Microbiome

RESEARCH

Open Access

# Modest functional diversity decline and pronounced composition shifts of microbial communities in a mixed waste-contaminated aquifer

Yupeng Fan[1,2†], Dongyu Wang[2†], Joy X. Yang[1], Daliang Ning[1,2], Zhili He[1,3], Ping Zhang[1,4], Andrea M. Rocha[5], Naijia Xiao[1,2], Jonathan P. Michael[1,2], Katie F. Walker[5,6], Dominique C. Joyner[5,6], Chongle Pan[2,7], Michael W. W. Adams[8], Matthew W. Fields[9], Eric J. Alm[10], David A. Stahl[11], Terry C. Hazen[5,6], Paul D. Adams[12,13], Adam P. Arkin[13,14] and Jizhong Zhou[1,2,7,15,16*]

## Abstract

**Background** Microbial taxonomic diversity declines with increased environmental stress. Yet, few studies have explored whether phylogenetic and functional diversities track taxonomic diversity along the stress gradient. Here, we investigated microbial communities within an aquifer in Oak Ridge, Tennessee, USA, which is characterized by a broad spectrum of stressors, including extremely high levels of nitrate, heavy metals like cadmium and chromium, radionuclides such as uranium, and extremely low pH (< 3).

**Results** Both taxonomic and phylogenetic α-diversities were reduced in the most impacted wells, while the decline in functional α-diversity was modest and statistically insignificant, indicating a more robust buffering capacity to environmental stress. Differences in functional gene composition (i.e., functional β-diversity) were pronounced in highly contaminated wells, while convergent functional gene composition was observed in uncontaminated wells. The relative abundances of most carbon degradation genes were decreased in contaminated wells, but genes associated with denitrification, adenylylsulfate reduction, and sulfite reduction were increased. Compared to taxonomic and phylogenetic compositions, environmental variables played a more significant role in shaping functional gene composition, suggesting that niche selection could be more closely related to microbial functionality than taxonomy.

**Conclusions** Overall, we demonstrated that despite a reduced taxonomic α-diversity, microbial communities under stress maintained functionality underpinned by environmental selection.

**Keywords** Taxonomic diversity, Functional diversity, Phylogenetic diversity, Contaminated aquifer

†Yupeng Fan and Dongyu Wang contributed equally to this work.

*Correspondence:
Jizhong Zhou
jzhou@ou.edu
Full list of author information is available at the end of the article

## Introduction

Microorganisms are adversely affected by environmental stressors such as pH [32], salinity [38], aridity [27], temperature [32], antibiotics [14], and heavy metals [10], leading to common observations that the number of microbial species declines under extreme conditions. Most of previous research, however, has predominantly focused on microbial taxonomy, neglecting to comprehensively assess the entire functional potentials of microbial communities. Consequently, it remains elusive whether microbial functional α-diversity mirrors the patterns observed in taxonomic α-diversity across various environmental stressors. A positive correlation between these two measures is reported, implying that functional α-diversity may increase with taxonomic α-diversity. This assumption received partial support from a study on the eastern Tibetan Plateau, where aridity stress concurrently reduced both functional and taxonomic α-diversities, albeit with a weak correlation [42]. Alternatively, functional α-diversity may increase rapidly with low α-taxonomic diversity but saturate with high taxonomic α-diversity, showing a non-linear relationship [28].

While microbial α-diversity is measured by the number of taxa and their abundance within a community, microbial β-diversity is defined as the variation in community composition between two communities, often expressed through pairwise dissimilarity [49]. Strong, positive linear correlations between taxonomic and functional gene β-diversity of microbial communities were observed in soil [12] and marine ecosystems [13]. Conversely, environmental conditions strongly affected the functional gene compositions of global marine microbial communities but only weakly affected taxonomic composition, indicating a decoupling of functionality from taxonomy [24, 26]. Similarly, the lack of correlation between microbial functionality and taxonomy was also observed in the soil mycobiome of the North American continent [43]. This discrepancy may be explained by functional redundancy, in which multiple microbial species perform the same ecological function, thereby leading to the observed inconsistencies between microbial functionality and taxonomy.

Despite significant advancements in environmental microbiome research, there remains a notable gap in generalizable insights into how microbial α- and β-diversities, particularly α- and β-functional diversities, react to various stressors [37]. The Anna Karenina Principle, which suggests that disease-associated microbial communities in hosts under stress of disease are more dissimilar than those of healthy ones, has recently been proposed as a framework for understanding microbial dynamics within the animal [51] or plant hosts [4]. However, it remains an open question whether the principle holds in aquifer microbial communities under stress.

While high levels of heavy metals restrict certain functional properties of bacterial communities [54], few research has yet quantitatively assessed functional diversity across a broad spectrum of heavy metal concentrations, where dramatic changes in species diversity are evident. We, therefore, selected a range of aquifer samples spanning from 0 to 17 mg/L in uranium concentrations, 0 to 9000 mg/L in nitrate concentrations, and 3.4 to 7.3 in pH (Fig. 1). These samples were collected from a legacy waste site with deposition of nitric acid-solubilized uranium waste between 1951 and 1983, along with mixed metal and organic wastes from other facilities of the US Department of Energy. We analyzed microbial taxonomic and phylogenetic diversities via 16S rRNA gene amplicon sequencing, and functional diversity or functional gene diversity via metagenome shotgun sequencing. We aimed to test the hypothesis that functional diversity mirrors taxonomic or phylogenetic diversities in response to environmental stressors. Specifically, we investigated whether microbial communities in contaminated wells exhibit distinct characteristics compared to those in uncontaminated wells, providing a testbed of the Anna Karenina Principle in aquifer microbial communities under stress.

## Results

### Environmental variables

The levels of conductivity, dissolved nitrous oxide ($N_2O$), chloride ($Cl^-$), manganese (Mn), and cadmium (Cd) were higher in high-contaminated wells compared to other wells, while the pH levels were lower ($p < 0.05$, Table S2). Additionally, there were higher concentrations of dissolved organic carbon, dissolved carbon dioxide ($CO_2$), nitrate ($NO_3^-$), sulfate ($SO_4^{2-}$), ferrous, potassium (K), calcium (Ca), barium (Ba), aluminum (Al), silver (Ag), iron (Fe), zinc (Zn), strontium (Sr), and uranium (U) along with lower dissolved nitrogen concentrations, dissolved oxygen concentrations, and dissolved methane concentrations in high-contaminated wells than in other wells, though some of these differences were statistically insignificant ($p > 0.05$). The dispersion of environmental variables significantly increased with increased contamination (Fig. S1).

Nitrite ($NO_2^-$) concentrations in the supernatant, sodium (Na), and magnesium (Mg) concentrations were higher in high-contaminated wells than in mid-contaminated wells, but not detectable in other wells. Nitrate ($NO_3^-$) concentrations in the supernatant, cobalt (Co), chromium (Cr), gallium (Ga), lithium (Li), and nickel (Ni) were higher in high-contaminated wells compared to

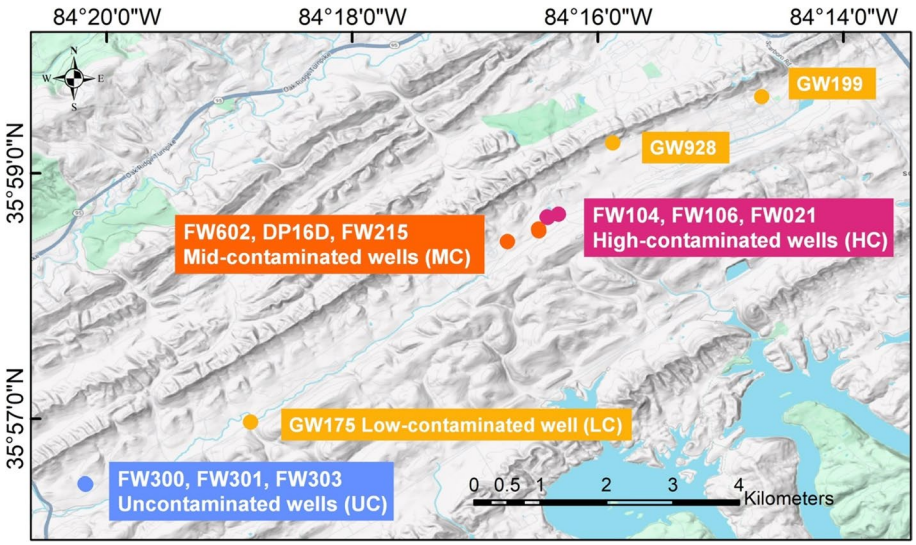


**Fig. 1** Geographical location of the study sites. Aquifer samples consist of uncontaminated wells (UC) FW300, FW301, and FW303; low-contaminated wells (LC) GW199, GW715, and GW928; mid-contaminated wells (MC) FW215, FW602, and DP16D; high-contaminated wells (HC) FW104, FW106, and FW021

mid- and low-contaminated wells, but not detectable in uncontaminated wells. Higher concentrations of arsenic (As), beryllium (Be), cesium (Cs), copper (Cu), and lead (Pb) were found in high-contaminated wells compared to low-contaminated wells, but not detected in other wells.

