Final Report

Project Title: Microbial diversity in arsenic-contaminated groundwater

By Prinpida Sonthiphand, Ph.D.

3 / 2020

Contract No. MRG6180127

Final Report

Project Title: Microbial diversity in arsenic-contaminated groundwater

Researcher	Institute
Prinpida Sonthiphand	Mahidol University
Saowarod Polart	Mahidol University
Supeerapat Kraidech	Chulalongkorn University
Srilert Chotpantarat	Chulalongkorn University
Ekawan Luepromchai	Chulalongkorn University
Kanthida Kusonmano	King Mongkut's University of Technology Thonburi
Pichahpuk Uthaipaisanwong	King Mongkut's University of Technology Thonburi
Chalida Rangsiwutisak	King Mongkut's University of Technology Thonburi

This Project Granted by the Thailand Research Fund

Office of the Higher Education Commission and Mahidol University

Abstract

Project Code : MRG6180127

Project Title : Microbial diversity in arsenic-contaminated groundwater

Investigator : Prinpida Sonthiphand

E-mail Address : prinpida.son@mahidol.ac.th

Project Period : 2 May 2018 – 1 May 2020

Abstract: The microbial community structures and arsenite-oxidizing bacteria in the groundwater, surface water, and soil within and surrounding a gold mining area were explored using high-throughput sequencing of the 16S rRNA gene and cloning-ddPCR of the aioA gene. Metagenomic analysis revealed that although *Proteobacteria* (13-99%) were commonly detected across all analyzed samples, the microbial diversity in the soil was relatively higher than that in the groundwater and surface water. The majority of the soil and surface water microbiomes were *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Chloroflexi*, whereas the groundwater microbiomes were dominated exclusively by the classes *Betaproteobacteria* and *Alphaproteobacteria*. Each sample harbored unique bacterial taxa, and cluster analyses showed that habitat primarily governed the community structures. Geochemical factors that potentially influenced the microbial structure in the groundwater were As, residence time, and groundwater flowrate, while those showing a positive correlation to the microbial structure in the surface water were TOC, ORP, and DO. The analysis of the aioA gene showed that arsenite-oxidizing bacteria retrieved from groundwater, surface water, and soil were associated with *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*. The ddPCR results indicated that arsenite-oxidizing bacteria were widespread at low abundance in all samples with low As concentrations. This study provides insights into the groundwater, surface water, and soil microbiomes within and surrounding a gold mine and expands the current understanding of the diversity and abundance of arsenite-oxidizing bacteria, playing a vital role in global As cycling.

Keywords : microbial community, groundwater, surface water, soil, gold mine, *aioA* gene, arsenite-oxidizing bacteria

บทคัดย่อ

รหัสโครงการ : MRG6180127
ชื่อโครงการ : ความหลากหลายของจุลินทรีย์ในน้ำใต้ดินปนเปื้อนสารหนู
นักวิจัย : พรินท์พิดา สนธิพันธ์
อีเมล์ : prinpida.son@mahidol.ac.th
ระยะเวลาโครงการ : 2 พฤษภาคม 2561 – 1 พฤษภาคม 2563

Abstract: งานวิจัยนี้ศึกษาโครงสร้างกลุ่มประชากรจุลินทรีย์และอาร์ซีไนต์ออกซิไดซิ่งแบคทีเรียในน้ำใต้ดิน น้ำผิวดิน และดิน บริเวณเหมืองทองและรอบ ๆ เหมืองทอง โดยใช้เทคโนโลยี high-throughput sequencing อ่านลำดับนิวคลีโอไทด์ของยีน 16S rRNA และเทคนิค PCR-cloning-sequencing ในการศึกษายีน *aioA* ผลจากการวิเคราะห์เมตาจีโนมิกส์แสดงให้เห็นว่า *Proteobacteria* (13-99%) เป็นกลุ่มจุลินทรีย์ที่ตรวจพบได้ทั่วไปในตัวอย่างที่ศึกษา และพบว่าตัวอย่างดินมีความหลากหลายของจุลินทรีย์มากกว่าตัวอย่างน้ำใต้ดินและน้ำผิวดิน กลุ่มประชากรจุลินทรีย์หลักที่พบในดินและน้ำผิวดิน คือ *Proteobacteria Actinobacteria Bacteroidetes* และ *Chloroflexi* ส่วนกลุ่มประชากรจุลินทรีย์หลักที่พบในน้ำใต้ดิน คือ *Betaproteobacteria* และ *Alphaproteobacteria* อีกทั้งพบกลุ่มจุลินทรีย์เฉพาะในตัวอย่างที่ศึกษาแต่ละชนิด และการวิเคราะห์คลัสเตอร์แสดงให้เห็นว่าชนิดตัวอย่างมีผลต่อโครงสร้างกลุ่มประชากรจุลินทรีย์ ปัจจัยทางธรณีเคมีที่มีผลต่อโครงสร้างกลุ่มประชากรจุลินทรีย์ในน้ำใต้ดิน คือ สารหนู ระยะเวลาคงอยู่ และอัตราการไหลของน้ำใต้ดิน สำหรับปัจจัยทางธรณีเคมีที่มีผลต่อโครงสร้างกลุ่มประชากรจุลินทรีย์ในน้ำผิวดิน คือ TOC ORP และ DO การศึกษายีน *aioA* แสดงให้เห็นว่าอาร์ซีไนต์ออกซิไดซิ่งแบคทีเรียที่พบในน้ำใต้ดิน น้ำผิวดิน และดิน มีความใกล้ชิดกับจุลินทรีย์ในกลุ่ม *Alphaproteobacteria Betaproteobacteria* และ *Gammaproteobacteria* ผล ddPCR แสดงให้เห็นว่าอาร์ซีไนต์ออกซิไดซิ่งแบคทีเรียมีปริมาณน้อยแต่พบได้ทั่วไปในตัวอย่างศึกษาที่มีความเข้มข้นสารหนูต่ำ งานวิจัยนี้สร้างฐานข้อมูลความหลากหลายทางชีวภาพของกลุ่มประชากรจุลินทรีย์ในน้ำใต้ดิน น้ำผิวดิน และดินบริเวณเหมืองทองและรอบ ๆ เหมืองทอง รวมทั้งฐานข้อมูลความหลากหลายของอาร์ซีไนต์ออกซิไดซิ่งแบคทีเรียที่มีบทบาทในวัฏจักรสารหนู

คำสำคัญ : กลุ่มประชากรจุลินทรีย์, น้ำใต้ดิน, น้ำผิวดิน, ดิน, เหมืองทอง, ยีน *aioA*, อาร์ซีไนต์ออกซิไดซิ่งแบคทีเรีย

Final report content:

1. Abstract

The microbial community structures and arsenite-oxidizing bacteria in the groundwater, surface water, and soil within and surrounding a gold mining area were explored using high-throughput sequencing of the 16S rRNA gene and cloning-ddPCR of the aioA gene. Metagenomic analysis revealed that although *Proteobacteria* (13-99%) were commonly detected across all analyzed samples, the microbial diversity in the soil was relatively higher than that in the groundwater and surface water. The majority of the soil and surface water microbiomes were *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Chloroflexi*, whereas the groundwater microbiomes were dominated exclusively by the classes *Betaproteobacteria* and *Alphaproteobacteria*. Each sample harbored unique bacterial taxa, and cluster analyses showed that habitat primarily governed the community structures. Geochemical factors that potentially influenced the microbial structure in the groundwater were As, residence time, and groundwater flowrate, while those showing a positive correlation to the microbial structure in the surface water were TOC, ORP, and DO. The analysis of the aioA gene showed that arsenite-oxidizing bacteria retrieved from groundwater, surface water, and soil were associated with *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*. The ddPCR results indicated that arsenite-oxidizing bacteria were widespread at low abundance in all samples with low As concentrations. This study provides insights into the groundwater, surface water, and soil microbiomes within and surrounding a gold mine and expands the current understanding of the diversity and abundance of arsenite-oxidizing bacteria, playing a vital role in global As cycling.

2. Executive summary

The microbial community structures in the groundwater, surface water, and soil from the gold mining area, including the upstream and downstream areas of the mine, were explored using MiSeq sequencing of the 16S rRNA gene. As concentrations were low across all sampling locations, possibly due to an effect of the wet season. The phylum Proteobacteria was highly represented in all analyzed samples. The soil and surface water microbiomes were mainly composed of the phyla *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Chloroflexi*. *Oxyphotobacteria*, members of the phylum *Cyanobacteria*, were dominant in surface waters but not in soil or groundwater. The groundwater microbiomes were dominated exclusively by *Proteobacteria*, especially the

classes *Betaproteobacteria* and *Alphaproteobacteria*. The heatmap of the most abundant OTUs revealed unique bacterial taxa in each sample. *Limnohabitans*, unclassified *Rhodocyclaceae*, *Hydrogenophaga*, and *Ferriphaselus* were dominant in L1-GW, L3-GW, L5-GW, and L6-GW, respectively. *Novosphingobium* was dominant in L2-GW and L4-GW. Cyanobium_PCC-6307 was dominant in L1-SW, whereas MWH-UniP1_aquatic_group was dominant in L2-SW and L6-SW. The three microbial assemblages predominant in L1-S, L2-S, and L6-S were *Pseudarthrobacter*, *Microvirga*, and uncultured *Hydrogenophilaceae*, respectively. These dominant bacterial taxa likely play roles in the carbon, nitrogen, and sulfur cycles. Cluster analyses showed that habitat primarily governed the microbial structures in the analyzed samples. The environmental factors that potentially influenced the microbial structure in the groundwater were As, residence time, and groundwater flowrate, while those affecting the microbial structure in the surface water were TOC, ORP, and DO. Arsenite-oxidizing bacteria in the groundwater, surface water, and soil samples were retrieved by cloning-sequencing of the *aioA* gene. All analyzed *aioA* sequences were closely related to *Proteobacteria*, indicating common arsenite-oxidizing bacterial taxa found in a mesophilic environment. The *aioA* sequences recovered from soils were related to *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*. The majority of *aioA* sequences that were detected in the soil and surface water samples were affiliated with *Alphaproteobacteria*. The groundwater-associated arenite-oxidizing bacterial taxa were *Alphaproteobacteria* and *Betaproteobacteria*. Together with the results of the ddPCR analysis, the results suggested that arsenite-oxidizing bacteria were widespread at low abundance across samples with low As concentrations. However, microbial community structure and arsenite-oxidizing bacteria were monitored only in a single season, and they should be further analyzed at different time points to better understand the effects of spatial and temporal variations. Overall, this study provides insights into the groundwater, surface water, and soil microbiomes within and surrounding a gold mine and expands the current knowledge of the diversity and abundance of arsenite-oxidizing bacteria, which could play a role in As transformation in impacted environments.

3. Result

3.1 Sample characteristics

Five groundwater samples (L1-GW to L5-GW) were collected from monitoring wells, whereas one sample (L6-GW) was collected from a pumping well. When groundwater

was consumed, the pumping system continuously pumped the groundwater into a storage tank. However, the groundwater in the monitoring wells was likely undisturbed. Therefore, the characteristics of groundwater wells likely affected the hydraulic retention time and groundwater flowrate. The groundwater residence time and groundwater flowrate of the monitoring wells were longer and slower, respectively, than those of the pumping wells. These characteristics may influence the indigenous groundwater microbiome in the monitoring and pumping wells.

