Uranium mining fuels evolution in deep groundwater microbiomes

Wei Xiu[1], Till L.V. Bornemann[2], Tianjing Zhang[1], André R. Soares[2], Luping Xie[1], Guoxi Lian[3], Haicheng Weng[4], Bing Yang[4], Ruixuan Gai[1], Zhipeng Gao[1], Di Zhang[1], Jonathan R Lloyd[5], Huaming Guo[1]*, and Alexander J. Probst[2]*

[1] *State Key Laboratory of Geomicrobiology and Environmental Changes (GMEC), China University of Geosciences, Beijing 100083, People's Republic of China*

[2] *Environmental Metagenomics, Research Center One Health Ruhr, University Alliance Ruhr, Faculty of Chemistry, University of Duisburg-Essen, Germany*

[3] *Nuclear and Radiation Safety Center, Ministry of Ecology and Environment, Beijing, 100082, China*

[4] *Beijing Research Institute of Chemical Engineering and Metallurgy CNNC, No.145 Jiukeshu, Tongzhou District, Beijing 101149, China*

[5] *Williamson Research Laboratories for Molecular Environmental Science, Department of Earth and Environmental Sciences, The University of Manchester, Manchester M13 9PL, United Kingdom*

* to whom correspondence should be addressed: Prof. Alexander J. Probst (alexander.probst@uni-due.de); Prof. Huaming Guo (hmguo@cugb.edu.cn)

Abstract

Microbial communities in the deep biosphere underpin global biogeochemical cycles, yet their evolutionary responses to extensive anthropogenic disturbances remain unclear. Uranium in-situ leaching (ISL), a dominant mining strategy, creates persistent radiochemical and redox gradients in aquifers, providing a unique system for studying subsurface microbial evolution. Here, by exploiting metagenomics and metatranscriptomics in examining microbial responses through various mining stages at neutral U ISL sites, we show that geochemical changes accompany significant alterations in microbial community composition, oxidative stress, biogeochemical cycling, and overexpression of DNA repair genes. Comparative metagenomics and metatranscriptomics revealed elevated non-synonymous to synonymous polymorphism ratios and altered nucleotide diversity correlated with radiochemical gradients, underscoring the influence of environmental stress in promoting microbial evolution. Our results demonstrate that neutral ISL drives adaptive genetic and functional diversification of subsurface microbiomes and extend core microbial evolutionary frameworks to anthropogenically modified subsurface environments, redefining the deep biosphere as a dynamic, responsive system.

Introduction

Microbial consortia in the deep subsurface adapted to nutrient-scarce, anoxic environments[1–4] mediate critical biogeochemical processes and help maintain groundwater quality[5–8]. However, anthropogenic stresses arising from mining, groundwater extraction, and energy production are perturbing these fragile environments at increasing rates[9]. Recent studies have shown that environmental stresses, including heavy metal contamination and altered pH, affect groundwater microbial communities[10,11]. Mining operations are particularly disruptive, as they rapidly subject pristine subsurface ecosystems to heavy metals, radioactive contaminants, altered redox conditions, and electrochemical gradients[9]. Notably, a recent study in the abyssal seafloor showed that deep-sea mining can cause biological disruptions persisting for decades[12], underscoring the need to assess long-term impacts of subsurface resource extraction on microbiomes. Although distinct from the deep subsurface examined here, this finding underscores how resource extraction can impose long-lasting impacts on microbial communities in Earth's hidden biomes.

In-situ leaching (ISL), a widely used uranium (U) mining technique, is based on injecting oxidizing agents into aquifers to mobilize U by oxidizing insoluble U(IV) to soluble U(VI). In addition to fundamentally altering groundwater chemistry[13], this activity imposes profound chemical and radiological stress on microbial communities that have evolved over eons under stable, reducing conditions. Whereas the environmental consequences of ISL are well documented mostly via 16S rRNA gene analysis[11,14–16], the evolutionary responses of microbial communities to these stressors remain poorly understood. Indeed, while mining-driven alterations to groundwater chemistry are expected to disrupt the microbial community structure,

66 functional potential, and genetic diversity[17–20], the question remains as to whether
67 anthropogenic stress can drive microbial evolution in deep aquifers.

68 To this end, we implement innovative metagenomic and metatranscriptomic
69 strategies to examine microbial community dynamics and evolutionary adaptations at
70 a neutral-pH ISL area (CO_2 + O_2 ISL). It involved sampling aquifers spanning mining
71 spanning pre-mining, active mining, and post-mining phases. Leveraging this rare
72 sampling opportunity to study microbial adaptation as a response to mining and
73 geochemical changes, we show that U exposure and radiochemical stress drive the
74 restructuring of microbial communities, oxidative stress response and DNA repair, and
75 alterations in genetic diversity. Elevated pN/pS ratios (the ratio of the rates of non-
76 synonymous (pN) and synonymous (pS) polymorphism) and shifts in nucleotide
77 diversity implicate mining-induced environmental gradients in significantly altering
78 microbial evolutionary processes. Our findings yield new insights into how deep
79 subsurface microbiomes structurally and evolutionally adapt to anthropogenic stress,
80 bearing implications for global biogeochemical cycles, ecosystem resilience, and
81 bioremediation strategies. As mining operations continue to intersect with fragile
82 groundwater ecosystems, understanding the evolutionary implications to subsurface
83 microbial communities will be of critical importance to predicting the long-term impacts
84 of resource extraction from Earth's deep biosphere. The findings of this investigation
85 underscore the consequences of uranium ISL on the evolution of subsurface life,
86 offering a novel framework for appraising microbial resilience and adaptation in
87 mining-impacted deep subsurface ecosystems.

88

89 **RESULTS**

90 **Geochemical and radiological alterations during and following Uranium ISL**

Neutral pH uranium *in-situ* leaching (U ISL) commences while the aquifer is in its natural state (pre-mining phase), proceeds through the injection of leaching solutions (mining phase), and concludes with natural restoration efforts upon cessation of mining activities (post-mining phase; *Fig. 1*). We monitored the all phases of neutral-pH ISL, having collected 80 groundwater samples from four distinct mining sites (S1, S2, S3a, and S3b; *Figs. 1b-c*). These sites underwent neutral ISL for varying durations, creating a unique time-resolved dataset with considerable variation in their redox conditions, from neutral-reducing pre-mining (BG, baseline groundwater) to weakly acid-oxidizing during mining (MT, mining groundwater), and back to neutral-reducing post-mining (PMI, PMIS, and PMS, representing initial, intermediate, and stable post-mining phases; *Figs. 1b–c*).

In the neutral-reducing pre-mining phase, low concentrations of total dissolved uranium (TU; 0.44 ± 0.31 mg/L*; Fig. 2b*) were detected, indicative of ongoing U mineralization under reducing conditions. This environment was rich in electron donors, including dissolved organic carbon (DOC, 5.26 ± 3.51 mg/L), Fe(II) (0.85 ± 1.11 mg/L), ammonium (0.73 ± 0.70 mg/L), and sulfide (0.92 ± 1.18 mg/L) (*Fig. 2* and *Table* S1). These sources likely fuelled the reduction of Fe(III), and sulfate, and enhanced the reductive immobilization of U in the aquifer.

During the oxidizing mining phase, the injection of CO_2 + O_2 mobilized U from ore deposits. Total dissolved uranium (TU) concentrations increased rapidly after *in-situ* leaching activities commenced (from MT0.25 to MT1), peaking after roughly four years (MT4), followed by a gradual decline thereafter (from MT5 to MT20) (*Fig. 2b*). The oxidation of U(IV) to U(VI) and the subsequent formation of soluble uranyl-carbonate complexes were the primary drivers of TU increase[21,22], as evidenced by strong positive correlations between TU and bicarbonate concentrations (Spearman

correlation, ρ = 0.43, n = 80, $p < 0.0001$) (*Fig. S1*). Significant increases in sulfate and nitrate were also observed during this phase, both positively correlated with TU (sulfate-TU, ρ = 0.65, n = 80, $p < 0.0001$; nitrate-TU, ρ = 0.69, n = 80, $p < 0.0001$) (*Fig. S1*), likely resulting from the oxidation of pyrite and ammonium, respectively[21,22].

Upon ISL decommissioning, the groundwater redox potential (Eh) decreased significantly, likely due to the degradation of dissolved organic matter (DOM; particularly protein-like and humic substances[23]). This transition from oxic to anoxic conditions was further evidenced by a decline in TU, sulfate, and sulfate/Cl ratios declined, while Fe(II) concentrations increased. This suggested the microbial reduction of Fe(III) and sulfate, alongside the reductive natural attenuation of U(VI)[23]. TU concentrations did not return to pre-mining levels (*Fig. 2*), however, indicative of incomplete restoration of the aquifers' original geochemical conditions.

As U and its decay products emit α and β radiation, changes in groundwater radioactivity can serve as indicators of U mobilization and contamination. Monitoring α and β activity therefore during mining operations, with elevated gross (α + β) values indicating radiogenic element mobilization, while declining levels after mining suggest the efficacy of natural attenuation processes in limiting further U migration. Here, gross (α + β) radiation levels increased significantly after mining activities commenced, but decreased following ISL decommissioning (Figs. *2c-e*). These changes in radioactivity correlated strongly with TU concentrations (ρ = 0.24, n = 80, $p < 0.0001$; Fig. *S1*), underscoring the coupling between uranium mobilization and radiogenic activity during mining and attenuation phases.

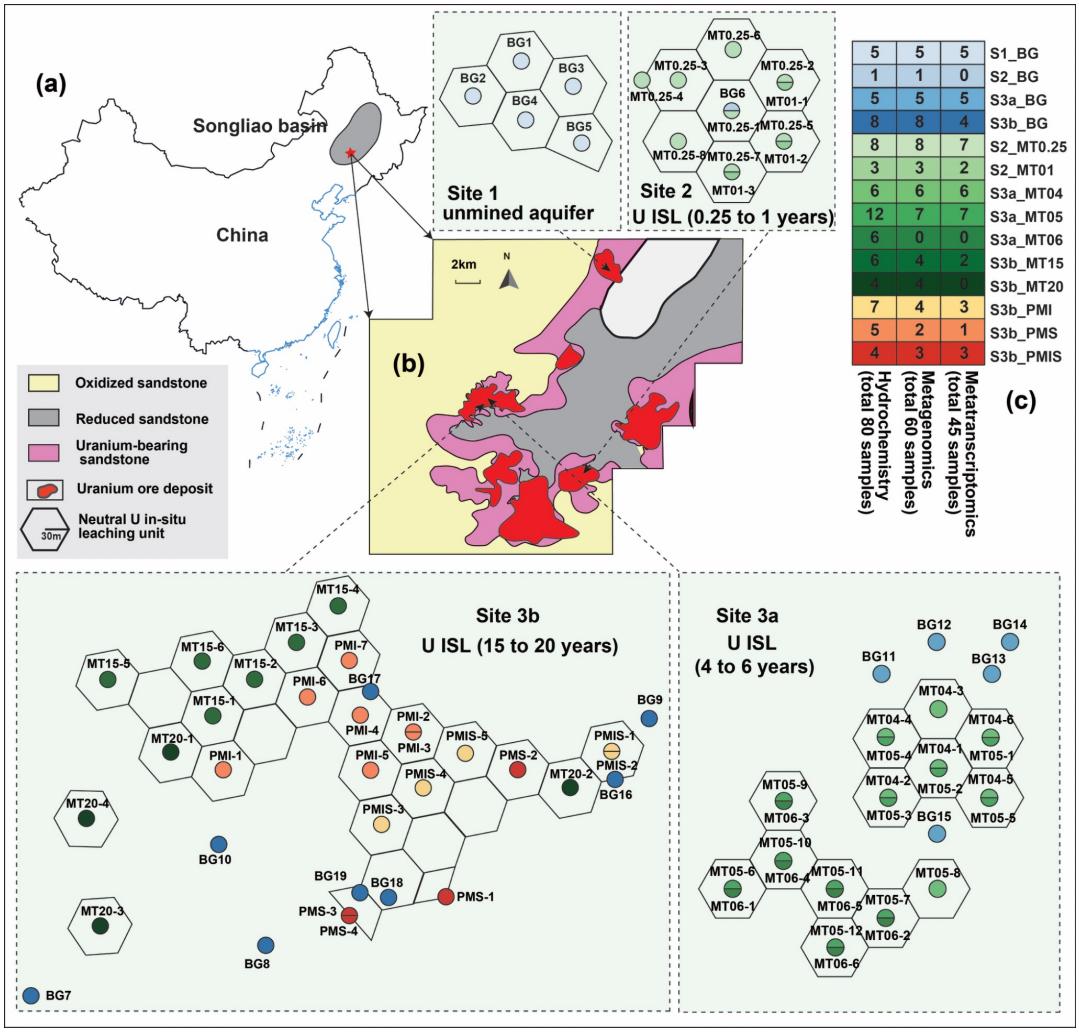


Figure 1 | Sampling locations and study overview. *(a) Geographic location of the Songliao Basin and broader regional context, modified from a previously published source[24]. (b) Schematic layout of study sites along a uranium ISL chronosequence: Site 1, unmined aquifer; Site 2, recent ISL (0.25–1 yr); Site 3a, intermediate ISL (4–6 yr); and Site 3b, long-term ISL (15–20 yr), which after decommissioning transitioned through post-mining phases of iron-reducing (PMI), mixed iron- and sulfate-reducing (PMIS), and sulfate-reducing (PMS) conditions. Wells shown with two names indicate that the same well was sampled at different time points. The background map shows oxidized sandstone (yellow), reduced sandstone (grey), uranium-bearing sandstone (pink), uranium ore deposits (red), and ISL units (arrow pointers). (c) Numbers of groundwater samples collected for hydrochemical, metagenomic, and*

metatranscriptomic analyses across stages of uranium ISL. Colors indicate groundwater type: blue, background aquifer (BG); green, mining phase (MT); orange, post-mining phases (PMI, PMIS, PMS). Sample codes (e.g., S2-BG or S3b-MT) combine the site number (S1–S3b) with the groundwater type: BG, background; MT, mining; PMI, post-mining initial; PMIS, post-mining intermediate; PMS, post-mining stable. Background wells were sampled at multiple sites throughout the chronosequence.

