Environmental Microbiome

RESEARCH

Open Access

Genome-resolved metagenomics and evolutionary analysis reveal conserved metabolic adaptations in extremophile communities from a copper mining tailing

Moises A. Rojas[1,2], Gladis Serrano[1,3], Jorge Torres[1,3], Jaime Ortega[1,3], Gabriel Gálvez[1,3], Emilio Vilches[1,4], Valentina Parra[1,5,6,8], Angélica Reyes-Jara[1,7], Vinicius Maracaja-Coutinho[1,6,8,9], Lorena Pizarro[1,10], Mauricio Latorre[1,3,4,11*] and Alex Di Genova[1,2,4*]

Abstract

Background Microbial communities in mining environments exhibit unique metabolic adaptations to extreme conditions, such as high metal concentrations and low pH. Their relatively low species complexity makes them an attractive model for fine-scale evolutionary analysis; nonetheless, genome-resolved metagenomic data from these environments are still scarce. Here, we employed genome-resolved metagenomics to analyze a high-quality Illumina-sequenced sample from the Cauquenes copper tailing in central Chile, one of the world's largest and oldest copper waste deposits. We aimed to uncover the taxonomic composition, metabolic potential, and evolutionary pressures shaping this extremophile community.

Results We reconstructed 44 medium- and high-quality metagenome-assembled genomes (MAGs), predominantly from the phyla *Actinomycetota*, *Pseudomonadota*, and *Acidobacteriota*. Taxonomic analysis revealed limited species-level classification, with only five MAGs assigned to known species, highlighting the challenges of characterizing extreme environments. Functional profiling identified enhanced metabolic capabilities in sulfur and copper pathways, critical for survival in mining ecosystems. Using evolutionary analysis on mining MAGs using dN/dS ratios, we uncoverd strong negative selection on genes involved in sulfur, copper, and iron metabolism, indicative of a conservative evolutionary state. In contrast, genes under positive selection were linked to motility, biofilm formation, and stress resistance, suggesting adaptive mechanisms for resource acquisition and survival.

Conclusions Our study provides a metagenome-wide evolutionary analysis of mining MAGs, demonstrating that microbial communities in copper tailings are highly specialized, with conserved metabolic pathways under strong purifying selection. At the same time, the recovery of previously unclassified species of extremophiles expands the

*Correspondence:
Mauricio Latorre
mauricio.latorre@uoh.cl
Alex Di Genova
alex.digenova@uoh.cl

Full list of author information is available at the end of the article

known biodiversity of mining ecosystems. These findings emphasise the challenges of leveraging these communities for biotechnological applications, such as biomining, due to their evolutionary constraints.

Keywords Metagenomics, MAGs, Extremophiles, Copper mining tailing, Mining metabolism, Selection pressure

Introduction

Metagenomics enables the study of genetic information and interactions within complex microbial communities in specific environments. Additionally, it enables the reconstruction of "metagenome-assembled genomes" (MAGs), in-silico genomes that are hard to obtain through the conventional isolation of microorganisms [1]. In recent years, metagenomics has been useful to map the complexity of heterogeneous microbial communities inhabiting extreme environments and their adaptations across various ecosystems characterized by low pH, wide range of temperatures, hypersalinity or high heavy metal concentrations [2–5]. Metagenomic analysis can inform taxonomic composition, metabolic-biogeochemical cycling functionalities, and selective pressure analysis of genetic evolution within these communities [6, 7].

Soil ecosystems are among the most diverse microbial habitats on Earth, harboring a vast array of metabolic capacities and genetic resources [8]. Despite high-throughput experiments, a significant portion of soil microbial diversity remains unexplored, particularly at finer taxonomic levels such as genus and species [9, 10]. This is particularly evident in extreme environments, where harsh conditions limit diversity and favor highly specialized taxa [11]. Mining ecosystems are characterized by extreme acidity, high metal concentrations, and relatively low microbial diversity [12]. These environments are dominated by bacterial communities belonging to phyla such as *Acidobacteriota*, *Actinomycetota* (synonym *Actinobacteriota*), and *Pseudomonadota* (former *Proteobacteria* phylum), which play specialized roles in biogeochemical cycles, particularly in iron and sulfur cycling, driven by unique genetic adaptations [13–15]. For example, acidophilic prokaryotes, such as *Acidithiobacillus thiooxidans*, have evolved specialized molecular pathways to tolerate heavy metals and acidic conditions, making them a model organism in mining microbiology [16–18]. Similarly, other acidophilic genera, including *Leptospirillum*, *Acidimicrobium*, *Ferromicrobium*, and *Sulfobacillus*, exhibit comparable capabilities and biotechnological potential [19, 20]. These microorganisms are particularly valued for their efficient copper bioleaching capabilities, a process especially critical to the Chilean mining industry [21–23]. Beyond bioleaching, they are increasingly being explored for broader applications, such as the bioremediation of heavy metals and pollutants in water and soils [24–26].

Over the past two decades, a series of seminal studies has built a solid foundation for understanding evolution in acid-mine-drainage (AMD) microbiomes [27–33]. The first genome-resolved comparison, between the isolate Ferroplasma acidarmanus (fer1) and its environmental counterpart [27], demonstrated that core genes remain under pervasive purifying selection. A curated community assembly for Leptospirillum Group II [28] extended these insights by revealing hypervariable plasmid/phage "islands" and confirming that strain variation is generally purged by selection. Time-series sampling of the same biofilm [29] then provided the field's first in-situ point-mutation rate estimates and showed that large homologous-recombination tracts can generate ecologically successful hybrids within years. Community-level relative-evolutionary-rate surveys across 40 metagenomes spanning six habitats [30] further established that AMD communities evolve faster than those in soils or oceans, attributing this to relaxed constraint in extreme environments. Subsequent work catalogued dozens of horizontal-gene-transfer (HGT) events among co-existing AMD taxa [31] and provided clues for virus-driven gene-content changes in the Richmond Mine system [32]. Most recently, comparative genomics of 16 *Acidithiobacillus ferrooxidans* isolates [33] identified a focused set of metal-resistance orthologues under recurrent positive selection, showing the adaptive potential of key biomining species.

Collectively, these studies have demonstrated the roles of purifying selection, recombination, HGT, and mobile elements in shaping AMD genomes. Nonetheless, they share common limitations: analyses often centred on one or two key species; few-gene markers; selection signals were obtained from orthologues rather than population-level SNPs; metagenome-assembled genomes (MAGs) were rarely used; and site replication remained scarce. Addressing these gaps requires genome-resolved, SNP-based selection tests across diverse taxa and multiple sites. Finally, the extent to which the selection patterns observed in iconic AMD systems and representative microorganisms generalise to geographically distinct mining sites or other less abundant community members has yet to be tested.

Here, we generated and analysed metagenomic data from sample S15 taken in the 12.5 km^2 Cauquenes copper tailings (in operation since 1936) in central Chile—an extreme habitat with millimolar copper, iron, and zinc and pH ≈ 4.0 (Table S1) [34]. Assembly and binning produced 44 medium/high-quality MAGs, giving genome-level insight that exceeds recent 16 S-based or pathway-specific surveys [35–37]. When benchmarked

against the 32,931-genome Soil MAGs catalogue [38], only 5 MAGs corresponded to known species-level bins and 31 to known genera, highlighting substantial novelty. Gene-dosage profiles showed marked enrichment of copper- and sulphur-oxidation pathways relative to non-mining soils [38]. Using 146,932 high-confidence SNPs and contamination-corrected dN/dS estimates for 39,438 genes, we found strong purifying selection (95.2% of genes dN/dS < 1) but identified a small set of positively selected genes in biomining functions. Applying the same pipeline to seven additional sites of Cauquenes tailing (250 dereplicated MAGs) revealed the same patterns, reinforcing the stability of evolutionary constraints in this environment.

Materials and methods

Sample collection and DNA extraction

Three replicated soil samples (100 g each) were collected in July 2022 (Chilean winter) under sterile conditions at 5 to 10 cm of the soil surface in a particular zone of the Cauquenes copper tailing belonging to the intermediate site and labeled as S15 sample (pH ~ 4.0, high copper concentration). The replicates were homogenized in situ before being stored on dry ice for further microbial analyses. DNA was extracted from the collected samples (each of the three soil samples) using the Zymo kit Quick-DNA Fecal/Soil Microbe MiniPrep™, combining the manufacturer's instructions and pre-treatment of the samples before following the kit protocol. 20 g of soil were resuspended in 20 ml PBS buffer. Then, the sample was homogenized using FastPrep-24 equipment (MP Biomedicals Cat. Number 6004500) in a cycle of 4 m/s for 20 s. The mixture was centrifuged (120 × g for 5 min) at room temperature and the supernatant fluid was transferred to a clean tube, and centrifuged (7,100 × g for 10 min) at room temperature, then the supernatant was discarded and the pellet was resuspended in 1 ml of PBS buffer. After, the samples were vortexed for 10 s, and the mixture was transferred into a DNA Fecal/Soil Microbe lysis tube to continue the kit protocol. Concentrations of DNA were evaluated with a Qubit 4 fluorometer.