Geochemical parameters indicated that temperature, TKN, and TP were comparable among the six groundwater and three surface samples (Table 1). Most of the analyzed water samples were neutral to slightly acidic, except for L5-GW (moderate acidic) and L6-SW (slightly basic). The concentrations of TOC were relatively low across almost all water samples; however, the TOC concentration was considerably high in L6-SW. DO values showed that the groundwater was less oxic than the surface water. The ORP measurements also indicated that the groundwater conditions were relatively reduced compared to the surface water conditions (Table 1). Although the concentrations of As were low and comparable in all water samples, the speciation of As heavily relied on pH and ORP (Smedley, 2008). In this study, the geochemical model PHREEQC was applied to predict the dominant species of As found in groundwater and surface water with different geochemical conditions.

Table 1. Geochemical parameters of 6 groundwater (GW) and 3 surface water (SW) samples

sample ID	pH	Temp. (°C)	DO (mg l^{-1})	ORP (mV)	TOC (mg l^{-1})	TKN (mg $l^{-1)}$	TP (mg l^{-1})	Total As (μg l^{-1})
L1-GW	6.50	29.20	3.87	212.20	1.24	0.30	0.02	3.97
L2-GW	6.87	32.30	2.84	59.40	1.34	0.40	0.02	3.52
L3-GW	6.39	30.50	3.29	115.40	3.24	0.10	0.05	4.45
L4-GW	6.70	31.80	2.95	88.90	1.40	0.30	0.06	3.50
L5-GW	4.90	31.00	2.92	296.90	1.22	0.10	0.01	3.12
L6-GW	7.21	32.30	4.79	149.50	1.43	0.50	0.04	2.05
L1-SW	6.34	31.00	7.63	390.30	4.19	0.10	0.05	2.29
L2-SW	6.98	33.40	7.14	308.10	2.92	0.50	0.02	2.31
L6-SW	8.22	34.00	7.19	269.10	16.62	0.30	0.03	2.66

3.2 Prediction of As speciation by using the geochemical model PHREEQC

The PHREEQC model revealed that As^{5+} was the dominant species found in all analyzed groundwater and surface water samples (Table 2). $H_2AsO_4^-$ was the dominant form detected in groundwater and surface water collected from locations L1 to L5, whereas $HAsO_4^{2-}$ was the major form of As^{5+} in samples taken from location L6 (Figure 1 and Table 2). These predicted forms are commonly found in natural surface and groundwater environments (Pal, 2015). The concentrations of DO across all analyzed water samples indicated slightly to highly oxic conditions, which potentially led to the predominant of As^{5+}.

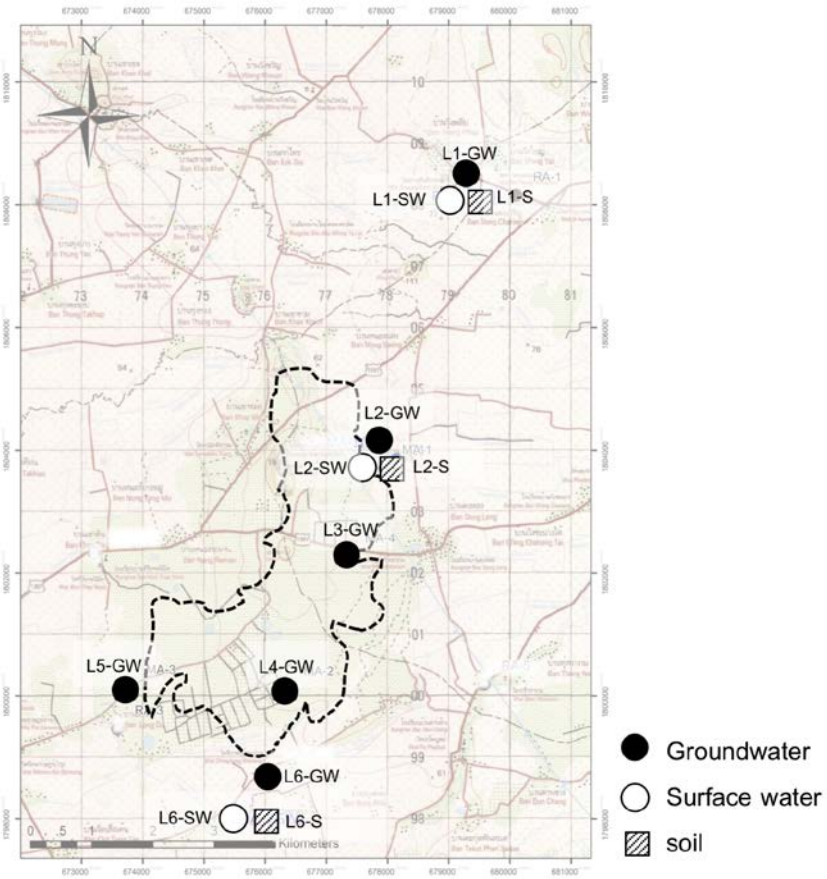

Figure 1. Map showing the sampling locations (L1-L6). GW, SW, and S refer to the groundwater, surface water, and soil samples, respectively.

Table 2. As speciation calculated by using the geochemical model PHREEQC

Sample ID	As^{3+} ($\mu g\ l^{-1}$)	As^{5+} ($\mu g\ l^{-1}$)	Dominant form (speciation model)	Dominant form (Hfo-As sorption model)
L1-GW	0.00	3.97	$H_2AsO_4^-$	$H_2AsO_4^-$
L2-GW	0.08	3.45	$H_2AsO_4^-$	$H_2AsO_4^-$
L3-GW	0.06	4.40	$H_2AsO_4^-$	$H_2AsO_4^-$
L4-GW	0.03	3.47	$H_2AsO_4^-$	$H_2AsO_4^-$
L5-GW	0.00	3.12	$H_2AsO_4^-$	$H_2AsO_4^-$
L6-GW	0.00	2.05	$HAsO_4^{2-}$	$hfo\_wOHAsO_4^{3-}$
L1-SW	0.00	2.29	$H_2AsO_4^-$	$hfo\_wOHAsO_4^{3-}$
L2-SW	0.00	2.31	$H_2AsO_4^-$	$hfo\_wOHAsO_4^{3-}$
L6-SW	0.00	2.66	$HAsO_4^{2-}$	$HAsO_4^{2-}$

Since sorption plays an important role in As speciation, a sorption model was also considered in this analysis. Due to the strong sorption capacities of As on amorphous iron oxides, the Hfo-As sorption model was selected for predicting whether the presence of iron oxides (As, Fe, Mn, and SO_4^{2-}) influenced As mobility in the analyzed samples (Smedley, 2008). Groundwater conditions associated with the presence of As species were illustrated by the Hfo-As sorption model. As with the resulting speciation model, $H_2AsO_4^-$ and $HAsO_4^{2-}$ were dominant in samples L1-GW to L5-GW and L6-SW, respectively (Table 2). However, $hfo\_wOHAsO_4^{3-}$ was found to be the major form detected in L6-GW, L1-SW, and L2-SW. The As forms under oxic conditions are related to electrostatic repulsion from the negative charge of the oxide surface, resulting in Hfo instability and dissolution (Smedley, 2008). The strong sorption tendency of iron oxides around a neutral pH range in oxic conditions is one of the major causes of low As concentrations in most natural groundwater (Smedley, 2008). However, Fe concentrations in this study were relatively similar; thus, Hfo sorption might not be the major factor controlling the As concentration in groundwater. On the other hand, dissolved As associated with S-species occurs in highly reducing conditions where the As-S mineral limits the dissolved As concentration in high sulfide concentrations (Smedley, 2008).

3.3 Microbial community structures in groundwater, surface water, and soil across low arsenic concentrations

The alpha diversity of the groundwater, surface water, and soil across low arsenic concentrations was revealed by the observed OTUs, Shannon, and Chao1. Rarefaction curves showed that microbial richness in soils was relatively higher than that in groundwater and surface water (Figure 2). Microbial richness was the highest in L1-S, followed by L2-S, and it was the lowest in L6-GW. Among the water samples, L2-GW and L3-GW showed higher microbial richness, which were comparable to L6-S (Figure 2). Microbial richness in other analyzed water samples, both groundwater and surface water, were comparable in the range of 100-500 observed OTUs.

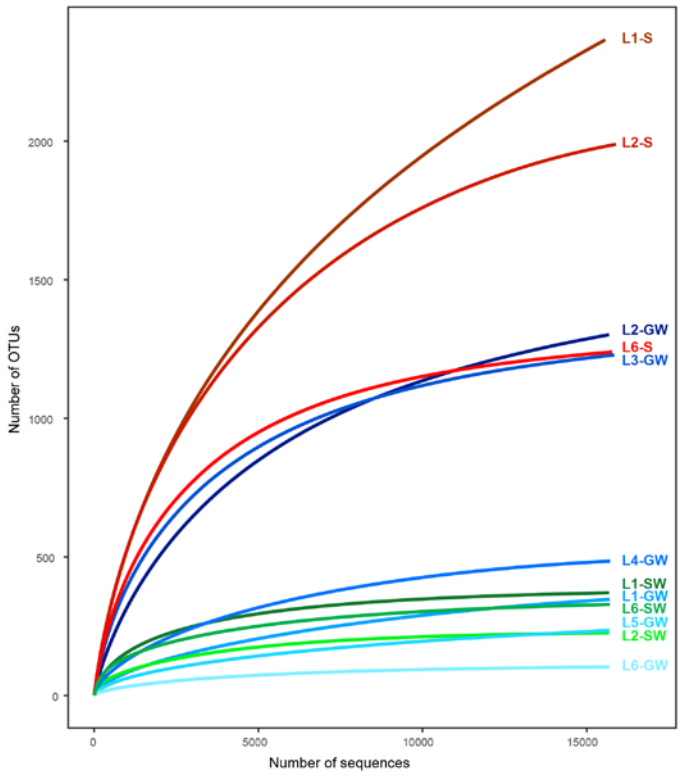

Figure 2. Rarefaction curves based on the number of OTUs of the bacterial 16S rRNA gene in groundwater (GW), surface water (SW), and soil (S) from 6 sampling locations (L1-L6).

Microbial compositions in groundwater, surface water, and soil from the gold mining area with low arsenic concentrations were determined by the MiSeq Illumina sequencing of the 16S rRNA gene. The resulting microbial abundance at the phylum level indicated that the phylum *Proteobacteria* was common in all samples, accounting for 13-99% of the total microbial abundance (Figure 3). The microbial community structure in groundwater,

L1-GW to L6-GW, was dominated by the phylum *Proteobacteria*, accounting for 65-95% of the total microbial abundance. The four major phyla that were found in the surface water, L1-SW, L2-SW, and L6-SW, were *Proteobacteria* (13-66%), *Cyanobacteria* (13-36%), *Actinobacteria* (7-28%), and *Bacteroidetes* (6-19%). Microbial compositions in soil samples, L1-S, L2-S, and L6-S, were mainly composed of the phya *Proteobacteria* (25-59%), *Actinobacteria* (9-35%), *Acidobacteria* (5-14%), and *Chloroflexi* (8-9%). The results indicated that microbial community structure was closely associated within each sample type, whereas it changed from one sample type to another (Figure 3).