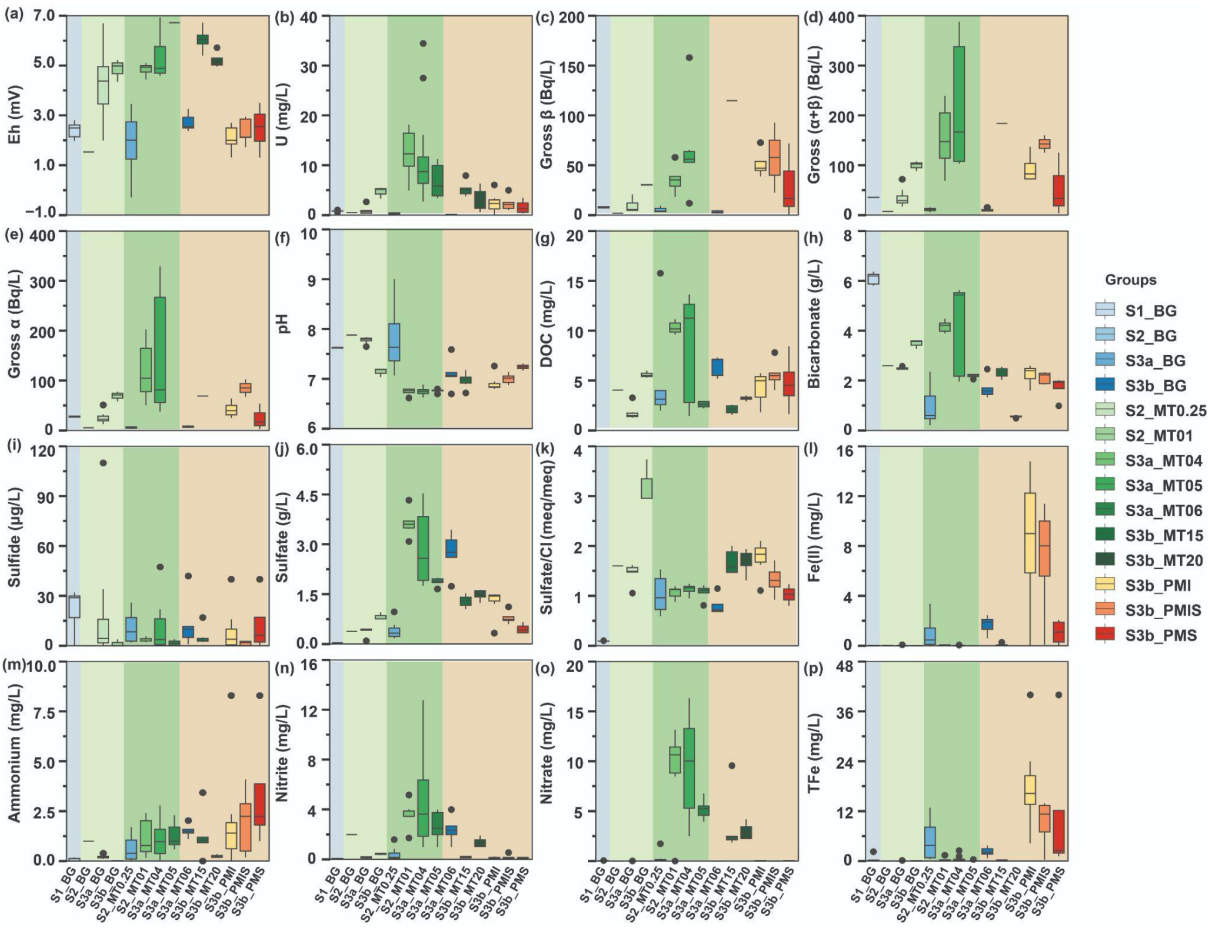


Figure 2 | Geochemical profiles of groundwater across neutral-pH ISL phases. *Boxplots show geochemical parameters through the stage of neutral-pH ISL, including pre-mining (BG), mining (MT), and post-mining (PM). Sampling groups include: Site 1 (unmined aquifer, S1_BG); Site 2 (unmined aquifer, S2_BG; recent ISL at 0.25-1 years: S2_MT0.25 and S2_MT01); Site 3a (unmined aquifer, S3a_BG; intermediate*

164 *ISL, 4-6 years: S3a_MT04, S3a_MT05, and S3a_MT06); and Site 3b (unmined*
165 *aquifer, S3b_BG; long-term ISL, 15-20 years: S3b_MT15 and S3b_MT20; and*
166 *decommissioned ISL under iron-reducing, S3b_PMI, mixed iron- and sulfate-reducing,*
167 *S3b_PMIS, and sulfate-reducing, S3b_PMS conditions). Panels represent (a – p) Eh,*
168 *gross (α+β), pH, DOC, bicarbonate, sulfide, sulfate, sulfate/Cl, Fe(II), ammonium,*
169 *nitrite, nitrate, and total Fe. Colored backgrounds denote different ISL phases. Colors*
170 *indicate study sites: light blue (Site 1), light green (Site 2), medium green (Site 3a),*
171 *and beige (Site 3b).*
172
173 **Thousands of metagenome-assembled genomes reveal microbiome changes**
174 **and enhanced co-occurence resulting from *in-situ* mining activity**
175 Sixty groundwater samples representative of the pre-mining, mining, and post-
176 mining conditions across more than 20 years of neutral-pH ISL were metagenomically
177 sequenced (*Fig. 1*). Nonpareil[25] analysis indicated that sequencing coverage,
178 exceeding 75%, was of sufficient depth to accurately reflect the aquifer microbiome
179 diversity (*Fig. S2*). To assess both bacterial and archaeal diversity, we focused on
180 analyzing the ribosomal protein S3 (*rpS3*) gene[26], a single-copy, universally conserved
181 marker that provides high phylogenetic resolution for both bacteria and archaea,
182 yielding 12,972 sequences that clustered (at 99% similarity) into 6,487 representative
183 sequences. Taxonomic assignments based on the Genome Taxonomy Database
184 (GTDB, r220) grouped 6,431 rpS3 sequences into 215 distinct phyla (21 archaeal, 194
185 bacterial; *Fig. S3*). Groundwater microbial community composition and alpha diversity,
186 based on *rpS3* gene profiles, were strongly influenced by geochemical gradients,
187 particularly uranium concentration and radiochemical activity, with higher diversity and

188 distinct community structures observed under elevated stress conditions (*Figs. S4 and*
189 *S5)*.

190 Through bioinformatic analysis, we identified 4,036 medium-quality
191 metagenome-assembled genomes (MAGs; completeness > 50%, contamination <
192 10%)[27], from which 2,255 strain-cluster representative MAGs (strain-cluster rMAGs)
193 were obtained via dereplicating at strain-level (99% ANI). These included 1,165 high-
194 quality and 1,090 medium-quality strain-cluster rMAGs bearing a median
195 completeness of 96.5%, contamination of 1.7%, and N50 of 50,891 (Fig. *3* and *Table*
196 S2). These MAGs represented 64 of the 215 phyla detected and accounted for roughly
197 35% of the total phylum-level diversity detected via *rpS3*, thus providing a robust
198 representation of the predominant microbial communities at Qianjiadian sites (*Figs.*
199 *S3, S6, and S7*). The phyla comprising the remaining 6.9% of rpS3-based abundance
200 were likely absent from the strain-cluster rMAGs because their low abundance
201 hindered successful genome reconstruction (*Fig. 2a*).

202 Neutral-pH ISL significantly altered microbial diversity in the mining aquifer, as
203 revealed by *rpS3* gene profiles. Microbial diversity showed strong positive correlations
204 with uranium concentration and radioactivity, with prolonged neutral-pH
205 ISL(S3a_MT04 and S3a_MT05) associated with increased Shannon and Invsimpson
206 indices (ANOVA, $p < 0.05$; Tukey HSD post-hoc test; *Fig. S4*). Microbial community
207 composition shifted markedly across the ISL chronosequence, as shown by non-
208 metric multidimensional scaling (NMDS) and sample-to-sample correlation analyses
209 of strain-cluster rMAGs (*Figs. 3e, 4a–d*). Communities from pre-mining, mining, and
210 post-mining stages formed distinct clusters, with ANOSIM indicating significant
211 separation between stages ($R = 0.72$, $p < 0.001$; Fig. 4d), reflecting divergent
212 community trajectories during and after ISL. Network analysis of relative abundances

213 of strain-cluster rMAGs revealed positive correlations, accounting for 90.4% of all
214 edges, defined as significant correlations between taxa, dominating microbial
215 interactions (*Fig. 4e*). Four distinct modules were identified, with microbial
216 compositions of Module 1 remaining stable throughout neutral-pH ISL, indicative of a
217 core microbial community. By contrast, the microbial compositions of Modules 2 and
218 4 were more prevalent during the early stages of neutral-pH ISL (indicative of
219 adaptation), while that of Module 3 associated primarily with post-mining (indicative of
220 recovery and/or attenuation; *Figs. 4f-i*). These results highlight the significant shifts in
221 microbial community structure, diversity, and ecological interactions imposed by
222 neutral-pH ISL. The observed tendency towards positive interactions emphasizes the
223 dynamic and evolving nature of subsurface microbiomes in response to acute
224 environmental stress, although such co-occurrence may arise from a variety of
225 ecological processes .