#### Microbial α-, β-, and γ-diversities

The taxonomic, phylogenetic, and functional α-diversities of richness were the lowest in high-contaminated wells, while they were the highest in mid-contaminated wells (Fig. 2A–C). When compared to uncontaminated wells, the taxonomic α-diversities in high-contaminated wells were reduced by 85% ($p = 0.025$), and the phylogenetic α-diversities were reduced by 81% ($p = 0.018$, Fig. 2A, B). In contrast, functional α-diversities were not significantly different between high-contaminated wells and uncontaminated wells, with a smaller decrease of 55% on average. Similar pattern were observed in Shannon index (Fig. 2D, E), whereas the functional α-diversities in high-contaminated wells were significantly lower than uncontaminated wells.

The taxonomic, phylogenetic, and functional compositions of microbial communities were well separated among uncontaminated, low-, mid-, and high-contaminated wells, as indicated by Non-metric Multidimensional Scaling (NMDS, Fig. 3A–C). To further explore these differences, three permutational tests of dissimilarity (Adonis, MRPP, and ANOSIM) were conducted, which revealed significant differences among the four groups of wells ($p < 0.001$, Table 1). Interestingly, the microbial taxonomic and phylogenetic compositions in high-contaminated wells had lower dispersion compared to the uncontaminated and low-contaminated wells (though not statistically significant, $p > 0.1$ by the permutational dispersion test, Fig. 3D), suggesting that they were more similar in high-contaminated wells. Conversely, microbial functional compositions in high-contaminated wells displayed the highest community dispersion values, indicating a pattern of microbial functional heterogeneity induced in high-contaminated wells ($p = 0.013$ by the permutational dispersion test, Fig. 3D).

When analyzing γ-diversities, the taxonomic, phylogenetic, and functional γ-diversities in high-contaminated wells were lower than uncontaminted wells, which were similar with those of α-diversities. However, a notable exception was found in the phylogenetic γ-diversity of mid-contaminated wells, which was higher than that in uncontaminated wells (Fig. S2).

#### Bacterial taxa

Proteobacteria was the most abundant phylum in high-contaminated wells, accounting for 74% of the relative abundance (Table S3A). In comparison, the average relative abundance of Proteobacteria in other wells was only 21%. Bacterial candidate phylum WPS- 2, also known as Eremiobacterota, was higher in high-contaminated wells than others, accounting for 12% of the relative abundance. Acidobacteria was also abundant in mid- and high-contaminated wells, accounting for 5–6% of the relative abundance.

*Rhodanobacter*, a Proteobacteria genus well known for denitrification [15], was the most abundant in

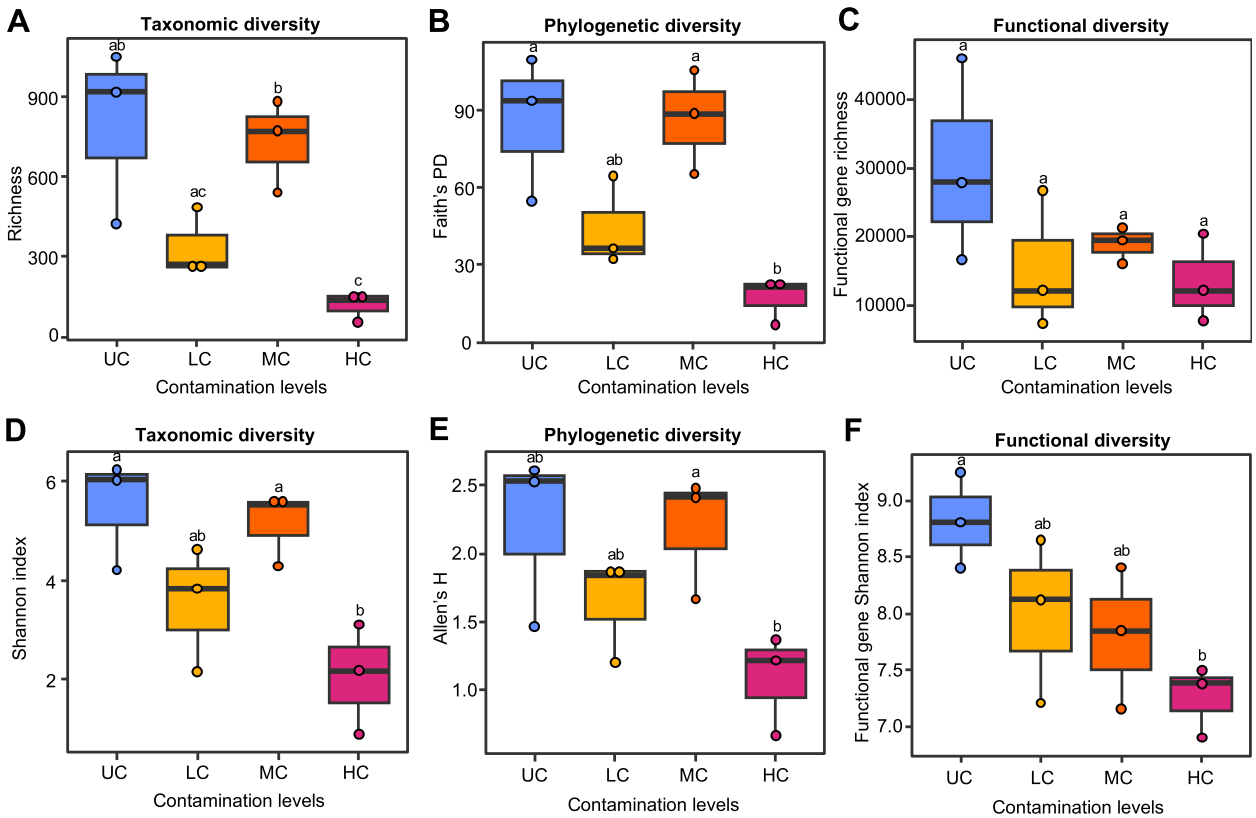


**Fig. 2** Microbial α-diversity of aquifer samples from uncontaminated wells (UC), low-contaminated wells (LC), mid-contaminated wells (MC), and high-contaminated wells (HC). Diversity metrics are represented as follows: **A** taxonomic diversity, **B** phylogenetic diversity, and **C** functional diversity, calculated by richness; and **D** taxonomic diversity, **E** phylogenetic diversity, and **F** functional diversity, calculated by the Shannon index. Statistical differences between groups were assessed using ANOVA followed by a post hoc test, with a significance threshold set at $p < 0.05$. Letters indicate significant differences between groups

high-contaminated wells (Table S3B), reaching an abundance of 80% in the FW106 well. In comparison, the relative abundance of *Rhodanobacter* in other wells was less than 1%. The second most abundant genus in high-contaminated wells belonged to *Candidatus* phylum Eremiobacterota. The third most abundant genus in high-contaminated wells was *Sulfurifustis*, a genus of sulfur-oxidizing bacteria affiliated with γ-Proteobacteria [22], accounting for 9% of the relative abundance. Other genera of sulfur-oxidizing bacteria, including *Sulfuricurvum* and *Sulfuritalea*, were detected in all wells except for high-contaminated ones.

