and Sapichu in Mexico (Brito et al., 2019), Socompa in the Andes (Costello et al., 2009), Surtsey in Iceland (Bergsten et al., 2021), and Mutnovsky and Gorely in Russia (Allaguvatova et al., 2022), all show high cyanobacterial diversity. In contrast, cyanobacteria appear to be rare in the steam vents of Los Azufres, Mexico (Marín-Paredes et al., 2023) and Sierra Negra in the Galapagos Islands (Mayhew et al., 2007), with <1% of sequencing reads assigned to cyanobacterial sequences in Los Azufres and only a single species isolated from the Sierra Negra.

Amplicon sequencing performed in this study largely agrees with earlier studies (Wall et al., 2015; Prescott et al., 2022) that highlight the abundance of cyanobacteria in the steam vents of Hawai'i (Figure 1a). Both amplicon sequencing and MAG assembly support that *Gloeobacter kilaueensis* is the dominant cyanobacterium in the pit-like environment with low-light conditions (Figures 1b, 2). Leptolyngbyaceae dominate the soil and wall environments, while Hapalosiphonaceae, Oculatellaceae, Scytonemataceae, and Chroococciopsidaceae constitute the majority of the cyanobacterial families in the open area subjected to high-light conditions (Figure 1c). The diversity documented here mirrors previous findings from Hawai'i's steam vents, including members of Gloeobacterales, Leptolyngbyales, and Nostocales (Wall et al., 2015; Prescott et al., 2022). Notably, Leptolyngbyales and Nostocales were also observed in the steam vents of Antarctica (Soo et al., 2009; Noell et al., 2025) and in Mutnovsky and Gorely, Russia (Allaguvatova et al., 2022), underscoring their wide distribution and resilience in high-temperature geothermal environments. The dominance of Cyanobacteria and Chloroflexota in our study suggests cooperative interactions that promote the formation of robust, light-responsive biofilms in hot spring mats (Bunbury et al., 2025), enabling colonization of new niches, such as the steam vents of Hawai'i. We note that our data were collected during a single expedition and do not account for seasonal or temporal shifts in the microbial community. Despite this limitation, our findings highlight the remarkable diversity of previously undescribed cyanobacteria inhabiting Hawai'i's steam vents and emphasize the importance of formally characterizing this hidden diversity. The SeqCode nomenclature system greatly accelerates this process by enabling valid species descriptions in the absence of cultivation.

Here, we describe eight novel cyanobacterial species and one novel cyanobacterial genus representing five orders—Chroococciopsidales, Leptolyngbyales, Nostocales, Oculatellales, and Oscillatoriales—based on high-quality genome sequences and following the SeqCode guidelines (Table 2). Except for *Kovackia ellensis* (W1-4\_bin.70<sup>TS</sup>) named here, none of the new species described here had previously been sequenced, despite the availability of ~8,000 cyanobacterial genomes in NCBI. This pattern is consistent with the hypothesis proposed by Wall et al. (2015) that steam vents act as a reservoir of yet-to-be-discovered cyanobacterial diversity.

### Recommended guideline for the description of new cyanobacterial species following SeqCode

The description of new cyanobacterial species remains challenging, largely due to the inherent complexity of cyanobacterial taxonomy (Kaštovský, 2024). Reliance on morphological characters alone is insufficient, as cyanobacteria exhibit pronounced phenotypic plasticity and morphological crypsis (Casamatta et al., 2003; Dvořák et al., 2025). Morphologically similar taxa are often revealed to be polyphyletic, as exemplified by *Synechococcus*-like cyanobacteria (Komárek et al., 2020). These limitations underscore the need to incorporate genome-based approaches into cyanobacterial taxonomy, particularly given the rapid increase in species descriptions—over 150 new cyanobacterial names were proposed in 2021 alone (Kaštovský, 2024).

