Environmental Microbiome

https://doi.org/10.1186/s40793-026-00931-6

Article in Press

Metagenomics identifies distinct functional and taxonomic profiles between freshwaters and associated sediments in alpine habitats

Received: 4 June 2025

Accepted: 1 July 2026

Published online: 22 July 2026

Cite this article as: Gattei A., Vettorazzo S., Pindo M. *et al.* Metagenomics identifies distinct functional and taxonomic profiles between freshwaters and associated sediments in alpine habitats. *Environmental Microbiome* (2026). https://doi.org/10.1186/s40793-026-00931-6

Anita Gattei, Sara Vettorazzo, Massimo Pindo, Nico Salmaso, Paolo Manghi, Stefano Larsen & Claudio Donati

Metagenomics identifies distinct functional and taxonomic profiles between freshwaters and associated sediments in alpine habitats

Anita Gattei[1, 3*], Sara Vettorazzo[2, 3], Massimo Pindo[1], Nico Salmaso[2, 3], Paolo Manghi[1], Stefano Larsen[1, 3], Claudio Donati[1, 3*]

[1:] Computational Biology Unit, Research and Innovation Centre, Fondazione Edmund Mach, via Mach 1, San Michele all'Adige 38098, Italy

[2:] Hydrobiology Unit, Research and Innovation Centre, Fondazione Edmund Mach, via Mach 1, San Michele all'Adige 38098, Italy

[3:] National Biodiversity Future Center (NBFC), Piazza Marina 61, 90133 Palermo, Italy

* (corresponding author)

Abstract

Background: Alpine freshwater habitats are biodiversity hotspots, home to uniquely adapted microbial communities. However, our understanding of how taxonomic composition and metabolic potential of microbial communities varies across habitat type and physiography remains limited.

Results: Using metagenomic sequencing, we examined microbial diversity and functions in sediment and water habitats across a physiographic gradient, including deep and shallow lakes, pasture ponds, ponds and peat bogs. Overall, sediments supported more diverse and even microbial communities. Conversely, waters displayed higher heterogeneity across the physiographic gradient in both taxonomic and functional composition. Notably, water communities from small water bodies, especially peat bogs and ponds, showed comparable diversity to their respective sediment communities, with higher taxonomic richness than aquatic communities from larger water bodies. Sediment microbial communities showed higher functional potential for nitrogen and methane metabolism, a pattern related to the likely anoxic conditions. By contrast, water bodies favored higher potential for carbohydrate

metabolism, carbon oxidation and photosynthesis. Lastly, we recovered 496 bacterial and archaeal metagenome-assembled genomes and found that 73% could not be assigned at the species level.

Conclusions: Our findings shed new light on the microbial diversity and metabolic potential of alpine freshwater ecosystems, highlighting the vast and unknown diversity harbored by these fragile alpine ecosystems.

Keywords

Freshwater microbial ecology

Metagenomics

Biodiversity monitoring

Metagenome-assembled genomes

Functional analysis

1. Background

Alpine freshwater bodies and wetlands drive essential ecological processes and services, including key biogeochemical cycles, water filtration, organic matter decomposition, carbon storage, and primary production [1-11]. Despite their significance for ecosystem multifunctionality [1], they are among the most fragile environments, threatened by multiple pressures resulting from human activity [2], including climate change and biodiversity loss. Alpine freshwater habitats, such as deep and shallow lakes, pasture ponds, ponds and peat bogs provide fundamental ecosystem services and are biodiversity hotspots [3], hosting species adapted to high altitudes [4]. As per other natural environments, alpine water bodies and wetlands are home to unique communities of microorganisms, whose ecological role is deeply intertwined with biogeochemical cycles [5].

Water and sediment habitats differ significantly in microbial composition due to variations in nutrient availability and physical and chemical conditions [6]. Sediment habitats, rich in organic matter and nutrients, support greater microbial diversity than aquatic habitats [7, 8], fostering processes like denitrification and methanogenesis due to lower oxygen levels [9]. In contrast, water habitats, directly influenced by temperature and light, exhibit higher oxygen levels that affect microbial community structure and metabolism [6]. Physiographic features of the water bodies further shape microbial communities, as shallow waters allow light penetration to the bottom, promoting autotrophic benthic communities [10]. Despite their ecological significance, alpine lakes, small water bodies, and peat bogs remain underexplored, particularly concerning their microbial diversity and role in ecosystem functioning [3, 11]. This knowledge gap limits our understanding of microbial community responses to environmental changes, hindering long-term monitoring and restoration efforts [3, 5, 12].

Here, we collected 19 water and 19 sediment samples across a broad physiographic gradient within the ‘Ledro Alps and Judicaria’ Biosphere Reserve (Italy), including deep-water alpine lakes, shallow-water lakes, pasture ponds, ponds and peat bogs (**Fig. 1**). These habitats

provide an opportunity to examine microbial diversity in a biodiverse yet anthropized Alpine context. Using high-depth shotgun metagenomics, we: (i) characterized the different components of microbial communities, including bacteria, archaea, fungi, microalgae and protists, understanding the factors that shape their diversity at the species-level; (ii) built a catalogue of metagenome-assembled genomes (MAGs) to characterize rare and novel bacterial and archaeal species; (iii) expanded the current knowledge over the metabolic potential of alpine environments' microbiome, identifying patterns related to nutrient cycling and photosynthesis. By integrating taxonomic, functional and genomic data across microbial domains, this study expands previous metabarcoding research [8], offering a more comprehensive understanding of microbial communities in alpine waters.

2. Methods

2.1 Study site and sample collection

We considered 19 locations within the 'Ledro Alps and Judicaria' MAB UNESCO Biosphere Reserve in Trentino, Italy, which extends for a total of about 47,000 ha [13]. Up to 34% of the habitats in the Reserve are protected due to their rarity, fragility, and high level of biodiversity. Sampling protocols were described previously [8]. Briefly, a total of 19 paired water and surface sediment samples were collected along a physiographic gradient (**Fig. 1**). The studied freshwater habitats included two large deep lakes (samples LL, LT), three shallow lakes (samples LA, LN, LV), three lake ponds (samples MST, TF1, PD), three pasture ponds (samples MT, BG, MS), four peat bogs (samples TF2, TF3, TL, L1) and four ponds (samples MC, MN, MM, MB). We divided the samples into two categories based on their size and depth. The category 'Lakes and shallow lakes' (L-SLs) included deep lakes (max depth 48 m), shallow lakes (4 m) and lake ponds (~0.5 m), all characterized by low connectivity between sediment and the water column. Conversely, the category 'Small water bodies and peat bogs' (SWB-PBs) included pasture ponds, ponds, and peat bogs that were consistently shallower than 1 m and featured high connectivity between water and sediment substrates. Within SWB-

PBs, we distinguished between ponds and pasture ponds, as ponds tended to be naturally emerging pools of water, which were sometimes ephemeral, whereas pasture ponds are continuously maintained by farmers as a waterhole for animals. Since the stability of the body of water is a substantial factor influencing the microbiome, we employed two separate categories. For the sampling of the water column, different methods were used depending on the depth. In particular, in lakes and shallow lakes samples were collected from the centre of the lake (always within the euphotic zone) in sterilized bottles using Hydrobios integrating water samplers, Niskin bottles, or sterilized buckets. In small water bodies, samples were collected using telescopic samplers from the shores. Surface sediments were collected in the littoral zone using sterilized Pasteur pipettes. Sediment samples were lyophilized and stored at -20°C prior to DNA extraction. Water samples were filtered within 24 hours of collection using sterile ISOPORE filters with a 0.22 µm pore size under a laminar flow hood, after which the filters were stored dry at −20 °C until DNA extraction.

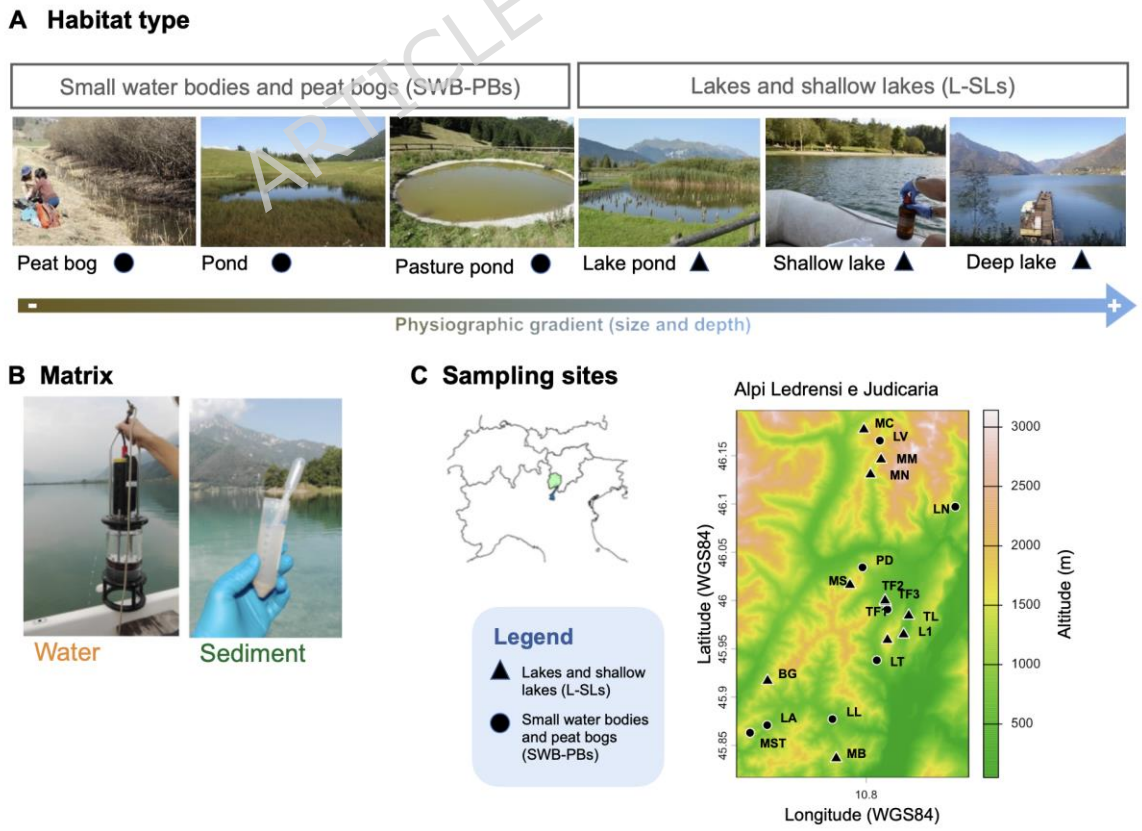


Figure 1. Study design and geographic map of the sampled sites in the Alpi Ledrensi and Judicaria UNESCO Biosphere Reserve. A) The sampled sites were classified into two categories based on their physiographic characteristics (water body size and depth). The category 'Lakes and shallow lakes' includes deep lakes (max depth 48 m), shallow lakes (4 m), and lake ponds (~0.5 m). The category 'Small water bodies and peat bogs' includes pasture ponds, ponds, and peat bogs with varying depths, but always shallower than 1 m. **B)** For each habitat, paired sampling of water and sediment was carried out. **C)** The sampled sites belong to the UNESCO MAB Reserve 'Alpi Ledrensi e Judicaria', located in the North-Eastern Alps, in the Trentino region, Italy. The sampled sites range over an altitudinal gradient from 600m to 1800m above sea level.

2.2 DNA extraction, library preparation and full shotgun sequencing

Environmental DNA (eDNA) from water and sediment samples was extracted using the DNeasy PowerWater® and DNeasy PowerSoil® kits (MO BIO Laboratories, a Qiagen Company, USA), respectively. DNA concentration and purity were assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific Inc., MA, USA). All samples yielded detectable DNA concentrations exceeding 2 ng μL^{-1}, with a mean ± SD of 28 ± 17 ng μL^{-1}. Starting from a total amount of 100 ng, total DNA was fragmented by enzymatic reaction at 37°C x 5 min producing fragments of 500bp. The paired-end library was prepared using the KAPA HyperPlus kit (Roche). KAPA Unique Dual-Indexed Adapter Kit (Roche) recommended for use with the KAPA HyperPlus Kit were ligated to the DNA fragments following the manufacturer's instructions. Libraries were quantified using the KAPA Library Quantification Kits (Roche) and were sequenced for 150 bp paired-end reads on the Illumina Novaseq 6000 platform (Illumina Inc., San Diego, CA, USA). A total of 4,026,914,640 paired-end reads were sequenced (average per sample: 105,971,437.9).

2.3 Metagenomes taxonomic profiling

Multiple bioinformatics analyses were conducted using the Metashot metagenomic suite, an integrated library of pipelines designed for shotgun metagenomics and bacterial genomics [14] and the nf-core suite [15] (**Supplementary Fig. 1**). Using the workflows metashot/kraken2 [16] and nf-core/taxprofiler [17], Illumina reads were first quality-filtered using BBDuk [18], which removed adapter sequences and bases with a quality < 10. Taxonomic classification of

the filtered reads was then performed with Kraken2 [19]. Specifically, bacterial and archaeal reads were classified using the GTDB R220 database, while eukaryotic reads were classified using the NCBI Reference Sequences PlusPFP database (version date: 12/28/2024) [20]. The abundance for all taxonomic levels, down to the species-level, was estimated with Bracken (Bayesian Re-estimation of Kraken output) [21].