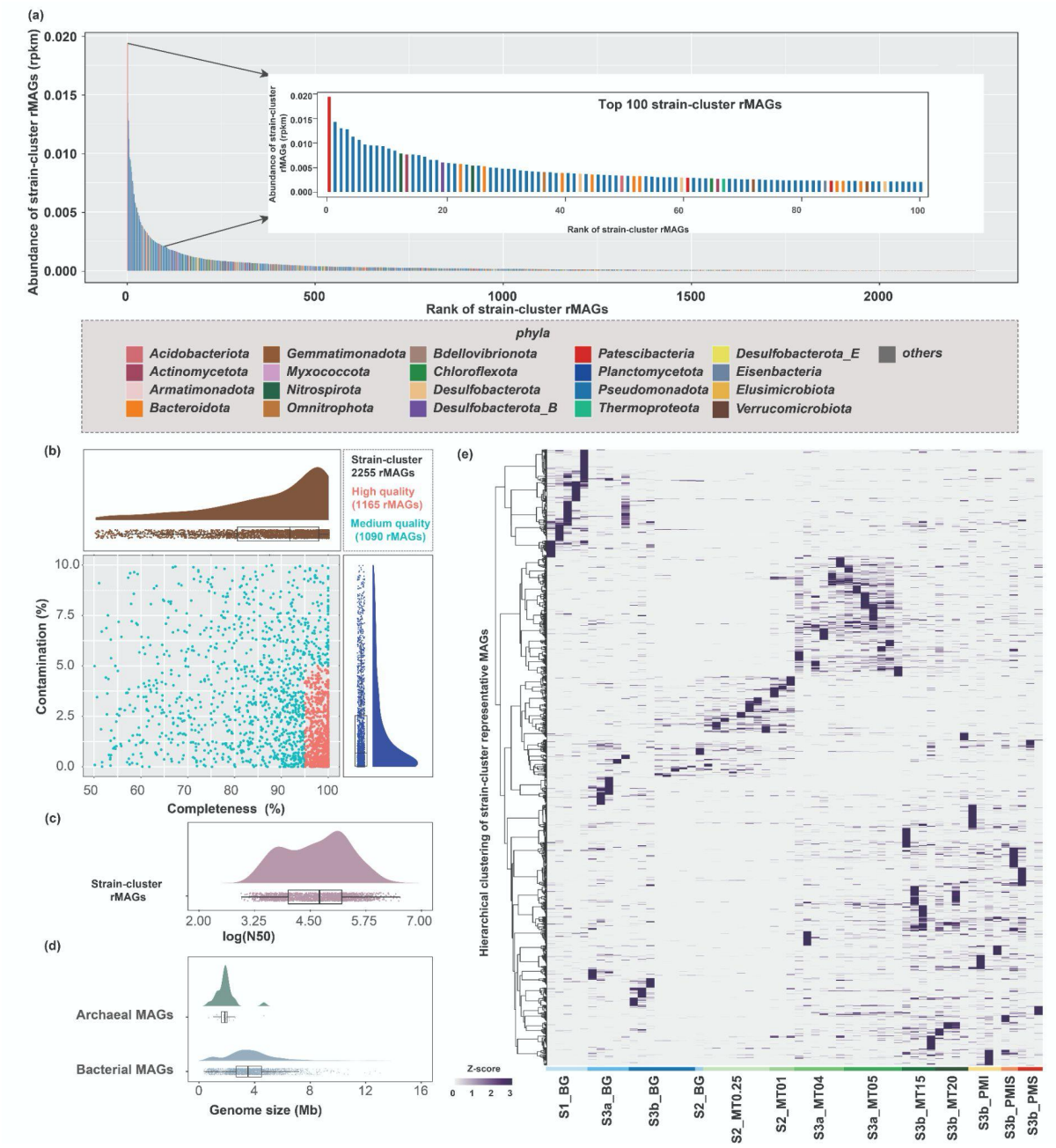


Figure 3 | Characteristics and distribution of strain-cluster representative metagenome-assembled genomes (strain-cluster rMAGs). *(a) Rank abundance of strain-cluster rMAGs across phyla, with the top 100 depicted in more detail. The displayed abundances represent the average strain-cluster rMAG abundances across all samples. (b) Strain-cluster rMAGs completeness and contamination by quality tier. (c) Distribution of log(N50) values of strain-cluster rMAGs. (d) Genome size comparison between archaeal and bacterial strain-cluster rMAGs. (e) Heatmap of*

strain-cluster rMAGs Z-score normalized relative abundance across sampling sites, clustered by similarity. Sampling groups include: Site 1 (unmined aquifer, S1_BG); Site 2 (unmined aquifer, S2_BG; recent neutral-pH ISL at 0.25-1 years: S2_MT0.25 and S2_MT01); Site 3a (unmined aquifer, S3a_BG; intermediate neutral-pH ISL, 4-6 years: S3a_MT04, S3a_MT05, and S3a_MT06); and Site 3b (unmined aquifer, S3b_BG; long-term neutral-pH ISL, 15-20 years: S3b_MT15 and S3b_MT20; and decommissioned neutral-pH ISL under iron-reducing, S3b_PMI, mixed iron- and sulfate-reducing, S3b_PMIS, and sulfate-reducing, S3b_PMS conditions).

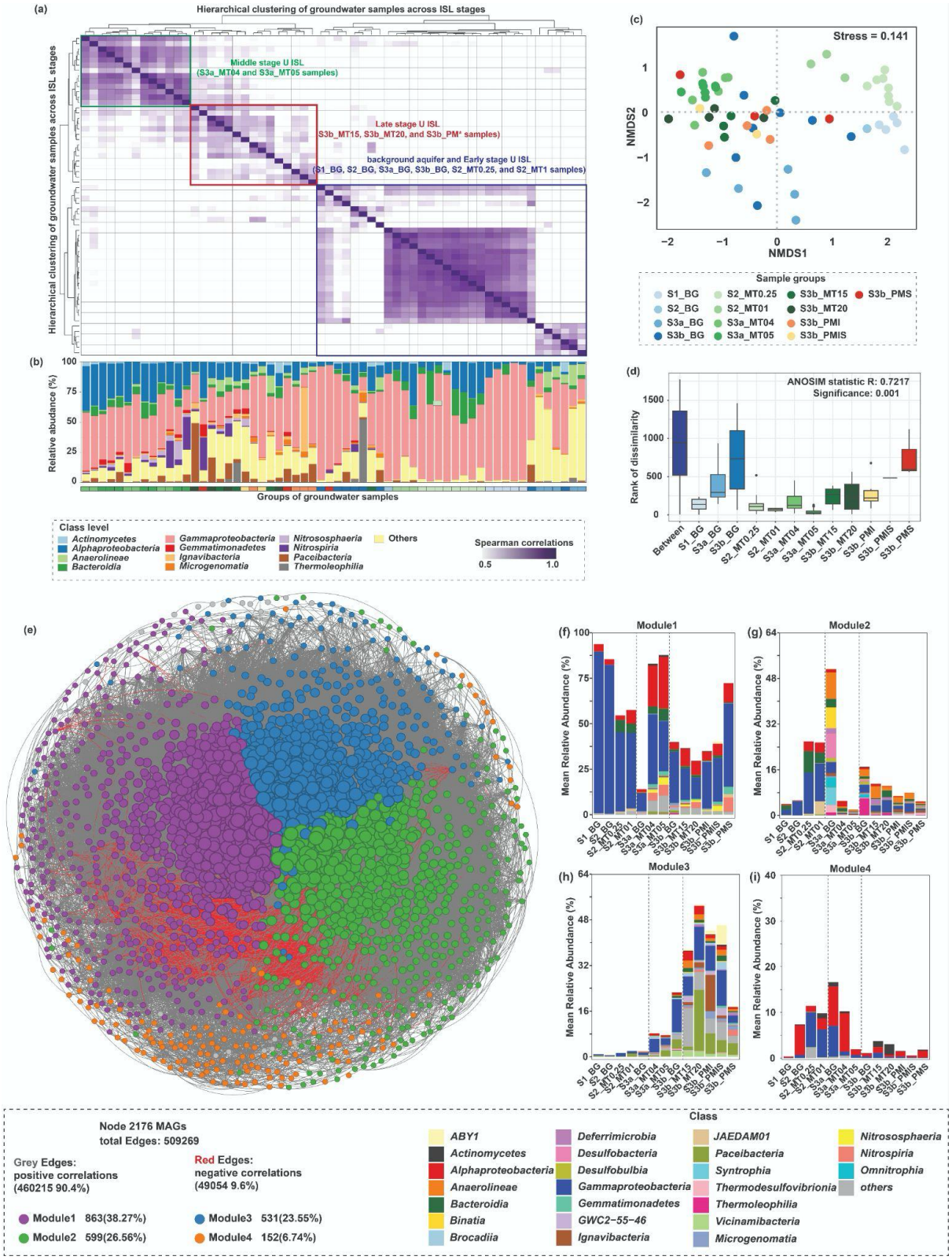


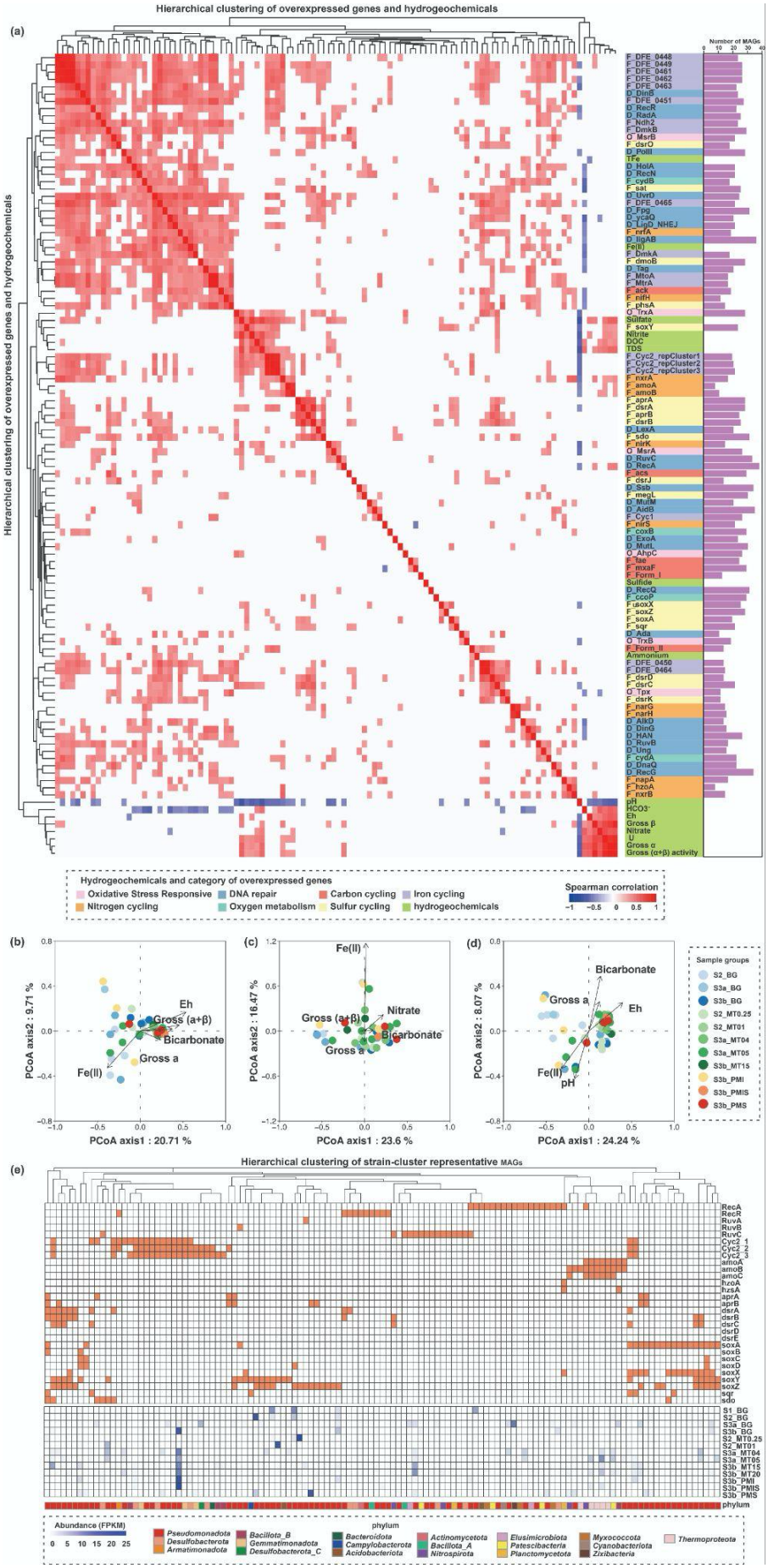
Figure 4 | Microbial community structure and modularity across neutral-pH ISL phases. *(a) Heatmap of sample-to-sample Spearman's rank correlations of strain-*

246 *cluster rMAG abundances across early (blue), mid (green), and late (red) neutral-pH*
247 *ISL stages. (b) Taxonomic composition of top 12 phyla across samples, with the*
248 *sample order consistent with that in the heatmap. (c) Non-metric multidimensional*
249 *scaling (NMDS) plot of community structure based on strain-cluster rMAGs (stress =*
250 *0.141). (d) Analysis of Similarity (ANOSIM) showing significant differences between*
251 *groups (R = 0.7217, p = 0.001). (e) Network of strain-cluster rMAGs co-occurrence,*
252 *with positive (grey) and negative (red) Spearman correlations grouped into four*
253 *modules. (f–i) Distribution of the abundance of module-specific taxa across neutral-pH*
254 *ISL phases.*
255
256 **Neutral-pH ISL impacts gene expression linked to DNA repair, oxidative stress,**
257 **and biogeochemical cycling**

258 To investigate microbial metabolic adaptations in response to long-term
259 neutral-pH ISL, we analyzed gene expression from 45 metatranscriptomic samples
260 representative of pre-mining, mining, and post-mining conditions over a 20-year
261 period. By integrating gene expression profiles with our suite of 2,255 de-replicated
262 and functionally annotated strain-cluster rMAGs, we identified widespread
263 overexpression (coverage increase at least 2-fold) of genes involved in DNA repair,
264 oxidative stress responses, and biogeochemical cycling (*Fig. S8-S14*). A
265 corresponding Spearman correlation heatmap revealed coherent expression patterns
266 and their associations with key environmental drivers (*Fig. 5a*). Principal coordinate
267 analysis (PCoA) identified key environmental variables, such as Fe(II), gross alpha
268 radioactivity, and bicarbonate, as major drivers of gene expression during different
269 neutral-pH ISL phases (*Figs. 5b-d*). Specifically, Fe(II), gross radioactivity (alpha and
270 beta) and redox potential (Eh) associated with the overexpression of DNA repair genes

(*Fig. 5b*), while Fe(II) and nitrate correlated with oxidative stress response gene expression (*Fig. 5c*). Changes in bicarbonate, Fe(II), Eh and pH affected genes involved in biogeochemical cycling (*Fig. 5d*). These findings indicate that both geochemical and radiochemical gradients jointly shape microbial metabolic responses during neutral-pH ISL.