Although Dvořák et al. (2025) provide clear guidelines for cyanobacterial species description, their exclusion of SeqCode is a notable limitation. While specimen deposition remains valuable for future studies, it is not universally feasible. Many cyanobacteria are difficult

to culture and maintain, and thus cannot be accommodated under traditional nomenclatural frameworks. Further, large-scale genomic analyses reveal that culture collections are strongly biased toward fast-growing organisms, leaving many slower-growing organisms (including cyanobacteria) underrepresented (Weissman et al., 2021). SeqCode directly addresses this gap by allowing genome sequences to serve as type material (Hedlund et al., 2022), making it essential for capturing the full extent of cyanobacterial diversity and ensuring that taxonomy keeps pace with genomic discovery.

The criteria currently used to determine whether a description of a new species applies, based on genomic data for prokaryotes—including cyanobacteria—is the 95% ANI threshold, which replaces the former 70% DNA–DNA hybridization standard (Jain et al., 2018). The same criterion is applied when determining whether a MAG represents a new species (Hedlund et al., 2022). Describing new species based on MAGs is extremely useful; however, the availability of biological material must always take priority over purely genome-based descriptions, as it allows for more applications. Therefore, in this first description of cyanobacteria under SeqCode, we propose specific guidelines that introduce three categories of biological material to refine the 95% ANI rule (Figure 3).

If a MAG shares more than 95% ANI with biological material, this does not necessarily exclude a SeqCode description, as it depends on the nature of that material (Figure 3, A part). In the case of a MAG matching a type strain, the rule is straightforward (Figure 3, A1). However, two other categories can be distinguished. A MAG may share more than 95% ANI with the closest genome corresponding to a strain that has not yet been taxonomically described but is available in a public culture collection. In this case, a taxonomic description is recommended, but priority is given to the biological material (Figure 3, A2). A third case may occur when a MAG shares more than 95% ANI with cultivated biological material that is not deposited in a public culture collection. Since the availability of this material is uncertain, a SeqCode description can be proposed, but it should include a recommendation to deposit the strain in a public culture collection (Figure 3, A3), thereby linking the new genome-based taxonomy to physical biological material. This is the case for MAG W1-4\_bin.70<sup>TS</sup>, described here as *Kovacikia ellensis*, which corresponds to a strain that has already been isolated but has not been deposited in a public culture collection.

When a MAG does not correspond to any biological material but instead matches *in silico* data at over 95% ANI, and those data are not yet described in SeqCode, a genome-based species description applies (Figure 3, A4,A5). A special case arises when a MAG lacks any GTDB match but has hits only in the NCBI database. Because the GTDB taxonomy is unavailable in this scenario, classification relies solely on the less curated NCBI taxonomy (Parks et al., 2022). We recommend ensuring congruence by considering the last common ancestor (LCA) of the three closest NCBI hits. An example is provided for three of our samples (HW-12\_bin.63, S2-4\_bin.10, and S29-31\_bin.283), where the taxonomy of the three top NCBI hits differed too greatly to allow a SeqCode description (Supplementary Table S3). When a MAG shares less than 95% ANI, it falls under the conventional framework commonly used for describing prokaryotic species (Figure 3).

In addition to species descriptions based on ANI, an AAI threshold of 65% should be used to determine the presence of a novel genus, which was not the case in this study. However, as illustrated by MAG W1-4\_bin.36<sup>TS</sup>, taxonomy of closely related ANI matches does not always allow for a clear genus assignment, since the closest genomes are not classified

with sufficient precision, here at the Nostocales order level. In such cases, phylogenetic placement provides the necessary resolution, as exemplified here by the establishment of the new genus *Paranostoc speluncaea*.

## Data availability statement

All sequencing reads generated in this study are available in the European Nucleotide Archive (ENA) under study accession number PRJEB52128 (run accession numbers: ERS11952021–ERS11952063 and ERS27818420–ERS27818420). The alignment file and list of cyanobacterial and melainabacterial used for fast ANI and fast AAI analyses are deposited in the FigShare repository (<https://doi.org/10.6084/m9.figshare.31123651>).