2.4 Metagenomic assemblies reconstruction

Metagenome-assembled genomes (MAGs) were reconstructed from quality-filtered reads using the metashot/mag-illumina v2.1.2 module [22]. Given the high sequencing depth of each sample, metagenomic assembly was performed independently for each sample, minimizing the risk of generating chimeric contigs. Briefly, reads were assembled into contigs using metaSPAdes [23] and binned into MAGs with MetaBAT2 [24]. Using the metashot/prok-quality v1.2.3 module [25], we performed quality assessment of the resulting MAGs, including the assessment of genome completeness and contamination estimates with CheckM [26]. Additional quality control steps included GUNC [27] to identify chimeric sequences, Barrnap [28] to annotate ribosomal RNA genes, and tRNAscan-SE [29] to detect transfer RNA genes, ensuring genome integrity and functional completeness. MAGs were considered of medium or high quality when showing a CheckM completeness estimate of > 50%, contamination estimate of < 10%, a quality score of *completeness - 5 * contamination* > *50*, > 40% presence of the bac120 or ar53 marker genes, < 1000 contigs, an N50 > 5kb, and < 100,000 ambiguous bases classified as “N”.

MAGs were further de-replicated at the species level (95% Average Nucleotide Identity) using dRep [30] and taxonomically classified using compmetagen/prok-classify [31], an updated version of metashot/prok-classify [32], which applies GTDB-Tk (v2.4.0) to place genomes within the GTDB taxonomic framework (v2.3.2). The phylogenetic tree of the taxonomically classified MAGs was generated using FastTree [33, 34].

To evaluate how representative the recovered MAGs were of the studied microbial communities, high-quality metagenomic reads were aligned against the dereplicated catalogue of MAGs using Bowtie2 (v2.5.4) [35]. The percentage of mapped reads was calculated as the number of reads aligned at least once over the total number of raw sequencing reads using `samtools flagstat` (v1.23) [36].

In parallel, to estimate the distribution of the representative MAGs within the sample sequences, we employed the metashot/containment v1.0.0 module [37], which utilizes Mash Screen [38]. Mash screen uses MinHash sketches, enabling the detection and quantification of reference sequences within sequencing data. Namely, quality-filtered reads were searched in non-redundant MAGs' sketches of size 10,000 and a match was defined as the query read sharing at least 95% of its k-mers with the corresponding sketch. The results were visualized by plotting the $\log_2$ of the 'match count' for each MAG against sample reads.

2.5 Statistical analysis

2.5.1 Taxonomic diversity comparison and differential abundance analysis

For downstream taxonomic analysis, the combined abundance table generated with Bracken was imported into the R statistical environment [39] at the species level using the `microeco` R package (v1.10.0). The proportions of classified and unclassified reads were compared between environments (sediment vs matrix and L-SLs vs SWB-PBs) and among domains (bacteria, archaea and eukaryotes) using a two-proportion Z test.

To reduce the influence of potential false-positive taxonomic assignments, taxa detected in fewer than two samples were removed as previously suggested [40]. Raw abundance counts were then rarefied to the minimum sequencing depth, using the function trans$rarefy (microeco package) (the rarefaction curve is shown in **Supplementary Fig. 2**). After rarefaction, the abundance table was disaggregated based on taxonomic domain, i.e. Bacteria, Archaea and Eukarya, keeping the three domains separate in the downstream

analysis. From the eukaryotes abundance table, plants were filtered out (i.e. all taxonomies classified as Kingdom *Viridiplantae,* except for *Chlorophyta*). To visualize the compositional structure of bacterial, archaeal and eukaryotic communities, compositional bar plots were generated using the `microeco` R package. The `trans_abund` function was used to show the compositional structure of each sample at the phylum level, considering the most abundant bacterial phyla, archaeal phyla and eukaryotic phyla (fungi, protists and microalgae only). The R package `ANCOMBC` v1.6.4 [41] was used to calculate the pairwise differential abundance between environmental groups at the phylum level. As mentioned previously, following Vettorazzo et al. [8], we classified samples into 'lakes and shallow lakes' (L-SLs) and 'small water bodies and peat bogs' (SWB-PBs) based on their depth and size profile to compare communities across the physiographic gradient. L-SLs included the lakes LL, LT, the shallow lakes LA, LN, LV and the lake ponds MST, PD, TF1. SWB-PBs included the pasture ponds MT, BG, MS, the ponds MC, MN, MM, MB and the peat bogs TF2, TF3, TL and L1. These classes have been applied throughout the study (**Supplementary Table 1**). Five pairwise comparisons were performed, i.e., L-SLs water vs SWB-PBs water, L-SLs water vs SWB-PB sediment, L-SLs water vs L-SLs sediment, SWB-PBs sediment vs L-SLs sediment, and SWB-PBs sediment vs SWB-PBs water. Phyla that were not significantly different between a minimum of two groups were filtered out (alpha = 0.05). All P-values, here and in the rest of the analyses, were corrected for multiple testing using the Holm-Bonferroni method.

Alpha diversity indices, including the Chao1 richness index and the Shannon diversity index, were calculated using the `cal_alphadiv` function (`microeco`) on Bracken-derived abundance tables. Statistical differences in alpha diversity indices between water and sediment samples were evaluated using the non-parametric Wilcoxon signed rank exact test, to account for paired non-normal data distributions.

Beta diversity was then assessed using the Bray-Curtis (BC) dissimilarity index, computed with the `cal_betadiv` function from `microeco`. This BC dissimilarity matrix was investigated with Principal Coordinates Analysis (PCoA) to visualize sample clustering based on

community composition. PERMANOVA was used to assess differences in composition between predefined sample groups, such as lake sediment (L-SLs) vs. small water bodies or peat bogs (SWB-PBs), in water vs. sediment environments. PERMANOVA was independently calculated for bacteria, archaea and eukarya using the `microeco` built-in function `cal_manova`.

To quantify the heterogeneity within and between sediment and water samples, we employed PERMDISP, a multivariate extension of Levene's test that evaluates the homogeneity of group variances. We first calculated the distance from each data point to its group centroid using the function `betadisper` from the R library `vegan`. We then tested whether the distances between the relative group centroids differed significantly using PERMANOVA using the function `adonis2` from `vegan`.

2.5.2 Functional annotation and analysis

The functional annotation of metagenomic scaffolds was conducted with the metashot/prok-annotate v3.0.1 workflow [42]. The pipeline uses Prokka [43] to identify open reading frames (ORFs) and eggNOG-mapper 2.0 [44] to assign orthology-based annotations using the eggNOG v5.0 database [45] and enable predictions of gene function and metabolic pathways. The `–metagenome` option was used when running the pipeline, to perform gene prediction directly from sample-specific scaffolds and not on the assembled MAGs. Based on eggNOG's predictions, genes were also annotated against the KEGG Orthology database (KEGG Orthology, Pathways, Modules and BRITE) [46]. Gene annotations were filtered with an *e*-value cutoff of 10^{-5}. For each of the 38 metagenomic samples, we considered a KEGG Orthology (KOs) profile and a KEGG Module profile (a higher-level organization of KEGG Orthologies).

We compared the functional diversity of each habitat, grouped by matrix and physiographic gradient (four habitats), by computing a Shannon Diversity Index on abundance matrices of the genes (KOs) and the functional KEGG Modules annotated within each sample. Of note,

the KEGG KOs and KEGG Module abundance matrix represents the number of times that a query protein sequence was matched to its best matching orthologous group in the eggNOG database using sequence similarity. We also computed the richness of KOs and KEGG Modules by counting the number of annotated functional genes in each sample. To compare samples uniquely by their known functional profile, unclassified genes were removed (the number of unclassified genes and modules for each sample can be found in **Supplementary Table 2** and **Supplementary Table 3**, respectively). We then compared the four habitats by their Shannon Index and Richness in a pairwise manner using the Wilcoxon Signed Rank test. Principal Component Analysis (PCA) was then applied to visualize the distances between samples based on these functional profiles. Specifically, PCA plots were generated using the non-scaled KEGG Orthology matrix and the KEGG Module matrix. To identify sample clusters based on KEGG Modules, we then performed hierarchical clustering on the PCA-transformed data. We calculated the Euclidean Distance between samples based on the top 5 principal components with the R function `dist()` and identified clusters with Ward's method using the R function `hclust()`.

Lastly, the distribution of orthologous genes associated with biogeochemical pathways and photosynthesis were compared between environments, i.e. water and sediment and L-SLs and SWB-PBs. First, the KEGG database [47] was queried to identify KEGG Modules related to carbohydrate and energy metabolism. These include modules involved in carbohydrate metabolism (organic carbon oxidation), carbon fixation, nitrogen, sulfur, and methane metabolism and photosynthesis. After this search, 92 Modules were selected for further testing, 78 of which were present in the matrix of annotated genes. The results of this search are listed in **Supplementary Table 4**. To account for library size, the counts of genes assigned to each KEGG Module of interest were scaled by dividing gene counts by sample library size (**Supplementary Table 5**). Following this, the normalized counts for each module of interest were divided by the total number of genes assigned to all KEGG Modules within the same sample. This step allowed us to obtain relative gene abundances for each sample, enabling

the comparison of functional composition across samples from markedly different environments (e.g., water column versus sediment). For each biogeochemical and photosynthesis module, a distribution of these relative proportions was plotted (not shown), comparing sediment samples to water samples and L-SL samples to SWB-PB samples. The statistical significance of the differences found between environments was tested using the non-parametric Wilcoxon signed-rank test (paired comparison, water vs sediment) and the Wilcoxon rank-sum test (L-SLs vs SWB-PBs). To account for multiple comparisons across the tested KEGG Modules, p-values were adjusted using the Benjamini-Hochberg method to control the false discovery rate.

Moreover, to connect taxonomic composition and functional potential, we integrated the metabolic annotations of recovered MAGs with read-based taxonomic abundance profiles derived from Kraken2/Bracken. First, KEGG Module counts identified in the MAGs were aggregated at the phylum level to establish the metabolic potential of each taxonomic group. Parallelly, sample-wise taxonomic abundances were aggregated at the phylum level. To capture the representative community structure of each habitat, we calculated the mean abundance of each phylum across all samples belonging to each of the four study environments (L-SL sediment, L-SL water, SWB-PB sediment, and SWB-PB water). This way, we obtained four environment-specific mean abundances for each phylum, which we then multiplied by the phylum-level KEGG module copy numbers to estimate environment-specific metabolic potentials. To quantify the functional importance of specific taxa to different biogeochemical pathways, we calculated, for each module, the percentage contribution of individual phyla to the total metabolic potential (defined as the number of module copies in one phylum divided by the sum of module copies across all phyla) within each environment. Finally, for the selected KEGG Modules (**Table 1**), we visualized the distribution of metabolic potential using heatmaps. We then followed the same procedure at the genus-level to obtain an overview of the functional contribution of the most abundant genera to each environment. To this end, we selected the 20 most abundant genera in each environment and plotted their

module abundance. Lastly, to characterize the contribution of unclassified taxa to ecosystemic functions, we produced a table focusing exclusively on the functional traits of such taxa, which are currently still lacking classification at the genus and species levels (**Supplementary Table 6**). The table links unclassified MAGs to their assigned functional modules, by reporting their weighted abundance in each of the four studied environments (L-SL sediment, L-SL water, SWB-PB sediment, SWB-PB water), calculated as explained above.

3. Results

A total of 4,026,914,640 paired-end reads were obtained from sequencing 19 paired water and sediment samples (38 total), yielding approximately 604.04 gigabases (Gb) of total bases sequenced. After quality filtering, 3,801,769,950 high-quality reads remained, corresponding to about 570.26 Gb. Each sample averaged 15.90 Gb of sequencing depth before filtering and 15.01 Gb after filtering. On average, 62.2% of the total reads were taxonomically classified across samples. Of those reads, 61.5% were bacterial, 0.45% archaeal and 0.20% eukaryotic. Significantly more reads were classified in sediment samples compared to water (mean sediment = 66%, mean water = 58%; two-proportion z-test, $P < 0.001$), while no differences were found between 'lakes and shallow lakes' (L-SLs) and 'small water bodies and peat bogs' (SWB-PBs). Sediments showed a higher percentage of total classified reads compared to water for bacteria (median in sediment samples = 66.44%; median in water samples = 55.32%, two-proportion z-test, $P < 0.001$), archaea (median sediment = 0.46%; median water = 0.28%, two-proportion z-test, $P < 0.001$) and eukaryotes (median sediment = 0.24%; median water = 0.15%, two-proportion z-test, $P < 0.001$). Moreover, applying a minimum threshold of 1,000 reads per species, we estimated that a total of 464,981,825 reads were classified at the species level, corresponding to 12.23% of the total number of high-quality reads.

3.1 Matrix and water body physiography determine microbial community composition and structure

3.1.1. Freshwater and sediment habitats host unique microbial communities

We analyzed the compositional and diversity patterns of bacterial, archaeal, and eukaryotic communities from freshwater and sediment habitats and detected 196 bacterial, 21 archaeal and 13 eukaryotic phyla.