Metagenome sequencing

Shotgun metagenomic sequencing was performed on DNA extracted from the S15 sample using the TruSeq™ DNA PCR-free Sample Prep Kit (Illumina, USA) following the manufacturer's instructions. Sequencing was performed on an Illumina HiSeq™ 2000 platform, generating a total of 103,790,393 paired-end raw reads (2 × 150 bp, Table S2).

Metagenomic assembly, binning, quality assessment, and classification

Raw reads were assembled, binned and annotated using the nf-core/mag v2.2.1 pipeline [39] available for the workflow management system Nextflow [40]. First, fastp v0.23.2 was used to perform the short reads trimming to remove adapters and low quality reads [41]. Quality control of the trimmed reads was assessed with FastQC (v0.11.9). The trimmed reads were assembled using metaSPAdes v3.15.3 [42, 43], binned with MetaBAT2 v2.15 [44] (min contig for binning of 1500 bp and a min size of a bin as output of 200,000 bp) and MaxBin2 v2.2.7 [45] (setting a min contig length of 1000 bp and a min bin size of 100,000 bp), then combined and refined using DASTool v1.1.4 [46] setting a min contig size of 1500 bp. This computational methodology allowed us to retrieve a total of 188 genome bins (MetaBAT2 and MaxBin2 results) reconstructed from the metagenomic sample (S15), but after DASTool refinement only 44 genomes were subsequently considered for downstream analysis. Additionally, we perform a comprehensive evaluation with the metashot/prok-quality pipeline running also in Nextflow [47], to estimate the genome quality of the refined bins by CheckM v1.1.2 [48]. The global average GC content is 62.14, N50 is 50,86 Kbp, and the number of contigs is 16,489. Gene prediction was performed with Prodigal v2.6.3 [49], and the prediction of CDS (coding sequences) and functional annotation of mining MAGs was performed with Bakta v1.8.1 running the full database (version 5.0) [50]. Taxonomic classification was performed using GTDB-Tk v2.1.1 (database r214) [51, 52], inferring MAG placements in a phylogenetic tree based on 120 bacterial single-copy genes (no archaeal markers were detected). The reference tree was built using the R packages ggplot2, ggtree v3.10.1 [53], and ggTreeExtra v1.12.0. Finally, MAGs abundances were computed by the nf-core/mag pipeline at contig and MAGs level.

Functional annotation of MAGs and statistical analyses

KEGG orthologs (KOs) were predicted using the KOfamScan annotation tool v1.3.0 [54] with default options. The protein sequences of the 44 mining MAGs and a representative subset of the SMAG catalog comprising 396 high-quality bins from the non-mining ecosystems were queried against the HMM models from KEGG database (release 108.1, 2023). Then, the raw results were merged in a matrix with MAGs in columns and KOs in rows, which was further processed to normalize the counts divided by the total genome size for each ecosystem in order to build the boxplots for targeted pathways (Cu, Fe, and S) with the help of the R package ggpubr v0.6.0. This matrix was also used for the bubble plot with relevant genes from the mining ecosystem. Moreover, to evaluate if the means are equal between ecosystems,

we performed an analysis of variance (ANOVA) in each intrinsic mining pathway. Next, we also evaluated the metabolic capacities of the mining MAGs with the Multigenomic Entropy-Based Score (MEBS) v1.2 [55], a tool that integrates curated pathways for known cycles (naming C, Fe, N, O, S) and protein families from the Pfam database [56], aimed to understand the metabolic potential within the mining metagenomes by using an entropy score for likelihood. The resulting dataset with abundances was used to generate the heatmap relating geochemical cycles across the 44 MAGs. PCA analyses aimed to explore and compare selected mining pathways were conducted using the *prcomp* function in R, and the results were plotted using ggplot2 and factoextra v1.0.7.

SNP calling and evolutionary analysis

To perform variant calling in the metagenomic mining sample (S15), first the reads were mapped to the previously assembled bins (in FASTA format) using the aligner bwa-mem2 v2.2.1 [57], then the resulting alignment was converted from SAM format to a BAM file with samtools v1.16.1 [58]. This BAM file was processed using elPrep tool [59] in the split/filter/merge mode (sfm) to mark duplicates and then sort by coordinate order (samtools sort command). The sorted BAM file and the reference genome of metagenomic bins were used as input for inStrain v1.9.0 to call variants [60] (setting a minimum read-pair ANI threshold at 95% and the mapQ score cutoff of 20 in the "profile" module). In order to refine the SNPs (Single Nucleotide Polymorphisms), we selected a minimal coverage of a position of five reads (read depth). Finally, synonymous and non-synonymous mutations were analyzed using the dNdScv R package (version 0.0.1.0), which employs maximum-likelihood methods to robustly quantify dN/dS ratios across datasets of varying sizes [61]. Unlike traditional methods, dNdScv improves accuracy by accounting for the non-equilibrium composition of substitutions and the context dependence of mutations, which is essential for metagenomic analyses. We adapted dNdScv to analyze SNPs called from reads mapped to metagenome-assembled genomes (MAGs). A custom reference genome was created by integrating all 44 MAGs using the *buildref* function, enabling a unified analysis that captures the full genomic diversity of the microbial community. SNPs identified with inStrain were then used to calculate dN/dS ratios for individual genes within the MAGs. This allowed parameter fitting and selection analysis at both the gene and MAG levels. The default substitution model in dNdScv, with 192 rate parameters to model trinucleotide context and two ω parameters for missense and nonsense mutations, was used. By combining local information (synonymous mutations within genes) with global mutation rate variation across genes, dNdScv computes accurate mutation rate estimates [61]. Its trinucleotide context-dependent substitution model minimizes biases from mutational hotspots and sequencing errors, ensuring accurate dN/dS estimates even with limited sample sizes [61]. Genes under positive selection (PS) have a dN/dS ratio >1, and genes under negative selection (NS) a ratio below 1. The false discovery rate (FDR) was adjusted at a p-value below 0.05.

Results

MAGs assembly from short-reads

The shotgun metagenomic sequencing of the S15 sample yielded 103,790,393 raw read pairs, with 99.23% passing the quality filters (see Table S2). These filtered reads were subsequently processed for MAGs *de novo* assembly using the standard nf-core/mag v2.2.1 pipeline [39], available in Nextflow [40]. After metagenomic binning and refinement, we successfully generated 44 MAGs. Of these, 36 were grouped into three categories following the standards of Minimum Information about Metagenome-Assembled Genomes (MIMAG) [62]. The quality control (QC) results (Fig. 1a), classified 30 MAGs as high-quality (≥ 90% completeness, ≤ 5% contamination), 5 as medium-quality (≥ 50% completeness, ≤ 10% contamination), and only one as low-quality (< 50% completeness, ≤ 10% contamination). The remaining 8 MAGs included 3 classified as almost high-quality (completeness >90%, contamination between 5 and 5.2%), and 5 MAGs with contamination levels ranging from 10 to 25.18% (Table S3).

Combined taxonomic composition of mining MAGs and the SMAG catalog

The taxonomic annotation of the 44 MAGs derived from the mining ecosystem was performed using the Genome Taxonomy Database Toolkit (GTDB-Tk) [51, 52]. All MAGs were classified within the bacterial domain, with the phylum *Actinomycetota* emerging as the most predominant clade (21 MAGs), followed by *Pseudomonadota* (10) and *Acidobacteriota* (6), collectively encompassing 84% of the annotated phyla (Fig. 1b and Table S4). The phylogenetic tree (Fig. 1c) aligns with findings from non-mining ecosystems (Fig. S1a) and is consistent with the globally reported predominant phyla in soil microbiomes [8, 38]. Moreover, all mining bins were assigned both at the class and at the order level, 43 at the family level, and 31 out of 44 were classified as kGGBs. Only 5 MAGs were assigned as kSGBs: *Acinetobacter johnsonii*, *Pseudomonas_E alloputida*, *Sphingobium yanoikuyae*, and two other species without nomenclature in GTDB (*JAJZVS01 sp021845555* and *JAKAYS01 sp021826695*) (Fig. S1b). The five identified species include three from the phylum *Pseudomonadota*, with two classified as medium quality and one as low quality. The remaining two belong to the phylum