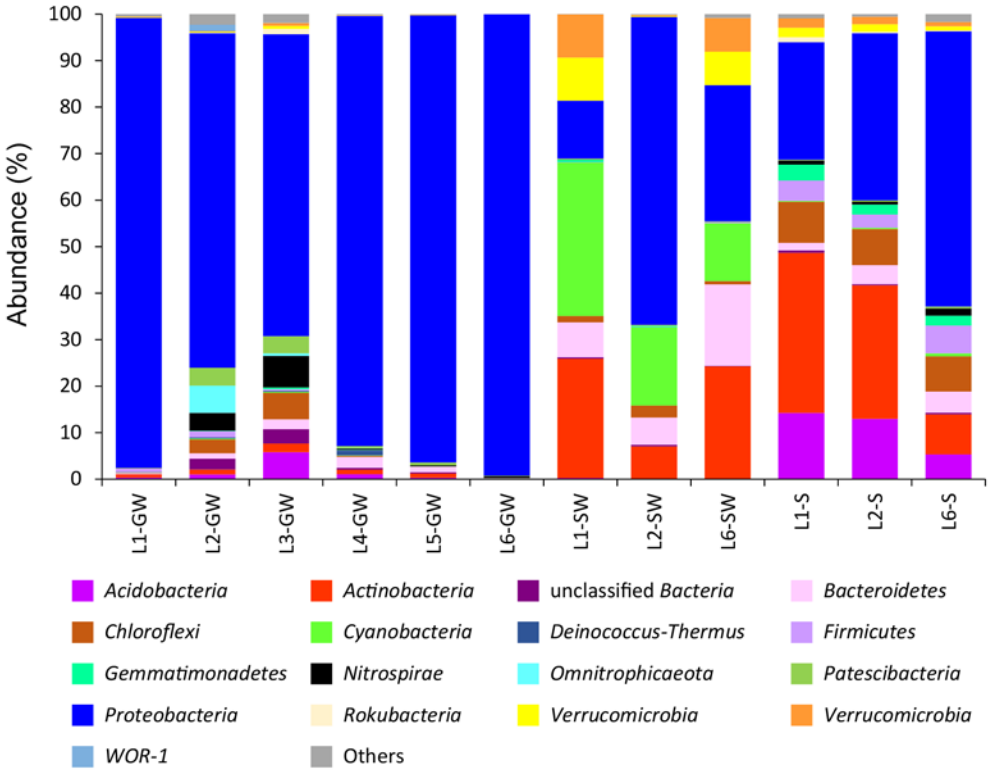

Figure 3 Abundance of microbial compositions at the phylum level in groundwater (GW), surface water (SW), and soil (S) from 6 sampling locations (L1-L6).

Overall, the resulting microbial abundance at the phylum level revealed six major phyla that were detected across all sample types: *Proteobacteria* (13-99%), *Actinobacteria* (0-35%), *Cyanobacteria* (0-36%), *Acidobacteria* (0-14%), *Bacteroidetes* (0-19%), and *Chloroflexi* (0-9%). To better understand the dynamics of microbial structure in groundwater, surface water, and soil, the 16S rRNA gene sequences associated with these six phyla were examined at the class level (Figure 4). The phylum *Proteobacteria* was noticeably dominant across the groundwater samples; it was also highly represented

in the surface water and soil samples. The five detected classes belonging to the phylum *Proteobacteria* were *Betaproteobacteria* (5-98%), *Alphaproteobacteria* (1-46%), *Gammaproteobacteria* (0-14%), *Deltaproteobacteria* (0-8%), and *Zetaproteobacteria* (0-3%) (Figure 4a). The phylum *Actinobacteria* was highly represented in the surface water and soil samples but not in groundwater. The detected phylum *Actinobacteria* was dominated by three classes: *Actinobacteria*, *Thermoleophilia*, and *Acidimicrobiia* (Figure 4b). The phylum *Cyanobacteria* was affiliated with the class *Oxyphotobacteria* and was dominant only in surface waters (Figure 4c). The phylum *Acidobacteria* was highly detected in the soil samples and was mainly composed of four classes: *Subgroup_6*, *Blastocatellia*, *Aminicenantia*, and *Holophagae* (Figure 4d). Although the phylum *Bacteroidetes* was highly represented in the surface water samples, it was slightly detected in the soil and groundwater samples. The three major classes belonging to the phylum *Bacteroidetes* were *Bacteroidia*, *Ignavibacteria*, and *Rhodothermia* (Figure 4e). The phylum *Chloroflexi* was highly represented in the soil samples and was composed of the classes *Anaerolineae*, *Chloroflexia*, *KD4-96*, *Ktedonobacteria*, *TK10*, and *Dehalococcoidia* (Figure 4f).

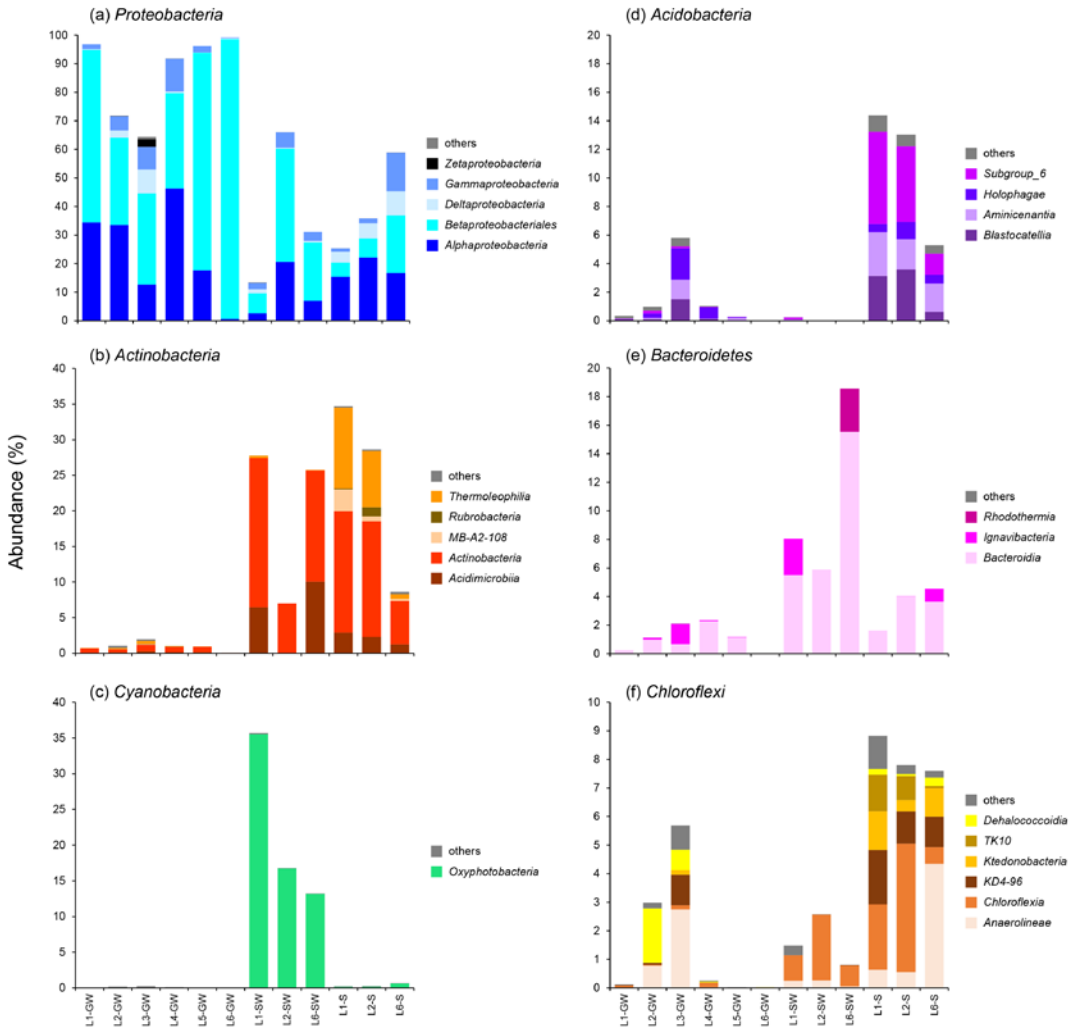

Figure 4. Abundance of microbial compositions at the class level associated with six major phyla (a-f) detected in groundwater (GW), surface water (SW), and soil (S) from 6 sampling locations (L1-L6).

To elucidate the dominant microbial assemblages detected in each sample, a heatmap was calculated based on the presence of OTUs that were more than 2% abundant (Figure 5). The resulting dendrogram revealed that the microbial structures in each sample type or habitat were closely related. The microbial assemblages that were highly associated with surface water were Cyanobium_PCC-6307, hgcI_clade, and MWH-UniP1_aquatic_group, belonging to the phyla *Cyanobacteria*, *Actinobacteria*, and *Proteobacteria*, respectively. Microbial community structures in L1-S and L2-S were relatively similar but were different from those in L6-S. *Microvirga* sp., uncultured *Xanthobacteraceae*, *Pseudarthrobacter* sp., and *Streptomyces* sp. were highly detected in L1-S and L2-S, while uncultured *Hydrogenophilaceae*, *Thiobacillus* sp., *Dyella* sp., and *Acidothermus* sp. were highly represented in L6-S (Figure 5). *Novosphingobium* sp. was

commonly found in the groundwater samples, except for L6-GW. Instead, the dominant microbial assemblage detected in L6-GW was related to *Ferriphaselus* sp. Those associated with *Limnohabitans* sp., *Rhodocyclaceae*, and *Hydrogenophaga* sp. were found at high proportions in L1-GW, L3-GW, and L5-GW, respectively (Figure 5).