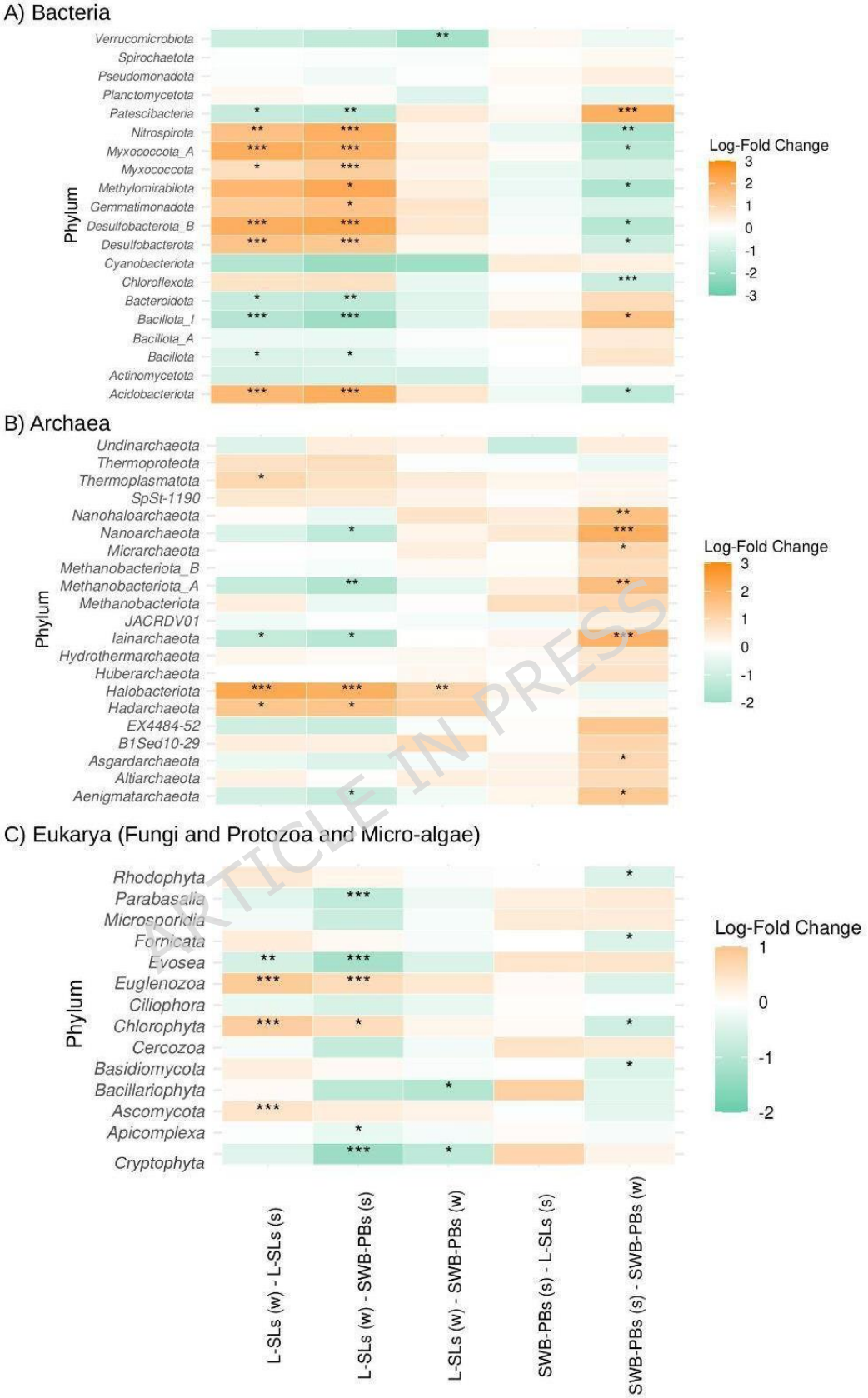
Bacterial communities exhibited distinct composition patterns between water and sediment samples. At the phylum level, water samples were dominated by *Pseudomonadota* (45%

mean abundance), *Actinomycetota* (19%), *Bacteroidota* (12%), *Verrucomicrobiota* (6%), and *Cyanobacteriota* (3%), while at the genus level, *Planktophila*, *Limnohabitans*, *Polynucleobacter*, *Flavobacterium*, emerged as the most abundant. In large lakes (including LL, LT, LA, and LV), *UBA953* (*Verrucomicrobiota*), *Fonsibacter* (*Pseudomonadota*), *Nanopelagicus* (*Actinomycetota*) and *Planktothrix* (*Cyanobacteriota*) showed particularly high relative abundances, while in the genus *JAIXTH01* (*Pseudomonadota*) was particularly high in the pasture pond MT. Sediment samples were dominated by *Pseudomonadota* (45%), *Actinomycetota* (16%), *Acidobacteriota* (6%), *Bacteroidota* (6%), *Chloroflexota* (4%), and *Verrucomicrobiota* (4%) (**Supplementary Table 7**, **Supplementary Fig. 3A**). At the genus level, sediments reported a higher relative abundance of *Bradyrhizobium* (*Psudomonadota*), *Streptomyces* (*Actinomycetota*) and *Rubrivivax* (*Psudomonadota*) (**Supplementary Fig. 4A**). Differential analysis using ANCOM-BC further highlighted these habitat-based differences. In L-SLs, we found a total of 4 phyla whose abundance was higher in water compared to L-SL sediments (Holm-Bonferroni $P<0.05$); viceversa, six phyla had a significantly higher abundance in sediments. Compared to L-SL sediments, L-SL water samples were enriched in *Patescibacteria* ($P = 0.012$, ANCOM-BC logFC = 1.17), *Bacteroidota* ($P = 0.03$, logFC = 1.17), *Bacillota* ($P = 0.02$, logFC = 0.73), and *Bacillota_I* ($P < 0.001$, logFC = 1.52). L-SL sediment samples showed higher abundance of *Myxococcota_A* ($P < 0.001$, logFC = 2.12), *Desulfobacterota* ($P < 0.001$, logFC = 1.44), and *Acidobacteriota* ($P < 0.001$, logFC = 1.81) compared to L-SL sediments, amongst others. A similar pattern was observed in SWB-PBs, where water samples showed higher abundances of *Patescibacteria* ($P < 0.001$, logFC = 2.07) and *Bacillota_I* ($P = 0.04$, log FC = 1.57) compared to sediments, while sediments were enriched in *Chloroflexota* ($P < 0.001$, logFC = 1.07), *Nitrospirota* ($P = 0.002$, logFC = 1.65), *Desulfobacterota_B* ($P = 0.01$, logFC = 1.48), and *Acidobacteriota* ($P = 0.02$, logFC = 1.32). The only significant difference between L-SLs and SWB-PBs was in waters, where *Verrucomicrobiota* showed higher abundance in L-SLs ($P = 0.005$, logFC = 1.83), while sediment communities showed no significant differences between physiographic groups (**Fig. 2A**, **Supplementary Table 8**).

Similarly to bacteria, archaea showed unique patterns in community composition in sediments and waters, though they were not affected by water body characteristics such as depth and size, showing minimal variation between L-SLs and SWB-PBs. Sediment samples were dominated by *Halobacteriota* (60% group mean abundance), *Thermoproteota* (17%), and *Thermoplasmatota* (12%), whereas water samples contained different ratios of *Halobacteriota* (32%), *Nanoarchaeota* (17%), *Thermoproteota* (14%), and *Thermoplasmatota* (14%, **Supplementary Table 9**, **Supplementary Fig. 3B**). At the genus level, the archaeal genera with highest relative abundance were *Methanoregula, Methanosarcina, Methanothirix, Halorubrum* (*Halobacteriota*) and *UBA184* (*Thermoplasmatota*), mostly high in sediments but also present in waters (**Supplementary Fig. 4B**). According to ANCOM-BC analysis, L-SL water samples were enriched in *Iainarchaeota* ($P = 0.02$, logFC = 1.23) compared to L-SL sediments. Conversely, L-SL sediments had higher *Halobacteriota* ($P < 0.001$, logFC = 2.26), *Hadarchaeota* ($P = 0.04$, logFC = 1.46), and *Thermoplasmatota* ($P = 0.04$, logFC = 1.02). In SWB-PBs, 7 phyla were enriched in water samples compared to sediments, including *Nanoarchaeota* ($P < 0.001$, logFC = 2.09), *Iainarchaeota* ($P < 0.001$, logFC = 2.01), *Nanohaloarchaeota* ($P = 0.003$, logFC = 1.63), *Methanobacteriota_A* ($P = 0.008$, logFC = 1.69). Only *Halobacteriota* ($P = 0.007$, higher in SWB-PB water samples, logFC = 1.15) differed significantly between the water samples of L-SLs and SWB-PBs (**Fig. 2B**, **Supplementary Table 10**).

Eukaryotic microbial communities were the least abundant but exhibited distinct patterns between sediment and water habitats, specifically in small water bodies and peat bogs. Across all samples, we identified three fungal phyla, seven non-photosynthetic protist phyla, and five photosynthetic microalgal phyla (**Supplementary Table 11**, **Supplementary Fig. 3C**). ANCOM-BC results showed that L-SL water samples were enriched in the phyla *Evosea* ($P < 0.01$), while sediments were enriched in *Euglenozoa* ($P < 0.001$), *Chlorophyta* ($P < 0.001$), and *Ascomycota* ($P < 0.001$). In SWB-PBs, sediments had a higher relative abundance of *Rhodophyta* ($P < 0.05$), *Microsporidia* ($P < 0.05$), *Chlorophyta* ($P < 0.05$), and *Basidiomycota*

340 ($P < 0.05$). Only *Bacillariophyta* ($P < 0.05$) differed significantly between L-SL and SWB-PB
341 water samples, while L-SL and SWB-PB sediment samples did not show significant
342 differences (**Fig. 2C, Supplementary Table 12**).


A) Bacteria
Verrucomicrobiota
Spirochaetota
Pseudomonadota
Planctomycetota
Patescibacteria
Nitrospirota
Myxococcota_A
Myxococcota
Methylomirabilota
Gemmatimonadota
Desulfobacterota_B
Desulfobacterota
Cyanobacteriota
Chloroflexota
Bacteroidota
Bacillota_I
Bacillota_A
Bacillota
Actinomycetota
Acidobacteriota
Phylum
Log-Fold Change
3
2
1
0
-1
-2
-3
B) Archaea
Undinarchaeota
Thermoproteota
Thermoplasmatota
SpSt-1190
Nanohaloarchaeota
Nanoarchaeota
Micrarchaeota
Methanobacteriota_B
Methanobacteriota_A
Methanobacteriota
JACRDV01
Iainarchaeota
Hydrothermarchaeota
Huberarchaeota
Halobacteriota
Hadarchaeota
EX4484-52
B1Sed10-29
Asgardarchaeota
Altiarchaeota
Aenigmatarchaeota
Phylum
Log-Fold Change
3
2
1
0
-1
-2
C) Eukarya (Fungi and Protozoa and Micro-algae)
Rhodophyta
Parabasalia
Microsporidia
Fornicata
Evosea
Euglenozoa
Ciliophora
Chlorophyta
Cercozoa
Basidiomycota
Bacillariophyta
Ascomycota
Apicomplexa
Cryptophyta
Phylum
Log-Fold Change
1
0
-1
-2
L-SLs (w) - L-SLs (s)
L-SLs (w) - SWB-PBs (s)
L-SLs (w) - SWB-PBs (w)
SWB-PBs (s) - L-SLs (s)
SWB-PBs (s) - SWB-PBs (w)

343

Figure 2. ANCOM-BC analysis reveals sample-type differential Bacteria, Archaea and Eukarya phyla. ANCOM-BC was used to perform pairwise comparisons between samples belonging to large lakes and shallow lakes (L-SLs) water samples, L-SLs sediment samples, small water bodies and peat bogs (SWB-PBs) water samples and SWB-PBs sediment samples. Note that the group reported to the left of the comparison (e.g. L-SLs (w), in the comparison L-SLs (w) vs L-SLs (s)) is the reference level for the calculation of Log-Fold Change. The analysis was performed at the phylum level, significance was considered as adjusted $P < 0.05$. Legend: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **(A)** Log-Fold Change values of bacterial phyla comparing water and sediment samples, within and between the L-SL and the SWB-PB groups. The 20 most abundant phyla are indicated. **(B)** Log-Fold Change values of 21 archaeal phyla comparing water and sediment samples, within and between the L-SL and the SWB-PB groups. **(C)** Log-Fold Change values of eukaryotic phyla (inclusive of fungi, non-photosynthetic protists and photosynthetic microalgae) comparing water and sediment samples, within and between the L-SL and the SWB-PB groups.

3.1.2 Matrix and water body size and depth influence species evenness and diversity

L-SL sediment samples exhibited significantly higher Inverse Simpson and Shannon indices than water samples (Wilcoxon signed-rank test, $P < 0.01$, **Fig. 3A** and **Fig. 3B**), indicating greater evenness and diversity. By contrast, SWB-PB sediments and water samples showed no significant differences in these indices, though Inverse Simpson and Shannon of water samples exhibited a wider range of values compared to sediment. These patterns suggest that sediments have structurally comparable bacterial communities, dominated by low-abundance species, while water samples, particularly in larger and deeper L-SLs, are more variable and often dominated by highly abundant taxa like *Planktothrix agardhii* and *Fonsibacter*. However, the water of the shallow Lake Nembia (LN) was an exception, showing a high diversity of lowly abundant species, a pattern typical in sediments. In SWB-PBs, the pasture ponds BG, MS and the ponds MB and MC were dominated by the genera *Limnohabitans* and *Flavobacterium*, while the ponds L1, MM, MN and the peat bogs TF2, TF3, TL resembled the pattern observed in sediments with their low-abundance, high-diversity species assemblages. For archaea, Inverse Simpson and Shannon indices showed no statistically significant differences between habitats (Wilcoxon signed-rank test, $P > 0.05$), though water samples tended to exhibit higher diversity values (**Fig. 3E** and **Fig. 3F**). In contrast to bacteria, eukaryotic diversity was significantly higher in water samples, particularly

in SWB-PBs, where Inverse Simpson and Shannon indices exceeded those of sediments (Wilcoxon signed-rank test, P < 0.01, **Fig. 3G** and **Fig. 3H**).