## Author contributions

ES: Writing – review & editing, Investigation, Formal analysis, Visualization, Data curation, Writing – original draft, Validation. MS: Formal analysis, Writing – review & editing, Investigation. BD: Writing – review & editing, Formal analysis, Writing – original draft, Software, Visualization, Investigation. MH: Writing – review & editing, Formal analysis, Writing – original draft, Investigation, Software. LC: Supervision, Project administration, Writing – review & editing, Writing – original draft, Conceptualization, Visualization, Funding acquisition. JS: Resources, Funding acquisition, Writing – review & editing, Project administration, Formal analysis, Investigation, Conceptualization, Methodology.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by a research grant (PDR T.0018.24 OR-OX-PHOT-IN-CYN) from the Belgian National Fund for Scientific Research (F. R. S.-FNRS). LC is supported by a mandate from the Belgian National Fund for Scientific Research (F. R. S.-FNRS). JS is supported by startup funds from The George Washington University and a US National Science Foundation (NSF) grant (award number 2442122).

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Allaguvatova, R. Z., Nikulin, A. Y., Nikulin, V. Y., Bagmet, V. B., and Gaysina, L. A. (2022). Study of biodiversity of algae and cyanobacteria of Mutnovsky and Gorely volcanoes soils (Kamchatka peninsula) using a polyphasic approach. *Diversity* 14:375. doi: 10.3390/d14050375
- Bergsten, P., Vannier, P., Klonowski, A. M., Knobloch, S., Gudmundsson, M. T., Jackson, M. D., et al. (2021). Basalt-hosted microbial communities in the subsurface of the young volcanic island of Surtsey, Iceland. *Front. Microbiol.* 12:728977. doi: 10.3389/fmicb.2021.728977
- Brito, E. M. S., Romero-Núñez, V. M., Caretta, C. A., Bertin, P., Valerdi-Negreros, J. C., Guyoneaud, R., et al. (2019). The bacterial diversity on steam vents from Paricutin and Sapichu volcanoes. *Extremophiles* 23, 249–263. doi: 10.1007/s00792-019-01078-8
- Bunbury, F., Rivas, C., Calatrava, V., Malkovskiy, A. V., Joubert, L.-M., Parvate, A. D., et al. (2025). Cyanobacteria and Chloroflexota cooperate to structure light-responsive biofilms. *Proc. Natl. Acad. Sci.* 122:e2423574122. doi: 10.1073/pnas.2423574122
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Casamatta, D. A., Vis, M. L., and Sheath, R. G. (2003). Cryptic species in cyanobacterial systematics: a case study of *Phormidium retzii* (Oscillatoriales) using RAPD molecular markers and 16S rDNA sequence data. *Aquat. Bot.* 77, 295–309. doi: 10.1016/j.aquabot.2003.08.005
- Castenholz, R. W. (1981). "Isolation and cultivation of thermophilic cyanobacteria," in *The Prokaryotes: A Handbook on Habitats, Isolation, and Identification of Bacteria*, eds. M. P. Starr, H. Stolp, H. G. Trüper, A. Balows and H. G. Schlegel (Berlin: Springer), 236–246.
- Cornet, L., Durieu, B., Baert, F., D'hooge, E., Colignon, D., Meunier, L., et al. (2023). The GEN-ERA toolbox: unified and reproducible workflows for research in microbial genomics. *GigaScience* 12:giad022. doi: 10.1093/gigascience/giad022
- Cornet, L., Wilmette, A., Javaux, E. J., and Baurain, D. (2018). A constrained SSU-rRNA phylogeny reveals the unsequenced diversity of photosynthetic cyanobacteria (Oxyphotobacteria). *BMC Res. Notes* 11:435. doi: 10.1186/s13104-018-3543-y
- Costello, E. K., Halloy, S. R. P., Reed, S. C., Sowell, P., and Schmidt, S. K. (2009). Fumarole-supported islands of biodiversity within a hyperarid, high-elevation landscape on Socompa volcano, Puna de Atacama Andes. *Appl. Environ. Microbiol.* 75, 735–747. doi: 10.1128/AEM.01469-08
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008. doi: 10.1093/gigascience/giab008
- Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 14, 927–930. doi: 10.1111/j.1654-1103.2003.tb02228.x
- Dvořák, P., Skoupy, S., Stanojković, A., Johansen, J. R., Villanueva, C., Jung, P., et al. (2025). A hitchhiker's guide to modern, practical cyanobacterial taxonomy. *J. Phycol.* 61, 1536–1552. doi: 10.1111/jpy.70102
- Edgar, R. C. (2022). Muscles: high-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* 13:6968. doi: 10.1038/s41467-022-34630-w