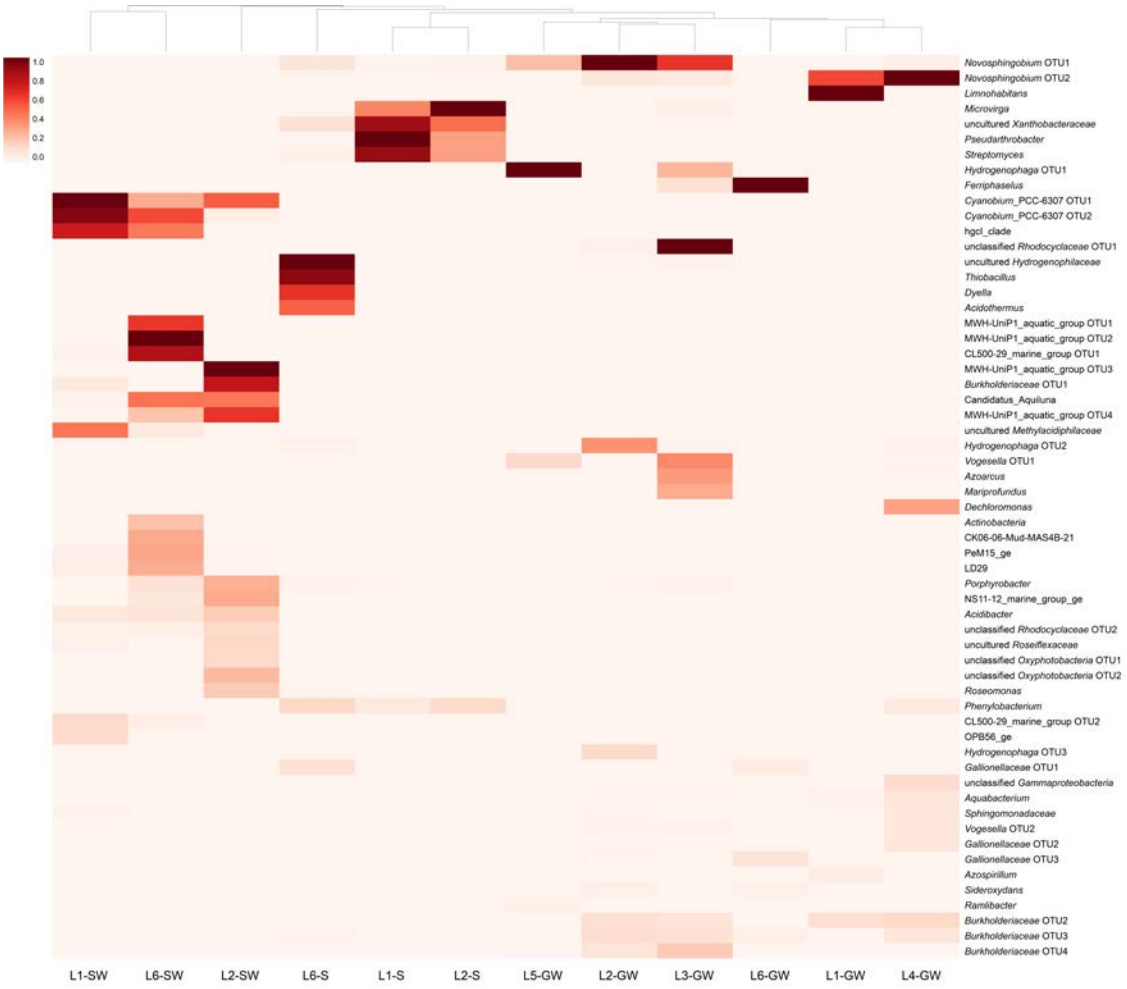

Figure 5. Clustering of the samples using a heatmap based on the presence of OTUs that were more than 2% abundant. The color intensity shows the proportion of a genus found in each sample.

3.4 Factors that impacted the microbial distribution patterns of the groundwater, surface water, and soil

Ordination analysis based on a Bray-Curtis distance matrix showed that the microbial communities in each habitat (groundwater, surface water, and soil) grouped together but were distinct from one another (Figure 6). Jaccard and Theta YC ordination analyses were in accordance with the Bray-Curtis ordination analysis. The microbial communities

in the six groundwater samples (L1-GW to L6-GW) grouped together but differed from those in the surface waters and soil samples. As with the groundwater, the microbial communities in the surface water (L1-SW, L2-SW, and L6-SW) were unique and closely related within their habitat (Figure 6), although the arrangement of microbial communities in the soil samples (L1-S, L2-S, and L6-S) were relatively similar. The microbial communities in soil samples L1-S and L2-S were closely related, but they were apart from L6-S.

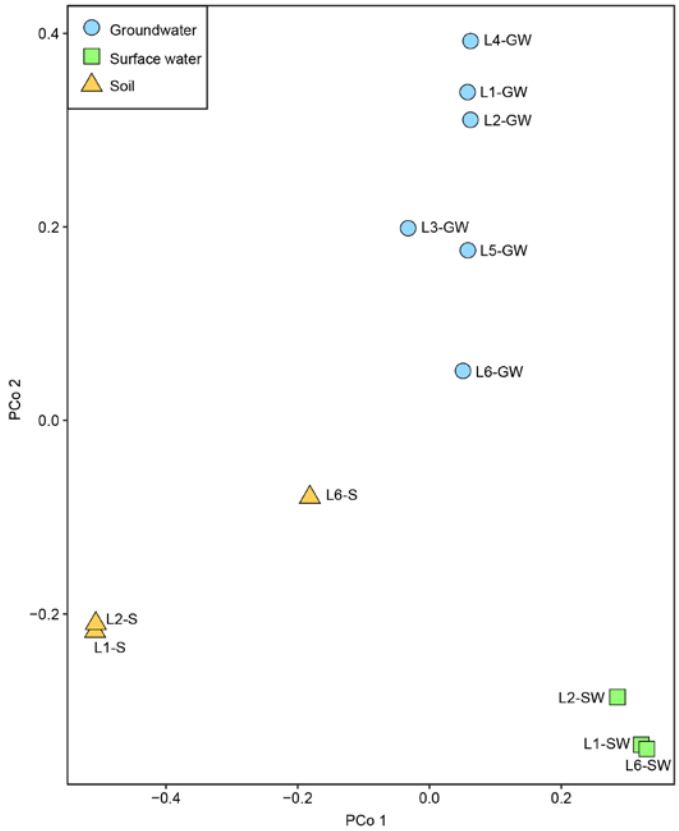

Figure 6. Principal coordinate analysis (PCoA) of microbial compositions in groundwater (GW), surface water (SW), and soil (S) across 6 sampling locations (L1-L6) based on the Bray-Curtis dissimilarity matrix.

A canonical correlation analysis (CCA) was conducted to investigate the relationships between geochemical factors (i.e., total As, TKN, TP, TOC, pH, temperature, DO, and ORP) and microbial distribution patterns. The results demonstrated that total As influenced the microbial community structures in groundwater samples L1-GW to L5-GW, with the exception of L6-GW (Figure 7). However, TOC, ORP, and DO showed a positive correlation with the microbial structures in the surface water samples (L1-SW, L2-SW, and L6-SW).

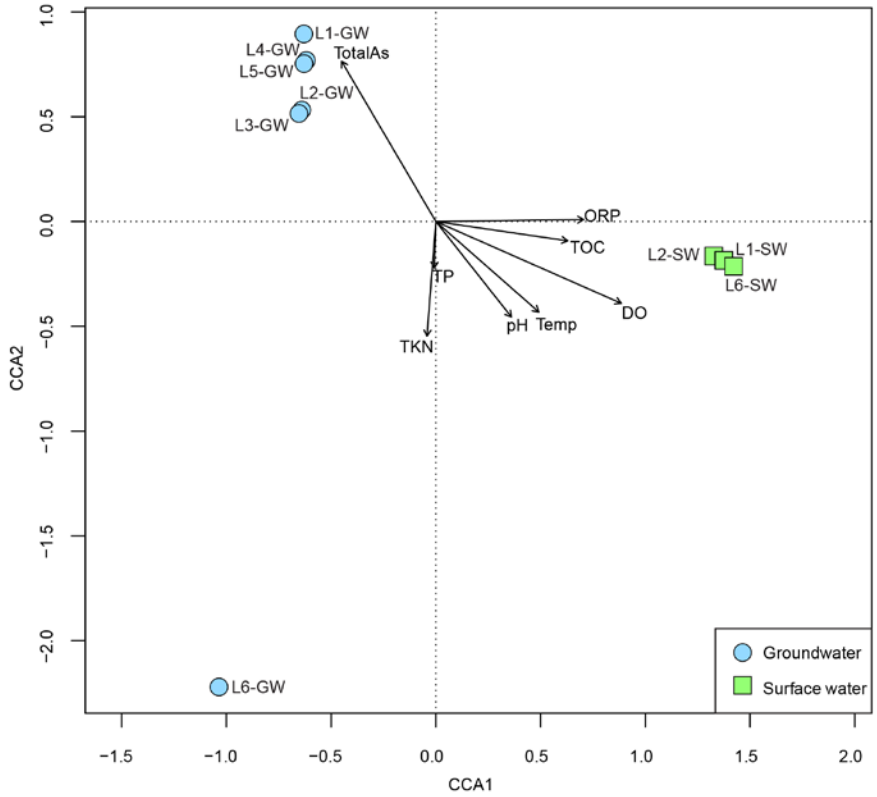

Figure 7. Canonical correspondence analysis (CCA) of microbial structures and geochemical parameters. Arrows indicate the correlation and magnitude of geochemical parameters associated with bacterial community structure.

3.5 Occurrence of arsenite-oxidizing bacteria in groundwater, surface water, and soil samples across low arsenic concentrations

By targeting the functional gene arsenite oxidase (*aioA*), the presence and abundance of arsenite-oxidizing bacteria were examined using PCR cloning-sequencing and ddPCR, respectively. All 12 samples showed positive amplifications of the *aioA* gene. After 257 clones from 12 libraries were sequenced, representative sequences of each library based on 97% OTU clustering were included in the phylogenetic analysis. The *aioA*-based phylogenetic tree was constructed using 86 representative OTUs from 12 libraries, including cultured and uncultured bacteria with *aioA* genes that were retrieved from the GenBank database as reference sequences (Figure 8). The results demonstrated that all analyzed *aioA* sequences found in this study showed 88-100% identity to previously reported *aioA* sequences recovered from As-contaminated environments, including groundwater (Hassan et al., 2015), surface water (Quéméneur et al., 2010), soils (Quéméneur et al., 2008; Hu et al., 2015; Sanyal et al., 2016), and sediments (Heinrich-Salmeron et al., 2011; Yamamura et al., 2014).

Phylogenetic analysis showed that all arsenite-oxidizing bacteria that were detected in the groundwater, surface water, and soil samples across low arsenic concentrations were closely related to *Proteobacteria* (Figure 8). All *aioA* sequences retrieved from L4-GW and L6-GW were phylogenetically related to only the Betaproteobacterial-like cluster; however, those recovered from L1-GW, L2-GW, L3-GW, and L5-GW fell into both the Betaproteobacterial-like and Alphaproteobacterial-like clusters. The majority of *aioA* sequences recovered from the surface water were affiliated with the Alphaproteobacterial-like cluster, although those closely related to the Betaproteobacterial-like cluster were only found in L6-SW (Figure 8). Among these three sample types, soil harbored the most diverse arsenite-oxidizing bacteria. Most of the *aioA* sequences retrieved from the soil were associated with the Alphaproteobacterial-like cluster, followed by the Betaproteobacterial-like and Gammaproteobacterial-like clusters, respectively. The Gammaproteobacterial-like cluster was only detected in L2-S and L6-S (Figure 8).