Microbial populations at the 20-year ISL site exhibited extensive DNA damage responses, as shown by the overexpression of genes involved in homologous recombination, nucleotide excision repair, and mismatch repair pathways (*Figs. 5a and 5e*), reflecting adaptation to the persistently oxidative, metal-rich, and radiological conditions maintained during prolonged neutral-pH ISL. The abundance of a key overexpressed gene (*recA*) involved in homologous DNA repair[28], was moderately correlated with gross (α + β) radioactivity ($\rho = 0.52$, n = 32, $p < 0.05$) (*Fig. 5e*). Similarly, U concentrations showed moderate associations with increased expression of *soxY*, a gene involved in the sulfur oxidation (Sox) system ($\rho = 0.44$, n = 26, $p < 0.05$), as well as with genes linked to ammonium oxidation, including *amoA* ($\rho = 0.43$, n = 8, $p < 0.05$) and *amoB* ($\rho = 0.43$, n = 10, $p < 0.05$) (*Fig. 5e*). Functional genes associated with iron, nitrogen, and sulfur redox transformations, processes that regulate electron flow and thereby influence U speciation and mobility, were widely distributed across 100 strain-cluster rMAGs. These included *Cyc2* cluster genes mediating Fe(II) oxidation in Gallionella spp. (potentially affecting U(IV) oxidation), *amoAB* genes for ammonia oxidation in *Nitrosarchaeum* spp., and *soxY* genes for sulfur oxidation in *Pseudomonadota* spp. (*Fig. 5e*). This functional consistency, potentially underpinned by redundancy and convergent adaptation, points to environmental filtering as a key mechanism structuring microbial communities under sustained geochemical stress[29–33].


Hierarchical clustering of overexpressed genes and hydrogeochemicals
(a)
Number of MAGs
Hydrogeochemicals and category of overexpressed genes
Oxidative Stress Responsive
DNA repair
Carbon cycling
Iron cycling
Nitrogen cycling
Oxygen metabolism
Sulfur cycling
hydrogeochemicals
Spearman correlation
(b)
PCoA axis2 : 9.71 %
PCoA axis1 : 20.71 %
Eh
Gross (a+β)
Bicarbonate
Gross a
Fe(II)
(c)
PCoA axis2 : 16.47 %
PCoA axis1 : 23.6 %
Fe(II)
Nitrate
Gross (a+β)
Bicarbonate
Gross a
(d)
PCoA axis2 : 8.07 %
PCoA axis1 : 24.24 %
Bicarbonate
Gross a
Eh
Fe(II)
pH
Sample groups
S2_BG
S3a_BG
S3b_BG
S2_MT0.25
S2_MT01
S3a_MT04
S3a_MT05
S3b_MT15
S3b_PMI
S3b_PMIS
S3b_PMS
(e)
Hierarchical clustering of strain-cluster representative MAGs
phylum
Abundance (FPKM)
Pseudomonadota
Desulfobacterota
Armatimonadota
Bacillota_B
Gemmatimonadota
Desulfobacterota_C
Bacteroidota
Campylobacterota
Acidobacteriota
Actinomycetota
Bacillota_A
Nitrospirota
Elusimicrobiota
Patescibacteria
Planctomycetota
Myxococcota
Cyanobacteriota
Zixibacteria
Thermoproteota

297 **Figure 5 | Correlation and environmental drivers of overexpressed genes in**
298 **metatranscriptomic datasets.** *(a) Spearman correlation heatmap illustrates*
299 *associations between hydrogeochemicals and the expression levels of overexpressed*
300 *genes (defined as transcript coverage > 2× the average 30S coverage), detected in at*
301 *least six samples across the dataset. Rows and columns are clustered by gene*
302 *expression patterns. Color scale denotes Spearman correlation coefficients (red =*
303 *positive, blue = negative), which the p values were adjusted using the BH method. The*
304 *right-side bar plot indicates the number of strain-cluster rMAGs expressing each gene.*
305 *Colored annotations indicate functional categories: oxidative stress response, DNA*
306 *repair, carbon cycling, nitrogen cycling, sulfur cycling, iron cycling, oxygen*
307 *metabolism, and hydrogeochemicals. Principal coordinate analysis (PCoA) of MAGs*
308 *based on expression of DNA repair genes (b), oxidative stress response genes (c),*
309 *genes involved in carbon, nitrogen, sulfur, and iron (C-N-S-Fe) cycling (d). Arrows*
310 *indicate environmental variables significantly correlated with expression patterns.*
311 *Points are colored and shaped by sample groups. (e) Heatmap showing presence or*
312 *absence of overexpressed genes across strain-cluster rMAGs. Gene abundance*
313 *(FPKM) is shown by color intensity, with taxa classified by phylum. The MAGs were*
314 *clustered based on the presence of the functional genes.*
315

316 **Uranium ISL-induced stress drives genetic diversification in subsurface**
317 **aquifers**

318 Genetic diversity and evolutionary pressures were assessed for species-cluster
319 rMAGs by calculating non-synonymous to synonymous polymorphism (pN/pS) ratios,
320 which indicate the balance between neutral variation and adaptive selection, and by
321 appraising nucleotide diversity via PCoA to capture overall genetic variation among

322 populations. Only genes with greater than 5× coverage in both metagenomic and
323 metatranscriptomic datasets were included, and analyses focused on dereplicated
324 species-cluster rMAGs (95% ANI) containing more than three clustered members
325 (1,733 species-cluster rMAGs). This ensured robust comparisons across neutral-pH
326 ISL phases. PCoA revealed that, based on metagenomic and metatranscriptomic
327 data, both nucleotide diversity (Fig. 6g, i) and pN/pS ratios (Fig. 6h, j) were influenced
328 by groundwater geochemistry. Nucleotide diversity was positively associated with
329 bicarbonate, gross (α + β) radioactivity, sulfate, DOC, and Eh, and negatively with
330 sulfide and pH (*bioenv* correlation analysis R^2, 0.27). Similar patterns were observed
331 for pN/pS ratios, although with slightly reduced explanatory strength (*bioenv*
332 correlation analysis R^2, 0.31). These results indicate that U concentrations, redox
333 conditions, and radiochemical stressors all contribute to alterations in microbiome
334 genetic diversity and evolutionary trajectory during and following neutral-pH ISL.

335 Across metagenomic datasets, no significant differences were observed in
336 pN/pS ratios or nucleotide diversity during neutral-pH ISL. By contrast,
337 metatranscriptomic data revealed elevated pN/pS ratios at both mining-impacted and
338 post-mining sites, with median values increasing from 0.585 in S3a_BG to 0.608 in
339 S3a_MT05, and from 0.380 in S3b_BG to 0.503 in S3b_MT15 and 0.491 in S3b_PMS,
340 respectively (*Figs. 6b, c*). Notably, non-synonymous SNPs were detected in 481 genes
341 that were transcriptionally overexpressed and functionally linked to oxidative stress
342 responses, DNA repair, and cycling of carbon, nitrogen, sulfur, and iron (*Fig. S15b*),
343 suggesting that positive selection preferentially targeted actively transcribed genes
344 under stress. Notably, pN/pS ratios exceeding 1 were predominantly found in the
345 metatranscriptomic datasets, with 7 species-cluster rMAGs in BG, 47 species-cluster
346 rMAGs in MT, and 5 species-cluster rMAGs in PM, while only 5 species-cluster rMAGs

347 (four in MT and one in PM) exhibited pN/pS ratios exceeding 1 in both metagenomic
348 and metatranscriptomic datasets (*Fig. S16*). These results suggest that elevated
349 genetic variation during ISL is linked to transcriptionally active taxa responding to
350 combined radiochemical and geochemical perturbations, including oxidative stress
351 and the influx of new substrates such as oxygen from leaching solutions, rather than
352 a simple genome-wide increase in mutation rates. This framework also explains why
353 post-mining communities, despite residual radioactivity, exhibit reduced diversity as
354 redox conditions stabilize and substrate inputs diminish. The fact that no significant
355 changes were detected in the early mining phase (*Fig. 6a*) suggests that the
356 extent/duration of mining plays a major role in the process of natural selection.
357 However, it should be noted that only one sample (BG6 in S2 site) from the
358 background aquifer was available during the early mining phase, which may limit the
359 conclusions drawn from these data and underscores the need for further sampling to
360 better understand evolutionary dynamics at this stage.

361 Metatranscriptomic data revealed distinct evolutionary responses to neutral-pH
362 ISL across mining stages, marked by dynamic shifts in nucleotide diversity and
363 species-level selection. During the mid-mining phase (MT in S3a site), nucleotide
364 diversity significantly increased relative to both background and other mining stages
365 (*Fig. 6d–f*), coinciding with peak uranium and radioactivity levels. This suggests that
366 sustained radiochemical and oxidative stress promoted elevated genetic variation,
367 potentially by enabling the persistence of diverse genotypes capable of tolerating
368 extreme conditions. By contrast, the early and late mining phases exhibited reduced
369 nucleotide diversity, likely reflecting genetic bottlenecks driven by acute oxidative
370 perturbations and the selective loss of more vulnerable populations. Notably, post-
371 mining samples also showed diminished nucleotide diversity, despite the fact that both

TU and radioactivity levels remained elevated. These findings suggest that both the intensity and duration of oxidative and radiochemical stress play a major role in shaping genetic diversity in mining-impacted aquifers.

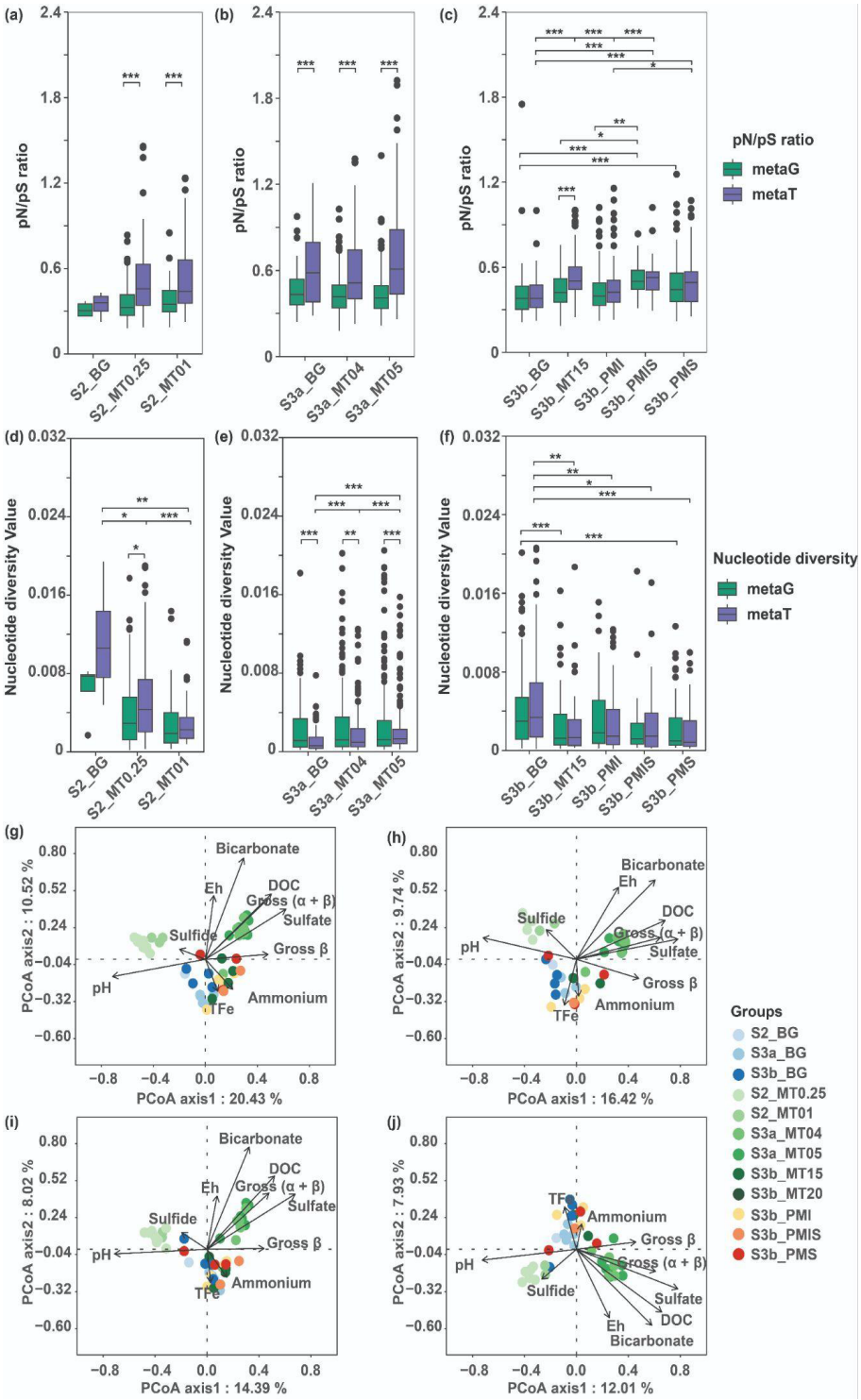


Figure 6 | pN/pS ratio and nucleotide diversity. *(a-c) Boxplots of pN/pS ratio from metagenomic (metaG) and metatranscriptomic (metaT) datasets spanning pre-mining,*