The genera of nitrifying bacteria and archaea were abundant in certain wells. *Nitrosarchaeum*, a genus of ammonia-oxidizing archaeon, comprised 19% of relative abundance in low-contaminated wells but was only 0.032 to 0.085% in uncontaminated and 0 to 0.364% in mid-contaminated wells. The ammonia-oxidizing bacteria GOUTA6 was detected in all wells, with a relative abundance of 5% in mid-contaminated wells. The nitrite-oxidizing bacteria (NOB) genus *Nitrospira* was present in all wells except high-contaminated wells and accounted for 5% of the relative abundance in low-contaminated wells. The α-Proteobacteria genus *Reyranella*, which produces acetic acid during respiration, was present in all wells, with the highest relative abundance (1.06 to 14.45%) in low-contaminated wells. The methane oxidation bacteria genus *Candidatus Methylomirabilis* was detected in contaminated wells, with the highest relative abundance (0.28 to 14.32%) in mid-contaminated wells.

### Functional gene categories

We used a shotgun metagenomic assembly approach to profile the functional genes. The assembled contigs were annotated by the pathway maps of the KEGG database, in which most annotated pathways in contaminated wells were not significantly different from uncontaminated wells (Fig. S3). Therefore, we used EcoFun-MAP to further annotate the shotgun metagenomic data.

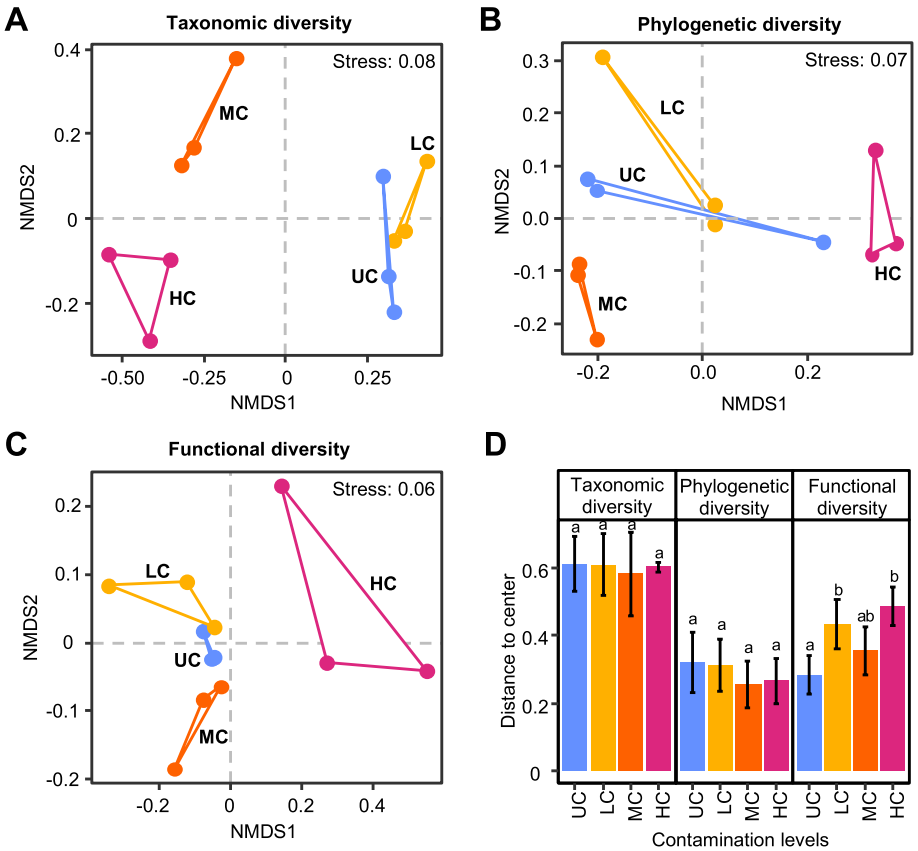


**Fig. 3** Microbial β-diversity of aquifer microbial communities for different diversity indices in uncontaminated wells (UC), low-contaminated wells (LC), mid-contaminated wells (MC), and high-contaminated wells (HC). Non-metric multidimensional scaling (NMDS) plots based on weighted Bray-Curties index for **A** taxonomic and **C** functional diversities, normalized weighted Unifrac (phylogenetic Bray-Curties) for phylogenetic diversity (**B**). Dispersion test (**D**) based on weighted Bray-Curties index for taxonomic and functional diversities, normalized weighted Unifrac (phylogenetic Bray-Curties) for phylogenetic diversity. Statistical differences between groups were assessed using ANOVA followed by a post hoc test, with a significance threshold set at $p < 0.05$. Letters indicate significant differences between groups

**Table 1** Significance tests of the groundwater microbial communities

| | Adonis | | ANOSIM | | MRPP | |
|---|---|---|---|---|---|---|
| | *F* | *p* | *R* | *p* | *δ* | *p* |
| Taxonomic diversity | 1.91 | 0.001 | 0.92 | 0.001 | 0.85 | 0.001 |
| Phylogenetic diversity | 2.43 | 0.001 | 0.57 | 0.001 | 0.33 | 0.001 |
| Functional diversity | 2.19 | 0.001 | 0.46 | 0.001 | 0.57 | 0.002 |

### Carbon degradation genes

We carried out response ratio analyses to reveal carbon degradation genes that were statistically different in relative abundances among uncontaminated, low-, mid-, and high-contaminated wells ($p < 0.05$, Fig. 4A, Fig. S4A). Most carbon degradation genes were decreased in contaminated wells, including *amyA* encoding α-amylase that hydrolyzes starch and glycogen, *xylA* encoding xylose isomerase that hydrolyzes hemicellulose, endochitinase and exochitinase genes that degrades chitin, *rgaE* encoding acetylesterase that hydrolyzes pectin, *vanA* encoding vanillate monooxygenase that degrades vanillin and lignin, *vdh* encoding vanillin dehydrogenase that degrades vanillin and lignin, and phenol oxidase gene that hydrolyzes lignin. Among them, *ara* encoding l-arabinofuranosidase that degrades hemicellulose decreased with increased