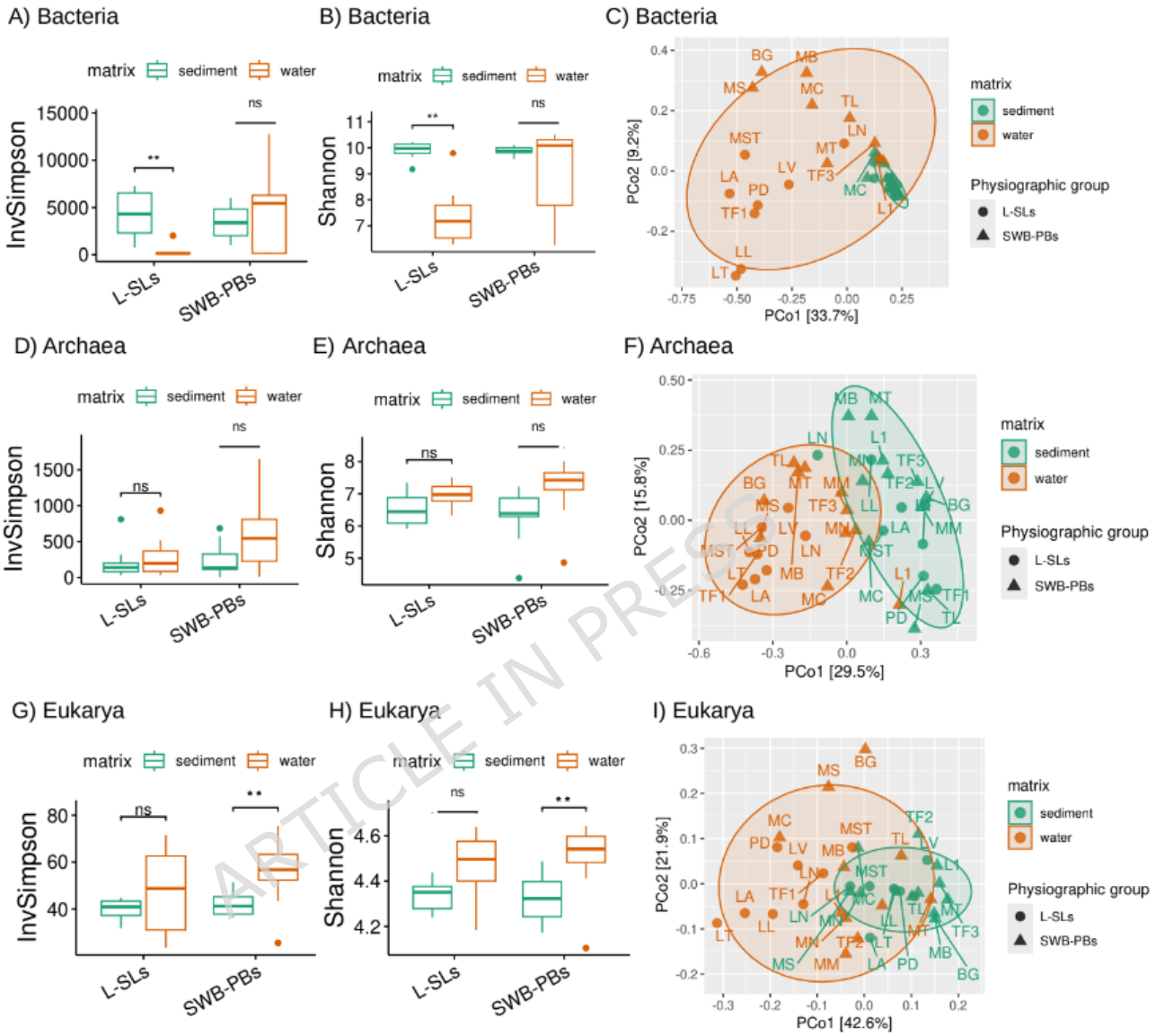


Figure 3. Water and sediment habitats differ in taxonomic diversity patterns (alpha and beta). All groups were compared with the Signed-rank Wilcoxon test, *ns: not significant, * P < 0.05, ** P < 0.01*). **(A, D and G)** Inverse Simpson Diversity Index of bacterial **(A)**, archaeal **(D)** and eukaryotic **(G)** species found in the samples (n = 19) divided by matrix (water or sediment) and paired by L-SLs or SWB-PBs. **(B, E and H)** Shannon Index distribution of bacterial **(B)**, archaeal **(E)** and eukaryotic **(H)** species, divided by matrix (water or sediment) and paired by L-SLs or SWB-PBs. **(C, F and I)** PCoA based on the species' Bray-Curtis distance between bacterial **(C)**, archaeal **(F)** and eukaryotic **(I)** communities across samples.

3.1.3 Water communities are highly heterogeneous, while sediment communities are more taxonomically similar across water bodies

To assess how habitat type and physiographic characteristics influence microbial community structure, we derived abundance matrices at the species level using Bracken and calculated Bray-Curtis (BC) dissimilarities between sediment and water samples from L-SLs and SWB-PBs, which were then visualized with PCoA.

For bacterial communities, the PCoA revealed a clear separation between water and sediment samples (PERMANOVA, $P < 0.001$, **Fig. 3C**). The distinction was also substantial between L-SLs and SWB-PBs ($P = 0.014$). The PERMDISP test showed that dispersion was significantly higher in water communities ($P < 0.0001$, **Supplementary Fig. 5A**), while it was lower in sediment communities due to their structural similarity.

Archaeal communities showed a significant difference in BC distances between water and sediment (PERMANOVA, $P < 0.001$, **Fig. 3F**), but no significant difference between L-SLs and SWB-PBs ($P = 0.063$), indicating that habitat type, not water body size or depth, dictates their structure. PERMDISP showed higher dispersion in sediment than water samples ($P < 0.04$, **Supplementary Fig. 5B**).

For eukaryotes, water samples showed greater variability, while sediments were more taxonomically similar. PERMANOVA revealed significant differences between water and sediment samples ($P < 0.001$) and between physiographic groups ($P = 0.015$, **Fig. 3I**). PERMDISP showed significantly higher dispersion in water than sediment samples ($P < 0.001$, **Supplementary Fig. 5C**). These findings highlight that water communities are more heterogeneous, while sediment communities are taxonomically more comparable. Differences between L-SLs and SWB-PBs were significant for bacteria and eukaryotes but not for archaea.

3.2 Water and sediment communities show markedly distinct functional potentials

We assembled 341,584,112 contigs for 103 Gb from the 38 metagenomes, predicting 131,865,881 genes. Of these, 31% were functionally annotated, while 69% had no functional analog in the EggNOg database. Since bacterial communities contributed over 90% of the classified reads, most of the observed functional potential can be attributed to bacterial genes.

Functional diversity across habitats was assessed using Shannon Diversity Index and richness metrics for KEGG Orthologues (KOs, **Fig. 4**) and KEGG Modules (**Supplementary Fig. 6A**, **B**). Significant differences in Shannon diversity based on KOs profiles were found between L-SL waters and L-SL sediments ($P = 0.008$) and L-SL waters and SWB-PB waters ($P = 0.0005$, **Fig. 4A**, **B**). No significant differences were found between SWB-PB waters and SWB-PB sediments ($P > 0.05$), nor between L-SL sediments and SWB-PB sediments ($P > 0.05$), or L-SL sediments and SWB-PB waters ($P > 0.05$). No differences in KEGG KOs richness were found across habitats. For KEGG Modules, richness was significantly different between L-SL sediments and L-SL waters ($P = 0.02$), but no differences were observed between groups for the Shannon Index (**Supplementary Fig. 6A, B**).

PCA of functional profiles (KEGG KOs, **Fig. 4C**; KEGG Modules, **Supplementary Fig. 6C**) showed that sediment and water habitats form distinct clusters. Some water samples, including peat bogs, ponds and the shallow lake LN, clustered closer to sediments (**Supplementary Fig. 6D, Supplementary Fig. 7**).

When analyzing the relative proportion of total genes in each habitat involved in carbon and energy metabolism, we found significant differences in metabolic potential between water and sediment, and L-SLs and SWB-PBs (**Table 1**). Sediment-associated bacterial communities showed a higher potential for methane metabolism and nitrogen cycling, such as methanogenesis, nitrification, denitrification, and dissimilatory nitrate reduction ($P < 0.001$). Water-associated communities exhibited stronger potential for carbon and sulfur metabolism, including organic carbon oxidation, carbon fixation, glycolysis, and assimilatory sulfate

reduction ($P < 0.001$). Water habitats also had a higher proportion of genes related to photosynthesis, including Photosystem I and II ($P < 0.01$; $P < 0.001$), as well as anoxygenic photosynthesis of type II (M00597, $P < 0.001$). Sediments had relatively more genes related to Anoxygenic System I (M00598, $P < 0.001$). When comparing L-SLs and SWB-PBs, the two groups differed mainly in assimilatory nitrate reduction ($P < 0.05$) and photosynthesis genes. L-SLs showed higher relative levels of genes related to photosystems I, II and anoxygenic photosystem II (M00161, M00163, M00597, $P < 0.05$). Overall, these findings illustrate the functional specialization of microbial communities, i.e. sediment microbiomes are more involved in anaerobic processes (methanogenesis, nitrogen cycling), while water microbiomes rely more on phototrophic pathways.

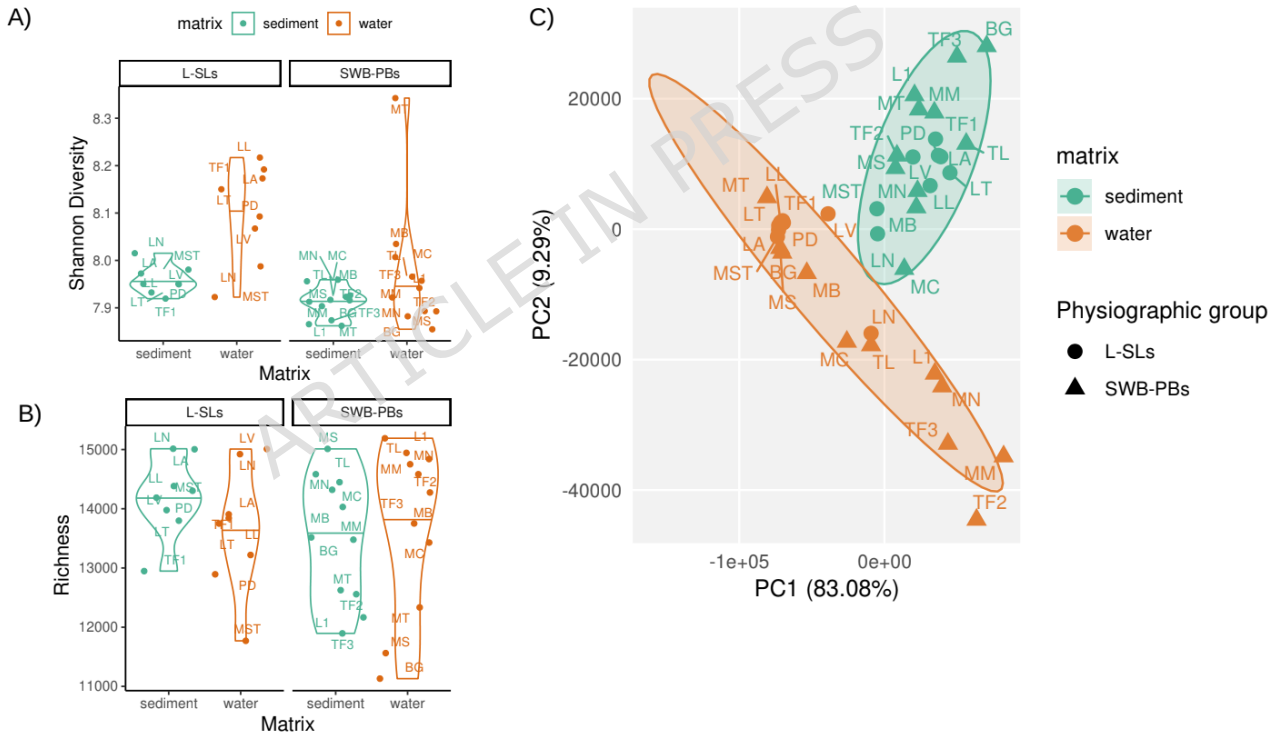


Figure 4. Habitats differ by their microbiome functional repertoire. (A) Shannon Index calculated on the annotated genes (KEGG KOs) based on the matrix of origin (sediment or water) and physiographic group (lakes and shallow lakes, L-SLs or small water bodies and peat bogs, SWB-PBs). **(B)** Total number of annotated genes (Richness) per sample, grouped by habitat type (sediment, water), and by water body type (L-SLs or SWB-PBs). **(C)** Beta-diversity analysis of functionally annotated samples based on the total KEGG KOs identified in each sample's metagenome, divided by sediment and water and by water body type (L-SLs or SWB-PBs).

See end of document for table 1. Table 1 should appear here.

Table 1. Comparison of the functional potential of sediment and water (matrix) and the physiographic groups L-SLs and SWB-PBs. Comparisons based on the relative proportions of genes dedicated to nutrient cycling (Carbohydrate-, Carbon-, Nitrogen-, Sulfur- and CH_4-related metabolism) and photosynthesis found in the metagenomes of these habitats. Measured by the relative number of genes (RPM) assigned to the KEGG Modules belonging to 'Carbohydrate metabolism' and 'Energy metabolism', relative to the total number of modules assigned for each sample. The 'matrix' and 'physiographic group' columns represent where the given Module was found significantly more abundant. The statistical significance of the differences found between environments (water vs sediment and L-SLs vs SWB-PBs) were tested using the non-parametric Wilcoxon signed-rank test (paired) or the Wilcoxon Rank-Sum test and P-values were adjusted using the Benjamini-Hochberg method. Legend: * for P ≤ 0.05, ** for P ≤ 0.01, *** for P ≤ 0.001.