*mining, and post-mining phases; (d-f) Boxplots of nucleotide diversity from metaG and metaT data across selected neutral-pH ISL stages; (g-j) PCoA plots showing the influence of geochemical parameters on pN/pS ratio and nucleotide diversity derived from metaG (g, h) and metaT (i, j) datasets.Statistical significance was assessed using the Kruskal–Wallis test followed by Dunn's post-hoc test with Benjamini–Hochberg (BH) correction for multiple comparisons (*p < 0.05, **p < 0.01, and ***p < 0.001). Differences in pN/pS ratios and nucleotide diversity between metaG and metaT for each group were assessed using two-sample t-tests (*p < 0.05, **p < 0.01, and ***p < 0.001).*

To better understand the genomic basis of adaptive responses, we identified 31 species-cluster rMAGs that exhibited consistent increases in nucleotide diversity from background (BG) to mining-impacted (MT) conditions (*Figs. 7a-b*). These rMAGs spanned diverse taxonomic groups, such as *Bacteroidales*, *Gallionella*, *Hyphomonas*, and *Kuenenia*, suggesting that the response to mining-related stress was phylogenetically widespread (*Fig. 7c*). Gene-level analysis showed that elevated nucleotide diversity was spread across many genes within each species-cluster rMAG, rather than being limited to a few specific regions, suggesting widespread mutational pressure under long-term environmental stress (*Fig. 7c*). Functional categorization of highly variable genomic regions showed enrichment in genes associated with information processing, regulatory networks, and cellular signaling, functions that are likely essential for coping with radiochemical and oxidative stress. Notably, the majority of these diversity increases were detected in metatranscriptomic data, reinforcing the interpretation that selection is preferentially acting on actively expressed genes. These patterns suggest that under prolonged mining stress, a

subset of microbial populations, those that remain transcriptionally active, exhibit elevated genetic diversity, potentially as a bet-hedging strategy to enhance adaptive flexibility. However, the absence of a community-wide increase in nucleotide diversity suggests that evolutionary responses to ISL are concentrated in metabolically active, redox-cycling taxa (e.g., iron-oxidizing Gallionella and nitrogen-oxidizing Nitrosarchaeum), while many other lineages remain genetically stable or enter dormancy under extreme stress. This indicates that anthropogenic disturbance does not uniformly accelerate microbial evolution, but instead drives targeted adaptation in functional groups central to biogeochemical cycling. This decoupling between metagenomic and metatranscriptomic signals highlights that evolutionary responses are concentrated within metabolically active fractions of the community, rather than reflecting broad community-wide genomic shifts. Notably, the functional categories enriched among highly variable regions, such as information processing, regulatory systems, and cellular signaling, further suggest that selection is acting on genes directly relevant to stress resilience. Taken together, these findings reveal elevated diversity among transcriptionally active strains in mining-impacted aquifers, underscoring how long-term ISL introduces radiochemical stress and geochemical perturbations, including altered redox conditions and new electron donors and acceptors, that collectively shape the active microbial fraction. While the overall genomic potential of the community may remain conserved, sustained environmental disturbance appears to activate and favor adaptive subpopulations, with responses driven not only by stress tolerance but also by shifts in substrate availability and redox gradients.

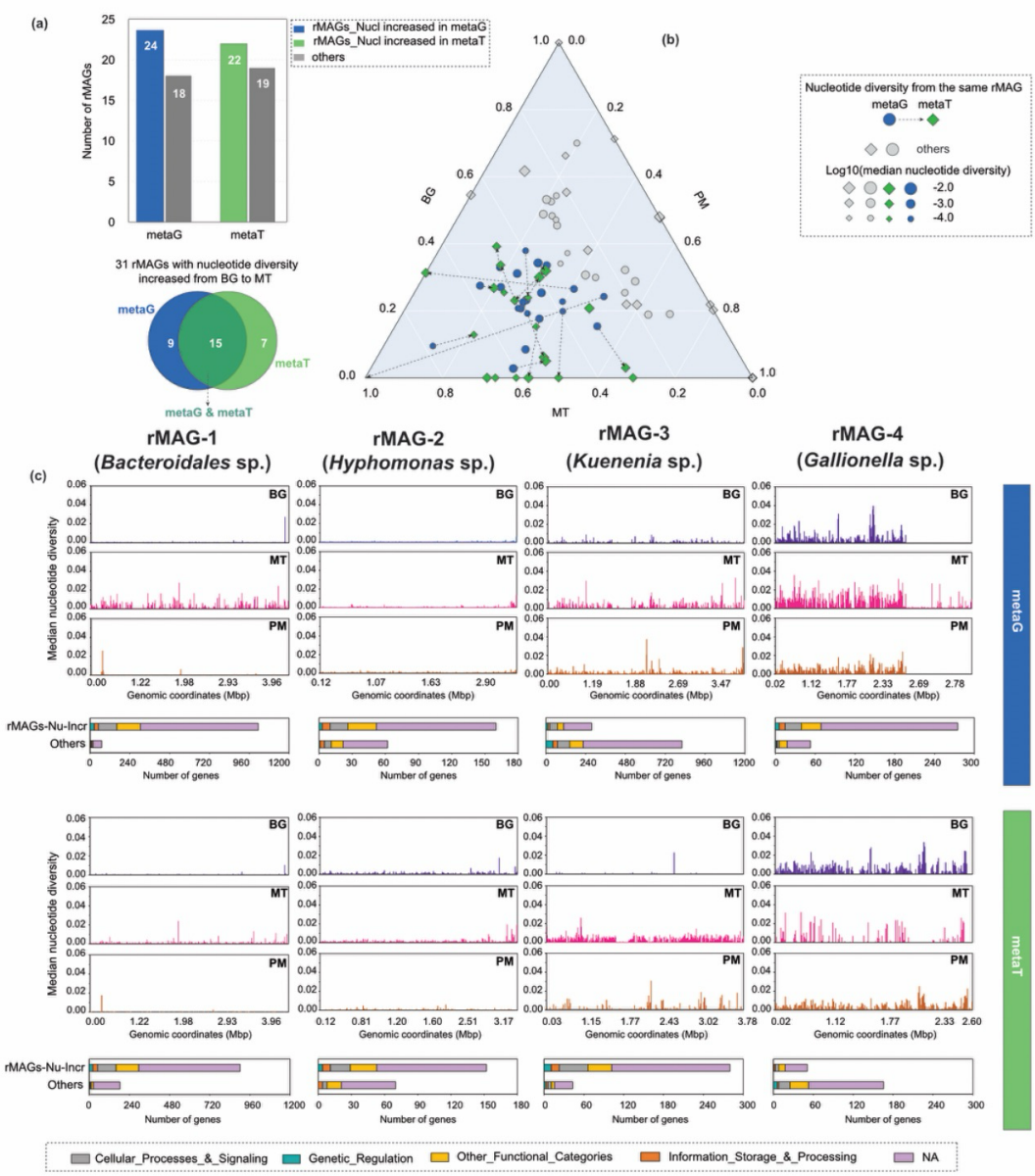


Figure 7 | Genome-wide nucleotide diversity patterns across mining stages in representative rMAGs. (a) Number of rMAGs with increased nucleotide diversity from background (BG) to mining-impacted (MT) conditions in metagenomic (metaG) and metatranscriptomic (metaT) datasets. (b) Ternary plot showing the distribution of rMAG-level nucleotide diversity across BG, MT, and post-mining (PM) samples, with points colored by dataset and sized by diversity magnitude. (c) Gene-level nucleotide diversity profiles for four representative rMAGs (*Bacteroidales*, *Hyphomonas*, *Kuenenia*, *Gallionella*), showing elevated and widespread diversity in MT, particularly

435 in transcriptionally active genes. Gene function distributions are indicated by functional
436 category bar plots.
437
438 **Discussion**
439 The global resurgence of nuclear energy has renewed interest in U ISL, a
440 minimally invasive yet chemically intensive extraction method poised to dominate
441 future uranium supply chains[34,35]. Given its operational efficiency and limited surface
442 disturbance, U ISL is expected to remain a dominant extraction method where
443 geological conditions permit in the near future[13]. While the geochemical and
444 biodiversity impacts of U ISL are well documented[13,17–21,23,36–38], its evolutionary
445 consequences for subsurface microbiomes remain underexplored. Here, we present
446 metagenomic and metatranscriptomic evidence that neutral-pH U ISL exerts sustained
447 selective pressure on microbial populations, driving *in situ* evolution in deep aquifers.
448 Neutral-pH ISL introduces persistent redox and geochemical gradients to
449 subsurface ecosystems by injecting oxidants into anoxic strata, mobilizing uranium,
450 radionuclides, and heavy metals. By examining a neutral-pH ISL area with aquifers
451 having undergone different time periods of leaching (up to 20-years), we reveal that
452 these sustained geochemical disturbances go beyond restructuring of microbial
453 community composition and alter transcriptional activity, along with an increased
454 genetic diversity illuminating long-term ecological and evolutionary consequences for
455 deep biosphere microbiomes. These transformations are shaped by sustained
456 selection pressures imposed by neutral-pH ISL, including oxidative stress, redox-
457 active metal toxicity, and low-dose radiation, stressors known to destabilize cellular
458 systems and promote adaptive traits linked to stress tolerance, DNA repair, and redox
459 regulation[39–41]. Correspondingly, we detected increased taxonomic diversity and

460 ecological modularity in neutral-pH ISL-impacted aquifers, signatures consistent with
461 stress-induced diversification[11,42–46]. Temporal succession patterns revealed dynamic
462 microbial responses, with oxidative stress-tolerant taxa dominating early neutral-pH
463 ISL stages and anaerobic fermenters and metal reducers emerging post-mining,
464 reflecting adaptation to evolving geochemical conditions.

465 Crucially, these ecological transitions were accompanied by molecular
466 signatures of adaptation. Genes associated with oxidative damage, heavy metal
467 toxicity, and radiation exposure were broadly overexpressed throughout all neutral-pH
468 ISL stages (*Figs. S8-S14*), suggesting a constitutive stress response. Yet, some
469 strains showed a substantial change in nucleotide diversification along with neutral U
470 ISL enabling the conclusion that repair mechanisms may have been insufficient to
471 counteract accumulated genomic damage. Consequently, elevated nucleotide
472 diversity, particularly in mid-mining metatranscriptomes, supports the hypothesis that
473 mutational input exceeded repair capacity, consistent with models of stress-induced
474 mutagenesis[47–50]. This model posits that environmentally triggered DNA damage,
475 coupled with error-prone repair, contributes to increased mutation rates. While DNA
476 repair pathways buffer deleterious mutations, they may also limit adaptive potential by
477 constraining the fixation of beneficial variants[51,52]. In our study, elevated nucleotide
478 diversity together with transcriptional activation of DNA repair and oxidative stress–
479 response genes indicates that adaptive evolution is concentrated in metabolically
480 active redox-cycling taxa rather than across the whole community. This resolves the
481 modest community-wide transcriptional changes yet strong gene-level positive
482 selection signals (Fig. S15b), highlighting that long-term ISL creates evolutionary
483 hotspots where radiochemical stress and altered substrate availability jointly promote
484 adaptation in key biogeochemical taxa.

485 These patterns are further supported by pN/pS ratios exceeding one, a
486 canonical signal of positive selection, detected predominantly in metatranscriptomes
487 (47 species-cluster rMAGs in mid-mining and 5 in post-mining), with only limited
488 overlap to respective metagenomes. Notably, the species-cluster rMAGs exhibiting
489 elevated nucleotide diversity spanned diverse taxa and showed widespread gene-
490 level variation in information processing and regulatory functions, underscoring that
491 selection acts primarily on transcriptionally active populations. This layered response,
492 genome-wide mutation accumulation coupled with transcription-linked selection,
493 highlights the dual role of environmental stress in both generating genetic variability
494 and filtering for functional adaptation, consistent with eco-evolutionary rescue
495 models[53–55], wherein microbial populations rebound from stress-induced bottlenecks
496 through the acquisition and retention of adaptive mutations.