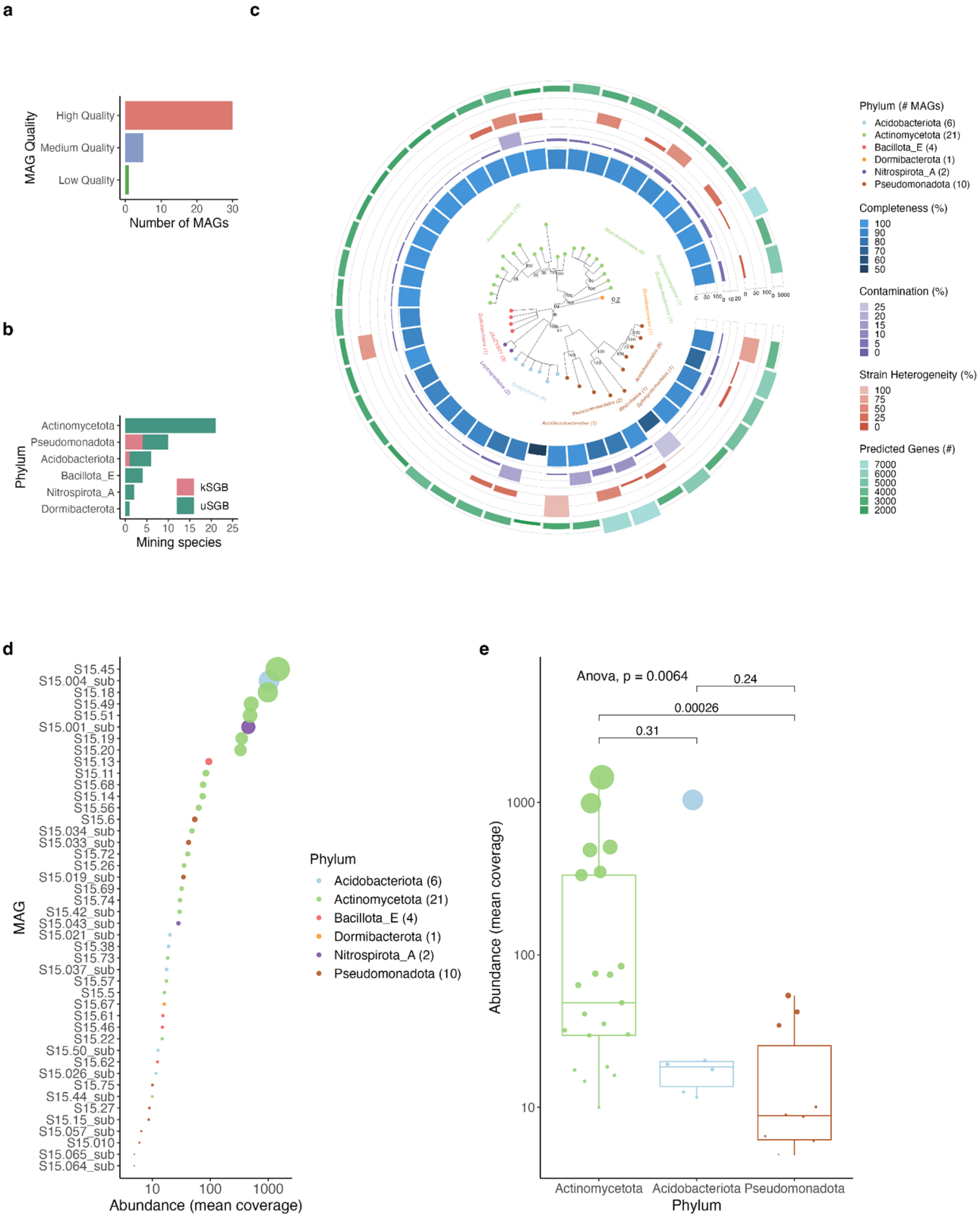

Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Quality metrics, taxonomic classification and abundances of the 44 MAGs obtained from the copper mining tailing (S15 sample). **a** MIMAG classification of MAGs according to their genome completeness and contamination. **b** Distribution of known/unknown species (kSGB/uSGB) at phylum level within the mining ecosystem. **c** Phylogenetic tree of the 44 mining MAGs including their phylum classification (number of MAGs) and individual metrics (from inner to outer ring: completeness, contamination, strain heterogeneity, and number of predicted genes). The tip for each MAG placement is colored by phyla and the label represents each order-level assignment. The tree scale and bootstrap values are also shown. **d** Distribution of log10-transformed abundances (mean coverage of mapped reads) across MAGs, colored by phyla. **e** Boxplot of abundances within the top three dominant phyla, sorted by median. ANOVA was applied to compare the means differences between the groups

Acidobacteriota and *Pseudomonadota*, with low and high quality, respectively.

Order-level taxonomic annotation of the 44 MAGs revealed a community dominated by core and auxiliary taxa linked to biomining processes. *Acidimicrobiales* (13 MAGs) dominated the microbiome, highlighting their critical role in iron cycling [63], essential for bioleaching [64]. *Terriglobales* (6 MAGs) and *Mycobacteriales* (6 MAGs) were also frequent, suggesting potential adaptations to acidic and metal-rich environments.

Abundances of mining MAGs are shown in Fig. 1d, where the mean coverage of MAGs is highly dominated by *Actinomycetota*. Among the most abundant MAGs affiliated to this phylum, we found four bins belonging to the order *Acidimicrobiales* and one bin of *Mycobacteriales*. Interestingly, among the three most representative phyla (Fig. 1e), *Pseudomonadota* showed significantly lower abundance compared to *Actinomycetota*, with a notable statistical difference (p-value = 0.00026).

Taxonomy of mining MAGs was compared to 32,931 soil MAGs from the SMAG catalog, revealing a consistent lack of species-level classification across soil environments, emphasizing the challenge of characterizing microorganisms from extreme ecosystems [65]. Unknown species bins (uSGBs) were dominated by *Pseudomonadota*, *Actinomycetota*, and *Acidobacteriota*, consistent with other soils (Fig. S2a). Compared to non-mining ecosystems, mining MAGs showed a relatively high known species ratio (5 species/sample), second only to artificial surfaces (Fig. S2b). Shared genus-level bins (kGGBs) included *Pseudomonas_E*, *Sphingobium*, and *Acinetobacter*, while *Ferrithrix* and *Acidithrix* were exclusive to mining (Fig. S2c). Its absence in the other analyzed environments indicates that this genus has primarily specialized in inhabiting mining settings [66, 67]. Despite similarities at the phylum level (i.e., the six different phyla we found in mining MAGs are contained within non-mining soil microbiomes), mining microbiomes exhibited unique and less diverse genus- and species-level taxonomy.

Metabolic capacities of mining MAGs

We compared the metabolic potential of 44 mining MAGs with 396 non-mining high-quality MAGs by annotating KEGG orthologs (KOs). Focusing on 247 KOs related to copper, iron, and sulfur metabolism (22, 68, and 157 marker genes, respectively; Table S5), we observed key differences in biogeochemical pathways between ecosystems.

Regarding copper metabolism (Fig. 2a) the mining ecosystem ranks second after the glacier and along with artificial surfaces are the only ecosystems above the mean. Mining MAGs highlight the markers for copper resistance proteins (*copB*, *copC* and *copD* genes) and *csoR* family transcriptional regulator that controls copper oxidation and homeostasis (Fig. S4a). In the case of the iron proteins (Fig. 2b), artificial surfaces are in the first position and proteins having transporter and binding functions are abundant across the other non-mining ecosystems, but mining MAGs are placed last having a few outlier genes. Considering that iron mechanisms have mainly specialized in metal uptake under high competition scenarios, it is expected that in a mining environment, where there is high bioavailability of this micronutrient, there would be a lower need in terms of the quantity and diversity of proteins involved in iron acquisition [65, 68]. Notably, mining MAGs exhibited statistically significant dominance in sulfur metabolism (Fig. 2c), with key enzymes such as sulfide-quinone oxidoreductase (EC: 1.8.5.4) and sulfurtransferase (EC: 4.4.1.37), along with the thiol-disulfide interchange protein (*dsbE* gene), and sulfur-carrier cysteine synthase and hydrolase proteins encoded by the *cysO* gene (EC: 2.5.1.113 and 3.13.1.6) that are associated to a mechanistic response of soil acidification [35, 69]. The Principal Component Analysis (PCA) highlights the specialization of the mining ecosystem for copper and sulfur metabolism, with PC1 explaining a high percentage of the variance (85.3% and 92%, respectively; Fig. 2, bottom). This specialization is further detailed by the top ten KO contributions for each pathway (Fig. S3a) and the differential distribution of the normalized counts for the top five KOs across ecosystems. Notably, sulfur metabolism shows a striking prominence in mining MAGs (Fig. S3b).