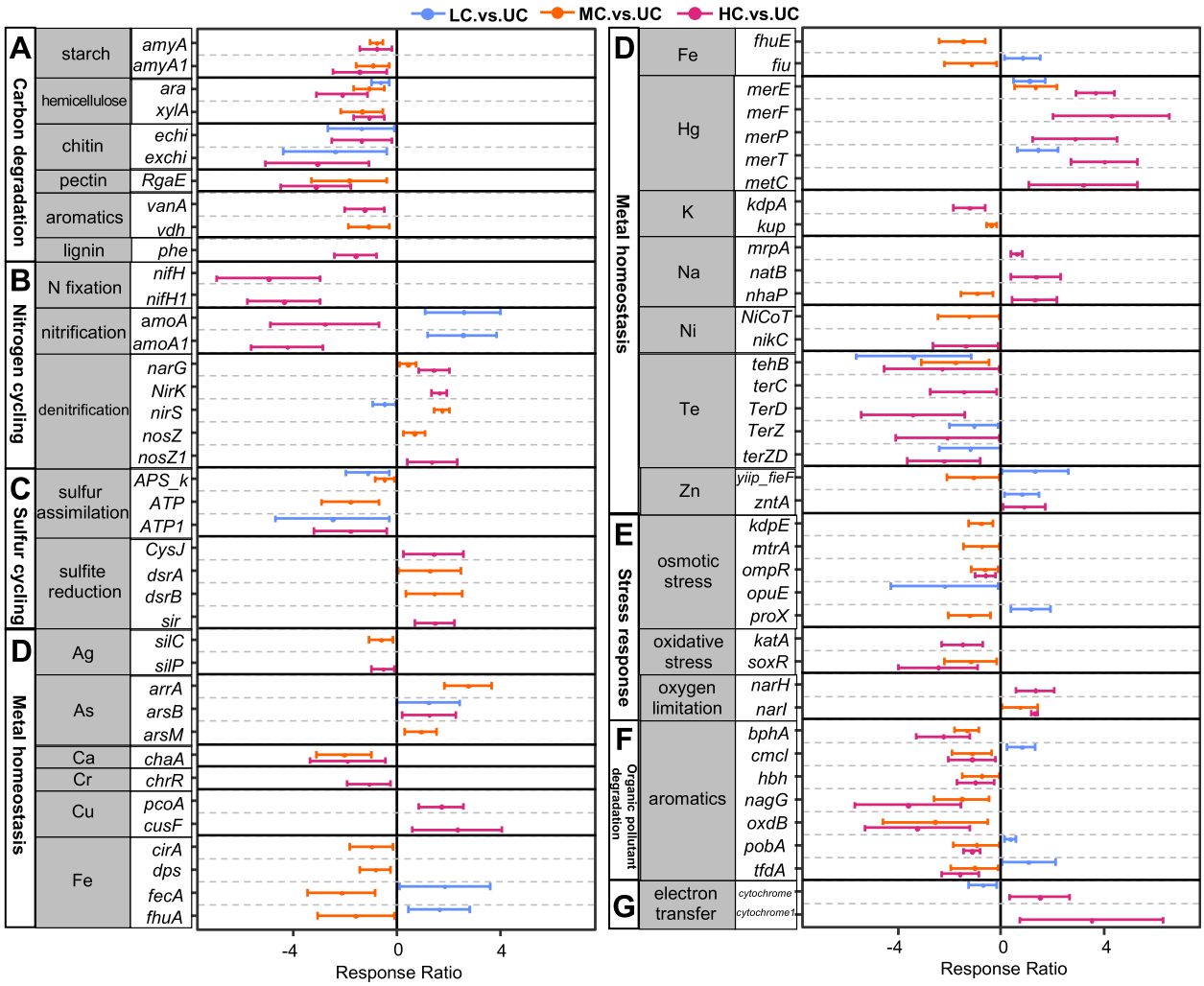


**Fig. 4** Differences in relative gene abundance of selected genes for 12 aquifer samples functional community based on response ratio. **A** Carbon degradation genes, **B** nitrogen cycling genes, **C** sulfur cycling genes, **D** metal homeostatic genes, **E** stress response genes, **F** organic pollutant degradation genes, and **G** electron transfer genes. All genes presented here are significantly different from those in unpolluted wells, judged by a 95% confidence interval

contamination, suggesting that its relative abundance was sensitive to environmental contamination.

### Nitrogen cycling genes

The relative abundances of most denitrification genes, including *narG* encoding nitrate reductase, *nirK* and *nirS* encoding nitrite reductase, and *nosZ* encoding nitrous oxide reductase, were increased in mid- and high-contaminated wells (Fig. 4B, Fig. S4B), which corresponded with high nitrate concentrations in those wells (Table S2). In contrast, biomarker genes of nitrogen fixation (*nifH* encoding the subunit of the Fe protein (Kp2) component of nitrogenase) and nitrification (*amoA* encoding ammonia monooxygenase subunit A) were decreased in high-contaminated wells (Fig. 4B, Fig. S4B), suggesting that functional potentials of nitrogen fixation and nitrification were reduced by high contamination.

### Sulfur cycling genes

Sulfite reduction genes, including *cysJ, dsrA, dsrB,* and *sir* encoding various sulfite reductase, were increased in mid- and high-contaminated wells (Fig. 4C, Fig. S4C), consistent with more anaerobic environments in those wells (Table S2). However, sulfur assimilation genes, including APS kinase, ATP sulfurylase in protists, and ATP sulfurylase, were decreased (Fig. 4C, Fig. S4C), which may suggest a stress response to maintain energy metabolism at the expense of growth.

### Metal homeostasis genes

Many metal homeostasis genes were increased in contaminated wells (Fig. 4D, Fig. S4D), including *merE*, *merF*, *merP*, *merT*, and *metC* encoding mercury transporter; *fiu*, *fhuA*, and *fecA* encoding TonB-dependent receptor that mediates substrate-specific transport across the outer membrane; *arrA* encoding arsenate respiratory reductase, *arsB* encoding an arsenical pump membrane protein, and *arsM* encoding arsenite S-adenosylmethyltransferase; *pcoA* encoding copper resistance protein A, *cusF* encoding Cu cation efflux system protein; *mrpA* and *nhaP* encoding $Na^+/H^+$ antiporter, *natB* encoding ABC transporter sodium permease; *yiip_fieF* encoding cation-efflux pump FieF, *zntA* encoding heavy metal translocating P-type ATPase. However, the relative abundances of several metal homeostasis genes were decreased in contaminated wells (Fig. 4D, Fig. S4D), including *silC* and *silP* encoding heavy metal RND efflux transporter; *chaA* encoding calcium/proton antiporter; *chrR* encoding chromate reductase; *cirA* encoding Colicin I receptor, *fhuE* encoding ferric-rhodotorulic acid outer membrane transporter, *dps* encoding DNA-binding ferritin-like protein; *kdpA* and *kup* encoding proteins in potassium transport system; *NiCoT* and *nikC* encoding nickel transport system proteins; *tehB, terC, terD, terZ*, and *terZD* encoding tellurite resistance protein (Fig. 4D, Fig. S4D).

### Stress response genes

Most osmotic stress genes, including *kdpE* encoding a transcriptional regulatory protein, *mtrA* encoding a DNA-binding response regulator, *ompR* encoding an osmolarity response regulator, *opuE* encoding osmoregulated proline transporter, and *proX* encoding glycine betaine transporter periplasmic subunit, were decreased in contaminated wells (Fig. 4E, Fig. S4E). Oxidative stress genes were decreased in contaminated wells, including *katA* encoding catalase that catalyzes the hydrogen peroxide and *soxR* encoding redox-sensitive transcriptional activator. Two oxygen limitation response genes, *narH* encoding the beta subunit of nitrate reductase and *narI* encoding the gamma subunit of nitrate reductase were increased in mid- and high-contaminated wells, which suggested microbial response to low dissolved oxygen concentration in mid- and high-contaminated wells (Fig. 4E, Fig. S4E).

### Organic pollutant degradation genes

The relative abundances of most organic pollutant degradation genes were decreased in mid- and high-contaminated wells (Fig. 4F, Fig. S4F), including *bphA* encoding biphenyl dioxygenase subunit alpha that catalyzes the oxygenation of biphenyl, *cmcI* encoding 3-carboxy-cis,cis-muconate cycloisomerase, *hbh* encoding 4-hydroxybenzoate hydroxylase that degrades aromatic compounds, *nagG* encoding salicylate 1-monooxygenase that catalyzes the decarboxylative hydroxylation of salicylate, *oxdB* encoding phenylacetaldoxime dehydratase that degrades styrene, *pobA* encoding 4-hydroxybenzoate 3-monooxygenase that catalyzes the hydroxylation of 4-hydroxybenzoate, *tfdA* encoding taurine catabolism dioxygenase that involves in taurine and hypotaurine metabolism.

### Electron transfer genes

The relative abundances of certain electron transfer genes were increased in high-contaminated wells (Fig. 4G, Fig. S4G), including C-type cytochrome genes encoding cytochrome c-type biogenesis protein CcmA and CcmF. The relative abundances of other electron transfer genes were decreased in low-contaminted wells, including some C-type cytochrome genes encoding cytochrome c class I.

### The linkages between microbial communities and environmental factors

To explore the relative importance of various factors in explaining microbial communities, we carried out partial least squares modeling (PLS) followed by variation partition analysis (VPA, Fig. 5A). Environmental variables and geographical distance explained considerable percentages of community variations for taxonomic compositions ($R^2 = 0.557$, $p = 0.001$) and phylogenetic compositions ($R^2 = 0.679$, $p = 0.001$), which were substantially lower than the explanatory power for functional compositions ($R^2 = 0.897$, $p = 0.001$). In addition, environmental variables were more important than the geographical distance in explaining the composition variations.

To evaluate the importance of individual environmental variables and geographic distance in PLS models, variable influence on projection (VIP) was calculated for taxonomic, phylogenetic, and functional diversity (Fig. 5B). DIC, DO, and Fe were important for all three dimensions of diversities (VIP > 0.10). Ba, Cu, U, and sulfide were important for taxonomic and phylogenetic diversities (VIP > 0.12), but not important for functional diversity. $CO_2$ and Cs were important for phylogenetic and function diversities (VIP > 0.66), but not important for taxonomic diversity. Furthermore, geographic distance contributed more to taxonomic and phylogenetic diversities than functional diversity.