3.3 Characterization of a metagenome-assembled genomes catalogue

A total of 566 quality-filtered MAGs were recovered from the sampled sites. After de-replication, we obtained 496 representative bacterial and archaeal MAGs (**Fig. 5**), with a majority of bacterial MAGs (n = 485, **Fig. 5A**) and a minority of archaeal MAGs (n = 11, **Fig. 5C**). Regarding completeness and contamination, 36% of bacterial MAGs scored a high-quality score (*Completeness > 90% and Contamination < 5%*, **Supplementary Fig. 8B**), while 64% scored a medium-quality score *(Completeness > 50% and Contamination < 10%)*. 18% of archaeal MAGs had a high-quality and 92% a medium-quality score (**Supplementary Fig. 9B**). Out of 496 MAGs, 133 bacterial and 3 archaeal MAGs were classified at the species level based on 95% whole-genome average nucleotide identity (ANI). In other words, 72% (n = 352) of the identified bacterial species and 73% (n = 8) of the archaeal species were novel, i.e. currently absent in the curated environmental bacterial database GTDB (RS220, **Fig. 5B, D**). In particular, for bacteria, the greatest number of novel species was found for the phyla *Pseudomonadota* (121 MAGs), *Actinomycetota* (75 MAGs), *Bacteroidota* (57 MAGs), *Verrucomicrobiota* (25 MAGs), *Patescibacteria* (23 MAGs), and *Cyanobacteriota* (10 MAGs) (**Fig. 5B**). In the Archaea domain, we found three novel *Nanoarcheota* species, two *Thermoplasmatota* species, two *Halobateriota* species and one *Thermoproteota* species (**Fig.**

485 **5D**). The uncovered bacterial genomes encode 1,128,681 protein sequences, while 15,940
486 genes were predicted from archaeal genomes (**Supplementary Tables 13** and **14,**
487 respectively). In particular, the MAG with the highest number of genes encoded 11600 genes
488 and was assigned to the genus *SZUA-265* (phylum *Myxococcota)*, The second most gene-
489 rich genome (with 6740 genes) was assigned to the genus *Runella* (phylum *Bacteroidota*).
490 Instead, amongst archaeal genomes, the MAGs with the highest number of annotated genes
491 were assigned to the genera *Methanosarcina* and *Methanotrix* (*Halobacteriota*) and had 2912
492 and 2280 genes, respectively.

493 Amongst the quality-filtered and de-replicated MAGs, 76% were reconstructed from water
494 habitats, while only 23% were from sediment habitats. Given the higher bacterial diversity
495 found in sediments compared to water (**Fig. 3B**), such result seems to be related to technical
496 difficulties during *de novo* assembly. Interestingly, bacterial MAGs isolated from sediment
497 samples were significantly bigger in genome size ($P < 0.001$, **Supplementary Fig. 8C**).
498 Moreover, when we mapped reads back to the dereplicated MAGs to gain a picture of how
499 representative MAGs were of the general community, we found that most reads mapped to
500 MAGs recovered from water samples, while a strikingly low percentage of high-quality reads
501 mapped to MAGs recovered from sediments (**Supplementary Table 15, Supplementary Fig.**
502 **10**).

503 To determine the distribution of the de-replicated MAGs (n = 496) within the 38 samples and
504 have a proxy measure of their abundance, we used the “MASH screen” containment analysis
505 with the “winner-take-all strategy”. Containment analysis showed that a higher percentage of
506 MAGs mapped to water habitats, especially to the L-SL group (**Supplementary Fig. 11,**
507 **Supplementary Table 16**). These MAGs were found in a non-uniform way across habitats
508 along the physiographic gradient and each water habitat showed a unique MAG fingerprint.
509 Conversely, species found in sediment habitats were mostly present in all samples, as shown,
510 for example, by species of the phylum *Acidobacteriota, Desulfobacterota, Methylomirabilota*
511 and *Pseudomonadota*. These results are in accordance with the results on community

composition (**Fig. 3**) that showed that sediments are overall more homogeneous at the species level compared to waters, which instead show a more diverse taxonomic profile between samples.

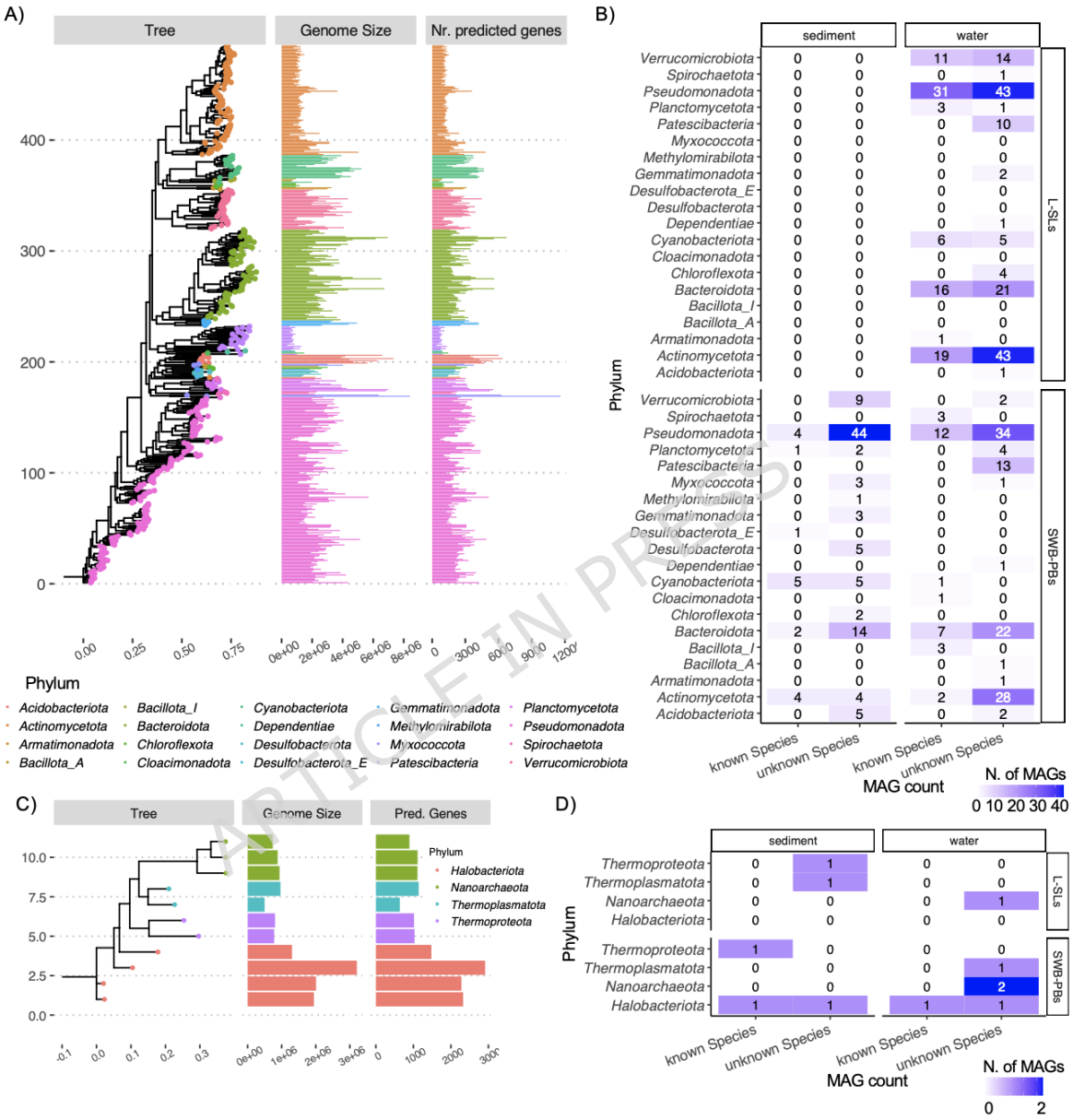


Figure 5. Filtered and de-replicated bacterial and archaeal Metagenome-Assembled Genomes overview. De-replicated and filtered bacterial and archaeal Metagenome-Assembled Genomes grouped by phylum. **(A)** Genome size distribution and number of predicted genes for each bacterial MAG (n = 485) ordered by phylogeny and colored by phylum. **(B)** The number of known and unknown bacterial species grouped by phylum, matrix (water vs sediment) and physiographic class (L-SLs and SWB-PBs). **(C)** The genome size distribution and number of predicted genes for each archaeal MAG (n = 11) ordered by phylogeny and colored by phylum. **(D)** The number of known and unknown archaeal species

grouped by phylum, matrix (water vs sediment) and physiographic class (L-SLs and SWB-PBs).

3.4 Functional traits of alpine lacustrine and sedimentary MAGs

Following the comparison of functional compositions across different environments (sediment vs. water and L-SLs vs. SWB-PBs), measured in terms of module abundance (**Table 1**), we examined the annotated genes and pathways from the most abundant MAGs and linked them to the environments in which they were uncovered. To this end, we multiplied the number of modules identified in bacterial and archaeal MAGs by their taxonomic abundance in the samples (obtained with Kraken2/Bracken). As a result, we obtained weighted gene annotation matrices of biogeochemical and photosynthetic pathways (KEGG Modules). We then plotted heatmaps showing the distribution of genomic functions across the 20 most abundant phyla in each environment (L-SL sediment, L-SL water, SWB-PB sediment, SWB-PB water) (**Fig. 6)**. We observed that, in all the studied environments, *Pseudomonadota* was the phylum that carried most ecosystemic functions, followed by *Actinomycetota, Bacteroidota, Cyanobacteriota* and *Halobacteriota*. In terms of methane metabolism, *Pseudomonadota* was the primary contributor to methanogenesis potential across all environments. Supporting pathways showed distinct taxonomic drivers: *Bacteroidota* largely accounted for Coenzyme M biosynthesis (particularly in water samples) and, together with *Desulfobacterota*, dominated the Acetyl-CoA pathway. Formaldehyde assimilation potential was broadly distributed among *Pseudomonadota, Bacteroidota,* and *Actinomycetota*. Finally, the archaeal phyla *Halobacteriota* and *Thermoplasmatota* were the main contributors to 2-Oxocarboxylic acid metabolism in all environments. The nitrogen cycle was mostly dominated by *Pseudomonadota*, which contributed with more than the 75% of genes involved in nitrification, denitrification, assimilatory and dissimilatory nitrate reduction. *Actinomycetota* and *Bacteroidota* contributed the remaining portions of the genes involved in nitrogen metabolism (<20%). A similar pattern was observed for sulfur metabolism, where *Pseudomonadota* had the highest potential (>75%) for assimilatory sulfate reduction, dissimilatory sulfate reduction

and sulfur oxidation (M00595). Compared to other studied environments, L-SL waters showed a decreased percentage contribution of genes from *Pseudomonadota* (between 50-75% contribution), and an increased proportion of genes contributed by *Actinomycetota* (~25%) and *Verrucomicrobiota* (~10%), particularly concerning carbohydrate metabolism, carbon fixation and sulfur metabolism. Interestingly, L-SL waters also showed *Cyanobacteriota*'s metabolic potential to perform functions involved in methanogenic metabolisms. Although photosynthesis (Photosystem I, and Photosystem II) was found to be significantly higher in water (M00163, $P < 0.006$; M00161, $P < 0.001$; **Table 1**), oxygenic photosynthesis modules appear abundant across all environments due to the widespread presence of the phyla *Cyanobacteriota* and *Bacteroidota,* which carried the highest photosynthesis potential. Anoxygenic photosynthesis I modules, more present in water than sediments, were mostly contributed by *Bacteroidota.*

To refine our understanding of the taxa important for lacustrine and sediment environments and their functional features, we visualized the abundance of modules in the MAGs of the top 20 most abundant taxa at the genus level (**Supplementary Fig. 12**). We observed that in the water column of both lake types, central carbon metabolism was largely sustained by the genera *Planktophila* (*Actinomycetota*), *Limnohabitans*, *Polynucleobacter*, *Rhodoluna, Rhodoferax C* and *Flavobacterium* (phylum *Pseudomonadota*), which exhibited the highest weighted abundance of glycolysis and citrate cycle modules. In contrast, sediments showed a more taxonomically varied functional landscape with lower overall weighted gene abundance. In sediments, *Rubrivivax* emerged as the primary driver across methane, nitrogen, sulfur metabolism and photosynthetic pathways.

Focusing more closely on the metabolic potential of unclassified species, we found that sediment environments were richer in taxa unclassified both at the genus and species level, indicating a more highly unexplored environment compared to waters. Waters also showed numerous unclassified taxa, but they were often classified at the genus level (see **Supplementary Table 6** for the full list of unclassified taxa found in this study with their

functional potential). Interestingly, we found a taxon in the family *Fimbriimonadaceae* (phylum *Armatimonadota*), which showed a high number of genes coding for Photosystem I (M00163), sulfur metabolism (M00176, M00595, M00596), both in the waters and sediments of SWB-PBs. Other unclassified taxa coding for photosynthetic genes belonged to the *Cyanobacteriaceae* and *Microcystaceae* (*Cyanobacteriota*) families, both in water and sediment environments. We also found MAGs belonging to the order *Opitutales* (*Verrucomicrobiota*) distributed across all environments, which encoded a high number of genes involved in the nitrogen cycle, specifically denitrification, assimilatory and dissimilatory nitrate reduction (M00529, M00530, M00531). In SWB-PB sediments, we found taxa from the order *Rhodospirillales*, still unclassified at the family level and particularly active in the metabolism of methane (M00344, M00346, M00356, M00357, M00563, M00567). Contributing to methanogenesis in sediments and belonging to the archaeal domain, we found 2 new species of *Methanothrix* (phylum *Halobacteriota*) in L-SL sediments and SWB-PB sediment and waters.