Finally, a specific metabolic-biogeochemical analysis was performed in the mining MAGs using MEBS (Multigenomic Entropy-Based Score) [55]. We found a similar response across the mining genomes (Fig. S4b) where the most likely reactions involved are sulfur oxidation (*dsrEFH*, *fccB*, *sox* genes), reduction (*sreABC*, *qmoABC*, *dsrABC*), disproportionation (thiosulfate), assimilation (*coB*/*coM*), and also nitrate reduction (denitrification, assimilatory). These findings are in accordance with

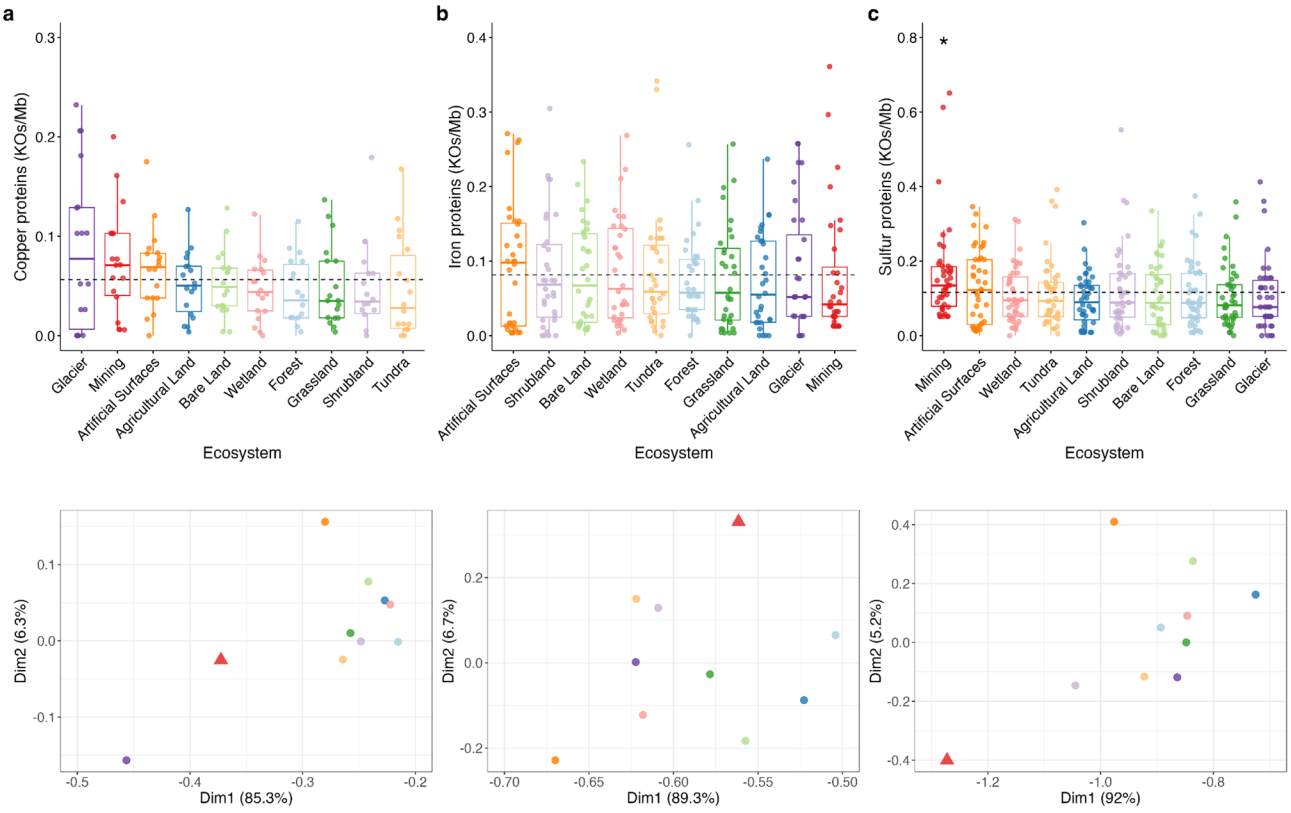

Fig. 2 Functional metabolic capacities of mining MAGs compared to non-mining ecosystems present in the SMAG catalog. The description of selected geochemical cycles and metabolic pathways potential was assessed using KOfamScan annotating the resulting KEGG orthology groups (KOs). Boxplots showing the KO definitions at selected **a** copper, **b** iron, and **c** sulfur proteins sorted by ecosystem. Each dot represents a KO annotation, where the counts were normalized using the length for all MAGs per ecosystem (KOs/Mb, y-axis). The *p*-value represents the statistical difference between ecosystems calculated with an ANOVA (* < 0.05 significance). The ANOVA showed no statistical differences between ecosystems among the three pathways, but only the mining ecosystem highlights the sulfur proteins in a paired *t*-test compared to the global mean and was marked with an asterisk (*). In all plots, the ecosystems (x-axis) are sorted by median in decreasing order. Below, PCA of each selected pathway depicting the ecosystems (mining the red triangle) based on the contributions of the annotated KOs. PCA highlights the specialization of mining MAGs in copper and sulfur metabolism

previous reports on the sulfur pathway detected in acidic microbiomes [15, 35, 69, 70].

Evolutionary analysis of mining MAGs

To assess evolutionary pressures on the 44 assembled mining MAGs, we identified 146,932 high-quality SNPs using inStrain [60] and estimated dN/dS ratios for 39,438 genes with dNdScv [61]. Quality control of SNPs (Supplementary Fig. 5A) confirms that mutation rates are not significantly influenced by completeness, contamination, GC content, or strain heterogeneity. While mean coverage, as expected, shows a weak positive association, it is not statistically significant, indicating that mutation calls are robust and unbiased by sequencing depth or other genomic factors. Additionally, a comparison of mutation rates between *Actinomycetota* and *Pseudomonadota* using an ANOVA test (p-value = 0.479) reveals no significant differences (Supplementary Fig. 5B), suggesting that mutation rates are not strongly influenced by phylum-level taxonomic classification.

Global dN/dS ratios for the 44 MAGs, normalized by contamination (Fig. 3A), reveal strong negative selection, implicating evolutionary constraints in this ecosystem. *Pseudomonadota* show the highest mean dN/dS ratio, followed by *Actinomycetota* and *Acidobacteriota*, though no significant differences were observed (Fig. 3B). Notably, the MAG S15.064_sub (classified as *Acinetobacter johnsonii*) exhibited the highest dN/dS ratio (0.52) and interestingly is the least abundant MAG (Fig. 1D). In contrast, a MAG from the *Acidobacteriaceae* family had the lowest dN/dS ratio (0.17), indicating strong negative selection acting on its gene pool.

At the gene level, 95.2% of genes (37,546) exhibit dN/dS < 1, indicating strong purifying selection. Among them, 274 genes involved in key metabolic pathways (identified via KOfamScan) show negative selection, with iron, sulfur, and copper-related proteins with 112, 106, and 56 genes, respectively. These genes are under strong purifying selection (mean dN/dS = 0.212), with iron metabolism being the most conserved, significantly more