## Discussion

Here, we explored three facets of microbial diversities, i.e., taxonomic, phylogenetic, and functional diversities along a broad spectrum of various contaminants. Consistent with the previous finding, [37], we revealed

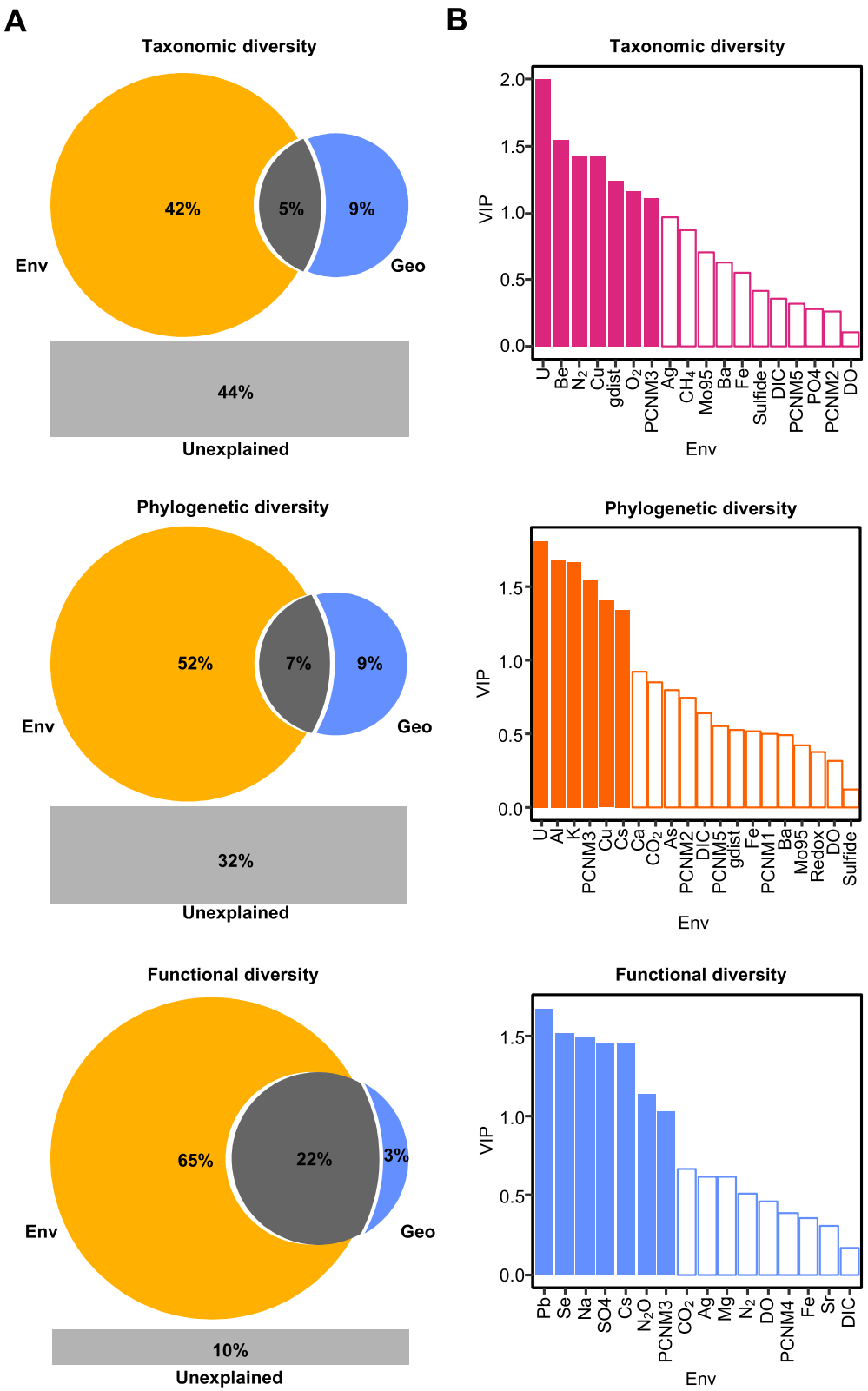


**Fig. 5** The linkage between aquifer microbial communities and environmental factors. **A** Variance partition analysis (VPA) showing relative contributions of geographical distance (Geo.) and environmental variables (Env.) to the different diversity indexes based on the PLS method. **B** Variable influence on projection (VIP) values based on the PLS model for different diversity indexes, where VIP value larger than 1 is filled, VIP value smaller than 1 is blank

a marked decrease in taxonomic and phylogenetic α-diversities in response to increasing contaminant levels (Fig. 2). In contrast, the reduction of functional α-diversities was much milder (Fig. 2). Our results also indicated an increase in functional composition heterogeneity correlating with environmental stressors, a pattern not mirrored in taxonomic and phylogenetic β-diversities (Fig. 3D). Consistently with findings in microbial communities associated with animal and plant hosts [4, 51], aquifer microbial functional compositions in high-contaminated wells diverged more substantially than those in uncontaminated wells (Fig. 3), leading to increased dispersion in microbial community composition. Therefore, the Anna Karenina Principle is not limited to host-associated microbial communities, but is also applicable to free-living microbial communities and functional diversity. This suggests that the principle, indicative of microbial responses to environmental stress, might be a more widespread phenomenon.

The dispersion of environmental variables increases under contamination (Fig. S1). As a result, these variables become more dissimilar, leading to greater heterogeneity in functional diversity. Our study revealed a significant increase in the relative abundance of functional genes associated with denitrification and sulfite reduction in mid- and high-contaminated wells (Fig. 4B and C, Fig. S4B and S4C), concurrent with the increased concentrations of nitrate and uranium. These findings align with previous research conducted at the same site [17, 50, 52], which suggests that these functional genes are critical for heavy metal reduction.

Similar to a previous observation [1], metabolic plasticity, involving various electron donors and acceptors, is a common trait in aquifer microorganisms. A wide metabolic repertoire is important in the face of the natural environmental perturbations that occur at the Oak Ridge site, where frequent storms and snows cause considerable water table fluctuations that move the oxic/anoxic interface. Consistent with earlier findings at our site [8, 15], the denitrifying *Rhodanobacter*, known for its ability to immobilize U(VI) under aerobic conditions by forming intracellular uranium–phosphate complexes but not for U(VI) reduction [15], was the dominant genus in the most contaminated wells characterized by low pH and high levels of nitrate and U concentrations (Table S2). Its dominance is likely due to its tolerance to NaCl and heavy metals [33]. A previous study has found that *Eremiobacterota* may be involved in Fe (II) oxidation [16], which could explain the high relative abundance of the genus WPS- 2 in high-contaminated wells with elevated ferrous levels (Table S3B).

*Sulfurifustis*, a genus of sulfur-oxidizing and glutathione-synthesizing bacteria, was enriched in high-contaminated wells (Table S3B) because glutathione synthesis serves as a mechanism for resisting cadmium toxicity [23]. An ammonia-oxidizing archaeon named *Nitrosoarchaeum* was the most abundant genus in low-contaminated wells (Table S3B), which was also verified by more abundant *amoA* genes in those wells (Fig. 4B, Fig. S4B). In contrast, an ammonia-oxidizing bacterium named *GOUTA6* was the most abundant genus in mid-contaminated wells (Table S3B). Methanotrophs, including the genus *Candidatus Methylomirabilis,* can use methane to transform heavy metals [20], whose unique methane oxidation pathway requires both nitrate and methane [46]. Accordingly, we found that the abundance of *Candidatus Methylomirabilis* in mid-contaminated wells was characterized by the concentrations of nitrate and methane (Table S2B). Our results revealed decreased calcium transporters in contaminated wells, suggesting a microbial strategy to mitigate uranium toxicity, similar to protective mechanisms in plants [39] and yeast [36]. Additionally, lower dissolved organic carbon (DOC) in low-contaminated wells may influence metal bioavailability and microbial stress responses, as DOC complexation reduces metal uptake [45].

Functional redundancy may also explain the Anna Karenina Principle. Functional redundancy was termed as the coexistence of different species capable of performing the same biochemical functions, which could explain why the taxonomic and phylogenetic α-diversities decreased more noticeably than functional α-diversity under stress (Fig. 2). Functional diversity showed the lowest dispersion in uncontaminated wells because of the functional redundancy. In contrast, the dispersion values of functional diversity increased as the functional redundancy decreased with the contamination levels, whereas the dispersion values of taxonomic and phylogenetic diversities remained relatively unchanged (Fig. 3).