The proportion of *aioA* to 16S rRNA genes was in the range of 0.07-4.49% (Table 3). This indicated that arsenite-oxidizing bacteria were a minor assemblage found in all samples. However, the proportion of *aioA* to 16S rRNA genes was relatively high in the groundwater from the mining area (L2-GW, L3-GW, and L4-GW). The resulting ddPCR was in accordance with the heatmap analysis, showing that arsenite-oxidizing bacteria, such as *Hydrogenophaga* sp., were detected at low abundance (Figure 5).

Table 3. The proportion of *aioA* to 16S rRNA genes estimated by ddPCR

Sample ID	*aioA* gene copies/16S rRNA gene copies (%)
L1-GW	0.18
L2-GW	4.49
L3-GW	1.71
L4-GW	2.34
L5-GW	0.09
L6-GW	0.61
L1-SW	0.12
L2-SW	0.07
L6-SW	0.27
L1-S	0.54
L2-S	0.22
L6-S	0.41

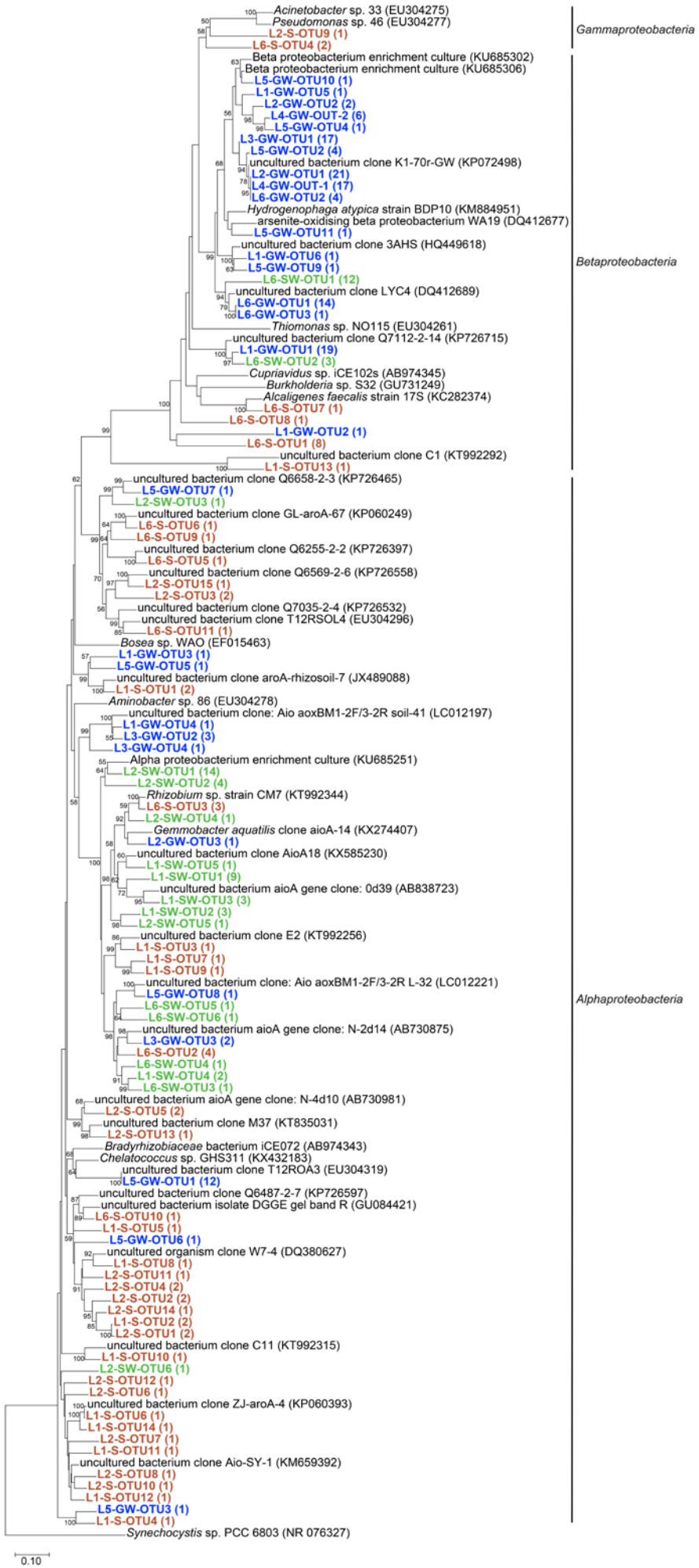
Acinetobacter sp. 33 (EU304275)
Pseudomonas sp. 46 (EU304277)
L2-S-OTU9 (1)
L6-S-OTU4 (2)
Gammaproteobacteria
Beta proteobacterium enrichment culture (KU685302)
Beta proteobacterium enrichment culture (KU685306)
L5-GW-OTU10 (1)
L1-GW-OTU5 (1)
L2-GW-OTU2 (2)
L4-GW-OUT-2 (6)
L5-GW-OTU4 (1)
L3-GW-OTU1 (17)
L5-GW-OTU2 (4)
uncultured bacterium clone K1-70r-GW (KP072498)
L2-GW-OTU1 (21)
L4-GW-OUT-1 (17)
L6-GW-OTU2 (4)
Hydrogenophaga atypica strain BDP10 (KM884951)
arsenite-oxidising beta proteobacterium WA19 (DQ412677)
L5-GW-OTU11 (1)
uncultured bacterium clone 3AHS (HQ449618)
L1-GW-OTU6 (1)
L5-GW-OTU9 (1)
L6-SW-OTU1 (12)
uncultured bacterium clone LYC4 (DQ412689)
L6-GW-OTU1 (14)
L6-GW-OTU3 (1)
Thiomonas sp. NO115 (EU304261)
uncultured bacterium clone Q7112-2-14 (KP726715)
L1-GW-OTU1 (19)
L6-SW-OTU2 (3)
Cupriavidus sp. iCE102s (AB974345)
Burkholderia sp. S32 (GU731249)
Alcaligenes faecalis strain 17S (KC282374)
L6-S-OTU7 (1)
L6-S-OTU8 (1)
L1-GW-OTU2 (1)
L6-S-OTU1 (8)
uncultured bacterium clone C1 (KT992292)
L1-S-OTU13 (1)
Betaproteobacteria
uncultured bacterium clone Q6658-2-3 (KP726465)
L5-GW-OTU7 (1)
L2-SW-OTU3 (1)
uncultured bacterium clone GL-aroA-67 (KP060249)
L6-S-OTU6 (1)
L6-S-OTU9 (1)
uncultured bacterium clone Q6255-2-2 (KP726397)
L6-S-OTU5 (1)
uncultured bacterium clone Q6569-2-6 (KP726558)
L2-S-OTU15 (1)
L2-S-OTU3 (2)
uncultured bacterium clone Q7035-2-4 (KP726532)
uncultured bacterium clone T12RSOL4 (EU304296)
L6-S-OTU11 (1)
Bosea sp. WAO (EF015463)
L1-GW-OTU3 (1)
L5-GW-OTU5 (1)
uncultured bacterium clone aroA-rhizosoil-7 (JX489088)
L1-S-OTU1 (2)
Aminobacter sp. 86 (EU304278)
uncultured bacterium clone: Aio aoxBM1-2F/3-2R soil-41 (LC012197)
L1-GW-OTU4 (1)
L3-GW-OTU2 (3)
L3-GW-OTU4 (1)
Alpha proteobacterium enrichment culture (KU685251)
L2-SW-OTU1 (14)
L2-SW-OTU2 (4)
Rhizobium sp. strain CM7 (KT992344)
L6-S-OTU3 (3)
L2-SW-OTU4 (1)
Gemmobacter aquatilis clone aioA-14 (KX274407)
L2-GW-OTU3 (1)
uncultured bacterium clone AioA18 (KX585230)
L1-SW-OTU5 (1)
L1-SW-OTU1 (9)
uncultured bacterium aioA gene clone: 0d39 (AB838723)
L1-SW-OTU3 (3)
L1-SW-OTU2 (3)
L2-SW-OTU5 (1)
uncultured bacterium clone E2 (KT992256)
L1-S-OTU3 (1)
L1-S-OTU7 (1)
L1-S-OTU9 (1)
uncultured bacterium clone: Aio aoxBM1-2F/3-2R L-32 (LC012221)
L5-GW-OTU8 (1)
L6-SW-OTU5 (1)
L6-SW-OTU6 (1)
uncultured bacterium aioA gene clone: N-2d14 (AB730875)
L3-GW-OTU3 (2)
L6-S-OTU2 (4)
L6-SW-OTU4 (1)
L1-SW-OTU4 (2)
L6-SW-OTU3 (1)
uncultured bacterium aioA gene clone: N-4d10 (AB730981)
L2-S-OTU5 (2)
uncultured bacterium clone M37 (KT835031)
L2-S-OTU13 (1)
Bradyrhizobiaceae bacterium iCE072 (AB974343)
Chelatococcus sp. GHS311 (KX432183)
uncultured bacterium clone T12ROA3 (EU304319)
L5-GW-OTU1 (12)
uncultured bacterium clone Q6487-2-7 (KP726597)
uncultured bacterium isolate DGGE gel band R (GU084421)
L6-S-OTU10 (1)
L1-S-OTU5 (1)
L5-GW-OTU6 (1)
uncultured organism clone W7-4 (DQ380627)
L1-S-OTU8 (1)
L2-S-OTU11 (1)
L2-S-OTU4 (2)
L2-S-OTU2 (2)
L2-S-OTU14 (1)
L1-S-OTU2 (2)
L2-S-OTU1 (2)
uncultured bacterium clone C11 (KT992315)
L1-S-OTU10 (1)
L2-SW-OTU6 (1)
L2-S-OTU12 (1)
L2-S-OTU6 (1)
uncultured bacterium clone ZJ-aroA-4 (KP060393)
L1-S-OTU6 (1)
L1-S-OTU14 (1)
L2-S-OTU7 (1)
L1-S-OTU11 (1)
uncultured bacterium clone Aio-SY-1 (KM659392)
L2-S-OTU8 (1)
L2-S-OTU10 (1)
L1-S-OTU12 (1)
L5-GW-OTU3 (1)
L1-S-OTU4 (1)
Alphaproteobacteria
Synechocystis sp. PCC 6803 (NR 076327)
0.10

Figure 8. Neighbor-joining phylogenetic tree of *aioA* gene sequences retrieved from groundwater (GW), surface water (SW), and soil (S) across 6 sampling locations (L1-L6). The tree was constructed with 1,000 bootstrap replicates, and only bootstrap values >50% are shown. Bold text and color indicate *aioA* sequences recovered from each habitat: groundwater (blue), surface water (green) and soil (brown).

4. Discussion

Groundwater, surface water, and soil samples were collected in the wet season from the gold mining area, including the upstream and downstream of the gold mine; however, As concentrations were low across all sampling locations. Low arsenic concentrations may result from the seasonal effect. The wet season could lead to the dilution of As in natural water. The seasonal effects on As concentration in the groundwater samples were observed; however, the relationship between the season and As concentration was unclear (Biswas et al., 2014; Savarimuthu et al., 2006; Thundiyil et al., 2007). The speciation and sorption models predicted that As^{5+} was dominant in all analyzed samples. Previous surveys reported that As^{5+} was a dominant species found in water affected by mines, including surface water and groundwater collected from mining areas in Thailand (Williams et al., 2001).