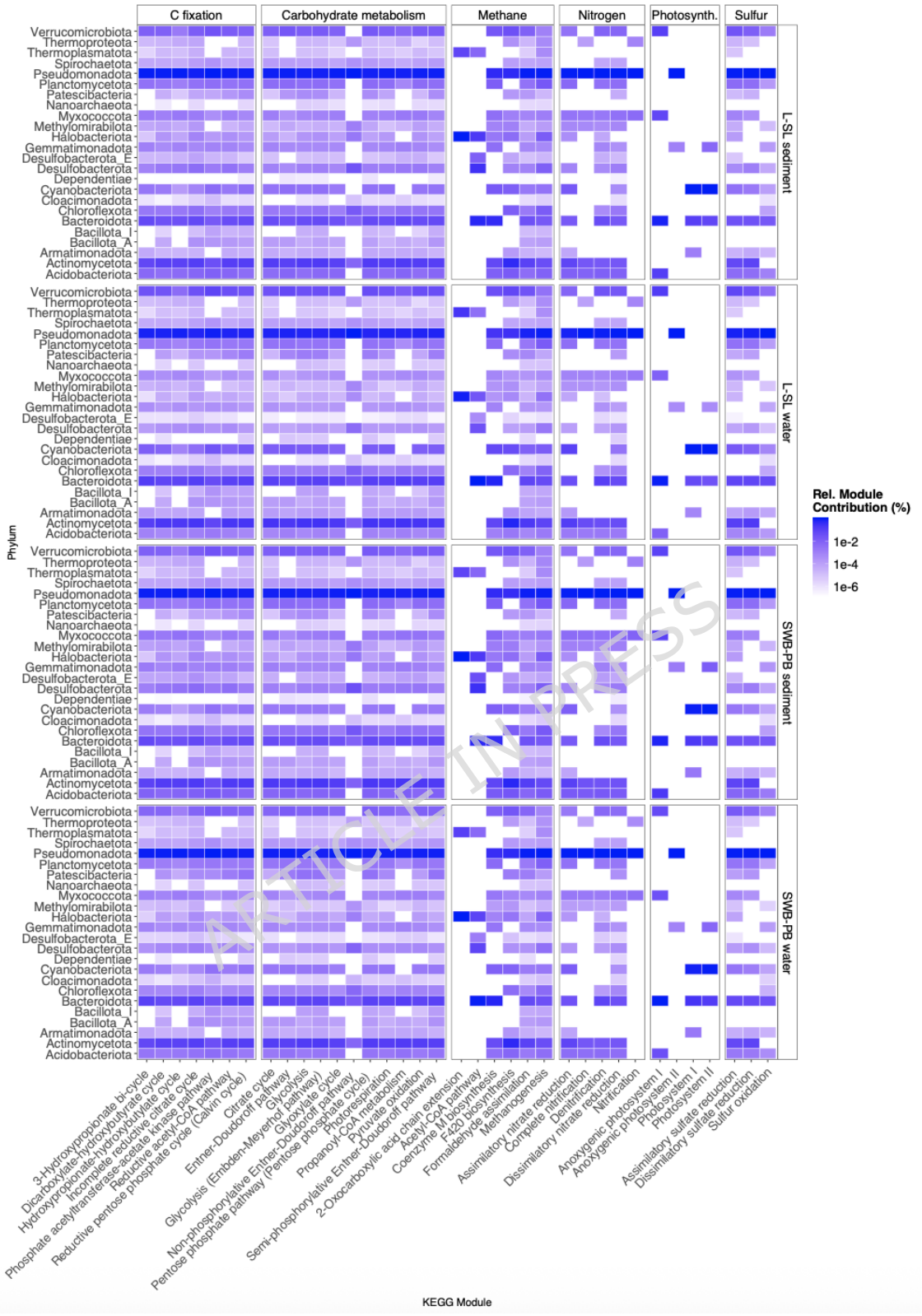


Fig. 6. Functional distribution (based on KEGG Modules) across the most abundant bacterial and archaeal phyla in the four studied environments. The heatmaps show the relationship between taxonomic composition and functional profiles relating to biogeochemical and photosynthetic metabolism. Functional profiles are based on the KEGG database (Modules). The KEGG module counts were weighted by multiplying the module copy number of each MAG (aggregated by phylum) by the mean abundance of phyla across samples

belonging to each of the four categories (L-SL sediment, L-SL water, SWB-PB sediment, SWB-PB water). The taxonomic abundance was estimated via Kraken2/Bracken. Percentage contribution was calculated separately for each module, considering the proportion of genes contributed to that function by individual phyla.The figure shows values between 0 and 1 for visualization purposes.

4. Discussion

4.1 Diversity and ecological roles of microbial communities

This study aimed to assess the diversity, taxonomic composition, and functional potential of microbial communities in alpine benthic and pelagic freshwater habitats across a gradient of water bodies varying in depth, size, and origin. The findings revealed distinct diversity patterns among bacterial and eukaryotic communities between sediment and water habitats, while archaeal diversity and community structure remained largely uniform. Both habitat type and water body characteristics, such as size and depth, significantly influenced microbial diversity and community composition. Sediments supported a diverse microbial community, which however appeared relatively homogeneous across the studied water bodies. Conversely, while communities in the water column displayed lower diversity, their composition varied greatly among water bodies. In small water bodies, benthic and pelagic communities tended to be more similar in composition and structure, while sediments and water in larger lakes supported clearly distinct assemblages. In particular, water habitats from small water bodies like peat bogs and ponds harbored higher taxonomic diversity compared to larger freshwater systems, featuring many low-abundant species.

Bacterial communities in pelagic habitats varied the most along the physiographic gradient. Notably, *Patescibacteria* were more abundant in water samples than in sediments, particularly in small water bodies. This superphylum is adapted to low-nutrient environments, exhibiting small genomes and cell sizes. Their limited metabolic capacity allows them to lead symbiotic lifestyles with different hosts, including eukaryotes and archaea [48, 49]. Waters showed higher proportions of *Bacillota I* compared to sediments, especially in large water bodies. Microbes from this phylum are known to be primary degraders, feeding on carbohydrates

627 (cellulose and chitin), proteins, nucleic acids and lipids. They simplify complex organic
628 molecules into simpler forms to be used by heterotrophic organisms, contributing significantly
629 to microbial food webs within aquatic ecosystems [50]. Additionally, *Bacteroidota*, known to
630 be involved in carbon cycling by decomposing complex polysaccharides [51], were more
631 abundant in waters compared to sediments in large water bodies but not small water bodies.
632 This phylum includes the family *Chlorobi*, mainly found in anoxic aquatic settings and known
633 for conducting anoxygenic photosynthesis, thereby contributing to primary production in
634 anoxic aquatic environments [52]. Moreover, large water bodies showed enrichment in
635 *Verrucomicrobiota* compared to small water bodies. *Verrucomicrobiota* are cosmopolitan
636 bacteria, as they have been ubiquitously found in soils and aquatic environments and have
637 been described to be polysaccharide degraders in freshwater habitats [53].

638 On the other hand, bacterial phyla such as *Nitrospirota*, *Myxococcota*, *Acidobacteria*, and
639 *Desulfobacterota* were enriched in sediment environments both in L-SLs and SWB-PBs. The
640 significantly higher abundance of these phyla in sedimentary environments can be largely
641 attributed to the anaerobic and nutrient-rich conditions characteristic of sediments. *Nitrospirota*
642 have been widely identified in hypoxic subsurface habitats and, while aerobic *Nitrospirota*
643 species exist, most species perform nitrogen and sulfur cycling under anaerobic conditions
644 [54]. *Myxococcota*, recognized for their predatory behavior, thrive in sediments rich in
645 microbial prey, thereby influencing nutrient turnover and organic matter decomposition [55].
646 *Acidobacteria*, abundantly and vastly present in a diversity of soil habitats, exhibit versatile
647 metabolic pathways, enabling them to adapt to various nutrient availability and soil acidities,
648 often flourishing in microaerophilic or anaerobic niches [56]. Members of the *Desulfobacterota*
649 are known for their role in sulfur cycling in anoxic environments such as sediments, utilizing
650 sulfate as a terminal electron acceptor in the absence of oxygen [57].

651 Additionally, *Methylomirabilota* and *Chloroflexota* were significantly higher in sediments than
652 waters in small water bodies, but not large water bodies. *Methylomirabilota*, anaerobic
653 methanotrophs, play an important role in greenhouse gas mitigation by sinking CH_4 [58],

654 highlighting one of the many ecosystem services provided by small water bodies.
655 *Chloroflexota* encompass diverse metabolic types, including anoxygenic photoautotrophs
656 such as the green nonsulfur bacteria *Chloroflexus* [59], making them highly adaptable to
657 sediment layers with light and oxygen gradients.

658 Archaeal communities did not differ significantly between pelagic and benthic habitats by their
659 richness but showed habitat-specific species enrichments, highlighting distinct ecological
660 roles. *Halobacteriota* were particularly enriched in sediments, while waters from small water
661 bodies showed higher abundances of this phylum compared to waters from larger lakes.
662 Within this phylum the predominant classes were *Methanosarcina* and *Metanobacteria*,
663 confirming their role in methanogenesis [60], as supported also by our functional analysis. In
664 addition, *Hadarcheota* and *Thermoplasmatota*, phyla that contribute to recycling of organic
665 matter, were enriched in the sediments of large and shallow lakes. Interestingly, *Hadarchaeota*
666 have been implicated in the anaerobic degradation of hydrocarbons, operating in syntropy with
667 methanogens [61]. In contrast, the waters of small water bodies showed a specific taxonomic
668 signature, with phyla from the DPANN superphylum (e.g. *Iainarchaeota* and
669 *Methanobacteriota,* amongst others), often characterized by small genomes and symbiotic
670 lifestyles [62]. These observations highlight the specialized ecosystemic roles that these
671 microorganisms play in biogeochemical processes across diverse freshwater ecosystems.

672 Lastly, eukaryotic communities showed unique compositions in each habitat, with generally
673 more abundant red and green algae in sediments, diatoms in waters and fungi uniquely spread
674 across the different environments. Nonetheless, diversity was particularly high in small water
675 bodies, which can be explained by their higher connectivity between adjacent pelagic, benthic,
676 littoral and riparian environments [10].

677 Overall, each habitat showed a unique taxonomic and functional composition of microbial
678 communities. Generally, biogeochemical and photosynthetic pathways were supported by a
679 wide array of phyla, amongst which the top contributors were *Pseudomonadota*,

Desulfobacterota, *Cyanobacteriota, Bacteroidota, Halobacteriota* and *Actinomycetota.* Water samples tended to be taxonomically and functionally more variable compared to benthic habitats. We found a clear functional separation between peat bogs and ponds and the rest of the analyzed habitats. While large deep and shallow lakes showed similar functional profiles to pasture ponds, peat bogs and ponds were characterized by a higher number of annotated genes, which indicates a higher functional richness. The higher genetic richness of these water bodies correlates with their lack of thermal stratification and generally higher volume of oxygens and nutrients, which provide a more suitable habitat for microbial life [63].

By comparing the relative abundance of each metabolic module in sediment and water, we observed that the distribution of metabolic pathways was coherent with the expected available substrates in water and sediment habitats. Sediment samples, which are typically anaerobic and rich in complex organic matter, supported pathways adapted to anaerobic conditions, revolving mostly around the cycling of nitrogen and methane, which are alternative electron acceptors to sustain energy production [64]. Interestingly, recent findings suggest that methanogenesis in freshwater habitats is also linked to pathways associated with chlorophyll metabolism and photosynthesis [65], justifying our finding of genes related to CH_4 production in cyanobacterial genomes. As our study revealed the presence of taxa from the families *Cyanobacteriaceae* and *Microcystaceae* unclassified at the species level, future work should focus on characterizing these species and understanding their role in recently proposed photosynthesis-driven CH_4 production pathways. By contrast, water samples, with higher oxygen levels, favored aerobic metabolic pathways such as glycolysis and the citrate cycle. For example, water column-residing methylothrophic species such as *Methylobacterium, Methylophilus, Methylococcus* (*Pseudomonadota*), and *Mycobacterium* (*Actinomycetota*) exhibited a high relative proportion of genes related to the assimilation of formaldehyde, which originates directly from the aerobic oxidation of one-carbon (C1) compounds. The prevalence of methylotrophic metabolic modules such as formaldehyde assimilation in water environments can be explained by the requirement for oxygen as a final electron acceptor [66]. Moreover, waters showed higher potential to perform assimilatory sulfate reduction, one

of the most dominant processes driving mineralization of organic matter in anoxic environments, such as deeper layers in stratified lakes [67]. Lastly, water environments, especially large and shallow lakes, harbored a significantly greater proportion of genes related to photosynthesis as well as anoxygenic photosynthesis of type II. This trend is consistent with the higher presence of aerobic photoautotrophic organisms in the waters of lakes compared to sediments, such as *Cyanobacteriota* in pelagic habitats. By contrast, the higher ratio of anoxygenic photosynthesis genes of type I in sediment habitats is consistent with the presence of green sulfur bacteria and purple photosynthetic bacteria such as the order *Rhodospirillales* (class *Alphaproteobacteria*) [68] in benthic habitats. An interesting finding in our study was the detection of *Armatimonadota* in the waters of shallow wetlands, including members of the family *Fimbriimonadaceae*. A recent study recovered *Armatimonadota* MAGs from sedimentary environments and revealed their involvement in carbon fixation, sulfur, and nitrogen cycling. Our finding aligns with expanding evidence that this phylum comprises globally distributed, understudied bacterial lineages with diverse ecological roles in oxic–anoxic interface habitats such as shallow wetland waters [69].