Both deterministic and stochastic processes contribute to the increased dissimilarity in stressed microbial communities, though the Anna Karenina Principle is mainly based on the important role of stochastic processes in disrupting normal community composition [4, 51]. The variability in key biogeochemical conditions (e.g., DIC, DO and Fe, Fig. 5B) emerged as important determinants of groundwater community compositions, which affected microbial fitness. A recent study revealed that stochastic processes, especially dispersal limitation, played an important role in shaping groundwater microbial communities [29], but the relative importance of stochastic processes decreased as contamination increased, which may explain why the Anna Karenina Principles was not observed in taxonomic and phylogenetic diversities. Other factors, including biotic interactions among community members and stochastic processes (e.g.,

ecological drift and dispersal limitation), could also play important roles in shaping community assembly [9, 55]).

Our findings emphasize the importance of intergrating functional gene into bioremediation strategies. Specific functional gene markers may serve as reliable fingerprints of contamination, providing a more accurate indication of pollutant types and degradation potential. These functional indicators could be leveraged to develop predictive models for contamination assessment and ecosystem restoration, improving the precision and efficacy of bioremediation efforts.

## Conclusion

In this study, we assayed biological and geochemical diversities in a mixed waste-contaminated aquifer at Oak Ridge, TN, USA. Our results showed that environmental stressors have significant impacts on microbial diversity, particularly on taxonomic and phylogenetic diversities. The observed decrease in functional α-diversity was modest, indicating that the functional traits of the microbial communities had a better buffering capacity against environmental stress. Our results of functional heterogeneity (Fig. 3) explained the often low efficacy in treating in situ groundwater contamination, which is costly and of large scale.

Understanding microbial functional responses in the stress environment is a central topic of microbial ecology. Therefore, our study is a useful asset for determining the critical factors linking community taxonomy to functions, which contribute to the development of accurate hydrogeochemical models that aid in assessing environmental treatments and evaluating risk management [21]. The functional composition may be a sensitive and informative metric for evaluating the responses of microbial communities to environmental stress, which could inform the development of more effective and efficient bioremediation strategies for contaminated sites by providing a better understanding of the functional traits of microbial communities effective in degrading specific contaminants. Furthermore, our study demonstrated the importance of considering microbial functionality when evaluating the health of ecosystems, as the functional traits of microbial communities play a crucial role in maintaining ecosystem processes. Overall, our study contributes to a growing body of research that seeks to understand the functional response of microbial communities to environmental stress, and showed that microbial functionality should be taken in account in environmental management and risk assessment.

## Methods

### Study site and sampling

We conducted this study at the Department of Energy's (DOE) Oak Ridge FRC site in Oak Ridge, Tennessee. The groundwater at this location is tainted with various contaminants including radionuclides (such as uranium and technetium), nitrate, sulfide, and others, predominantly originating from the erstwhile S- 3 waste disposal ponds. Groundwater samples were obtained from 12 representative wells along a gradient of various contaminants during winter 2012 and spring 2013 (Fig. 1, Table. S1): uncontaminated wells (UC) including FW300, FW301, and FW303 (FW305 for metagenome shotgun sequencing due to FW303 was not easy to access); low-contaminated wells (LC) including GW199, GW715, and GW928 with nitrate concentrations less than 2 mg/L, uranium concentrations less than 0.01 mg/L, and neutral pH (6.5–7.2); mid-contaminated wells (MC) including FW215, FW602, and DP16D with nitrate concentrations between 5.5 and 1471 mg/L, uranium concentrations between 0.1 and 1.5 mg/L, and neutral pH (6.5–6.8); high-contaminated wells (HC) including FW104, FW106, FW021 with nitrate concentrations between 2692 and 11,648 mg/L, uranium concentrations between 3.8 and 55 mg/L, and low pH (3–5.2).

### Geophysical and geochemical analyses

The groundwater properties of temperature, pH, dissolved oxygen (DO), conductivity, and redox were measured by an In-Situ Troll 9500 system (In-Situ Inc., CO, USA). U.S. EPA methylene blue method (Hach; EPA Method 8131) and the 1,10-phenanthroline method (Hach; EPA Method 8146) were used to measure sulfide and ferrous iron concentrations, respectively. The concentrations of dissolved gases ($N_2$, $O_2$, $CO_2$, $CH_4$, and $N_2O$) were determined using an SRI 8610 C gas chromatograph (GC) with argon as the carrier gas. The method was derived from EPA RSK- 175 and United States Geological Survey (USGS) Reston Chlorofluorocarbon Laboratory protocols. Concentrations of dissolved organic carbon (DOC) and dissolved inorganic carbon (DIC) were ascertained using a Shimadzu TOC-V CSH analyzer (Tokyo, Japan), following the EPA Method 415.1. Anion concentrations, including bromide, chloride, nitrate, phosphate, and sulfate, were quantified using a Dionex 2100 system equipped with an AS9 column and a carbonate eluent, in accordance with U.S. EPA Methods 300.1 and 317.0. The levels of metals and trace elements present in the groundwater were assessed using an inductively coupled plasma/mass spectrometry (ICP-MS) instrument (Elan 6100), employing a technique akin to EPA Method 200.7.

### Amplicon and metagenomic sequencing

The DNA extraction method was described in a previous study [41]. The phasing amplicon sequencing (PAS) method [48] was used to amplify the V4 region of 16S rRNA genes and the samples were sequenced on an Illumina MiSeq platform. The primers are 515 F (5′-GTG CCAGCMGCCGCGGTAA- 3′) and 806R (5′-GGACTA CHVGGGTWTCTAAT- 3′). The sequencing data were processed by Qiime2 (version 2019.7). After barcode and primer sequences were trimmed with zero maximum error, sequencing data were processed by DADA2 to identify exact amplicon sequence variants (ASV), followed by the default settings. The ASVs were identified taxonomically based on the silva- 132–99–515–806-nb-classifier. The ASV sequences were then used to build a phylogenetic tree by FastTree [34, 35]. The ASV table was processed by removing reads classified at the Order level as "Chloroplast" and "Mitochondria." To ensure comparability, all samples were rarefied to 28,322 sequences, standardizing sequencing depth for downstream analyses.

We used the KAPA Hyper Prep Kit (KR0961) to construct the metagenomic sequencing libraries following the manufacturer's instructions, and the samples were sequenced using an Illumina HiSeq 3000 sequencer. The read-based metagenomic data analysis was performed using an internal pipeline (http://iegst1.rccc.ou.edu:8080/ecofunmap/) following the guidelines from [40]. For assembly-based analysis, metagenomic reads were preprocessed using BBTools for removing adaptor, trimming and filtering reads, and sequencing error correction [6]. The pre-processed reads were assembled with Metaspades [30]. Genes were predicted from scaffolds >1 kbp using the Prodigal [18]. The gene abundance was estimated as TPM [53]. Genes functions were annotated using the Kofamsan [19]. Species-level quantitative taxonomic profiling was performed using MetaPhlAn4 [5].

### EcoFun-MAP

EcoFun-MAP (Ecological Function Oriented Metagenomic Analysis Pipeline) is a bioinformatics tool designed for the functional analysis of shotgun metagenomic sequencing data with an ecological focus. It provides an efficient framework for annotating functional genes relevant to key biogeochemical processes, such as carbon, nitrogen, sulfur, and phosphorus cycling, as well as antibiotic resistance, metal homeostasis, and microbial stress responses. The pipeline is based on a gene-centric paradigm, utilizing a hierarchical reference database that categorizes functional genes according to their ecological roles. It employs a combination of Hidden Markov Models (HMMs), DIAMOND-based sequence alignment, and probabilistic modeling to enhance both speed and accuracy in gene annotation [40].

### Statistical analyses

We used richness and Shannon index to represent microbial α-diversity and analysis of variance (ANOVA) to compare the difference between each group. The assumptions of normality and homogeneity of variances were validated prior to ANOVA tests using the Shapiro–Wilk test and Levene's test, respectively. Post hoc pairwise comparisons were conducted using Tukey's Honest Significant Difference (HSD) test to identify specific group differences when ANOVA results were significant ($p < 0.05$). To control for multiple comparisons and reduce the risk of Type I errors, Bonferroni corrections were applied where appropriate. Non-metric multidimensional (NMDS) and permutation test of multivariate homogeneity of groups dispersions [2] were used for microbial β-diversity, taxonomic and functional β-diversity were measured by Bray–Curtis dissimilarity [7] and phylogenetic β-diversity were measured by UniFrac [25]. We used permutation test of multivariate homogeneity of groups dispersions [2] based on the Euclidean distance for environmental variables. The statistical significance of the effects of contaminations on β-diversity was tested by multi response permutation procedure (MRPP) [47], permutational multivariate analysis of variance (Adonis) [3], and analysis of similarities (Anosim) [11, 47] by using function "mrpp," "adonis," and "anosim" in R package "vegan," respectively [31]. The response ratio was calculated using an internal R packcage "ieggr" based on 95% confidence intervals. We used a partial least squares (PLS) model to detect the relationships between environmental variables and geographical distance for each diversity index. Basically, each optimal PLS model is selected through a forward selection process from all factors that could influence the dependent variable. This selection is based on predictive performance, taking into account the proportion of variation explained ($R^2Y$) and the statistical significance of the model (*P* values for $R^2Y$ and $Q^2Y$ less than 0.05). Notably, a significant $Q^2Y$ aids in preventing overfitting of the model. To visualize the relevant associations, we used the variance partition analysis (VPA) and the software Inkscape 1.3 (https://inkscape.org/). The PLS-related analysis was performed using the R package "ropls" [44].