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2026.1824103/full#supplementary-material>

### SUPPLEMENTARY TABLE S1.

Relative abundance (%) of the taxa, at different taxonomic levels (Phylum, Family and Genus) in the four sampling sites.

### SUPPLEMENTARY TABLE S2.

Statistics and taxonomic classifications of 40 high-quality cyanobacterial and melainabacterial MAG assemblies.

### SUPPLEMENTARY TABLE S3.

The number of detected gene sequences in the nine newly characterized MAGs, categorized by functions related to the phycobilisome, photosynthesis, and respiratory machinery.

Eitel, E. M., Utter, D. R., Connon, S. A., Orphan, V. J., and Murali, R. (2025). CABO-16S-a combined Archaea, Bacteria, organelle 16S rRNA database framework for amplicon analysis of prokaryotes and eukaryotes in environmental samples. *NAR Genomics Bioinform.* 7:lqaf061. doi: 10.1093/nargab/lqaf061

Gerhardt, K., Ruiz-Perez, C. A., Rodriguez-R, L. M., Jain, C., Tiedje, J. M., Cole, J. R., et al. (2025). FastAAI: efficient estimation of genome average amino acid identity and phylum-level relationships using tetramers of universal proteins. *Nucl. Acids Res.* 53:gkaf348. doi: 10.1093/nar/gkaf348

Hadland, N., Hamilton, C. W., and Duhamel, S. (2024). Young volcanic terrains are windows into early microbial colonization. *Commun. Earth Environ.* 5:114. doi: 10.1038/s43247-024-01280-3

Harmel, M., Durieu, B., Savaglia, V., Tytgat, B., Baurain, D., Wilmette, A., et al. (2025). A FAIR amplicon sequencing workflow for long-term environmental monitoring. *bioRxiv*. [Preprint]. doi: 10.64898/2025.12.04.692289

Hedlund, B. P., Chuvochina, M., Hugenholtz, P., Konstantinidis, K. T., Murray, A. E., Palmer, M., et al. (2022). SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat. Microbiol.* 7, 1702–1708. doi: 10.1038/s41564-022-01214-9

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y., and Ishiguro-Watanabe, M. (2025). KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* 53, D672–D677. doi: 10.1093/nar/gkae909

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. doi: 10.7717/peerj.7359

Kans, J. (2025). "Entrez Direct: E-utilities on the Unix Command Line," in Entrez Programming Utilities Help [Internet], (National Center for Biotechnology Information (US)). Available online at: <https://www.ncbi.nlm.nih.gov/books/NBK179288/> (Accessed December 16, 2025).

Kaštovský, J. (2024). Welcome to the jungle!: an overview of modern taxonomy of cyanobacteria. *Hydrobiologia* 851, 1063–1077. doi: 10.1007/s10750-023-05356-7

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

Lee, M. D. (2019). GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35, 4162–4164. doi: 10.1093/bioinformatics/btz188

Lefler, F. W., Berthold, D. E., and Laughinghouse, H. D. IV (2023). Cyanoseq: a database of cyanobacterial 16S rRNA gene sequences with curated taxonomy. *J. Phycol.* 59, 470–480. doi: 10.1111/jpy.13335

Letunic, I., and Bork, P. (2024). Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucl. Acids Res.* 52, W78–W82. doi: 10.1093/nar/gkae268

Lifschitz, S., Haeusler, E. H., Catanho, M., de Miranda, A. B., Molina de Armas, E., Heine, A., et al. (2022). Bio-strings: a relational database data-type for dealing with large biosequences. *Biotech* 11:31. doi: 10.3390/biotech11030031