4.2 Technical limitations of the study

While metagenomics provides exceptional tools for the assembly of environmental genomes, the field is still facing technical hurdles reflecting the inherent challenges in assembling eDNA from soil-rich environments. MAG assembly generally benefits from a higher proportion of mapped reads in low-diversity samples, whereas the elevated microbial diversity in sediments necessitates deeper sequencing efforts. Sediments and soils contain substantial amounts of “relic DNA,” referring to extracellular DNA from dead microorganisms, along with repetitive genetic elements, both of which complicate assembly by introducing fragmented and incomplete sequences [70]. Increased Shannon entropy in these environments further exacerbates assembly difficulties, leading to a higher prevalence of incomplete and contaminated sequence bins. In our study, sediment-derived MAGs were numerically inferior to those from water samples. In particular, the L-SL group yielded a greater number of

assembled and representative MAGs, which is likely due to the lower diversity of upper water columns. In contrast, MAG recovery from sediments was proved to be more challenging, with higher contamination and lower completeness levels. The intrinsic obstacles of MAG recovery from soil samples could have been improved using multiple binning tools in combination with a bin-refinement tool, such as MAGScoT [71] or DAS Tool [72]. Relying on a single binning tool is known to be prone to introducing systemic biases, e.g., missing strain variations or producing chimeras, overall reducing the observed taxonomic diversity. Nonetheless, even when combining multiple binning tools, recovering high-quality genomes from a complex and heterogeneous matrix such as soil or sediment remains challenging to date [73].

5. Conclusions

Our study provides the taxonomic identities and functional evidence of the metabolic potential of microorganisms residing in alpine freshwater bodies of different sizes and depths and their relative sediment habitats. Our analysis of biogeochemical cycling and photosynthetic potential demonstrates that biogeochemical processes are mostly shaped by habitat matrix, likely as a direct consequence of differing oxygen and other nutrient levels, and that shallow water-logged habitats such as peat bogs and ponds share taxonomic diversity and functional patterns with those found in sediments. The rare and protected habitats analyzed in this study were found to harbor a remarkable yet unexplored bacterial and archaeal genomic diversity. Most assembled genomes could not be classified into known species, revealing the importance of exploring, in a non-invasive way, biodiversity-rich protected habitats such as nature reserves.

6. References

1. Delgado-Baquerizo M, Maestre FT, Reich PB, et al. Microbial diversity drives multifunctionality in terrestrial ecosystems. Nat Commun. 2016;7:10541. http://dx.doi.org/10.1038/ncomms10541

2. Grizzetti B, Poikane S. The importance of inland waters. In: Wetzel's Limnology (Fourth Edition). Elsevier; 2024:7–13.
3. Biggs J, von Fumetti S, Kelly-Quinn M. The importance of small water bodies for biodiversity and ecosystem services: implications for policy makers. Hydrobiologia. 2017;793:3–39. http://dx.doi.org/10.1007/s10750-016-3007-0
4. Fait P, Demierre E, Ilg C, Oertli B. Small mountain reservoirs in the Alps: new habitats for alpine freshwater biodiversity? Aquatic Conservation. 2020;30:617–630. https://doi.org/10.1002/aqc.3306
5. Page SE, Baird AJ. Peatlands and global change: response and resilience. Annu Rev Environ Resour. 2016;41:35–57. http://dx.doi.org/10.1146/annurev-environ-110615-085520
6. Hahn MW. The microbial diversity of inland waters. Curr Opin Biotechnol. 2006;17:256–261. https://doi.org/10.1016/j.copbio.2006.05.006
7. Wetzel RG. Limnology: Lake and River Ecosystems. Gulf Professional Publishing; 2001.
8. Vettorazzo S, Boscaini A, Cervasino L, Salmaso N. From small water bodies to lakes: exploring the diversity of freshwater bacteria in an Alpine Biosphere Reserve. Sci Total Environ. 2024;954:160081. https://doi.org/10.1016/j.scitotenv.2024.176495
9. Dušek J, Dařenová E, Pavelka M, Marek MV. Methane and carbon dioxide release from wetland ecosystems. In: Lal R, ed. Climate Change and Soil Interactions. Elsevier; 2020:509–553. https://doi.org/10.1016/B978-0-12-818032-7.00019-9
10. Meerhoff M, Beklioğlu M. Shallow lakes and ponds. In: Wetzel's Limnology. Elsevier; 2024:859–892. https://doi.org/10.1016/B978-0-12-822701-5.00026-4
11. Grossart HP, Massana R, McMahon KD, Walsh DA. Linking metagenomics to aquatic microbial ecology and biogeochemical cycles. Limnol Oceanogr. 2020;65. http://dx.doi.org/10.1002/lno.11382

12. Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. Science. 2008;320:1034–1039. http://dx.doi.org/10.1126/science.1153213
13. UNESCO. Ledro Alps and Judicaria Biosphere Reserve. Man and the Biosphere Programme. Available at: https://www.unesco.org/en/mab/ledro-alps-and-judicaria (Accessed: 18 May 2025).
14. MetaShot. MetaShot: Curated Docker Images and Nextflow Workflows for Metagenomics and Microbiome Genomics. Available at: https://metashot.github.io/ (Accessed: 18 May 2025).
15. Ewels PA, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020;38:276–278. http://dx.doi.org/10.1038/s41587-020-0439-x
16. MetaShot. metashot/kraken2: Taxonomic Classification of Reads and Abundance Estimation of Species in Metagenomic Samples. GitHub repository. Available at: https://github.com/metashot/kraken2 (Accessed: 18 May 2025).
17. nf-core. nf-core/taxprofiler: Highly Parallelised Multi-Taxonomic Profiling of Shotgun Short- and Long-Read Metagenomic Data. GitHub repository. Available at: https://github.com/nf-core/taxprofiler (Accessed: 18 May 2025).
18. Joint Genome Institute. BBDuk Guide: Quality Trimming and Filtering of NGS Reads. Available at: https://archive.jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbduk-guide/ (Accessed: 18 May 2025).
19. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20:257. http://dx.doi.org/10.1186/s13059-019-1891-0
20. Langmead B. Kraken 2 and Bracken Indexes. AWS Public Datasets. Available at: https://benlangmead.github.io/aws-indexes/k2 (Accessed: 18 May 2025).
21. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. PeerJ Comput Sci. 2017;3:e104. http://dx.doi.org/10.7717/peerj-cs.104

22. MetaShot. metashot/mag-illumina: Workflow for Metagenome-Assembled Genomes from Illumina Data. GitHub repository. Available at: https://github.com/metashot/mag-illumina (Accessed: 18 May 2025).
23. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–834. http://dx.doi.org/10.1101/gr.213959.116
24. Kang DD, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359. http://dx.doi.org/10.7717/peerj.7359
25. Albanese D and Donati C. Large-scale quality assessment of prokaryotic genomes with metashot/prok-quality [version 1; peer review: 2 approved]. F1000Research 2021, 10:822. https://doi.org/10.12688/f1000research.54418.1
26. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–1055. http://dx.doi.org/10.1101/gr.186072.114
27. Orakov A, et al. GUNC: detection of chimerism and contamination in prokaryotic genomes. Genome Biol. 2021;22:178. http://dx.doi.org/10.1186/s13059-021-02393-0
28. Seemann T. tseemann/barrnap: Rapid Ribosomal RNA Prediction. GitHub repository. Available at: https://github.com/tseemann/barrnap (Accessed: 18 May 2025).
29. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. In: Kollmar M, ed. Gene Prediction. Vol. 1962. Springer New York; 2019:1–14. http://dx.doi.org/10.1007/978-1-4939-9173-0_1
30. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11:2864–2868. http://dx.doi.org/10.1038/ismej.2017.126

31. Computational Metagenomics Unit at Fondazione Edmund Mach. compmetagen/prok-classify: Taxonomic Classification of Prokaryotic Genomes. GitHub repository. Available at: https://github.com/compmetagen/prok-classify (Accessed: 18 May 2025).
32. MetaShot. metashot/prok-classify: Workflow for Prokaryotic Genome Classification. GitHub repository. Available at: https://github.com/metashot/prok-classify (Accessed: 18 May 2025).
33. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26:1641–1650. http://dx.doi.org/10.1093/molbev/msp077
34. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5:e9490. http://dx.doi.org/10.1371/journal.pone.0009490
35. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–359. https://doi.org/10.1038/nmeth.1923
36. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools, *GigaScience*, 2021; 10:2, giab008, https://doi.org/10.1093/gigascience/giab008.
37. MetaShot. metashot/containment: Sequence Containment Estimation Tools. GitHub repository. Available at: https://github.com/metashot/containment (Accessed: 18 May 2025).
38. Ondov BD, et al. Mash Screen: high-throughput sequence containment estimation for genome discovery. Genome Biol. 2019;20:232. http://dx.doi.org/10.1186/s13059-019-1841-x
39. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2021). https://www.R-project.org/.

40. Bokulich NA, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 2013;10:57–59. http://dx.doi.org/10.1038/nmeth.2276
41. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. Nat Commun. 2020;11:3514. http://dx.doi.org/10.1038/s41467-020-17041-7
42. MetaShot. metashot/prok-annotate v3.0.1: Workflow for Prokaryotic Genome Annotation. GitHub repository. Available at: https://github.com/metashot/prok-annotate (Accessed: 18 May 2025).
43. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153
44. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38:5825–5829. http://dx.doi.org/10.1093/molbev/msab293
45. Huerta-Cepas J, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309–D314. http://dx.doi.org/10.1093/nar/gky1085
46. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457–D462. http://dx.doi.org/10.1093/nar/gkv1070
47. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Available at: https://www.kegg.jp/ (Accessed: 18 May 2025).
48. Tian R, et al. Small and mighty: adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. Microbiome. 2020;8:51. https://doi.org/10.1186/s40168-020-00825-w

49. Hu H, et al. Global abundance patterns, diversity, and ecology of Patescibacteria in wastewater treatment plants. Microbiome. 2024;12:55. https://doi.org/10.1186/s40168-024-01769-1
50. Yadav S, et al. Organic matter degradation in the deep, sulfidic waters of the Black Sea: insights into the ecophysiology of novel anaerobic bacteria. Microbiome. 2024;12:98. https://doi.org/10.1186/s40168-024-01816-x
51. Thomas F, Hehemann JH, Rebuffet E, Czjzek M, Michel G. Environmental and Gut Bacteroidetes: The Food Connection. Front Microbiol. 2011;2. http://dx.doi.org/10.3389/fmicb.2011.00093
52. Bryant DA, Frigaard NU. Prokaryotic photosynthesis and phototrophy illuminated. Trends Microbiol. 2006;14(11):488–496. http://dx.doi.org/10.1016/j.tim.2006.09.001
53. He S, et al. Ecophysiology of Freshwater Verrucomicrobia Inferred from Metagenome-Assembled Genomes. mSphere. 2017;2:e00277-17. http://dx.doi.org/10.1128/mSphere.00277-17
54. Mosley OE, Gios E, Handley KM. Implications for nitrogen and sulphur cycles: phylogeny and niche-range of Nitrospirota in terrestrial aquifers. ISME Commun. 2024;4:ycae047. https://doi.org/10.1093/ismeco/ycae047
55. Zou D, Zhang C, Liu Y, et al. Biogeographical distribution and community assembly of Myxococcota in mangrove sediments. Environ Microbiome. 2024;19:47. https://doi.org/10.1186/s40793-024-00593-2
56. Kielak AM, Barreto CC, Kowalchuk GA, Van Veen JA, Kuramae EE. The Ecology of Acidobacteria: Moving beyond Genes and Genomes. Front Microbiol. 2016;7. http://dx.doi.org/10.3389/fmicb.2016.00744
57. Dyksma S, Pester M. Growth of sulfate-reducing Desulfobacterota and Bacillota at periodic oxygen stress of 50% air-O_2 saturation. Microbiome. 2024;12:191. https://doi.org/10.1186/s40168-024-01909-7
58. Roldán DM, Carrizo D, Sánchez-García L, Menes RJ. Diversity and Effect of Increasing Temperature on the Activity of Methanotrophs in Sediments of Fildes

Peninsula Freshwater Lakes, King George Island, Antarctica. Front Microbiol. 2022;13:822552. http://dx.doi.org/10.3389/fmicb.2022.822552

59. Brock TD. The Genus Chloroflexus. In: Thermophilic Microorganisms and Life at High Temperatures. Springer, New York, NY; 1978:180–216. http://dx.doi.org/10.1007/978-1-4612-6284-8_7
60. Zhou J, Holmes DE, Tang H-Y, Lovley DR. Correlation of Key Physiological Properties of Methanosarcina Isolates with Environment of Origin. Appl Environ Microbiol. 2021;87:e00731-21. http://dx.doi.org/10.1128/AEM.00731-21
61. Yu T, et al. Thermophilic Hadarchaeota grow on long-chain alkanes in syntrophy with methanogens. Nat Commun. 2024;15:6560. https://doi.org/10.1038/s41467-024-50883-z
62. Qi Y-L, Zhang H-T, Li M, Li W-J, Hua Z-S. Recovery of nearly 3,000 archaeal genomes from 152 terrestrial geothermal spring metagenomes. Sci Data. 2025;12:151. https://doi.org/10.1038/s41597-025-04493-z
63. Lew S. The effects of environmental parameters on the microbial activity in peat-bog lakes. PLoS ONE. 2019;14(9):e0223134. https://doi.org/10.1371/journal.pone.0224441
64. Sigee DC. Freshwater Microbiology: Biodiversity and Dynamic Interactions of Microorganisms in the Aquatic Environment. Microbial Diversity and Freshwater Ecosystems. Wiley; 2004:1–46. http://dx.doi.org/10.1002/0470011254
65. Perez-Coronel, E., Michael Beman, J. Multiple sources of aerobic methane production in aquatic ecosystems include bacterial photosynthesis. *Nat Commun.* 2022; 13, 6454. https://doi.org/10.1038/s41467-022-34105-y
66. Klein VJ, Irla M, Gil López M, Brautaset T, Fernandes Brito L. Unravelling Formaldehyde Metabolism in Bacteria: Road towards Synthetic Methylotrophy. *Microorganisms*. 2022;10(2):220. https://doi.org/10.3390/microorganisms10020220
67. Diao M, Dyksma S, Koeksoy E, Ngugi DK, Anantharaman K, Loy A, Pester M. Global diversity and inferred ecophysiology of microorganisms with the potential for