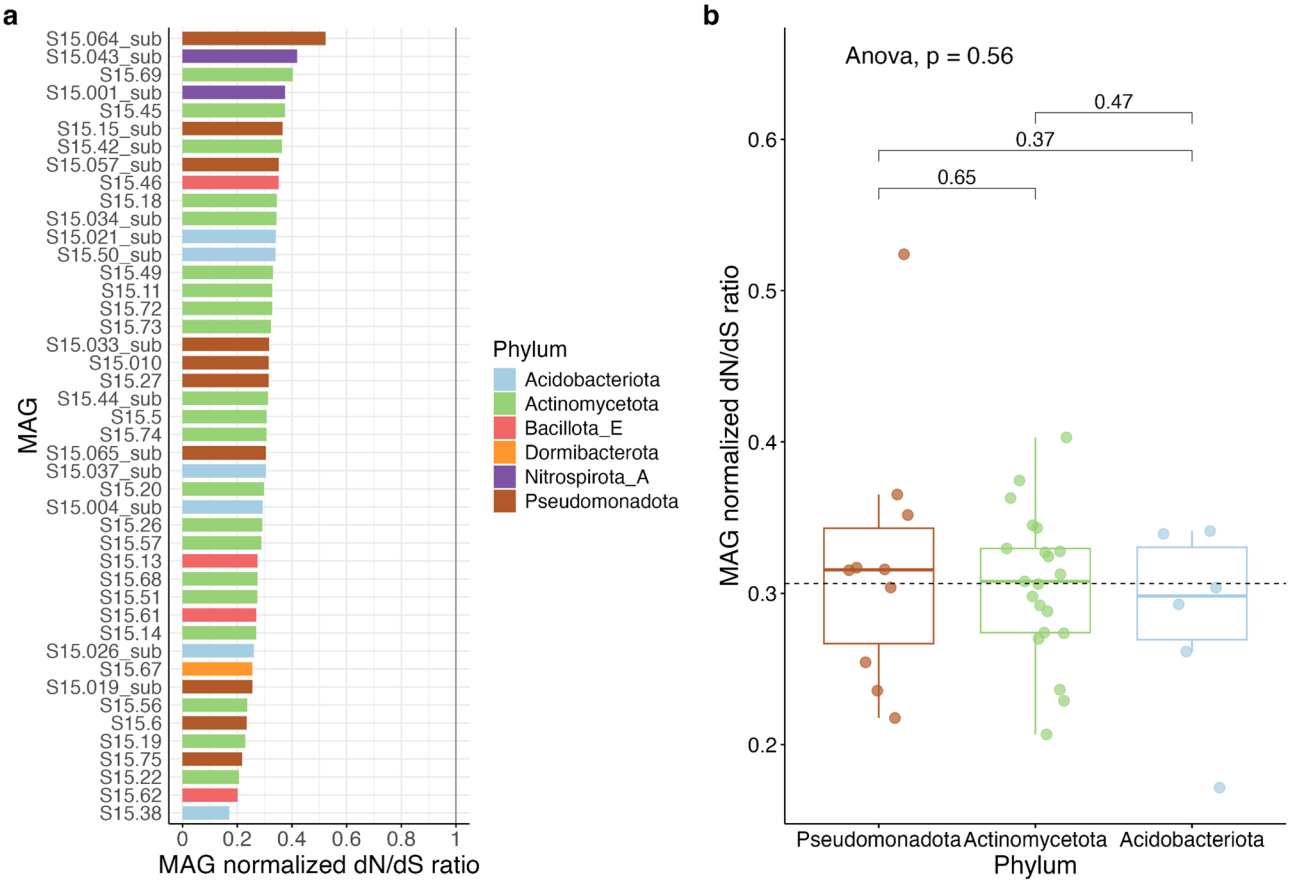

Fig. 3 Global dN/dS ratios of mining MAGs and its distribution across phyla. **a** Barplot for global estimation of dN/dS ratios (normalized by MAG contamination) accumulated in the 44 MAG bins, sorted by descending strength and colored by phyla. **b** Boxplot of global estimation of dN/dS ratios within the top three representative phyla, where each dot represents a MAG. Ratios were normalized by MAG contamination, then grouped and sorted by median. ANOVA was applied to compare the means differences between the groups

than sulfur ($p = 0.0083$, Fig. 4A). The mean dN/dS values per pathway were 0.245 (sulfur), 0.206 (copper), and 0.184 (iron). Additionally, these 274 genes were grouped by gene products and ranked by dN/dS ratios, revealing 66 unique gene products (Fig. 4B), with two of the top three involved in sulfur metabolism (*thiS* and *cysO*) and the copper chaperone *copZ*.

A total of 1892 genes (4.8%) exhibit dN/dS >1, indicating positive selection (PS), though they are vastly outnumbered by genes under purifying selection. Among them, 202 genes pass a 5% false discovery rate, and 1764 (93.2%) are annotated as hypothetical proteins. Only 96 genes have at least 5 SNPs, which were selected for further analysis (Fig. 4C). Notably, the *purK* gene (involved in purine biosynthesis) shows the strongest selection signal (dN/dS = 4.34), potentially aiding bacterial adaptation in extreme mining environments, where nucleotide metabolism plays a crucial role [71]. Among the PS genes linked to intrinsic mining pathways, *msrB*, involved in oxidative stress resistance, was found in *Actinomycetota* MAGs [72], while *cysQ*, related to sulfur assimilation and sulfur-oxidizing bacteria, was detected in *Actinomycetota* and *Pseudomonadota* MAGs [73]. Additionally, *csoR*, a copper-sensing transcriptional repressor involved in copper resistance, was identified in mining bacteria [74].

At the species level, the distribution of PS genes using KEGG KO annotations (Fig. 4D), highlighting core survival functions, including flagellar body proteins (*fliE*), cell cycle regulators (*ftsH*), and biogenesis proteins (*tsf*). Notably, *mshA* biofilm motor proteins [75] suggest bacterial adaptation through biofilm formation, a key mechanism for avoiding metal toxicity in mining environments [76].

To confirms that this trend is consistent across multiple mining samples, we computed SNPs and dN/dS ratios across seven additionally mining samples (Supplementary Fig. 6). Specifically, global dN/dS ratios indicate strong negative selection across all samples (Supplementary Fig. 6A), reinforcing the stability of evolutionary constraints in this environment.

Overall, our genome-resolved dN/dS scan map the small subset of genes permissive to adaptive variation within an otherwise strongly constrained biomining

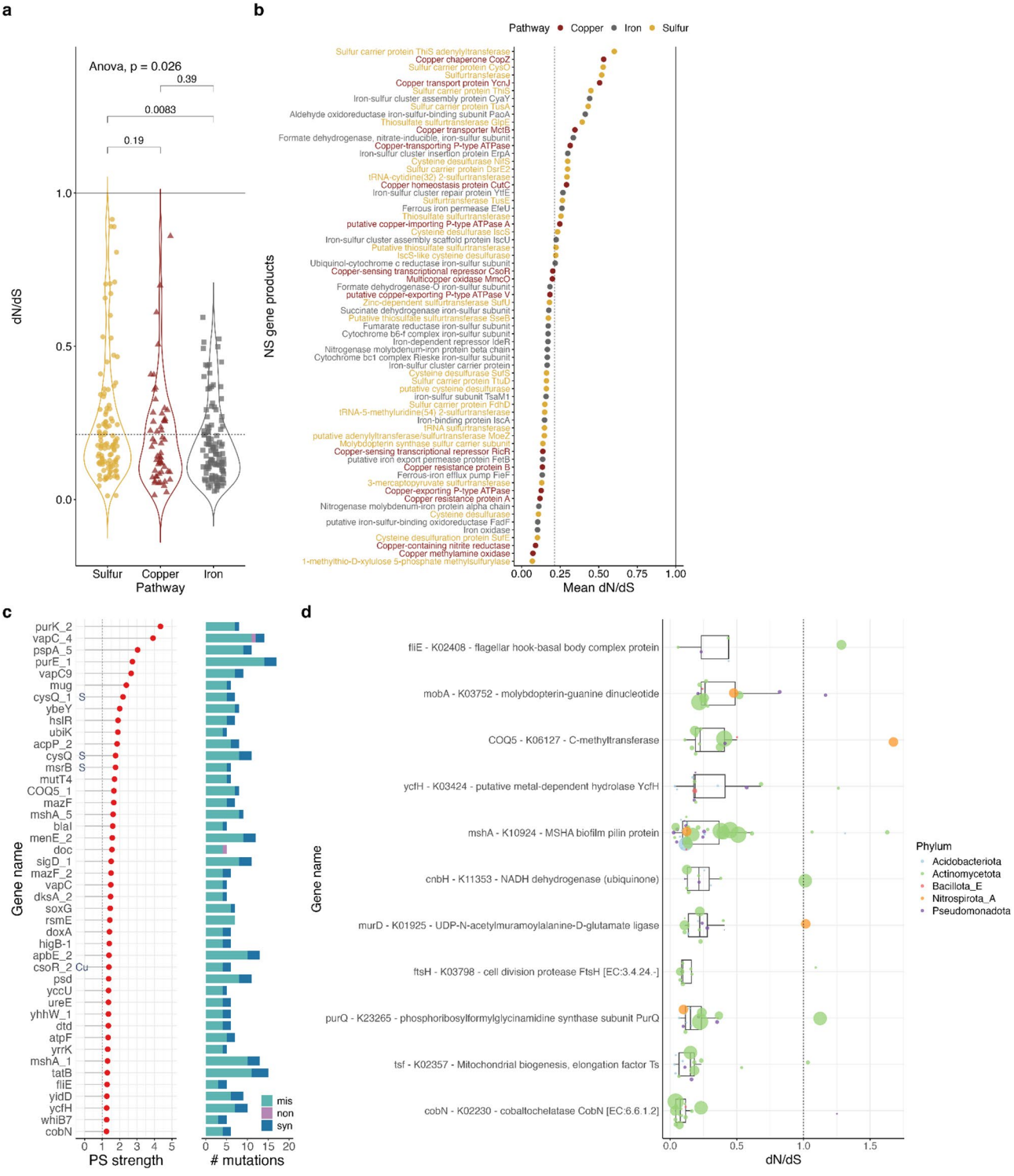

Fig. 4 Evolutionary analysis on mining MAGs, summarizing the genes under positive (PS) and negative selection (NS). **a** Violin plot for genes under NS focused on intrinsic mining pathways, ordered by decreasing dN/dS ratios (the dashed line represents the overall mean). An ANOVA was applied to differentiate between groups. **b** Dotplot of the annotated gene products for selected NS genes, colored by pathway and sorted by selection strength. Each dot depicts the mean dN/dS among unique products (the dashed line represents the overall mean). **c** Ranking of genes under PS and number of mutations per type, discarding hypothetical proteins and genes holding below 5 SNPs. Only for visualization purposes, a dN/dS ratio cutoff was set to 1.25. (legend: S = sulfur-related proteins; Cu = copper protein; mis = missense mutations; non = nonsense; syn = synonymous). **d** List of aggregated genes from those under PS and their dN/dS ratios (n = 11), containing at least 10 aggregated genes (except for *fliE* and *ftsH*). Genes are sorted by mean, colored by phyla and the size represents the MAG abundance

microbiome, thereby identifying concrete genetic targets for the optimisation of the biomining consortia.