Metagenomic analysis revealed that *Proteobacteria* (13-99%) were commonly detected at high abundance in the groundwater, surface water, and soil samples analyzed in this study. Previous studies also reported that *Proteobacteria* were dominant in the groundwater of Assam, India (62.6%); groundwater of Rayong, Thailand (37-93%); groundwater of the Nakdong River Bank, South Korea (64-98%); soils from Bangladesh, China, and the United Kingdom (29-38%); heavy metal-contaminated soils from a Pb-Zn mine in Iran (38-49%); and water and sediments from the Apies River in South Africa (69.8%) (Das et al., 2017; Sonthiphand et al., 2019; Lee et al., 2018; Gu et al., 2017; Hemmat-Jou et al., 2018; Abia et al., 2018).

Among these three sample types, the microbial diversity in the soil was the highest, while the diversities in the groundwater and surface water were comparable (Figure 2). Alpha diversity calculations of samples collected from Alaska, USA, showed that bacterial diversities in soil waters and headwater streams were the highest, whereas those in lake water were the lowest (Crump et al., 2012). Metagenomic analysis of samples from a cold lake in China revealed that microbial diversity in sediments was much higher than that in water (Fang et al., 2015). The study of the microbial diversity in sediments and

water from the Apies River in South Africa demonstrated the higher microbial diversity in sediments (Abia et al., 2018). The analysis of microbial communities of samples from groundwater aquifers in China also showed that the microbial diversity in high As-contaminated sediments was higher than that in high As-contaminated groundwaters (Li et al., 2015). The higher microbial diversity in the soils/sediments, compared to the diversity in the surface water/groundwater, was possibly due to the higher biomass and higher sample complexity.

The soil microbiomes harbored a variety of taxa, such as *Proteobacteria*, *Actinobacteria*, *Acidobacteria*, *Chloroflexi*, and *Bacteroidetes* (Figure 4). The relative abundances of these detected taxa were comparable to those of soils from various environments such as forests, vineyards, and grasslands (Castañeda & Barbosa, 2017; Janssen, 2006).

The microbiomes from the surface water samples mainly contained *Proteobacteria*, *Cyanobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Chloroflexi* (Figure 4). These detected taxa were also found in other surface water environments (Crump, et al., 2012; Fang et al., 2015). *Oxyphotobacteria*, members of the phylum *Cyanobacteria*, were highly detected, particularly in surface water (Figure 4c). A high relative abundance of *Oxyphotobacteria* was found in water from a water purification facility (Stamps et al., 2018). Members of the class *Oxyphotobacteria* are oxygenic phototrophs (Shih et al., 2016). Together with the DO concentrations of the surface water analyzed in this study, the results of the class *Oxyphotobacteria* show that it may play an important role in photosynthesis within surface water. Overall, except for the high abundance of *Cyanobacteria*, the major microbial compositions of the surface waters were almost similar to those of the soils. It is possible that surface water microbiomes may be built upon initial inocula from soil microbiomes (Crump et al., 2012).

The groundwater microbiomes were dominated solely by *Proteobacteria* (65-95%), particularly the classes *Betaproteobacteria* and *Alphaproteobacteria* (Figure 4a). Using metagenomic analysis of the 16 rRNA gene, the classes *Betaproteobacteria*, *Alphaproteobacteria*, and *Gammaproteobacteria* were highly represented in alluvial groundwater in South Korea (Lee et al., 2018) and As-contaminated groundwater in Thailand and India (Sonthiphand et al., 2019; Das et al., 2017).

The heatmap of the most abundant OTUs showed that the most dominant taxa of each sample were unique (Figure 5). These findings imply that different geochemical parameters and environmental conditions may shape the indigenous microbial structures

in each habitat. Heatmap and PCoA analyses revealed that sample types or habitats primarily governed the microbial structures in the analyzed samples. The microbial structures of the same habitat were closely clustered but were distant from one another (Figure 6). Using metagenomics and statistical analyses, previous studies demonstrated that microbial community structures in groundwaters and sediments were significantly different (Li et al., 2015). Distinct microbial clusters were also observed when the microbial structures of waters and sediments were compared (Fang et al., 2015; Abia et al., 2018). Clustering analysis revealed that the microbial community compositions in soils clustered together, distant from lake waters (Monard et al., 2016).

Total As concentrations influenced all the groundwater microbiomes (L1-GW to L5-GW) collected from the monitoring wells. However, one groundwater microbiome (L6-GW) that was collected from the pumping well showed a unique pattern that was distinct from those retrieved from the monitoring stations (Figure 7). The groundwater residence time and flowrate of the pumping well were shorter and higher, respectively, compared to the monitoring well. These characteristics possibly accounted for the unique pattern of the groundwater microbiome in L6-GW. Hydrological characteristics, including residence time, flow system, and pumping, influenced the microbial community structure in the groundwater (Ben Maamar et al., 2015). The analysis of the 16S rRNA gene sequences revealed that intensive groundwater pumping was responsible for the absence of taxa in the analyzed sample (Ben Maamar et al., 2015). Flowrate and water residence time also affected the microbial community structures in river water (Read et al. 2015; Wang et al., 2016a). In addition to the effect of hydrodynamics on the groundwater microbial community, As concentration appeared to be a major factor that differentiated the groundwater microbial community in the low-As and high-As assemblages (Wang et al., 2016b). The As concentration was one of the geochemical parameters that influenced the microbial community in the groundwater of Inner Mongolia, China (Li et al., 2015).

While As, residence time, and groundwater flowrate potentially influenced the microbial structure in the groundwater, TOC, ORP, and DO showed positive correlations with the microbial structure in the surface waters (Figure 7). TOC was found to be one of the geochemical factors affecting the microbial communities in groundwaters (Li et al., 2015; Wang et al., 2016b). By denaturing gradient gel electrophoresis (DGGE) and 16S rRNA gene sequencing, the microbial community in surface water of the Jiulong River, China, was correlated with water temperature, conductivity, PO_4-P, and TN/TP instead of with TOC (Liu et al., 2013). By next generation sequencing analysis of the 16S rRNA

gene, it was found that ORP and DO impacted the microbial community in the subsurface groundwater of the Nakdong River Bank in South Korea (Lee et al., 2018). In addition, ORP was one of the geochemical parameters influencing the microbial community in the groundwater of the Hetao Plain in Inner Mongolia (Wang et al., 2016b).

Although the total As concentrations were relatively low across all sampling stations, arsenite-oxidizing bacteria were detected. All retrieved arsenite-oxidizing bacteria were closely related to *Proteobacteria* (Figure 8). Arsenite-oxidizing bacteria belonging to *Proteobacteria* have been exclusively detected in mesophilic environments, including groundwater, surface water, and soil (i.e., Hassan et al., 2015; Inskeep et al., 2007; Quéméneur et al., 2008; Quéméneur et al., 2010). A mesophilic environment may enhance the presence of arsenite-oxidizing bacteria belonging to *Proteobacteria,* and it may be considered a common niche for this group of bacteria. As with the alpha diversity analysis, the soil samples harbored the most diverse arsenite-oxidizing bacteria that were affiliated with *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria*. Although arsenite-oxidizing bacteria associated with *Alphaproteobacteria* were commonly found in all analyzed samples, they were dominant in the soil and surface water samples. The *aioA* sequences related to *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria* were previously discovered in As-contaminated soil and surface water (Quéméneur et al., 2008; Quéméneur et al., 2010). The most dominant arsenite-oxidizing bacteria found in As-contaminated soil belonged to the *Alphaproteobacteria* (Sun et al., 2017). However, the arsenite-oxidizing bacteria found in the groundwaters analyzed in this study were dominated by *Alphaproteobacteria* and *Betaproteobacteria.* The cooccurrence of arsenite-oxidizing bacteria related to *Alphaproteobacteria* and *Betaproteobacteria* was observed in other groundwaters in Bangladesh (Hassan et al., 2015) and Taiwan (Das et al., 2013).

The quantitative data showed a low abundance of arsenite-oxidizing bacteria across all samples. The numbers of *aioA* genes ranged from 0.80-49.56 copies per ng of genomic DNA. Although the information on the abundance of *aioA* gene copies detected in natural environments is limited, the resulting qPCR survey showed that the abundances of the *aioA* gene in surface water and groundwater upstream and downstream of disused mines in France were in the range of 8.1×10^3 to 1.6×10^5 copies per ng of genomic DNA (Quéméneur et al., 2010). The abundance of the *aioA* gene in samples from geothermal areas in China ranged from 1.63×10^1 to 7.08×10^3 copies per ng of genomic DNA (Jiang et al., 2014). To account for biases caused by

different biomass, background bacterial abundance, and DNA extraction, the abundance of *aioA* gene copies was normalized to that of the bacterial 16S rRNA gene copies. The ratios of *aioA* to total bacterial 16S rRNA gene copies across samples analyzed in this study ranged from 0.07-4.49. Previous studies reported the ratios of *aioA* to total bacterial 16S rRNA gene copies in surface water and groundwater ranging from 0.01 to 0.14 (Quéméneur et al., 2010). Another study showed that the ratios of *aioA* to total bacterial 16S rRNA gene copies across samples from geothermal areas ranged from less than 0.10-15.10 (Jiang et al., 2014). Although the numbers of *aioA* genes detected across the samples analyzed in this study were relatively low, the ratios of *aioA* to total bacterial 16S rRNA gene copies were comparable to those of previously reported studies. The ratios of *aioA* to total 16S rRNA gene copies were relatively high in groundwaters from the mining area, potentially indicating that the gold mine may enhance the presence of arsenite-oxidizing bacteria in the analyzed groundwater.

References

Abia, A. L. K., Alisoltani, A., Keshri, J., & Ubomba-Jaswa, E. (2018). Metagenomic analysis of the bacterial communities and their functional profiles in water and sediments of the Apies River, South Africa, as a function of land use. *Science of The Total Environment*, 616, 326-334.

Bhattacharya, P., Welch, A. H., Stollenwerk, K. G., McLaughlin, M. J., Bundschuh, J., & Panaullah, G. (2007). Arsenic in environment: biology and chemistry. *Science of The Total Environment*, 379, 109-120. https://doi.org/10.1016/j.scitotenv.2007.02.037

Ben Maamar, S., Aquilina, L., Quaiser, A., Pauwels, H., Michon-Coudouel, S., Vergnaud-Ayraud, V., ... & Dufresne, A. (2015). Groundwater isolation governs chemistry and microbial community structure along hydrologic flowpaths. *Frontiers in Microbiology*, 6, 1457.

Biswas, A., Biswas, S., & Santra, S. C. (2014). Arsenic in irrigated water, soil, and rice: perspective of the cropping seasons. *Paddy and Water Environment*, 12, 407-412.

Boonsrang, A., Chotpantarat, S., & Sutthirat, C. (2017). Factors controlling the release of metals and a metalloid from the tailings of a gold mine in Thailand. *Geochemistry: Exploration, Environment, Analysis*, 18, 109-119.