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-025-02105-x.

Additional file 1: Tables S1–S3 and Figs. S1–S4.

**Acknowledgements**
This study by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies; http://enigma.lbl.gov), a Science Focus Area Program at Lawrence Berkeley National Laboratory, is based on work supported by the US Department of Energy, Office of Science, Biological and Environmental Research Program under contract number DE-AC02-05CH11231 to Lawrence Berkeley National Lanoratory.

**Authors' contributions**
All authors contributed intellectual input and assistance to this study and the paper preparation. J. Z.. conceived the research question. Z. H., M. W. W. A., M. W. F., E. J. A., T. C. H., P. D. A., A. P. A., and J. Z. designed and organized the experiment. P. Z., A. M. R., J. D. V. N., and D. C. J. collected or generated the data. Y.F., D. W., J. X. Y. and J. P. M. intergrated the data and performed statistical analyses with the assistance of D.N. and C. P.; Y.F. and J. X. Y. wrote the paper with inputs from D.N. and J.Z.

**Funding**
This work is supported by the US Department of Energy, Office of Science, Biological and Environmental Research Program under contract number DE-AC02-05CH11231 to Lawrence Berkeley National Lanoratory.

**Data availability**
The 16S rRNA gene amplicon sequences were submitted to NCBI database SRA under the project PRJNA514085 with accession SRR8427255. The shotgun metagenomic sequences were submitted to NCBI database SRA under the project PRJNA513876 with accession SRR8426587 - SRR8426598.

## 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]Institute for Environmental Genomics, University of Oklahoma, 101 David L Boren Blvd, Norman, OK 73019, USA. [2]School of Biological Sciences, University of Oklahoma, Norman, OK 73019, USA. [3]Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai, Guangdong 519080, China. [4]Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX 77030, USA. [5]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA. [6]Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996, USA. [7]School of Computer Science, University of Oklahoma, Norman, OK 73019, USA. [8]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA. [9]Center for Biofilm Engineering and Department of Microbiology and Cell Biology, Montana State University, Bozeman, MT 59717, USA. [10]Department of Biological Engineering, Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [11]Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA. [12]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [13]Department of Bioengineering, University of California, Berkeley, CA 94720, USA. [14]Division of Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [15]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [16]School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK 73019, USA.

Received: 14 August 2024 Accepted: 2 April 2025
Published online: 28 April 2025

**References**

1. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun. 2016;7:13219.
2. Anderson MJ. Distance-based tests for homogeneity of multivariate dispersions. Biometrics. 2006;62(1):245–53.
3. Anderson MJ. Permutational Multivariate Analysis of Variance (PERMANOVA). In: Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F, Teugels JL, editors. 'Wiley StatsRef: Statistics Reference Online'. Chichester: Wiley; 2017. pp. 1–15
4. Arnault G, Mony C, Vandenkoornhuyse P. Plant microbiota dysbiosis and the Anna Karenina Principle. Trends Plant Sci. 2023;28(1):18–30.
5. Blanco-Miguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, Manghi P, Dubois L, Huang KD, Thomas AM, Nickols WA, Piccinno G, Piperni E, Puncochar M, Valles-Colomer M, Tett A, Giordano F, Davies R, Wolf J, Berry SE, Spector TD, Franzosa EA, Pasolli E, Asnicar F, Huttenhower C, Segata N. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat Biotechnol. 2023;41(11):1633–44.
6. Bushnell B. BBTools software package. http://bbtools.jgi.doe.gov. 2014.
7. Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr. 1957;27(4):325–49.
8. Carlson HK, Price MN, Callaghan M, Aaring A, Chakraborty R, Liu H, Kuehl JV, Arkin AP, Deutschbauer AM. The selective pressures on the microbial community in a metal-contaminated aquifer. ISME J. 2019;13(4):937–49.
9. Chen H, Chen Z, Chu X, Deng Y, Qing S, Sun C, Wang Q, Zhou H, Cheng H, Zhan W, Wang Y. Temperature mediated the balance between stochastic and deterministic processes and reoccurrence of microbial community during treating aniline wastewater. Water Res. 2022;221: 118741.
10. Chen J, He F, Zhang X, Sun X, Zheng J, Zheng J. Heavy metal pollution decreases microbial abundance, diversity and activity within particle-size fractions of a paddy soil. FEMS Microbiol Ecol. 2014;87(1):164–81.
11. Clarke KR. Non-parametric multivariate analyses of changes in community structure. Aust J Ecol. 1993;18(1):117–43.
12. Fierer N, Ladau J, Clemente JC, Leff JW, Owens SM, Pollard KS, Knight R, Gilbert JA, McCulley RL. Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. Science. 2013;342(6158):621–4.
13. Galand PE, Pereira O, Hochart C, Auguet JC, Debroas D. A strong link between marine microbial community composition and function challenges the idea of functional redundancy. ISME J. 2018;12(10):2470–8.
14. Gao Q, Dong Q, Wu L, Yang Y, Hale L, Qin Z, Xie C, Zhang Q, Van Nostrand JD, Zhou J. Environmental antibiotics drives the genetic functions of resistome dynamics. Environ Int. 2020;135: 105398.
15. Green SJ, Prakash O, Jasrotia P, Overholt WA, Cardenas E, Hubbard D, Tiedje JM, Watson DB, Schadt CW, Brooks SC, Kostka JE. Denitrifying bacteria from the genus Rhodanobacter dominate bacterial communities in the highly contaminated subsurface of a nuclear legacy waste site. Appl Environ Microbiol. 2012;78(4):1039–47.
16. Grettenberger CL, Hamilton TL. Metagenome-assembled genomes of novel taxa from an acid mine drainage environment. Appl Environ Microbiol. 2021;87(17):e0077221.
17. He Z, Zhang P, Wu L, Rocha AM, Tu Q, Shi Z, Wu B, Qin Y, Wang J, Yan Q, Curtis D, Ning D, Van Nostrand JD, Wu L, Yang Y, Elias DA, Watson DB, Adams MWW, Fields MW, Alm EJ, Hazen TC, Adams PD, Arkin AP, Zhou J. Microbial functional gene diversity predicts groundwater contamination and ecosystem functioning. mBio. 2018;9(1):10–1128.
18. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
19. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci. 2020;29(1):28–35.
20. Karthikeyan OP, Smith TJ, Dandare SU, Parwin KS, Singh H, Loh HX, Cunningham MR, Williams PN, Nichol T, Subramanian A, Ramasamy K, Kumaresan D. Metal(loid) speciation and transformation by aerobic methanotrophs. Microbiome. 2021;9(1):156.
21. King AJ, Preheim SP, Bailey KL, Robeson MS 2nd, Roy Chowdhury T, Crable BR, Hurt RA Jr, Mehlhorn T, Lowe KA, Phelps TJ, Palumbo AV, Brandt CC, Brown SD, Podar M, Zhang P, Lancaster WA, Poole F, Watson DB, M, W.F.,