- Lumian, J. E., Jungblut, A. D., Dillion, M. L., Hawes, I., Doran, P. T., Mackey, T. J., et al. (2021). Metabolic capacity of the Antarctic cyanobacterium *Phormidium pseudopriestleyi* that sustains oxygenic photosynthesis in the presence of hydrogen sulfide. *Genes* 12:426. doi: 10.3390/genes12030426
- Marín-Paredes, R., Bolívar-Torres, H. H., Coronel-Gaytán, A., Martínez-Romero, E., and Servín-Garcidueñas, L. E. (2023). A metagenome from a steam vent in Los Azufres geothermal field shows an abundance of *Thermoplasmatales archaea* and Bacteria from the Phyla Actinomycetota and Pseudomonadota. *Curr. Issues Mol. Biol.* 45, 5849–5864. doi: 10.3390/cimb45070370
- Mayhew, L. E., Geist, D. J., Childers, S. E., and Pierson, J. D. (2007). Microbial community comparisons as a function of the physical and geochemical conditions of Galápagos Island fumaroles. *Geomicrobiol. J.* 24, 615–625. doi: 10.1080/01490450701672133
- McMurdie, P. J., and Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Noell, S. E., Abbaszadeh, J., Richards, H., Labat Saint Vincent, M., Lee, C. K., Herbold, C. W., et al. (2025). Antarctic geothermal soils exhibit an absence of regional habitat generalist microorganisms. *Environ. Microbiol.* 27:e70032. doi: 10.1111/1462-2920.70032
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116
- O’Leary, N. A., Cox, E., Holmes, J. B., Anderson, W. R., Falk, R., Hem, V., et al. (2024). Exploring and retrieving sequence and metadata for species across the tree of life with NCBI datasets. *Sci. Data* 11:732. doi: 10.1038/s41597-024-03571-y
- Obbels, D., Verleyen, E., Mano, M.-J., Namsaraev, Z., Sweetlove, M., and Tytgat, B. (2016). Bacterial and eukaryotic biodiversity patterns in terrestrial and aquatic habitats in the Sor Rondane Mountains, Dronning Maud land, East Antarctica. *FEMS Microbiol. Ecol.* 92:fiw041. doi: 10.1093/femsec/fiw041
- Orakov, A., Fullam, A., Coelho, L. P., Khedkar, S., Szklarczyk, D., Mende, D. R., et al. (2021). GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biol.* 22:178. doi: 10.1186/s13059-021-02393-0
- Oren, A. (2020). Three alternative proposals to emend the Rules of the International Code of Nomenclature of Prokaryotes to resolve the status of the Cyanobacteria in the prokaryotic nomenclature. *International Journal of Systematic and Evolutionary Microbiology* 70, 4406–4408. doi: 10.1099/ijsem.0.004268
- Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., and Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50, D785–D794. doi: 10.1093/nar/gkab776
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Pinevich, A. V., and Averina, S. G. (2024). Taxonomy of cyanobacteria: the era of change. *Microbiology* 93, 521–536. doi: 10.1134/S0026261724605724
- Prescott, R. D., Chan, Y. L., Tong, E. J., Bunn, E., Onouye, C. T., Handel, C., et al. (2023). Bridging place-based astrobiology education with genomics, including descriptions of three novel bacterial species Isolated from Mars analog sites of cultural relevance. *Astrobiology* 23, 1348–1367. doi: 10.1089/ast.2023.0072
- Prescott, R. D., Zamkovaya, T., Donachie, S. P., Northup, D. E., Medley, J. J., Monsalve, N., et al. (2022). Islands within islands: bacterial phylogenetic structure and consortia in Hawaiian lava caves and fumaroles. *Front. Microbiol.* 13:934708. doi: 10.3389/fmicb.2022.934708
- Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., and Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Microbiology* 111, 1–61. doi: 10.1099/00221287-111-1-1