Castañeda, L. E., & Barbosa, O. (2017). Metagenomic analysis exploring taxonomic and functional diversity of soil microbial communities in Chilean vineyards and surrounding native forests. *PeerJ*, 5, e3098. https://doi.org/10.7717/peerj.3098

Crump, B. C., Amaral-Zettler, L. A., & Kling, G. W. (2012). Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *The ISME Journal*, 6, 1629. https://doi.org/10.1038/ismej.2012.9

Das, S., Kar, S., Jean, J. S., Rathod, J., Chakraborty, S., Liu, H. S., & Bundschuh, J. (2013). Depth-resolved abundance and diversity of arsenite-oxidizing bacteria in the groundwater of Beimen, a blackfoot disease endemic area of southwestern Taiwan. *Water Research*, 47, 6983-6991.

Das, S., Bora, S. S., Yadav, R. N. S., & Barooah, M. (2017). A metagenomic approach to decipher the indigenous microbial communities of arsenic contaminated groundwater of Assam. *Genomics Data*, 12, 89-96.

Epelde, L., Lanzen, A., Blanco, F., Urich, T., & Garbisu, C. (2015). Adaptation of soil microbial community structure and function to chronic metal contamination at an abandoned Pb-Zn mine. *FEMS Microbiology Ecology*, 91, 1-11.

Fang, L., Chen, L., Liu, Y., Tao, W., Zhang, Z., Liu, H., & Tang, Y. (2015). Planktonic and sedimentary bacterial diversity of Lake Sayram in summer. *MicrobiologyOpen*, 4, 814-825.

Fashola, M., Ngole-Jeme, V., & Babalola, O. (2016). Heavy metal pollution from gold mines: environmental effects and bacterial strategies for resistance. *International Journal of Environmental Research and Public Health*, 13, 1047.

Gu, Y., Van Nostrand, J. D., Wu, L., He, Z., Qin, Y., Zhao, F. J., & Zhou, J. (2017). Bacterial community and arsenic functional genes diversity in arsenic contaminated soils from different geographic locations. *PLOS ONE*, 12, e0176696. https://doi.org/10.1371/journal.pone.0176696

Guo, H., Nasir, M., Lv, J., Dai, Y., & Gao, J. (2017). Understanding the variation of microbial community in heavy metals contaminated soil using high throughput sequencing. *Ecotoxicology and Environmental Safety*, 144, 300-306.

Hassan, Z., Sultana, M., van Breukelen, B. M., Khan, S. I., & Röling, W. F. (2015). Diverse arsenic-and iron-cycling microbial communities in arsenic-contaminated aquifers used for drinking water in Bangladesh. *FEMS Microbiology Ecology*, 91, fiv026. https://doi.org/10.1093/femsec/fiv026

Heinrich-Salmeron, A., Cordi, A., Brochier-Armanet, C., Halter, D., Pagnout, C., Abbaszadeh-Fard, E., ... & Arsène-Ploetze, F. (2011). Unsuspected diversity of arsenite-oxidizing bacteria as revealed by widespread distribution of the *aoxB* gene in prokaryotes. *Applied and Environmental Microbiology*, 77, 4685-4692.

Hemmat-Jou, M. H., Safari-Sinegani, A. A., Mirzaie-Asl, A., & Tahmourespour, A. (2018). Analysis of microbial communities in heavy metals-contaminated soils using the metagenomic approach. *Ecotoxicology*, 27, 1281-1291.

Hu, M., Li, F., Liu, C., & Wu, W. (2015). The diversity and abundance of As (III) oxidizers on root iron plaque is critical for arsenic bioavailability to rice. Scientific Reports, 5, 13611. https://doi.org/10.1038/srep13611

Inskeep, W. P., McDermott, T. R., Fendorf, S., & Frankenberger, W. (2001). Arsenic (V)/(lll) cycling in soils and natural waters: Chemical and microbiological processes. *Environmental Chemistry of Arsenic*, 183-215.

Inskeep, W. P., Macur, R. E., Hamamura, N., Warelow, T. P., Ward, S. A., & Santini, J. M. (2007). Detection, diversity and expression of aerobic bacterial arsenite oxidase genes. *Environmental Microbiology*, 9, 934-943.

Jackson, C. R., Jackson, E. F., Dugas, S. L., Gamble, K., & Williams, S. E. (2003). Microbial transformations of arsenite and arsenate in natural environments. *Recent Research Developments in Microbiology*, 7, 103-118.

Jackson, T. A., Vlaar, S., Nguyen, N., Leppard, G. G., & Finan, T. M. (2015). Effects of bioavailable heavy metal species, arsenic, and acid drainage from mine tailings on a microbial community sampled along a pollution gradient in a freshwater ecosystem. *Geomicrobiology Journal*, 32, 724-750.

Janssen, P. H. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and Environmental Microbiology*, 72, 1719-1728.

Jiang, Z., Li, P., Jiang, D., Wu, G., Dong, H., Wang, Y., ... & Guo, Q. (2014). Diversity and abundance of the arsenite oxidase gene *aioA* in geothermal areas of Tengchong, Yunnan, China. *Extremophiles*, 18, 161-170.

Kromkhun, K. (2005). Geological setting, mineralogy, alteration, and nature of ore fluids of the H zone, the Chatree deposit, Thailand. Unpublished MSc thesis, Centre of Ore Deposit Research, University of Tasmania, Hobart, Tasmania, 232 pp.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33, 1870-1874.

Lee, J. H., Unno, T., & Ha, K. (2018). Vertical and horizontal distribution of bacterial communities in alluvial groundwater of the Nakdong River bank. *Geomicrobiology Journal*, 35, 74-80.

Li, P., Wang, Y., Dai, X., Zhang, R., Jiang, Z., Jiang, D., ... & Dong, H. (2015). Microbial community in high arsenic shallow groundwater aquifers in Hetao Basin of Inner Mongolia, China. *PLOS ONE*, 10, e0125844. https://doi.org/10.1371/journal.pone.0125844

Li, P., Jiang, Z., Wang, Y., Deng, Y., Van Nostrand, J. D., Yuan, T., ... & Zhou, J. (2017). Analysis of the functional gene structure and metabolic potential of microbial community in high arsenic groundwater. *Water Research*, 123, 268-276.

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22, 1658-1659.

Liu, L., Yang, J., Yu, X., Chen, G., & Yu, Z. (2013). Patterns in the composition of microbial communities from a subtropical river: effects of environmental, spatial and temporal factors. *PLOS ONE*, 8, e81232. https://doi.org/10.1371/journal.pone.0081232

Monard, C., Gantner, S., Bertilsson, S., Hallin, S., & Stenlid, J. (2016). Habitat generalists and specialists in microbial communities across a terrestrial-freshwater gradient. *Scientific Reports*, 6, 37719. http://dx.doi.org/10.1038/srep37719

Nordstrom, D. K., & Archer, D. G. (2003). *Arsenic thermodynamic data and environmental geochemistry*. In Arsenic in ground water (pp. 1-25): Springer.

Pal, P. (2015). *Groundwater arsenic remediation: treatment technology and scale up*. Oxford, UK: Elsevier Inc.

Parkhurst, D. L., & Appelo, C. A. J. (1999). User's guide to PHREEQC (Version 2): A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. *Water-Resources Investigations Report, 99*(4259), 312.

Parkhurst, D. L., & Appelo, C. A. J. (2013). Description of input and examples for PHREEQC version 3: a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations (2328-7055). Retrieved from US Geological Survey. https://doi.org/10.3133/tm6A43

Quéméneur, M., Heinrich-Salmeron, A., Muller, D., Lièvremont, D., Jauzein, M., Bertin, P. N., ... & Joulian, C. (2008). Diversity surveys and evolutionary relationships of *aoxB* genes in aerobic arsenite-oxidizing bacteria. *Applied and Environmental Microbiology*, 74, 4567-4573.

Quéméneur, M., Cébron, A., Billard, P., Battaglia-Brunet, F., Garrido, F., Leyval, C., & Joulian, C. (2010). Population structure and abundance of arsenite-oxidizing bacteria

along an arsenic pollution gradient in waters of the Upper Isle River Basin, France. *Applied and Environmental Microbiology*, 76, 4566-4570.

Read, D. S., Gweon, H. S., Bowes, M. J., Newbold, L. K., Field, D., Bailey, M. J., & Griffiths, R. I. (2015). Catchment-scale biogeography of riverine bacterioplankton. *The ISME Journal*, 9, 516-526.

Sanyal, S. K., Mou, T. J., Chakrabarty, R. P., Hoque, S., Hossain, M. A., & Sultana, M. (2016). Diversity of arsenite oxidase gene and arsenotrophic bacteria in arsenic affected Bangladesh soils. *AMB Express*, 6, 21. https://doi.org/10.1186/s13568-016-0193-0

Santini, J. M., Sly, L. I., Schnagl, R. D., & Macy, J. M. (2000). A new chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine: phylogenetic, physiological, and preliminary biochemical studies. *Applied and Environmental Microbiology*, 66, 92-97.

Savarimuthu, X., Hira-Smith, M. M., Yuan, Y., von Ehrenstein, O. S., Das, S., Ghosh, N., ... & Smith, A. H. (2006). Seasonal variation of arsenic concentrations in tubewells in West Bengal, India. *Journal of Health, Population, and Nutrition*, 24, 277–281.

Shih, P. M., Hemp, J., Ward, L. M., Matzke, N. J., & Fischer, W. W. (2016). Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobiology*, 15, 19-29.

Smedley, P. L. (2008). Sources and distribution of arsenic in groundwater and aquifers. In T. Appelo (Ed.), *Arsenic in Groundwater: a World Problem* (pp. 4-32). Utrecht, the Netherlands: International Association of Hydrogeologists.

Sonthiphand, P., Ruangroengkulrith, S., Mhuantong, W., Charoensawan, V., Chotpantarat, S., & Boonkaewwan, S. (2019). Metagenomic insights into microbial diversity in a groundwater basin impacted by a variety of anthropogenic activities. *Environmental Science and Pollution Research*.

Stamps, B. W., Leddy, M. B., Plumlee, M. H., Hasan, N. A., Colwell, R. R., & Spear, J. R. (2018). Characterization of the Microbiome at the World's Largest Potable Water Reuse Facility. *Frontiers in Microbiology*, 9. https://doi.org/10.3389/fmicb.2018.02435

Sun, W., Xiao, E., Xiao, T., Krumins, V., Wang, Q., Häggblom, M., ... & Xia, B. (2017). Response of soil microbial communities to elevated antimony and arsenic contamination indicates the relationship between the innate microbiota and contaminant fractions. *Environmental Science & Technology*, 51, 9165-9175.

Thundiyil, J. G., Yuan, Y., Smith, A. H., & Steinmaus, C. (2007). Seasonal variation of arsenic concentration in wells in Nevada. *Environmental Research*, 104, 367-373.