497 Our study provides rare empirical evidence that stress-induced mutagenesis
498 and transcription-linked selection operate in natural deep subsurface microbiomes
499 over decadal timescales. Although such mechanisms are well characterized in
500 laboratory systems[56–62] and diverse surface environments[47,63–66], their manifestation in
501 oligotrophic, energy-limited subsurface ecosystems[67–70] has remained largely
502 speculative due to assumptions of metabolic quiescence and evolutionary stasis[71–74].
503 Our findings challenge this paradigm, demonstrating that chronic anthropogenic stress
504 in engineered aquifers can generate sustained mutational input and facilitate
505 directional selection. The convergence of elevated nucleotide diversity, DNA repair
506 activation, and positive selection in expressed genes reveals a coordinated
507 evolutionary response to long-term environmental disturbance. These observations
508 extend core theoretical frameworks, including stress-induced mutagenesis and eco-
509 evolutionary rescue, from controlled systems to complex, anthropogenically modified

510 subsurface environments, redefining the deep biosphere as a dynamic, evolutionarily
511 responsive system.
512
513 **MATERIAL AND METHODS**
514 ***Study area, sampling strategy, and sample collection.*** The study area located in
515 the Qianjiadian U ore deposit within the QJD sag of the Kailu Depression in the
516 southwestern Songliao Basin, northeast China, represents a typical sandstone-hosted
517 uranium system[75]. The stratigraphy comprises the Qingshankou (K_2qn), Yaojia (K_2y),
518 and Nenjiang (K_2n) formations, overlain by Neoproterozoic and Quaternary alluvial
519 deposits. The uranium mineralization is primarily hosted in the K_2y Formation, which
520 consists of red, yellow, gray, and mineralized gray sandstones, with widespread
521 carbonate cementation (dolomite, siderite, ankerite, and calcite). Pitchblende and
522 coffinite are the dominant uranium minerals, while Fe-Ti oxides, natural organic matter
523 (NOM), clay minerals, and pyrite serve as additional uranium hosts[36,76–78].
524 Groundwater flow follows a southwest-to-northeast trajectory within a confined
525 Cretaceous sandstone aquifer (~298 m depth, 35 m thickness), with permeability
526 coefficients averaging 0.145 m/day[21]. The aquifer is hydraulically isolated by a low-
527 permeability mudstone-dominated aquitard, with limited connectivity to adjacent
528 hydrogeological units.
529 The Qianjiadian uranium deposit has been a key site for advancing in-situ recovery
530 (ISR) technology in China, serving as the first large-scale demonstration of CO_2 + O_2
531 leaching in a deep, low-permeability sandstone-hosted system. During the neutral-pH
532 ISL process, the mixed reagent of dissolved O_2 (oxidant) and CO_2 (complexing agent)
533 was injected into the ore-bearing aquifer, and subsequently, U-containing groundwater
534 was pumped out and U was retrieved using ion exchange resins[21,79]. The effluent from

the reverse osmosis system was reused as a leaching solution after being mixed with CO_2 and O_2 to maintain carbonate buffering and oxidizing conditions[80]. The orebody, located within confined sandstone aquifers at depths ranging from 220 to 460 m, presents complex hydrogeological conditions that influence solute transport, aquifer reactivity, and microbial community dynamics. The unique geological and hydrogeochemical setting of the Qianjiadian deposit, combined with its prolonged history of in situ leaching (ISR) operations, establishes it as a natural laboratory for investigating subsurface fluid-rock-microbe interactions, uranium mobilization processes, and the long-term geochemical and microbial responses to neutral-pH ISL as well as post-mining restoration.

To investigate the hydrogeochemical and microbial responses to neutral-pH ISL mining, groundwater samples were collected from the Qianjiadian (QJD) sandstone-hosted uranium deposit during two field campaigns in October-November 2021 and August–October 2022 (Figure 1). A total of 80 groundwater samples were obtained from sites representing different stages of neutral-pH ISL, spanning ore-bearing aquifers at depths of 220–460 m. Sampling locations encompassed neutral-pH ISL sites spanning multiple operational durations: 0.25 years (MT0.25, 8 samples), 1 year (MT1, 3 samples), 4 years (MT4, 6 samples), 5 years (MT5, 7 samples), 6 years (MT6, 6 samples), 15 years (MT15, 6 samples), and 20 years (MT20, 4 samples). Post-mining phases were sampled under distinct redox conditions, including iron-reducing (PMI, 4 samples), mixed iron- and sulfate-reducing (PMIS, 2 samples), and sulfate-reducing (PMS, 3 samples) conditions. Background aquifer samples (BG, 19 total) were collected across the chronosequence, including non-mining areas (5 samples), early-stage sites (MT0.25 and MT1, 1 sample), mid-stage sites (MT4 and MT5, 5 samples), and late-stage/post-mining sites (MT15, MT20, and PM, 8 samples).

Background wells were present at all sites and sampled concurrently with mining and post-mining wells, providing site-specific baseline controls. Detailed hydrogeochemical data for MT20, MT4, MT5, PMI, PMIS, PMS, and BGPM are reported in our previous studies[21–23,36].

Prior to sample collection using a deep groundwater sampler (Geosub 2, Geotech, USA), monitoring and post-mining wells were pumped for at least 20 minutes until water temperature, electrical conductivity (EC), and oxidation-reduction potential (ORP) stabilized. Groundwater was then pumped after well flushing and approximately 45 ± 0.5 L of water was filtered through 0.2-μm sterile cellulose nitrate membrane filters (Sartorius, Göttingen, Germany) to capture microbial organisms from each groundwater well. All filtered membranes were frozen at -80 °C for high-throughput sequencing. Groundwater samples for physicochemical analysis were collected in 5 L sterile PET bottles and stored at -20 °C.

Chemical Analysis. Major anions (Cl^-, NO_3^-, and SO_4^{2-}) were measured using ion chromatography (Dionex ICS-600, Thermo Fisher Scientific, Massachusetts, USA), while major cations (Na^+, K^+, Mg^{2+}, and Ca^{2+}) were determined using inductively coupled plasma optical emission spectrometry (ICP-OES, Optima 8300, PerkinElmer, Massachusetts, USA). Trace elements, including U, Fe, and other transition metals, were quantified using inductively coupled plasma mass spectrometry (ICP-MS, 7900 ICP-MS, Agilent Technologies, California, USA). Alkalinity was measured by Gran titration using a Metrohm 888 Titrando (Metrohm, Herisau, Switzerland). The total organic carbon (TOC) and dissolved organic carbon (DOC) concentrations were determined using a TOC analyzer (TOC-L, Shimadzu, Kyoto, Japan). The gross alpha and gross beta radioactivity of groundwater samples were measured using a low-background α/β meter (LB6008, Beijing Yida Measurement Technology Co., Ltd.,

585 Beijing, China) following the thick source method for radioactivity analysis.
586 Measurements of gross alpha and gross beta radioactivity were conducted at the
587 China Nuclear Mining Science and Technology Corporation, Beijing, China, to assess
588 radiological characteristics associated with uranium mobilization.

589 **DNA extraction and sequencing.** Total metagenomic DNA was extracted from
590 biomass retained on 0.22-μm sterile cellulose nitrate membrane from 60 groundwater
591 samples, selected from a total of 80 samples retrieved from different neutral-pH ISL
592 stages in ore-bearing aquifers (see above and *Figure 1*). DNA extraction was
593 performed using the FastDNA® SPIN Kit for Soil (MP Biomedicals, Solon, OH, USA)
594 following the manufacturer's protocol. The concentration and purity of the DNA
595 extracts were measured using a TBS-380 fluorometer (Turner Biosystem, USA) and a
596 NanoDrop 2000 spectrophotometer (Thermo Scientific, USA), respectively. The
597 quality of the DNA extracts was checked using 1%-agarose gel electrophoresis.
598 Paired-end sequencing was performed on the Illumina HiSeq XTen platform (Illumina,
599 Inc., San Diego, CA, USA) at Majorbio Bio-Pharm Technology Co., Ltd., China. DNA
600 extracts were fragmented to an average size of approximately 400 bp using a Covaris
601 M220 ultrasonicator (Gene Company Limited, China). Paired-end libraries were
602 constructed using the NEXTFLEX Rapid DNA-Seq Kit (Bioo Scientific, Austin, TX,
603 USA), with adapters containing the full complement of sequencing primer hybridization
604 sites ligated to the blunt end of fragments.

605 **Metagenome processing and MAG recovery.** Raw reads were trimmed and quality-
606 filtered using fastp v0.19.7[81]. The microbial diversity and its coverage based on quality-
607 filtered raw metagenomic reads was estimated by Nonpareil v3.4.1[25] in "kmer" mode.
608 High-quality reads from individual samples were assembled into scaffolds using
609 metaSPAdes version 3.15.5 (--k-list 21, 33, 55, 77, 99, and 127)[82]. Open-Reading

610 Frames on scaffolds were predicted using prodigal (v2.6.3)[83] in “meta” mode. MAGs
611 were recovered from each assembly using the binning module (parameters: -metabat2
612 -maxbin2 -concoct) within metaWRAP v1.3 (metabat2 v2.12.1, MaxBin2 v.2.2.6, and
613 concoct v0.4.0)[84–87], which bins based on reads mapped only from the corresponding
614 sample. Quality of genome bins (completeness and contamination) was evaluated
615 using CheckM2 (v2.0.11)[88]. The best MAG representative for each sample was
616 determined using DASTool v1.1.7 with default parameters[89]. Finally, a total of 4036
617 MAGs were obtained with completeness >50% and contamination <10%. Bins were
618 further consolidated via dereplication at strain level using dRep (v.3.2.2) at 99% ANI[90]
619 resulting in 2,255 strain-cluster representative MAGs (strain-cluster rMAGs).
620 **Taxonomic classification of strain-cluster rMAGs and phylogenetic tree**
621 **generation.** Taxonomies were assigned to each strain-cluster rMAG using GTDB-Tk
622 v2.1.1 (https://github.com/Ecogenomics/GtdbTk, database release 220)[91]. Conserved
623 marker genes (ribosomal proteins and RNA polymerase subunits) were identified
624 using MarkerFinder (-m ribo_rnap, -c) (https://github.com/faylward/markerfinder) and
625 concatenated into a supermatrix, which were used for constructing a maximum-
626 likelihood phylogenetic tree via IQ-TREE 2 (v 2.2.0.3)[92] with 1,000 ultrafast
627 bootstrapping iterations. ModelFinder[93] determined Qpfam+I+I+R10 to be the best
628 model. The phylogenetic tree was visualized using the Interactive Tree of Life (iTOL)[94].
629 **Coverage estimation and metabolic potential prediction.** Clean reads from each
630 sample were mapped to strain-cluster rMAGs using strobealign (v0.13.0)[95] in CoverM
631 (v0.7.0; default parameters)[96]. The metabolic potential of strain-cluster rMAGs was
632 determined using METABOLIC v4.0[97]. We constructed two databases, one containing
633 96 genes associated with DNA repair mechanisms and one containing 32 genes
634 implicated in microbial responses to oxidative stress, with all genes being sourced

635 from NCBI (https://www.ncbi.nlm.nih.gov/). Genes of strain-cluster rMAGs were
636 annotated against these databases using DIAMOND BLAST[98] with an E-value
637 threshold of $1e^{-5}$ and the parameter -k 1 to retain only the top scoring hit. To ensure
638 the specificity of these annotations, a reverse BLAST was subsequently conducted by
639 aligning database hits back against the original amino acid sequences of the
640 corresponding strain-cluster rMAGs using the same E-value threshold. Only reciprocal
641 hits that aligned to the same ORFs as the original query were retained, ensuring one-
642 to-one correspondence. Cases with multiple high-scoring matches were excluded to
643 avoid ambiguous annotations. Final hits were filtered based on a similarity threshold
644 of >50%, ensuring high-confidence functional assignments. These curated datasets
645 facilitated a refined analysis of microbial resilience and adaptation strategies in diverse
646 environmental contexts.