Chandonia, J.M., Alm, E.J., Zhou, J., Adams, M.W., Hazen, T.C., Arkin, A.P. and Elias, D.A. Temporal dynamics of in-field bioreactor populations reflect the groundwater system and respond predictably to perturbation. Environ Sci Technol. 2017;51(5):2879–89.
22. Kojima H, Shinohara A, Fukui M. Sulfurifustis variabilis gen. nov., sp. nov., a sulfur oxidizer isolated from a lake, and proposal of Acidiferrobacteraceae fam. nov. and Acidiferrobacterales ord. nov. Int J Syst Evol Microbiol. 2015;65(10):3709–13.
23. Liang T, Ding H, Wang G, Kang J, Pang H, Lv J. Sulfur decreases cadmium translocation and enhances cadmium tolerance by promoting sulfur assimilation and glutathione metabolism in Brassica chinensis L. Ecotoxicol Environ Saf. 2016;124:129–37.
24. Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean microbiome. Science. 2016;353(6305):1272–7.
25. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. ISME J. 2011;5(2):169–72.
26. Ma X, Zhang Q, Zheng M, Gao Y, Yuan T, Hale L, Van Nostrand JD, Zhou J, Wan S, Yang Y. Microbial functional traits are sensitive indicators of mild disturbance by lamb grazing. ISME J. 2019;13(5):1370–3.
27. Maestre FT, Delgado-Baquerizo M, Jeffries TC, Eldridge DJ, Ochoa V, Gozalo B, Quero JL, Garcia-Gomez M, Gallardo A, Ulrich W, Bowker MA, Arredondo T, Barraza-Zepeda C, Bran D, Florentino A, Gaitan J, Gutierrez JR, Huber-Sannwald E, Jankju M, Mau RL, Miriti M, Naseri K, Ospina A, Stavi I, Wang D, Woods NN, Yuan X, Zaady E, Singh BK. Increasing aridity reduces soil microbial diversity and abundance in global drylands. Proc Natl Acad Sci U S A. 2015;112(51):15684–9.
28. Nielsen UN, Ayres E, Wall DH, Bardgett RD. Soil biodiversity and carbon cycling: a review and synthesis of studies examining diversity–function relationships. Eur J Soil Sci. 2011;62(1):105–16.
29. Ning D, Wang Y, Fan Y, Wang J, Van Nostrand JD, Wu L, Zhang P, Curtis DJ, Tian R, Lui L, Hazen TC, Alm EJ, Fields MW, Poole F, Adams MWW, Chakraborty R, Stahl DA, Adams PD, Arkin AP, He Z, Zhou J. Environmental stress mediates groundwater microbial community assembly. Nat Microbiol. 2024;9(2):490–501.
30. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27(5):824–34.
31. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Szoecs E, Wagner H. Vegan: community ecology package. R package version 2.5–7. 2020. Available at https://CRAN.Rproject.org/package=vegan.
32. Power JF, Carere CR, Lee CK, Wakerley GLJ, Evans DW, Button M, White D, Climo MD, Hinze AM, Morgan XC, McDonald IR, Cary SC, Stott MB. Microbial biogeography of 925 geothermal springs in New Zealand. Nat Commun. 2018;9(1):2876.
33. Prakash O, Green SJ, Singh P, Jasrotia P, Kostka JE. Stress-related ecophysiology of members of the genus Rhodanobacter isolated from a mixed waste contaminated subsurface. Front Environ Sci Eng. 2021;15(2):1–9.
34. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26(7):1641–50.
35. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5(3): e9490.
36. Revel B, Catty P, Ravanel S, Bourguignon J, Alban C. High-affinity iron and calcium transport pathways are involved in U(VI) uptake in the budding yeast Saccharomyces cerevisiae. J Hazard Mater. 2022;422: 126894.
37. Rocca JD, Simonin M, Blaszczak JR, Ernakovich JG, Gibbons SM, Midani FS, Washburne AD. The Microbiome Stress Project: Toward a Global Meta-Analysis of Environmental Stressors and Their Effects on Microbial Communities. Front Microbiol. 2018;9:3272.
38. Ruhl IA, Grasby SE, Haupt ES, Dunfield PF. Analysis of microbial communities in natural halite springs reveals a domain-dependent relationship of species diversity to osmotic stress. Environ Microbiol Rep. 2018;10(6):695–703.
39. Sarthou MCM, Devime F, Baggio C, Figuet S, Alban C, Bourguignon J, Ravanel S. Calcium-permeable cation channels are involved in uranium uptake in Arabidopsis thaliana. J Hazard Mater. 2022;424: 127436.
40. Shi ZJ, Xiao N, Ning D, Tian R, Zhang P, Curtis D, Van Nostrand JD, Wu L, Hazen TC, Rocha AM, He Z, Arkin AP, Firestone MK, Zhou J. EcoFun-MAP: An Ecological Function Oriented Metagenomic Analysis Pipeline. bioRxiv. 2022. https://doi.org/10.1101/2022.04.05.481366.
41. Smith MB, Rocha AM, Smillie CS, Olesen SW, Paradis C, Wu L, Campbell JH, Fortney JL, Mehlhorn TL, Lowe KA, Earles JE, Phillips J, Techtmann SM, Joyner DC, Elias DA, Bailey KL, Hurt RA, Preheim SP Jr, Sanders MC, Yang J, Mueller MA, Brooks S, Watson DB, Zhang P, He Z, Dubinsky EA, Adams PD, Arkin AP, Fields MW, Zhou J, Alm EJ, Hazen TC. Natural bacterial communities serve as quantitative geochemical biosensors. mBio. 2015;6(3):e00326-00315.
42. Song HK, Shi Y, Yang T, Chu H, He JS, Kim H, Jablonski P, Adams JM. Environmental filtering of bacterial functional diversity along an aridity gradient. Sci Rep. 2019;9(1):866.
43. Talbot JM, Bruns TD, Taylor JW, Smith DP, Branco S, Glassman SI, Erlandson S, Vilgalys R, Liao HL, Smith ME, Peay KG. Endemism and functional convergence across the North American soil mycobiome. Proc Natl Acad Sci U S A. 2014;111(17):6341–6.
44. Thevenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. J Proteome Res. 2015;14(8):3322–35.
45. Trenfield MA, Ng JC, Noller B, Markich SJ, van Dam RA. Dissolved organic carbon reduces uranium toxicity to the unicellular eukaryote Euglena gracilis. Ecotoxicolog. 2012;21(4):1013–23.
46. Versantvoort W, Guerrero-Cruz S, Speth DR, Frank J, Gambelli L, Cremers G, van Alen T, Jetten MSM, Kartal B, Op den Camp HJM, Reimann J. Comparative genomics of Candidatus Methylomirabilis species and description of Ca. Methylomirabilis Lanthanidiphila Front Microbiol. 2018;9:1672.
47. Warton DI, Wright ST, Wang Y. Distance-based multivariate analyses confound location and dispersion effects. Methods Ecol Evol. 2012;3(1):89–101.
48. Wu L, Wen C, Qin Y, Yin H, Tu Q, Van Nostrand JD, Yuan T, Yuan M, Deng Y, Zhou J. Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. BMC Microbiol. 2015;15:125.
49. Whittaker RH. Evolution and measurement of species diversity. Taxon. 1972;21(2–3):213–51.
50. Xu M, Wu WM, Wu L, He Z, Van Nostrand JD, Deng Y, Luo J, Carley J, Ginder-Vogel M, Gentry TJ, Gu B, Watson D, Jardine PM, Marsh TL, Tiedje JM, Hazen T, Criddle CS, Zhou J. Responses of microbial community functional structures to pilot-scale uranium in situ bioremediation. ISME J. 2010;4(8):1060–70.
51. Zaneveld JR, McMinds R, Vega Thurber R. Stress and stability: applying the Anna Karenina principle to animal microbiomes. Nat Microbiol. 2017;2:17121.
52. Zhang P, Wu WM, Van Nostrand JD, Deng Y, He Z, Gihring T, Zhang G, Schadt CW, Watson D, Jardine P, Criddle CS, Brooks S, Marsh TL, Tiedje JM, Arkin AP, Zhou J. Dynamic succession of groundwater functional microbial communities in response to emulsified vegetable oil amendment during sustained In Situ U(VI) Reduction. Appl Environ Microbiol. 2015;81(12):4164–72.
53. Zhao S, Ye Z, Stanton R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. RNA. 2020;26(8):903–9.
54. Zhou J, Deng Y, Zhang P, Xue K, Liang Y, Van Nostrand JD, Yang Y, He Z, Wu L, Stahl DA, Hazen TC, Tiedje JM, Arkin AP. Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. Proc Natl Acad Sci U S A. 2014;111(9):E836-845.
55. Zhou J, Ning D. Stochastic community assembly: does it matter in microbial ecology? Microbiol Mol Biol Rev. 2017;81(4).

## Publisher's Note