947 dissimilatory sulfate/sulfite reduction. *FEMS Microbiol Rev*. 2023;47(5):fuad058.
948 https://doi.org/10.1093/femsre/fuad058
949 68. Asif A, Mohsin H, Rehman Y. Purple nonsulfur bacteria: An important versatile tool in
950 biotechnology. In: *Recent Advancement in Microbial Biotechnology*. Elsevier;
951 2021:309–337. http://dx.doi.org/10.1016/B978-0-12-822098-6.00003-3
952 69. Carlton, J.D., Langwig, M.V., Gong, X. *et al.* Expansion of Armatimonadota through
953 marine sediment sequencing describes two classes with unique ecological roles. *ISME*
954 *COMMUN.* 2023; 3, 64. https://doi.org/10.1038/s43705-023-00269-x
955 70. Anthony WE, Allison SD, Broderick CM, et al. From soil to sequence: filling the critical
956 gap in genome-resolved metagenomics is essential to the future of soil microbial
957 ecology. *Environmental Microbiome*. 2024;19:56. https://doi.org/10.1186/s40793-024-
958 00599-w

959 71. Rühlemann MC, Wacker EM, Ellinghaus D, Franke A. MAGScoT: a fast, lightweight
960 and accurate bin-refinement tool. *Bioinformatics.* 2022; 38, 5430–5433.
961 https://doi.org/10.1093/bioinformatics/btac694

962 72. Sieber CMK, Probst AJ, Sharrar A. *et al.* Recovery of genomes from metagenomes via
963 a dereplication, aggregation and scoring strategy. *Nat Microbiol.* 2018; 3, 836–843.
964 https://doi.org/10.1038/s41564-018-0171-1

965 73. Kim J, Kim N, Cha JH. *et al.* Comprehensive benchmarking of metagenomic binning
966 tools reveals key factors for improved genome recovery. *Nat Commun* 2026; 17, 3467.
967 https://doi.org/10.1038/s41467-026-71521-w

7. Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Sequences were deposited to the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB86943 (https://www.ebi.ac.uk/ena/browser/view/PRJEB86943). The sequences of the metagenome-assembled genomes, the relative datasets with their taxonomic classification and other metadata have been deposited in Zenodo with the DOI: 10.5281/zenodo.15111114.

The datasets supporting the taxonomic and functional analyses and the conclusions of this article are included within the article and its additional files.

Code Availability

The code to reproduce the analyses can be found in the following GitHub repositories: MetaShot metagenomic suite (https://metashot.github.io/); compmetagen/prok-classify (https://github.com/compmetagen/prok-classify); R code for the taxonomic, functional, ANCOM-BC and MAG analyses (https://github.com/anitagat/freshwater_metagenomics).

Competing interests

The authors declare that they have no competing interests.

Funding

The research was financially supported by Alpi Ledrensi e Judicaria, Riserva di Biosfera UNESCO, Programma sull'uomo e la biosfera (MAB). The analysis of the data was carried out in the framework of the National Biodiversity Future Centre (NBFC) funded by the Italian Ministry of University and Research and received funding from the European Union Next-GenerationEU (PNRR, Mission 4 Component 2, Investment 1.4, D.D. 1034 17/06/2022, Project CN00000033). This presentation reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Authors' contributions

AG: Formal analysis, Methodology, Data curation, Writing - original draft, Writing - Review & Editing. SV: Conceptualization, Methodology, Review & Editing. PM: Methodology, Writing – Review & Editing. SL: Review & Editing. NS: Conceptualization, Study Design, Funding Acquisition, Writing – Review & Editing. MP: Sequencing, Methodology. CD: Conceptualization, Methodology, Writing – Review & Editing.

Acknowledgements

The authors want to thank the staff of MUSE - Museo delle Scienze (Trento) and of Agenzia Provinciale per la Protezione dell'Ambiente (APPA, Trento) for the support in field sampling, and the staff of the Sequencing Platforms at FEM, for the water chemistry and sequencing analysis.

Supplemental information

Document S1: Figures S1–S11
Table S1: Samples metadata, with variables describing habitats' characteristics and sampling date.
Table S2: KEGG KO abundance table, generated with EggNOG-mapper. Within each sample, the abundance of a KEGG Ortholog is recorded as the frequency of assignment of such ortholog to a gene.

Table S3: KEGG Module abundance table, generated with EggNOG-mapper. Within each sample, the abundance of a KEGG Module is recorded as the frequency of assignment of such module to a gene annotation.
Table S4: Selected KEGG Modules and KEGG KOs included in the functional analysis.
Table S5: Library size per sample: reads number, scaffold length and contig number.
Table S6: Functional analysis of unclassified taxa. Contribution is calculated as the number of genes found in each taxon/MAG weighted by the mean taxonomic abundance of that taxon in each of the four environments.
Table S7: Taxonomic abundance table of bacterial species, generated with Bracken-Kraken2 (Rarefied).
Table S8: Differential abundance table of bacterial species with p.adj values and log2-fold changes, generated with ANCOMBC.
Table S9: Taxonomic abundance table of archaeal species, generated with Bracken-Kraken2 (Rarefied)
Table S10: Differential abundance table of archaeal species with p.adj values and log2-fold changes, generated with ANCOMBC.
Table S11: Taxonomic abundance table of eukaryotic species, generated with Bracken-Kraken2 (Rarefied).
Table S12: Differential abundance table of eukaryotic species with p.adj values and log2-fold changes, generated with ANCOMBC**.**
Table S13: Taxonomic annotation of bacterial metagenome-assembled metagenomes (MAGs) and MAG quality data.
Table S14: Taxonomic annotation of archaeal metagenome-assembled metagenomes (MAGs) and MAG quality data.
Table S15: Number and percentage of original shotgun reads mapped to MAG sequences per sample.
Table S16: Containment analysis with Mash screen. Values represent the number of matches between MAGs and sample reads.
Table S17: Taxonomic abundance table of bacterial, archaeal and eukaryotic species generated with Bracken-Kraken2. Not rarefied.
Table S18: Taxonomic abundance table of bacterial, archaeal and eukaryotic species, where the names of Phyla containing clade information have been aggregated to facilitate comparison with NCBI nomenclature.
Table S19: Environmental data, reporting information on the sampled sites.
Table S20: KEGG Modules assigned within each MAG.
Table S21: KEGG KOs assigned within each MAG.

1050

Metabolism	Module	Module name	p-value	Adj. p-value	Significance	Matrix
Carbohydrate metabolism	**M00001**	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	0.000004	0.00002	***	Water
Carbohydrate metabolism	**M00002**	Glycolysis, core module involving three-carbon compounds	0.000004	0.00002	***	Water
Carbohydrate metabolism	**M00307**	Pyruvate oxidation, pyruvate => acetyl-CoA	0.000004	0.00002	***	Sediment
Carbohydrate metabolism	**M00010**	Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate	0.001171	0.00218	***	Water
Carbohydrate metabolism	**M00004**	Pentose phosphate pathway (Pentose phosphate cycle)	0.000038	0.00011	***	Water
Carbohydrate metabolism	**M00308**	Semi-phosphorylative Entner-Doudoroff pathway, gluconate => glycerate-3P	0.000004	0.00002	***	Water
Carbohydrate metabolism	**M00309**	Non-phosphorylative Entner-Doudoroff pathway, gluconate/galactonate => glycerate	0.000053	0.00015	***	Sediment
Carbohydrate metabolism	**M00012**	Glyoxylate cycle	0.014069	0.01960	**	Water
Carbohydrate metabolism	**M00532**	Photorespiration	0.000126	0.00030	***	Water
Carbohydrate metabolism	**M00741**	Propanoyl-CoA metabolism, propanoyl-CoA => succinyl-CoA	0.000645	0.00123	***	Sediment
Carbon fixation	**M00165**	Reductive pentose phosphate cycle (Calvin cycle)	0.000004	0.00003	***	Water
Carbon fixation	**M00376**	3-Hydroxypropionate bi-cycle	0.001171	0.00202	***	Water
Carbon fixation	**M00375**	Hydroxypropionate-hydroxybutylate cycle	0.005329	0.00723	***	Sediment
Carbon fixation	**M00374**	Dicarboxylate-hydroxybutyrate cycle	0.005329	0.00723	***	Water
Carbon fixation	**M00377**	Reductive acetyl-CoA pathway	0.018082	0.02290	*	Sediment
Carbon fixation	**M00579**	Phosphate acetyltransferase-acetate kinase pathway, acetyl-CoA => acetate	0.001171	0.00202	***	Sediment
Carbon fixation	**M00620**	Incomplete reductive citrate cycle, acetyl-CoA => oxoglutarate	0.000011	0.00006	***	Sediment
Methane metabolism	**M00567**	Methanogenesis, CO2 => methane	0.000027	0.00013	***	Sediment
Methane metabolism	**M00357**	Methanogenesis, acetate => methane	0.000164	0.00062	***	Sediment
Methane metabolism	**M00356**	Methanogenesis, methanol => methane	0.000210	0.00072	***	Sediment
Methane metabolism	**M00563**	Methanogenesis, methylamine/dimethylamine/trimethylamine => methane	0.000126	0.00053	***	Sediment
Methane	**M00358**	Coenzyme M biosynthesis	0.00532	0.0072	***	Sediment

metabolism			9	3		
Methane metabolism	**M00608**	2-Oxocarboxylic acid chain extension, 2-oxoglutarate => 2-oxoadipate => 2-oxopimelate => 2-oxosuberate	0.000523	0.0132	***	Sediment
Methane metabolism	**M00346**	Formaldehyde assimilation, serine pathway	0.000004	0.00003	***	Water
Methane metabolism	**M00345**	Formaldehyde assimilation, ribulose monophosphate pathway	0.000336	0.0098	***	Water
Methane metabolism	**M00344**	Formaldehyde assimilation, xylulose monophosphate pathway	0.000004	0.00003	***	Water
Methane metabolism	**M00378**	F420 biosynthesis, archaea	0.008232	0.01079	**	Sediment
Methane metabolism	**M00422**	Acetyl-CoA pathway, CO2 => acetyl-CoA	0.000645	0.0144	***	Sediment
Nitrogen metabolism	**M00530**	Dissimilatory nitrate reduction, nitrate => ammonia	0.000008	0.00005	***	Sediment
Nitrogen metabolism	**M00529**	Denitrification, nitrate => nitroge	0.000004	0.00003	***	Sediment
Nitrogen metabolism	**M00528**	Nitrification, ammonia => nitrite	0.000965	0.0183	***	Sediment
Nitrogen metabolism	**M00804**	Complete nitrification, comammox, ammonia => nitrite => nitrate	0.000004	0.00003	***	Sediment
Sulfur metabolism	**M00176**	Assimilatory sulfate reduction, sulfate => H2S	0.000790	0.0167	***	Water
Photosynthesis	**M00161**	Photosystem II	0.000645	0.0144	***	Water
Photosynthesis	**M00163**	Photosystem I	0.003918	0.00595	**	Water
Photosynthesis	**M00597**	Anoxygenic photosystem II	0.003342	0.0529	***	Water
Photosynthesis	**M00598**	Anoxygenic photosystem I	0.001477	0.0244	***	Sediment
Metabolism	**Module**	**Module name**	**p-value**	**Adj. p-value**	**Significance**	**Physiographic group**
Nitrogen metabolism	**M00531**	Assimilatory nitrate reduction, nitrate => ammonia	0.003852	0.04293	*	SWB-PBs
Carbohydrate metabolism	**M00008**	Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate	0.000785	0.01530	*	SWB-PBs
Photosynthesis	**M00161**	Photosystem II	0.002149	0.02619	*	L-SLs
Photosynthesis	**M00163**	Photosystem I	0.000610	0.01530	*	L-SLs
Photosynthesis	**M00597**	Anoxynogenic photosystem II	0.001430	0.02231	*	L-SLs

Table 1. Comparison of the functional potential of sediment and water (matrix) and the physiographic groups L-SLs and SWB-PBs. Comparisons based on the relative proportions of genes dedicated to nutrient cycling (Carbohydrate-, Carbon-, Nitrogen-, Sulfur-

and CH_4-related metabolism) and photosynthesis found in the metagenomes of these habitats. Measured by the relative number of genes (RPM) assigned to the KEGG Modules belonging to 'Carbohydrate metabolism' and 'Energy metabolism' relative to the total number of modules assigned for each sample. The 'matrix' and 'physiographic group' columns represent where the given Module was found significantly more abundant. The statistical significance of the differences found between environments (water vs sediment and L-SLs vs SWB-PBs) were tested using the non-parametric Wilcoxon signed-rank test (paired) or the Wilcoxon Rank-Sum test and P-values were adjusted using the Benjamini-Hochberg method. Legend: * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$.