647 **Community profiling based on the *rpS3* marker gene.** ORFs representing the *rpS3*
648 marker gene were identified using hmmsearch[99] against custom Hidden Markov
649 Models (HMMs) for Archaea and Bacteria (E-value 1e-28;
650 https://github.com/AJProbst/rpS3_trckr). Sequences were pooled and clustered at
651 99% similarity using vsearch[100], representing species-level clusters[101]. Representative
652 sequences were selected following a tiered approach: (1) centroids that could be
653 extended by 1000 bp in both directions, (2) non-centroids that were similarly
654 extendable, and (3) the longest available sequence when neither of the first two criteria
655 were met. To improve abundance estimates, *rpS3* gene sequences of representative
656 *rpS3* genes were extended by up to 1,000 bp in both directions on their scaffold of
657 origin if possible. This was done as mapping on just the approximately 1,000 bp long
658 *rpS3* gene results in poor mapping on the gene edges and this has progressively more
659 impact when the total length of the sequence is short. The resulting *rpS3* gene

660 sequences, ideally extended in both directions by 1,000 bp, are henceforth called
661 rpS3-extended. The workflow for clustering, extracting, and calculating coverage is
662 available at https://github.com/ProbstLab/publication-
663 associated_scripts/tree/main/FigueroaGonzalez_Bornemann_etal. The rpS3 genes
664 in rpS3-extended sequences (in amino acid format) were taxonomically classified
665 using DIAMOND[98] (E-value ≤ $1e^{-5}$) against the GTDB r220 protein database,
666 assigning the best hit to be the taxonomy. If no hit meeting the search criteria was
667 found, "Unclassified" was assigned. Abundance was calculated by mapping reads of
668 all metagenomic samples to rpS3-extended sequences and calculating the average
669 coverage per sample. Additionally, the breadth, i.e., the number of nucleotide positions
670 in rpS3-extended having a coverage of at least 1, was determined. The coverage of
671 an rpS3-extended sequence in a given sample was set to zero if the breadth of the
672 sequence in that sample was less than 95%, indicating spurious mapping instead of
673 actual presence of the rpS3-extended-represented species. Relative coverage values
674 were transformed into percentages (%).
675 ***RNA extraction, sequencing, and metatranscriptomics.*** Total RNA was isolated
676 and purified using Soil RNA Extraction Kit (Majorbio, Shanghai, China) following the
677 manufacturer's procedure. The integrity and quantity of the extracted RNA were
678 measured with a NanoDrop 2000 spectrophotometer (Thermo Scientific, MA, USA)
679 and an Agilent 5300 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). The RNA
680 was subjected to standard Illumina library preparation with an Illumina® Stranded
681 mRNA Prep, Ligation (Illumina, San Diego, CA, USA), and rRNA was depleted using
682 a RiboCop rRNA Depletion Kit for Mixed Bacterial Samples (Lexogen, USA). Depleted
683 RNA was extracted from biomass retained on 0.22 μm sterile cellulose nitrate
684 membranes from 60 groundwater samples (same to metagenomic groundwater

685 samples) and finally library preparation of RNA extraction from 45 groundwater
686 samples were successfully constructed (Figure 1). Sequencing was performed on an
687 Illumina Novaseq6000 sequencer (Illumina, San Diego, CA, USA) with paired-end
688 sequencing method at Majorbio Bio-Pharm Technology Co., Ltd. (Shanghai, China).
689 Quality control of the raw data generated by sequencing was carried out with fastp
690 v0.19.7[81]. SortMeRNA (v4.3.6) was used to remove non-coding RNA sequences
691 (tRNA, tmRNA, 5S, 16S, 18S, 23S, and 28S)[102]. The proportion of reads removed per
692 sample is summarized in Supplementary Table S3 and Figure S17, providing an
693 overview of rRNA/tRNA content prior to filtering.

694 ***Gene overexpression in each metatranscriptome.*** InStrain (v1.8.1)[103] with the
695 parameter (--min_cov 0) was used on strain-cluster rMAGs recovered from the same
696 sample to get the coverage of each gene using the clean reads from each
697 metagenome and each metatranscriptome. The gene coverage in each strain-cluster
698 rMAG was received by using the cluster database generated when dereplicate
699 genomes at 99% ANI using dRep (v3.2.2)[90]. The mean coverage of 30S ribosomal
700 proteins were used to calculate the normalization factor of each gene from the strain-
701 cluster rMAGs in each metatranscriptome sample. The genes with normalization factor
702 greater than two were defined as overexpressed genes.

703 **Species-level clustering and calculation of evolutionary metrics*.*** Filtered reads
704 from each metagenome and metatranscriptome were mapped to the best MAGs
705 output from DAStool of the same sample using strobealign (v0.13.0)[95] in CoverM
706 (v0.7.0; default parameters)[96]. Genetic annotation of MAGs was performed with
707 Prodigal (v2.6.3; –p meta)[83] for the gene module of inStrain (v1.8.1)[103]. Population
708 statistics and nucleotide metrics, including nucleotide diversity (SNVs/kbp) and
709 nonsynonymous to synonymous mutation ratios (pN/pS) were calculated from these

710 mappings using the profile module of the inStrain program (v1.8.1; –database mode;
711 default parameters)[103] at genome and gene levels. The 4036 MAGs reconstructed
712 from all individual assemblies (n = 60) were combined and dereplicated for species-
713 level clustering using dRep (v3.2.2)[90] at 95% ANI. A total of 1733 MAGs with the
714 highest genome quality from each species cluster were designated as the
715 representative species (species-cluster rMAGs). The species-cluster rMAGs with over
716 three members in the cluster database generated when dereplicate genomes at 95%
717 ANI were selected for the final calculation of microdiversity metrics such as coverage,
718 breadth, SNPs density, nucleotide diversity, and pN/pS ratios.
719 For SNP calling, the number of quality-filtered reads mapping to the position had to be
720 at least 5× coverage and 5% SNP frequency, and the variant based on reads had to
721 be lower than the expected sequencing error rate ($1e^{-6}$). Genome-level evolutionary
722 metrics based on both metagenomic and metatranscriptomic data from mining sites
723 were compared with those from background aquifer samples at the same location
724 (*Table S4, S5*). To evaluate the impact of radioactivity and chemical stress on
725 microbial genomes, 41 species rMAGs present in at least two groups (BG, MT, and
726 PM) and in a minimum of three samples from each group were selected. Gene
727 annotations for these 41 species rMAGs were performed (*Table S6*) using
728 MetaCerberus v1.3.1[104]. Gene-level nucleotide diversity of microbial genes within
729 these 41 rMAGs was then analyzed using both metagenomic and metatranscriptomic
730 data across the BG, MT, and PM groups (*Table S7,S8*).
731 **Statistical analyses and data visualization.** Spearman correlation analyses were
732 performed in the R platform (http://cran.r-project.org), with p-values adjusted using the
733 Benjamini-Hochberg procedure at a minimum significance threshold of 0.05[105]. Co-
734 occurrence networks were constructed using the *igraph* package with default

735 parameters[106]. Only Spearman correlations with p-values < 0.001 and |ρ|> 0.6
736 between pairs of rMAGs were considered statistically significant, contributing to the
737 construction of a reliable network. The resulting correlation network was imported into
738 Gephi (v0.9.2) for visualization[107]. The Spearman correlations between over-
739 expressed genes from rMAGs and hydrogeochemical variables were visualized using
740 the *pheatmap* package in R (v4.4.2)[108]. To assess sample similarity and identify key
741 factors driving clustering patterns, Principal Coordinates Analysis (PCoA) was
742 conducted based on the Bray-Curtis dissimilarity of microbial community
743 compositions. Analysis of Similarity (ANOSIM) was employed to evaluate significant
744 differences among sample groups.

745 ACKNOWLEDGEMENTS

746 This study was supported by the National Natural Science Foundation of China grant
747 nos. 42130509 (to H.G.) and 42072273 (to W.X.), the Deep Earth Probe and Mineral
748 Resources Exploration–National Science and Technology Major Project grant no.
749 2024ZD1000601 (to W.X.), the Fundamental Research Funds for the Central
750 Universities grant no. 590223006 (to W.X.) and the 111 Project grant no. B20010 (to
751 H.G.); and by the European Research Council Synergy Grant “Archean Park” under
752 grant no. 101118631 (to A.J.P.) and the German Research Foundation grant no. PR
753 1603/4-1 (to A.J.P.). We thank Prof. Rizlan Bernier-Latmani (Environmental
754 Microbiology Laboratory, École Polytechnique Fédérale de Lausanne, Switzerland) for
755 valuable comments on the manuscript and for providing access to laboratory
756 resources.

757 CONFLICT OF INTEREST

The authors declare no conflict of interest.

Ethics Statement

All sampling and sample collection were performed with permission from the local environmental authorities and in compliance with all applicable regulations governing environmental research in China. No human subjects or animal samples were involved in this study.

CODE AVAILABILITY STATEMENT

All code used in this manuscript is publicly available and referenced in the MAterials and Methods section.

References:

1. Xie, X. *et al.* Groundwater Quality and Public Health. *Annual Review of Environment and Resources* **48**, 395–418 (2023).
2. Orsi, W. D., Edgcomb, V. P., Christman, G. D. & Biddle, J. F. Gene expression in the deep biosphere. *Nature* **499**, 205–208 (2013).
3. Magnabosco, C. *et al.* The biomass and biodiversity of the continental subsurface. *Nature Geoscience* **11**, 707–717 (2018).
4. Beaver, R. C. & Neufeld, J. D. Microbial ecology of the deep terrestrial subsurface. *The ISME Journal* **18**, wrae091 (2024).
5. Bornemann, T. L. V. *et al.* Genetic diversity in terrestrial subsurface ecosystems impacted by geological degassing. *Nature Communications* **13**, 1–12 (2022).
6. Patsis, A. C., Schuler, C. J., Toner, B. M., Santelli, C. M. & Sheik, C. S. The potential for coupled organic and inorganic sulfur cycles across the terrestrial deep subsurface biosphere. *Nat Commun* **16**, 3827 (2025).

7. Probst, A. J. *et al.* Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nature Microbiology* **3**, 328–336 (2018).
8. He, C. *et al.* Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. *Nature Microbiology* **6**, 354–365 (2021).
9. Amundson, K. K., Borton, M. A. & Wilkins, M. J. Anthropogenic impacts on the terrestrial subsurface biosphere. *Nat Rev Microbiol* **23**, 147–161 (2025).
10. Vigneron, A. *et al.* Succession in the petroleum reservoir microbiome through an oil field production lifecycle. *The ISME Journal* **11**, 2141–2154 (2017).
11. Ning, D. *et al.* Environmental stress mediates groundwater microbial community assembly. *Nature Microbiology* **9**, 490–501 (2024).
12. Jones, D. O. B. *et al.* Long-term impact and biological recovery in a deep-sea mining track. *Nature* (2025) doi:10.1038/s41586-025-08921-3.
13. Li, G. & Yao, J. A Review of In Situ Leaching (ISL) for Uranium Mining. *Mining* **4**, 120–148 (2024).
14. Jroundi, F. *et al.* Profiling native aquifer bacteria in a uranium roll-front deposit and their role in biogeochemical cycle dynamics: Insights regarding in situ recovery mining. *Science of the Total Environment* **721**, (2020).
15. Vrionis, H. A. *et al.* Microbiological and geochemical heterogeneity in an in situ uranium bioremediation field site. *Applied and Environmental Microbiology* **71**, 6308–6318 (2005).
16. Chen, H. *et al.* Uranium contamination mediating soil and ore microbial community assembly at four mining sites, South China. *Front. Microbiol.* **16**, 1553072 (2025).

17. Zhu, F., Zhao, B., Min, W. & Li, J. Characteristics of groundwater microbial communities and the correlation with the environmental factors in a decommissioned acid in-situ uranium mine. *Front. Microbiol.* **13**, 1078393 (2023).
18. Coral, T. *et al.* Microbial communities associated with uranium in-situ recovery mining process are related to acid mine drainage assemblages. *Science of The Total Environment* **628–629**, 26–35 (2018).
19. Van Nostrand, J. D. *et al.* Dynamics of Microbial Community Composition and Function during In Situ Bioremediation of a Uranium-Contaminated Aquifer. *Appl Environ Microbiol* **77**, 3860–3869 (2011).
20. Li, B. *et al.* Bacterial Community Shift and Coexisting/Coexcluding Patterns Revealed by Network Analysis in a Uranium-Contaminated Site after Bioreduction Followed by Reoxidation. *Applied and Environmental Microbiology* **84**, e02885-17 (2018).
21. Lu, C. *et al.* Multi-Isotope Based Identification and Quantification of Oxygen Consuming Processes in Uranium Hosting Aquifers With CO_2 + O_2 In Situ Leaching. *Water Resources Research* **59**, e2022WR033980 (2023).
22. Lian, G. *et al.* Genome-Resolved Metagenomic Insights into Aquifer Microbiome Adaptations Following Two Decades of Neutral In-Situ Leaching. *Authorea Preprints* (2024).
23. Lu, C. *et al.* Natural Attenuation of Groundwater Uranium in Post-Neutral-Mining Sites Evidenced from Multiple Isotopes and Dissolved Organic Matter. *Environ. Sci. Technol.* **58**, 12674–12684 (2024).
24. Wang, Q. *et al.* Mineral Composition and Full-Scale Pore Structure of Qianjiadian Sandstone-Type Uranium Deposits: Application for In Situ Leaching Mining. *Geofluids* **2022**, 1–15 (2022).

25. Rodriguez-R, L. M., Gunturu, S., Tiedje, J. M., Cole, J. R. & Konstantinidis, K. T. Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. *mSystems* **3**, e00039-18 (2018).
26. Sharon, I. *et al.* Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. *Genome Res.* **25**, 534–543 (2015).
27. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology* **35**, 725–731 (2017).
28. Lalande, E. & El Sayyed, H. Break-ups and make-ups: DNA search and repair. *Nature Reviews Microbiology* **20**, 66–66 (2022).
29. Louca, S. *et al.* Function and functional redundancy in microbial systems. *Nature Ecology and Evolution* **2**, 936–943 (2018).
30. Allison, S. D. & Martiny, J. B. H. Resistance, resilience, and redundancy in microbial communities. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11512–11519 (2008).
31. Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. H. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology* **10**, 497–506 (2012).
32. Krause, S. *et al.* Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Front. Microbiol.* **5**, (2014).
33. Yang, Y. Emerging Patterns of Microbial Functional Traits. *Trends in Microbiology* **29**, 874–882 (2021).
34. *Energy, Electricity and Nuclear Power Estimates for the Period up to 2050*. (INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna, 2024).
35. Monnet, A., Gabriel, S. & Percebois, J. Long-term availability of global uranium resources. *Resources Policy* **53**, 394–407 (2017).