Discussion

Shotgun metagenomic sequencing of the Cauquenes copper mining tailing (S15) produced 44 MAGs, revealing a community dominated by *Actinomycetota*, *Pseudomonadota*, and *Acidobacteriota*. Within these, *Acidimicrobiales*, *Terriglobales*, and *Mycobacteriales* were the most abundant orders. Comparison with the SMAG soil catalog showed that the mining microbiome displays a distinct and less diverse taxonomic structure but a higher proportion of known species, indicating strong ecological filtering under extreme conditions. Functional annotation highlighted specialization in sulfur and copper metabolism, while evolutionary analysis demonstrated pervasive purifying selection, suggesting long-term genomic stability and functional conservation across copper mining tailings sites.

The taxonomic composition of the Cauquenes tailing microbiome reflects both the selective pressure of metal-rich environments and the persistence of functionally specialized lineages. In particulalry, MAGs belonging to *Acidimicrobiales* (phylum *Actinomycetota*) have been previously identified in acidic mine environments comprising bacteria involved in nitrogen oxidative metabolism [77]. Interestingly, MAGs from the order *Mycobacteriales* (*Actinomycetota*) have been recently characterized in non-mining soil microbiomes involved in carbon fixation, and in the metabolism of nitrogen and sulfur [78]. The other predominant orders, *Terriglobales* (*Acidobacteriota*) have been reported in other non-mining extreme ecosystems, uncovering complex diversity in tundra [79], whereas *Acetobacterales* (*Pseudomonadota*, 5 bins) have been recently described with high abundances in extremophilic copper mining communities [80], and also may contribute to nitrogen-fixing processes associated to non-mining soils [81]. Furthermore, known mining taxa such as *Leptospirillales* (2 MAGs), *Sulfobacillales* (1 MAG), and *Acidiferrobacterales* (1 MAG) were identified, which are important in metal solubilization and sulfur oxidation [82, 83]. Novel lineages, including *JAJZYS01* (3 MAGs), and less-characterized groups like *Dormibacterales* (1 MAG) and *Solirubrobacterales* (1 MAG), suggest unique adaptations and unexplored functional roles. Additionally, auxiliary taxa such as *Pseudomonadales* (2 MAGs) may contribute to nutrient cycling and biofilm formation [84].

The functional annotation of these MAGs show that the copper mining tailings provide a great habitat for sulfur-oxidizing microorganisms that require, for instance, the enzyme molybdopterin sulfur reductase (*sreABC* gene), which is needed for elemental sulfur (S^0) reduction to H_2S for energy production [85]. This gene was particularly found in the majority of mining MAGs with medium to high likelihood. On the contrary, marker genes belonging to nitrogen metabolism are annotated with low abundances, as they are expected in other ecosystems like forests or cultivable lands [86].

The evolutionary analysis provides additional insight into the long-term stability of these metabolic systems. The dN/dS ratios on seven additional samples (Supplementary Fig. 6B) show that abundant phyla are predominantly under negative selection, mirroring our findings in S15. The correlation analysis (Supplementary Fig. 6C) further supports the robustness of these patterns, with a significant Spearman correlation ($\rho = 0.612$, p-value = 1.58e−05) between dN/dS values in S15 and those computed using a dereplicated reference of 250 MAGs, confirm that the evolutionary forces observed in S15 are not an isolated case but rather a generalizable trend across mining sites. Furthermore, genes associated with sulfur, iron, and copper metabolism exhibit strong negative selection (Supplementary Fig. 6D) across all samples, suggesting long-term evolutionary constraints on these pathways. These findings collectively validate our methodology and main result in S15 by demonstrating that the evolutionary pressures shaping mining microbiomes are consistent across eight independent samples.

Our selection analysis relies on short-read assemblies and MAG reconstructions. Even with stringent QC and contamination-normalized dN/dS, fragmented genes, strain heterogeneity, and uneven coverage can affect variant calls and, in turn, selection estimates. To resolve these limitations, time-series and gradient sampling, long-read assemblies, and isolate-based validation will be essential to directly assess recombination, mobile elements, and within-species variation in mining tailings—processes that are only indirectly captured here.

Conclusions

We assembled and annotated 44 medium and high-quality metagenome-assembled genomes (MAGs) from a metagenomic sample of the Cauquenes copper tailing (S15). Taxonomically, we observed a prevalence of bacterial species belonging to the phylum *Actinomycetota*, *Pseudomonadota*, and *Acidobacteriota*, which were also found to be dominant in the SMAG catalog containing bacterial MAGs from nine different non-mining soils. At the species level, only five out of the 44 MAGs were annotated: *Acinetobacter johnsonii*, *Pseudomonas_E alloputida*, *Sphingobium yanoikuyae*, and two species without nomenclature in GTDB (*JAJZVS01 sp.* and *JAKAYS01 sp.*), remarking the fact that in extreme scenarios the annotation at the finest levels is still challenging.

The metabolic machinery of the mining MAGs emphasise the importance of copper and sulfur pathways

proteins, highlighted by key genes, such as those involved in copper resistance and carrier/transporter proteins specific to acidic environments like copper mining tailings. In contrast, iron metabolism showed no relevant activity on mining compared to the other ecosystems. The evolutionary analysis using dN/dS allowed us to investigate genes under positive selection and negative selection. Notably, only a few genes related to the metabolism of sulfur and copper were identified to be under PS (*msrB*, *cysQ*, and *csoR* genes), indicating a clear trend toward a conservative evolution in the intrinsic mining pathways. We also identified a high-quality MAG belonging to the genus *Acidithrix*, uniquely established within the mining ecosystem.

We confirm that biomining communities are dominated by strong purifying selection and—by resolving dN/dS at both gene, pathway and genome scale—identify exactly where that constraint falls. Sulfur- and copper-metabolism genes are tightly conserved, and iron-oxidation proteins display the lowest dN/dS values, marking them as the most evolutionarily constrained. In contrast, the few genes under positive selection cluster in motility and cell-surface biogenesis modules, suggesting selection for more effective resource allocation in metal-rich niches. Finally, we demonstrated that the evolutionary pressures shaping mining microbiomes are consistent across eight independent samples of the Cauquenes tailings; however, without time-series data our analisis remains a snapshot. Future time-series or samples with physicochemical gradients will be invaluable for capturing the fine-scale evolutionary dynamics of AMD communities.

Our study provides a detailed evolutionary analysis of mining species at MAGs level, demonstrating that microbial communities in copper tailings are highly specialized, with conserved metabolic pathways under strong purifying selection. These findings emphasize the challenges of optimizing these communities for biotechnological applications, such as biomining, due to their evolutionary constraints.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40793-025-00811-5.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
This work was supported by Proyecto Anillo Regular ANID ACT210004; Center for Mathematical Modeling; Apoyo a Centros de Excelencia ACE210010; ANID Millennium Science Initiative Program ICN2021_044; Fondo Basal FB210005; ANID FONDECYT 1190742, 1221029, 1230195, 1211731, and 1230194; Fondo Interdisciplinario y Proyecto Núcleo UOH; Proyecto Postdoctoral ANID 3220080. The supercomputing infrastructure of the NLHPC (ECM-02); the supercomputing infrastructure of the High-Performance Computing UOH laboratory (FIC 40059065-0) of University of O'Higgins; and Minera Valle Central company (Requinoa, O'Higgins, Chile).

Author contributions
M.A.R. contributed to the data collection, processing, analysis and interpretation, literature review, writing, and critical review of the manuscript. G.S., J.T., J.O., and G.G. were involved in data collection and processing, and materials. E.V., V.P., A.R.-J., V.M.-C., and L.P. contributed to funding acquisition and critically reviewed the manuscript. L.P was also involved in data collection and processing. M.L. was responsible for the conception, supervision, funding acquisition, materials, data collection and processing, writing and critical review of the manuscript. A.D.G. contributed to the conception, design, supervision, funding acquisition, analysis and interpretation, literature review, writing, and critical review of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding
This work was supported by Proyecto Anillo Regular ANID ACT210004; Center for Mathematical Modeling; Apoyo a Centros de Excelencia ACE210010; ANID Millennium Science Initiative Program ICN2021_044; Fondo Basal FB210005; ANID FONDECYT 1190742, 1221029, 1230195, 1211731, and 1230194; Fondo Interdisciplinario y Proyecto Núcleo UOH; Centro UOH de Bioingeniería (CUBI); Proyecto Postdoctoral ANID 3220080.