- Rippka, R., Waterbury, J. B., and Stanier, R. Y. (1981). “Isolation and purification of cyanobacteria: some general principles,” in *The Prokaryotes: A Handbook on Habitats, Isolation, and Identification of Bacteria*, eds. M. P. Starr, H. Stolp, H. G. Trüper, A. Balows and H. G. Schlegel (Berlin: Springer), 212–220.
- Saw, J. H. W., Schatz, M., Brown, M. V., Kunkel, D. D., Foster, J. S., Shick, H., et al. (2013). Cultivation and complete genome sequencing of *Gloeobacter kilauensis* sp. nov., from a lava cave in Kilauea caldera, Hawai‘i. *PLoS One* 8:e76376. doi: 10.1371/journal.pone.0076376
- Saw, J. H., Shlafstein, M. D., Pavlouidi, C., Monsalve, N., Prescott, R. D., Chain, P. S. G., et al. (2026). Amplicon and metagenomic data from fumarole-associated geothermal features of Hawai‘i. *bioRxiv*. doi: 10.64898/2026.02.26.708227
- Soo, R. M., Wood, S. A., Grzymalski, J. J., McDonald, I. R., and Cary, S. C. (2009). Microbial biodiversity of thermophilic communities in hot mineral soils of tramway ridge, mount Erebus, Antarctica. *Environ. Microbiol.* 11, 715–728. doi: 10.1111/j.1462-2920.2009.01859.x
- Stal, L. J. (1995). Physiological ecology of cyanobacteria in microbial mats and other communities. *New Phytol.* 131, 1–32. doi: 10.1111/j.1469-8137.1995.tb03051.x
- Strunecký, O., Ivanova, A. P., and Mareš, J. (2023). An updated classification of cyanobacterial orders and families based on phylogenomic and polyphasic analysis. *Journal of Phycology* 59, 12–51. doi: 10.1111/jpy.13304
- Sudianto, E., Baurain, D., and Cornet, L. (2025). Horizontal gene transfers underpin ribose heterotrophy and central carbon metabolism remodeling in gloeobacteraceae. *bioRxiv*. doi: 10.1101/2025.11.10.686926
- Taufiq, A., Singh, S., and Singh, T. (2026). “An overview of microbes present in the extreme environment,” in *Polar Microbes and Climate Change: A Molecular Understanding for Sustainable Future*, eds. A. W. Fernandes Duarte, M. R. Z. Passarini, J. Elster, P. K. Singh and A. Kumar (Cham: Springer), 331–350. doi: 10.1007/978-3-032-07096-8\_12
- Wall, K., Cornell, J., Bizzoco, R. W., and Kelley, S. T. (2015). Biodiversity hot spot on a hot spot: novel extremophile diversity in Hawaiian fumaroles. *MicrobiologyOpen* 4, 267–281. doi: 10.1002/mbo3.236
- Wang, Q., and Cole, J. R. (2024). Updated RDP taxonomy and RDP classifier for more accurate taxonomic classification. *Microbiol. Resour. Announc.* 13, e0106323–e0106323. doi: 10.1128/mra.01063-23
- Waterbury, J. B. (2006). “The cyanobacteria— isolation, purification and identification,” in *The Prokaryotes*, (New York: Springer), 1053–1073.
- Weissman, J. L., Hou, S., and Fuhrman, J. A. (2021). Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc. Natl. Acad. Sci.* 118:e2016810118. doi: 10.1073/pnas.2016810118
- Whitman, W. B., Chuvochina, M., Hedlund, B. P., Konstantinidis, K. T., Palmer, M., Rodriguez-R, L. M., et al. (2024). Why and how to use the SeqCode. *mLife* 3, 1–13. doi: 10.1002/mlf2.12092
- Whitton, B. A., and Potts, M. (2012). “Introduction to the cyanobacteria,” in *Ecology of Cyanobacteria II: Their Diversity in Space and Time*, ed. B. A. Whitton (Dordrecht: Springer), 1–13.
- Wickham, H. (2016). *ggplot2*. Cham: Springer International Publishing doi: 10.1007/978-3-319-24277-4.
- Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., Software, P., et al. (2023). *dplyr: A Grammar of Data Manipulation*. Available at: <https://cran.r-project.org/web/packages/dplyr/index.html> (Accessed December 16, 2025).