36. Lu, C. *et al.* Characteristics of Dissolved Organic Matter in Uranium Hosting Aquifers and Potential Molecular Transformation During Neutral In Situ Leaching. *JGR Biogeosciences* **129**, e2023JG007851 (2024).
37. Asghar, F. *et al.* Geochemical characteristics and uranium neutral leaching through a co2 + o2 system—an example from uranium ore of the elzpa ore deposit in Pakistan. *Metals* **10**, 1–19 (2020).
38. Ruiz, O., Thomson, B., Cerrato, J. M. & Rodriguez-Freire, L. Groundwater restoration following in-situ recovery (ISR) mining of uranium. *Applied Geochemistry* **109**, 104418 (2019).
39. Wani, A. K., Akhtar, N., Sher, F., Navarrete, A. A. & Américo-Pinheiro, J. H. P. Microbial adaptation to different environmental conditions: molecular perspective of evolved genetic and cellular systems. *Arch Microbiol* **204**, 144 (2022).
40. Newsome, L., Morris, K. & Lloyd, J. R. The biogeochemistry and bioremediation of uranium and other priority radionuclides. *Chemical Geology* **363**, 164–184 (2014).
41. Carvalho, F. *et al.* The environmental behaviour of uranium. *International Atomic Energy Agency* (2023) doi:Retrieved from: https://www.iaea.org/publications/14688/the-environmental-behaviour-of-uranium.
42. Da Cruz Nizer, W. S. *et al.* Oxidative stress responses in biofilms. *Biofilm* **7**, 100203 (2024).
43. Nguyen, J., Lara-Gutiérrez, J. & Stocker, R. Environmental fluctuations and their effects on microbial communities, populations and individuals. *FEMS Microbiology Reviews* **45**, fuaa068 (2021).
44. Galhardo, R. S., Hastings ,P. J. & and Rosenberg, S. M. Mutation as a Stress

881 Response and the Regulation of Evolvability. *Critical Reviews in Biochemistry*
882 *and Molecular Biology* **42**, 399–435 (2007).
883 45. Evans, S., Martiny, J. B. H. & Allison, S. D. Effects of dispersal and selection on
884 stochastic assembly in microbial communities. *The ISME Journal* **11**, 176–185
885 (2017).
886 46. Zhou, J. *et al.* Stochasticity, succession, and environmental perturbations in a
887 fluidic ecosystem. *Proc. Natl. Acad. Sci. U.S.A.* **111**, (2014).
888 47. MacLean, R. C., Torres-Barceló, C. & Moxon, R. Evaluating evolutionary models
889 of stress-induced mutagenesis in bacteria. *Nat Rev Genet* **14**, 221–227 (2013).
890 48. Mojica, E. A. & Kültz, D. Physiological mechanisms of stress-induced evolution.
891 *Journal of Experimental Biology* **225**, jeb243264 (2022).
892 49. Pal, A., Bhattacharjee, S., Saha, J., Sarkar, M. & Mandal, P. Bacterial survival
893 strategies and responses under heavy metal stress: a comprehensive overview.
894 *Critical Reviews in Microbiology* **48**, 327–355 (2022).
895 50. Couce, A. *et al.* Changing fitness effects of mutations through long-term bacterial
896 evolution. *Science* **383**, eadd1417 (2024).
897 51. Denamur, E. & Matic, I. Evolution of mutation rates in bacteria. *Molecular*
898 *Microbiology* **60**, 820–827 (2006).
899 52. Maharjan, R. P. *et al.* Mutation accumulation and fitness in mutator
900 subpopulations of *Escherichia coli*. *Biol. Lett.* **9**, 20120961 (2013).
901 53. Custer, G. F., Bresciani, L. & Dini-Andreote, F. Toward an integrative framework
902 for microbial community coalescence. *Trends in Microbiology* **32**, 241–251
903 (2024).
904 54. Shade, A. Microbiome rescue: directing resilience of environmental microbial
905 communities. *Current Opinion in Microbiology* **72**, 102263 (2023).

906 55. Yamamichi, M., Letten, A. D. & Schreiber, S. J. Eco-evolutionary maintenance of
907 diversity in fluctuating environments. *Ecology Letters* **26**, (2023).
908 56. Jinks-Robertson, S. & Bhagwat, A. S. Transcription-Associated Mutagenesis.
909 *Annu. Rev. Genet.* **48**, 341–359 (2014).
910 57. Tenaillon, O. *et al.* Tempo and mode of genome evolution in a 50,000-generation
911 experiment. *Nature* **536**, 165–170 (2016).
912 58. Wei, W. *et al.* Rapid evolution of mutation rate and spectrum in response to
913 environmental and population-genetic challenges. *Nat Commun* **13**, 4752 (2022).
914 59. Ferrare, J. T. & Good, B. H. Evolution of evolvability in rapidly adapting
915 populations. *Nat Ecol Evol* **8**, 2085–2096 (2024).
916 60. Bruckbauer, S. T. & Cox, M. M. Experimental evolution of extremophile resistance
917 to ionizing radiation. *Trends in Genetics* **37**, 830–845 (2021).
918 61. Daly, M. J. A new perspective on radiation resistance based on Deinococcus
919 radiodurans. *Nature Reviews Microbiology* **7**, 237–245 (2009).
920 62. Pavlopoulou, A. *et al.* Unraveling the mechanisms of extreme radioresistance in
921 prokaryotes: Lessons from nature. *Mutation Research/Reviews in Mutation*
922 *Research* **767**, 92–107 (2016).
923 63. Martincorena, I., Seshasayee, A. S. N. & Luscombe, N. M. Evidence of non-
924 random mutation rates suggests an evolutionary risk management strategy.
925 *Nature* **485**, 95–98 (2012).
926 64. Ferenci, T. Irregularities in genetic variation and mutation rates with
927 environmental stresses. *Environmental Microbiology* **21**, 3979–3988 (2019).
928 65. Foster, P. L. Stress-Induced Mutagenesis in Bacteria. *Critical Reviews in*
929 *Biochemistry and Molecular Biology* **42**, 373–397 (2007).
930 66. Woodcroft, B. J. *et al.* Genome-centric view of carbon processing in thawing

permafrost. *Nature* **560**, 49–54 (2018).

67. Mehrshad, M. *et al.* Energy efficiency and biological interactions define the core microbiome of deep oligotrophic groundwater. *Nature Communications* **12**, 1–12 (2021).

68. Jones, R. M., Goordial, J. M. & Orcutt, B. N. Low Energy Subsurface Environments as Extraterrestrial Analogs. *Front. Microbiol.* **9**, 1605 (2018).

69. Hoehler, T. M. & Jørgensen, B. B. Microbial life under extreme energy limitation. *Nature Reviews Microbiology* **11**, 83–94 (2013).

70. Orsi, W. D. Ecology and evolution of seafloor and subseafloor microbial communities. *Nature Reviews Microbiology* **16**, 671–683 (2018).

71. Jørgensen, B. B. & Marshall, I. P. G. Slow Microbial Life in the Seabed. *Annu. Rev. Mar. Sci.* **8**, 311–332 (2016).

72. Becraft, E. D. *et al.* Evolutionary stasis of a deep subsurface microbial lineage. *The ISME Journal* **15**, 2830–2842 (2021).

73. Garber, A. I., Ramírez, G. A. & D'Hondt, S. Genomic stasis over millions of years in subseafloor sediment. *Environmental Microbiology* **26**, e16674 (2024).

74. Schopf, J. W. *et al.* Sulfur-cycling fossil bacteria from the 1.8-Ga Duck Creek Formation provide promising evidence of evolution's null hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2087–2092 (2015).

75. Jia, J. *et al.* Mineralogy and geochemistry of carbonate cement in sandstone and implications for mineralization of the Qianjiadian sandstone-hosted uranium deposit, southern Songliao Basin, China. *Ore Geology Reviews* **123**, 103590 (2020).

76. Fan, Y. *et al.* The behavior of pyrite during in-situ leaching of uranium by CO2 + O2: A case study of the Qianjiadian uranium deposit in the Songliao Basin,

northeastern China. *Ore Geology Reviews* **169**, 106085 (2024).

77. Rong, H. *et al.* Effects of igneous intrusions on diagenesis and reservoir quality of sandstone in the Songliao Basin, China. *Marine and Petroleum Geology* **127**, 104980 (2021).

78. Rong, H. *et al.* Effects of basic intrusions on REE mobility of sandstones and their geological significance: A case study from the Qianjiadian sandstone-hosted uranium deposit in the Songliao Basin. *Applied Geochemistry* **120**, 104665 (2020).

79. Qiu, W. *et al.* What chemical reaction dominates the CO2 and O2 in-situ uranium leaching? Insights from a three-dimensional multicomponent reactive transport model at the field scale. *Applied Geochemistry* **148**, 105522 (2023).

80. Xu, L., Wang, H., Liu, N. & Du, Z. Waste Water Treatment of CO2 + O2 in-situ Leaching Uranium. *Uranium Mining and Metallurgy* **31**, 96–99 (2012).

81. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

82. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**, 455–477 (2012).

83. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

84. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nature Methods* **11**, 1144–1146 (2014).

85. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **2015**, 1–15 (2015).

981 86. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for
982 genome-resolved metagenomic data analysis. *Microbiome* **6**, 1–13 (2018).
983 87. Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: An automated binning
984 algorithm to recover genomes from multiple metagenomic datasets.
985 *Bioinformatics* **32**, 605–607 (2016).
986 88. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid,
987 scalable and accurate tool for assessing microbial genome quality using machine
988 learning. *Nat Methods* **20**, 1203–1212 (2023).
989 89. Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a
990 dereplication, aggregation and scoring strategy. *Nature Microbiology* **3**, 836–843
991 (2018).
992 90. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. DRep: A tool for fast and
993 accurate genomic comparisons that enables improved genome recovery from
994 metagenomes through de-replication. *ISME Journal* **11**, 2864–2868 (2017).
995 91. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2:
996 memory friendly classification with the genome taxonomy database.
997 *Bioinformatics* **38**, 5315–5316 (2022).
998 92. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
999 Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534
1000 (2020).
1001 93. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L.
1002 S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat*
1003 *Methods* **14**, 587–589 (2017).
1004 94. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: An online tool for
1005 phylogenetic tree display and annotation. *Nucleic Acids Research* **49**, W293–

1006 W296 (2021).
1007 95. Sahlin, K. Strobealign: flexible seed size enables ultra-fast and accurate read
1008 alignment. *Genome Biol* **23**, 260 (2022).
1009 96. Aroney, S. T. N. *et al.* CoverM: Read alignment statistics for metagenomics.
1010 Preprint at https://doi.org/10.48550/arXiv.2501.11217 (2025).
1011 97. Zhou, Z. *et al.* METABOLIC: high-throughput profiling of microbial genomes for
1012 functional traits, metabolism, biogeochemistry, and community-scale functional
1013 networks. *Microbiome* **10**, 1–23 (2022).
1014 98. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using
1015 DIAMOND. *Nature Methods* **12**, 59–60 (2014).
1016 99. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in
1017 homology search: HMMER3 and convergent evolution of coiled-coil regions.
1018 *Nucleic Acids Research* **41**, e121–e121 (2013).
1019 100. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile
1020 open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
1021 101. Ruff, S. E. *et al.* A global comparison of surface and subsurface microbiomes
1022 reveals large-scale biodiversity gradients, and a marine-terrestrial divide. *Sci.*
1023 *Adv.* **10**, eadq0645 (2024).
1024 102. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of
1025 ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217
1026 (2012).
1027 103. Olm, M. R. *et al.* inStrain profiles population microdiversity from metagenomic
1028 data and sensitively detects shared microbial strains. *Nat Biotechnol* **39**, 727–736
1029 (2021).
1030 104. Figueroa Iii, J. L., Dhungel, E., Bellanger, M., Brouwer, C. R. & White Iii, R. A.

MetaCerberus: distributed highly parallelized HMM-based processing for robust functional annotation across the tree of life. *Bioinformatics* **40**, btae119 (2024).

105. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **57**, 289–300 (1995).

106. Csádi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, complex systems* **1695**, 1–9 (2006).

107. Bastian, M., Heymann, S. & Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *ICWSM* **3**, 361–362 (2009).

108. Kolde, R. *Pheatmap: Pretty Heatmaps*. (2025).