Data availability
The main datasets and code used in the analyses, along with the figures generated for this work, are available on GitHub (https://github.com/digenoma-lab/biomining_metagenomes). Genomic reads of sample S15 have been submitted to NCBI SRA under bioproject PRJNA1167620 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1167620).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Systemix Center, ANID Anillo ACT210004, Universidad de O'Higgins, Rancagua, Chile
[2]Computational Biology Lab, Instituto de Ciencias de la Ingeniería, Universidad de O'Higgins, Rancagua, Chile
[3]Laboratorio de Bioingeniería, Instituto de Ciencias de La Ingeniería, Universidad de O'Higgins, Rancagua, Chile
[4]Centro de Modelamiento Matemático UMI-CNRS 2807, Universidad de Chile, Santiago, Chile
[5]Laboratorio de Diferenciación Celular y Metabolismo, Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile
[6]Advanced Center of Chronic Diseases (ACCDiS), Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile
[7]Laboratorio de Microbiología y Probióticos, Instituto de Nutrición y Tecnología de los Alimentos (INTA), Universidad de Chile, Santiago, Chile
[8]Unidad de Genómica Avanzada (UGA), Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile
[9]Laboratório de Medicina e Saúde Pública de Precisão (MESP2), Fundação Oswaldo Cruz, Fiocruz, Salvador, Brazil
[10]Laboratorio de Inmunidad Vegetal, Instituto de Ciencias Agroalimentarias, Animales y Ambientales (ICA3), Universidad de O'Higgins, San Fernando, Chile
[11]Laboratorio de Bioinformática y Expresión Génica, Unidad de Nutrición Básica, INTA, Universidad de Chile, Santiago, Chile

Received: 9 September 2024 / Accepted: 25 October 2025
Published online: 19 December 2025

References

1. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from uncultivated genomes of the global human gut microbiome. Nature. 2019;568:505–10.
2. Cowan DA, Ramond J-B, Makhalanyane TP, De Maayer P. Metagenomics of extreme environments. Curr Opin Microbiol. 2015;25:97–102.
3. Sierra MA, Ryon KA, Tierney BT, Foox J, Bhattacharya C, Afshin E, et al. Microbiome and metagenomic analysis of Lake Hillier Australia reveals pigment-rich polyextremophiles and wide-ranging metabolic adaptations. Environ Microbiome. 2022;17:60.
4. Waschulin V, Borsetto C, James R, Newsham KK, Donadio S, Corre C, et al. Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. ISME J. 2022;16:101–11.
5. Li S, Lian W-H, Han J-R, Ali M, Lin Z-L, Liu Y-H, et al. Capturing the microbial dark matter in desert soils using culturomics-based metagenomics and high-resolution analysis. NPJ Biofilms Microbiomes. 2023;9:67.
6. Haro-Moreno JM, Cabello-Yeves PJ, Garcillán-Barcia MP, Zakharenko A, Zemskaya TI, Rodriguez-Valera F. A novel and diverse group of Candidatus Patescibacteria from bathypelagic Lake Baikal revealed through long-read metagenomics. Environ Microbiome. 2023;18:12.
7. Liang J-L, Feng S-W, Lu J-L, Wang X-N, Li F-L, Guo Y-Q, et al. Hidden diversity and potential ecological function of phosphorus acquisition genes in widespread terrestrial bacteriophages. Nat Commun. 2024;15:2827.
8. Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol. 2017;15:579–90.
9. Daniel R. The metagenomics of soil. Nat Rev Microbiol. 2005;3:470–8.
10. Banerjee S, van der Heijden MGA. Soil microbiomes and one health. Nat Rev Microbiol. 2023;21:6–20.
11. Quatrini R, Johnson DB. Microbiomes in extremely acidic environments: functionalities and interactions that allow survival and growth of prokaryotes at low pH. Curr Opin Microbiol. 2018;43:139–47.
12. Valenzuela L, Chi A, Beard S, Orell A, Guiliani N, Shabanowitz J, et al. Genomics, metagenomics and proteomics in biomining microorganisms. Biotechnol Adv. 2006;24:197–211.
13. Zhan Y, Yang M, Zhang S, Zhao D, Duan J, Wang W, et al. Iron and sulfur oxidation pathways of *Acidithiobacillus ferrooxidans*. World J Microbiol Biotechnol. 2019;35:60.
14. Schlatter DC, Kahl K, Carlson B, Huggins DR, Paulitz T. Soil acidification modifies soil depth-microbiome relationships in a no-till wheat cropping system. Soil Biol Biochem. 2020;149:107939.
15. Kumar J, Sharma N, Singh SP. Genome-resolved metagenomics inferred novel insights into the microbial community, metabolic pathways, and biomining potential of Malanjkhand acidic copper mine tailings. Environ Sci Pollut Res Int. 2023;30:50864–82.
16. Valdés J, Pedroso I, Quatrini R, Dodson RJ, Tettelin H, Blake R II, et al. *Acidithiobacillus ferrooxidans* metabolism: from genome sequence to industrial applications. BMC Genomics. 2008;9:597.
17. Travisany D, Cortés MP, Latorre M, Di Genova A, Budinich M, Bobadilla-Fazzini RA, et al. A new genome of *Acidithiobacillus thiooxidans* provides insights into adaptation to a bioleaching environment. Res Microbiol. 2014;165:743–52.
18. Zhang X, Feng X, Tao J, Ma L, Xiao Y, Liang Y, et al. Comparative genomics of the extreme acidophile *Acidithiobacillus thiooxidans* reveals intraspecific divergence and niche adaptation. Int J Mol Sci. 2016;17:1355. https://doi.org/10.3390/ijms17081355.
19. Norris PR, Clark DA, Owen JP, Waterhouse S. Characteristics of *Sulfobacillus acidophilus* sp. nov. and other moderately thermophilic mineral-sulphide-oxidizing bacteria. Microbiology. 1996;142(Pt 4):775–83.
20. Anderson I, Chertkov O, Chen A, Saunders E, Lapidus A, Nolan M, et al. Complete genome sequence of the moderately thermophilic mineral-sulfide-oxidizing firmicute *Sulfobacillus acidophilus* type strain (NAL(T)). Stand Genomic Sci. 2012;6:1–13.
21. Latorre M, Cortés MP, Travisany D, Di Genova A, Budinich M, Reyes-Jara A, et al. The bioleaching potential of a bacterial consortium. Bioresour Technol. 2016;218:659–66.
22. Gentina J, Acevedo F. Copper bioleaching in Chile. Minerals (Basel). 2016;6:23.
23. Cárdenas JP, Quatrini R, Holmes DS. Genomic and metagenomic challenges and opportunities for bioleaching: a mini-review. Res Microbiol. 2016;167:529–38.
24. Yang L, Zhao D, Yang J, Wang W, Chen P, Zhang S, et al. *Acidithiobacillus thiooxidans* and its potential application. Appl Microbiol Biotechnol. 2019;103:7819–33.
25. Sharma P, Kumar S, Pandey A. Bioremediated techniques for remediation of metal pollutants using metagenomics approaches: a review. J Environ Chem Eng. 2021;9:105684.
26. Huang W-L, Wu P-C, Chiang T-Y. Metagenomics: potential for bioremediation of soil contaminated with heavy metals. Ecol Genet Genom. 2022;22:100111.
27. Allen EE, Tyson GW, Whitaker RJ, Detter JC, Richardson PM, Banfield JF. Genome dynamics in a natural archaeal population. Proc Natl Acad Sci USA. 2007;104:1883–8.
28. Simmons SL, Dibartolo G, Denef VJ, Goltsman DSA, Thelen MP, Banfield JF. Population genomic analysis of strain variation in Leptospirillum group II bacteria involved in acid mine drainage formation. PLoS Biol. 2008;6:e177.
29. Denef VJ, Banfield JF. In situ evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. Science. 2012;336:462–6.
30. Li S-J, Hua Z-S, Huang L-N, Li J, Shi S-H, Chen L-X, et al. Microbial communities evolve faster in extreme environments. Sci Rep. 2014;4:6205.
31. Guo J, Wang Q, Wang X, Wang F, Yao J, Zhu H. Horizontal gene transfer in an acid mine drainage microbial community. BMC Genomics. 2015;16:496.
32. Huang L-N, Kuang J-L, Shu W-S. Microbial ecology and evolution in the acid mine drainage model system. Trends Microbiol. 2016;24:581–93.
33. Liu R, Ma L, Wang H, Liu D, Lu X, Huang X, et al. Comparative genomics reveals intraspecific divergence of *Acidithiobacillus ferrooxidans*: insights from evolutionary adaptation. Microb Genom. 2023;9:001038. https://doi.org/10.1099/mgen.0.001038.
34. Galvez G, Ortega J, Fredericksen F, Aliaga-Tobar V, Parra V, Reyes-Jara A, et al. Co-occurrence interaction networks of extremophile species living in a copper mining tailing. Front Microbiol. 2021;12:791127.
35. Li J-T, Jia P, Wang X-J, Ou S-N, Yang T-T, Feng S-W, et al. Metagenomic and metatranscriptomic insights into sulfate-reducing bacteria in a revegetated acidic mine wasteland. NPJ Biofilms Microbiomes. 2022;8:71.
36. Chen M, Grégoire DS, Bain JG, Blowes DW, Hug LA. Legacy copper/nickel mine tailings potentially harbor novel iron/sulfur cycling microorganisms within highly variable communities. Appl Environ Microbiol. 2024;90:e0014324.
37. Chakraborty S, Ghosh S, Banerjee S, Kumar S, Bhattacharyya P. Elucidating the synergistic effect of acidity and metalloid poisoning on the microbiome through metagenomics and machine learning approaches. Environ Res. 2024;243:117885.
38. Ma B, Lu C, Wang Y, Yu J, Zhao K, Xue R, et al. A genomic catalogue of soil microbiomes boosts mining of biodiversity and genetic resources. Nat Commun. 2023;14:7318.
39. Krakau S, Straub D, Gourlé H, Gabernet G, Nahnsen S. nf-core/mag: a best-practice pipeline for metagenome hybrid assembly and binning. NAR Genom Bioinform. 2022;4:lqac007.
40. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol. 2017;35:316–9.
41. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90.
42. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
43. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr Protoc Bioinf. 2020;70:e102.
44. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive Binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359.
45. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated Binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32:605–7.
46. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018;3:836–43.
47. Albanese D, Donati C. Large-scale quality assessment of prokaryotic genomes with metashot/prok-quality. F1000Res. 2021;10:822.
48. 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;25:1043–55.
49. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 2010;11:119.
50. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via

alignment-free sequence identification. Microb Genom. 2021;7:000685. https://doi.org/10.1099/mgen.0.000685.
51. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36:996–1004.
52. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics. 2019;36:1925–7.
53. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36.
54. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36:2251–2.
55. De Anda V, Zapata-Peñasco I, Poot-Hernandez AC, Eguiarte LE, Contreras-Moreira B, Souza V. MEBS, a software platform to evaluate large (meta) genomic collections according to their metabolic machinery: unraveling the sulfur cycle. Gigascience. 2017;6:1–17.
56. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9.
57. Vasimuddin M, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE international parallel and distributed processing symposium (IPDPS). IEEE; 2019. https://doi.org/10.1109/ipdps.2019.00041
58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. Bioinformatics. 2009;25:2078–9.
59. Herzeel C, Costanza P, Decap D, Fostier J, Wuyts R, Verachtert W. Multi-threaded variant calling in ElPrep 5. PLoS ONE. 2021;16:e0244471.
60. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. InStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat Biotechnol. 2021;39:727–36.
61. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in cancer and somatic tissues. Cell. 2017;171:1029–41.
62. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;35:725–31.
63. Stackebrandt E. The family acidimicrobiaceae. In: The prokaryotes. Berlin, Heidelberg: Springer; 2014. pp. 5–12.
64. Sun W, Xiao E, Krumins V, Dong Y, Li B, Deng J, et al. Comparative analyses of the microbial communities inhabiting coal mining waste dump and an adjacent acid mine drainage creek. Microb Ecol. 2019;78:651–64.
65. Li M, Wen J. Recent progress in the application of omics technologies in the study of bio-mining microorganisms from extreme environments. Microb Cell Fact. 2021;20:178.
66. Vidal P, Martínez-Martínez M, Fernandez-Lopez L, Roda S, Méndez-García C, Golyshina OV, et al. Metagenomic mining for esterases in the microbial community of los rueldos acid mine drainage formation. Front Microbiol. 2022;13:868839.
67. She Z, Wang J, Pan X, Ma D, Gao Y, Wang S, et al. Multi-omics insights into biogeochemical responses to organic matter addition in an acidic pit lake: implications for bioremediation. Water Res. 2024;254:121404.
68. Méndez-García C, Peláez AI, Mesa V, Sánchez J, Golyshina OV, Ferrer M. Microbial diversity and metabolic networks in acid mine drainage habitats. Front Microbiol. 2015;6:475.
69. Li X, Chen D, Carrión VJ, Revillini D, Yin S, Dong Y, et al. Acidification suppresses the natural capacity of soil microbiome to fight pathogenic *Fusarium* infections. Nat Commun. 2023;14:5090.
70. Levett A, Gleeson SA, Kallmeyer J. From exploration to remediation: a microbial perspective for innovation in mining. Earth Sci Rev. 2021;216:103563.
71. Sajjad W, Ilahi N, Kang S, Bahadur A, Banerjee A, Zada S, et al. Microbial diversity and community structure dynamics in acid mine drainage: acidic fire with dissolved heavy metals. Sci Total Environ. 2024;909:168635.
72. Grimaud R, Ezraty B, Mitchell JK, Lafitte D, Briand C, Derrick PJ, et al. Repair of oxidized proteins. Identification of a new methionine sulfoxide reductase. J Biol Chem. 2001;276:48915–20.
73. Malviya D, Varma A, Singh UB, Singh S, Singh HV, Saxena AK. Sulfur-oxidizing bacteria from coal mine enhance sulfur nutrition in pigeonpea (*Cajanus cajan* L). Front Environ Sci. 2022;10:932402. https://doi.org/10.3389/fenvs.2022.932402.
74. Huo X, Yang H, Feng S. Insights into the defensive mechanism of bioleaching microorganisms under extreme environmental copper stress. Rev Environ Sci Biotechnol. 2022;22:79–103. https://doi.org/10.1007/s11157-022-09640-1.
75. Sun J, Liu L, Zhan Y, Song J, Zhang W, Wang B, et al. Comparative genomics provides new insights into pathogenesis of spotting disease causative bacteria in farmed *Strongylocentrotus intermedius*. Aquaculture. 2021;539:736564.
76. Harrison JJ, Ceri H, Turner RJ. Multimetal resistance and tolerance in microbial biofilms. Nat Rev Microbiol. 2007;5:928–38.
77. Grettenberger CL, Hamilton TL. Metagenome-assembled genomes of novel taxa from an acid mine drainage environment. Appl Environ Microbiol. 2021;87:e0077221.
78. Huang J, Wu Y, Gao Q, Li X, Zeng Y, Guo Y, et al. Metagenomic exploration of the rhizosphere soil microbial community and their significance in facilitating the development of wild-simulated ginseng. Appl Environ Microbiol. 2024;90:e0233523.
79. Messyasz A, Männistö MK, Kerkhof LJ, Häggblom MM. Genome analysis and description of *Tunturibacter* gen. nov. Expands the diversity of terriglobia in tundra soils. Environ Microbiol. 2024;26:e16640.
80. Distaso MA, Bargiela R, Johnson B, McIntosh OA, Williams GB, Jones DL, et al. Microbial diversity of a disused copper mine site (Parys Mountain, UK), dominated by intensive eukaryotic filamentous growth. Microorganisms. 2022;10:1694. https://doi.org/10.3390/microorganisms10091694.
81. Reis VM, Teixeira KR dos S. Nitrogen fixing bacteria in the family acetobacteraceae and their role in agriculture. J Basic Microbiol. 2015;55:931–49.
82. Rattray JE, Chakraborty A, Elizondo G, Ellefson E, Bernard B, Brooks J, et al. Endospores associated with deep seabed geofluid features in the eastern Gulf of Mexico. Geobiology. 2022;20:823–36.
83. D'Angeli IM, Ghezzi D, Leuko S, Firrincieli A, Parise M, Fiorucci A, et al. Geomicrobiology of a seawater-influenced active sulfuric acid cave. PLoS ONE. 2019;14:e0220706.
84. Sushmitha TJ, Rajeev M, Sriyutha Murthy P, Ganesh S, Toleti SR, Karutha Pandian S. Bacterial community structure of early-stage biofilms is dictated by temporal succession rather than substrate types in the southern coastal seawater of India. PLoS ONE. 2021;16:e0257961.
85. Ugwuanyi IR, Fogel ML, Bowden R, Steele A, De Natale G, Troise C, et al. Comparative metagenomics at Solfatara and pisciarelli hydrothermal systems in Italy reveal that ecological differences across substrates are not ubiquitous. Front Microbiol. 2023;14:1066406.
86. Levy-Booth DJ, Prescott CE, Grayston SJ. Microbial functional genes involved in nitrogen fixation, nitrification and denitrification in forest ecosystems. Soil Biol Biochem. 2014;75:11–25.

Publisher's note