Wang, P., Chen, B., Yuan, R., Li, C., & Li, Y. (2016a). Characteristics of aquatic bacterial community and the influencing factors in an urban river. *Science of The Total Environment*, 569, 382-389. http://dx.doi.org/10.1016/j.scitotenv.2016.06.130

Wang, Y., Li, P., Jiang, Z., Sinkkonen, A., Wang, S., Tu, J., ... & Wang, Y. (2016b). Microbial community of high arsenic groundwater in agricultural irrigation area of Hetao Plain, Inner Mongolia. *Frontiers in Microbiology*, 7, 1917. https://doi.org/10.3389/fmicb.2016.01917

Williams, M. (2001). Arsenic in mine waters: an international study. *Environmental Geology*, 40, 267-278.

Yamamura, S., Watanabe, K., Suda, W., Tsuboi, S., & Watanabe, M. (2014). Effect of antibiotics on redox transformations of arsenic and diversity of arsenite-oxidizing bacteria in sediment microbial communities. *Environmental Science & Technology*, 48, 350-357.

5. Output (Acknowledge the Thailand Research Fund)

5.1 International Journal Publication

This study has produced a manuscript entitled “Arsenic speciation, abundance of arsenite-oxidizing bacteria and microbial community structures in groundwater, surface water and soil from a gold mine” and it has been submitted to Scientific Reports. The Submission ID is fab33221-9004-43fe-a6f6-d882420d77b7. Please see Appendix for more details

The Thailand Research Fund is acknowledged in the submitted manuscript.

6. Appendix

A manuscript entitled “Arsenic speciation, abundance of arsenite-oxidizing bacteria and microbial community structures in groundwater, surface water and soil from a gold mine”, which has been submitted for publication in Scientific Reports, is attached.

Receipt of Manuscript "Arsenic speciation, abundance..."

Scientific Reports <srep@nature.com>

Sat 3/14/2020 1:58 PM

To:Prinpida Sonthiphand <prinpida.son@mahidol.ac.th>;

Ref: Submission ID fab33221-9004-43fe-a6f6-d882420d77b7

Dear Dr Sonthiphand,

Thank you for submitting your manuscript to Scientific Reports.

Your manuscript is now at our initial Quality Check stage, where we look for adherence to the journal's submission guidelines, including any relevant editorial and publishing policies. If there are any points that need to be addressed prior to progressing we will send you a detailed email. Otherwise, your manuscript will proceed into peer review.

Please note you have submitted to a new peer review system which does not yet offer the ability to track your manuscript status.

Kind regards,

Peer Review Advisors
Scientific Reports

SPRINGER NATURE | SCIENTIFIC REPORTS

Review

Arsenic speciation, abundance of arsenite-oxidizing bacteria and microbial community structures in groundwater, surface water and soil from a gold mine

Success

Submission received

Thank you for submitting your manuscript to Scientific Reports.

Please note: you have submitted to our new peer-review system that does not currently have a manuscript tracking facility. However, we will contact you if we need you to take any action.

Arsenic speciation, abundance of arsenite-oxidizing bacteria and microbial community structures in groundwater, surface water and soil from a gold mine

Prinpida Sonthiphand[1*], Supeerapat Kraidech[2], Saowarod Polart[1], Srilert Chotpantarat[3,4,5], Kanthida Kusonmano[6,7], Pichahpuk Uthaipaisanwong[6], Chalida Rangsiwutisak[6], Ekawan Luepromchai[8,9]

[1]Department of Biology, Faculty of Science, Mahidol University, 272 Rama VI Road, Ratchathewi, Bangkok 10400 Thailand

[2]International Postgraduate Program in Hazardous Substance and Environmental Management, Chulalongkorn University, 9th Floor, CU Research Building, Phayathai Road, Bangkok 10330 Thailand

[3]Department of Geology, Faculty of Science, Chulalongkorn University, Thailand

[4]Research Program on Controls of Hazardous Contaminants in Raw Water Resources for Water Scarcity Resilience, Center of Excellence on Hazardous Substance Management (HSM), Chulalongkorn University, Thailand

[5]Research Unit of Site Remediation on Metals Management from Industry and Mining (Site Rem), Chulalongkorn University, Bangkok, Thailand

[6]Bioinformatics and Systems Biology Program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok, 10150, Thailand

[7]Systems Biology and Bioinformatics Research Laboratory, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok, 10150, Thailand

[8]Research Program on Remediation Technologies for Petroleum Contamination, Center of Excellence on Hazardous Substance Management (HSM), Chulalongkorn University, Bangkok, Thailand

[9]Microbial Technology for Marine Pollution Treatment Research Unit, Department of Microbiology, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

*To whom correspondence should be addressed.

E-mail: prinpida.son@mahidol.ac.th

Phone: (+66) 2201-5250

Fax: (+66) 2354-7161

Abstract

The arsenic speciation, the abundance of arsenite-oxidizing bacteria, and microbial community structures in the groundwater, surface water, and soil from a gold mining area were explored using the PHREEQC model, cloning-ddPCR of the *aioA* gene, and high-throughput sequencing of the 16S rRNA gene, respectively. The analysis of the *aioA* gene showed that arsenite-oxidizing bacteria retrieved from groundwater, surface water, and soil were associated with *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*. In groundwaters from the mining area, there were relatively high ratios of *aioA*/total 16S rRNA gene copies and the dominance of As^{5+}, which indicated the presence and activity of arsenite-oxidizing bacteria. Metagenomic analysis revealed that the majority of the soil and surface water microbiomes were *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Chloroflexi*, whereas the groundwater microbiomes were dominated exclusively by *Betaproteobacteria* and *Alphaproteobacteria*. Geochemical factors influencing the microbial structure in the groundwater were As, residence time, and groundwater flowrate, while those showing a positive correlation to the microbial structure in the surface water were TOC, ORP, and DO. This study provides insights into the groundwater, surface water, and soil microbiomes from a gold mine and expands the current understanding of the diversity and abundance of arsenite-oxidizing bacteria, playing a vital role in global As cycling.

Keywords: microbial community, gold mine, arsenic, *aioA* gene, arsenite-oxidizing bacteria

Introduction

Microbial communities in natural environments play important roles in biogeochemical cycles and in degrading pollutants; however, they are directly impacted by anthropogenic activities such as mining, agriculture, and wastewater treatment operations [1,2,3]. Among anthropogenic activities, gold mining operations have generated significant amounts of pollutants in the surrounding environments [4,5]. Tailings that contain high levels of heavy metals (i.e., As, Pb, and Cu) are considered the major source of wastes released from gold extraction. These heavy metals possibly leach out and subsequently contact surrounding groundwater, surface water, and soil. Leaching of heavy metals from gold mine tailings not only adversely affects human health but also influences the diversity, abundance, and activity of microorganisms [5]. Waste pollution from gold mine tailings affects microbial dynamics in sediment and water samples by enhancing the growth of *Holophagal Acidobacteria*, green sulfur bacteria, and *Alphaproteobacteria* and decreases the overall productivity, biomass, and microbial biodiversity [6].

Since high-throughput sequencing technologies have been developed, the understanding of microbial community structures in various environments has been widely extended. However, the microbial community structures in environments affected by a gold mine have not yet been extensively studied. Metagenomic analyses of the microbial community structures in freshwater habitats showed that *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, *Planctomycetes*, *Actinobacteria*, *Verrucomicrobia*, and *Nitrospirae* were commonly detected [1,3,7,8,9]. A variety of microbial taxa, such as *Chloroflexi*, *Actinobacteria*, *Acidobacteria*, and *Ktedonobacteria*, was found in soils [2,10]. Previous studies reported that heavy metals

90 such as Cd, Pb, Zn, As, and Fe influenced microbial community structures in impacted
91 environments [2,11,12,13].

92 Although arsenic (As) is released from gold mining activities, As release can also occur
93 naturally. Exposure to As, even at low concentrations, can cause serious health problems [14].
94 There is evidence suggesting that microorganisms play key roles in As redox transformations.
95 Arsenite oxidation is the microbial process responsible for reducing the toxicity, solubility and
96 mobility of As in contaminated environments [15]. Arsenite oxidation is mediated by arsenite-
97 oxidizing bacteria using the key enzyme arsenite oxidase (aio) [16]. Molecular surveys of the
98 *aioA* gene have revealed that arsenite-oxidizing bacteria are diverse and have been detected in
99 various environments, including soils [17], gold mines [18], groundwater [19], and geothermal
100 regions [20].

101 The study site was located in the vicinity of a gold mine in Phetchabun and Pichit
102 Provinces, Thailand. The majority of the land in this area was used for paddy cultivation. This
103 study aimed to examine a potential role of arsenite-oxidizing bacteria in mediating arsenite
104 oxidation within a gold mining area through analyzing the microbial community and abundance
105 of arsenite-oxidizing bacteria using PCR-cloning-sequencing and droplet digital PCR (ddPCR)
106 of the *aioA* gene, respectively, together with the prediction of arsenate (As^{5+}), the product of
107 arsenite oxidation, using the geochemical PHREEQC model. In addition, this study also aimed
108 to investigate the microbial community structures of groundwater, surface water, and soil
109 within and surrounding the gold mining area and to reveal environmental factors that
110 potentially influenced the analyzed microbiomes. The microbiomes of groundwater, surface
111 water, and soil within and surrounding the gold mining area were explored using an Illumina
112 MiSeq platform targeting the V3 and V4 regions of the 16S rRNA gene. Together with insight

into the microbial communities, the correlations of geochemical factors and the microbial structures of groundwater, surface water, and soil were elucidated.

Results

Sample characteristics

Five groundwater samples (L1-GW to L5-GW) were collected from monitoring wells, whereas one sample (L6-GW) was collected from a pumping well (Fig. S1). When groundwater was consumed, the pumping system continuously pumped the groundwater into a storage tank. However, the groundwater in the monitoring wells was likely undisturbed. Therefore, the characteristics of groundwater wells likely affected the hydraulic retention time and groundwater flowrate. The groundwater residence time and groundwater flowrate of the monitoring wells were longer and slower, respectively, than those of the pumping wells. These characteristics may influence the indigenous groundwater microbiome in the monitoring and pumping wells.

Geochemical parameters indicated that temperature, TKN, and TP were comparable among the six groundwater and three surface samples (Table 1). Most of the analyzed water samples were neutral to slightly acidic, except for L5-GW (moderate acidic) and L6-SW (slightly basic). The concentrations of TOC were relatively low across almost all water samples; however, the TOC concentration was considerably high in L6-SW. DO values showed that the groundwater was less oxic than the surface water. The ORP measurements also indicated that the groundwater conditions were relatively reduced compared to the surface water conditions (Table 1). Although the concentrations of As were low and comparable in all water samples, the speciation of As heavily relied on pH and ORP [21]. In this study, the

geochemical model PHREEQC was applied to predict the dominant species of As found in groundwater and surface water with different geochemical conditions.

Prediction of As speciation by using the geochemical model PHREEQC

The PHREEQC model revealed that As^{5+} was the dominant species found in all analyzed groundwater and surface water samples (Table 2). $H_2AsO_4^-$ was the dominant form detected in groundwater and surface water collected from locations L1 to L5, whereas $HAsO_4^{2-}$ was the major form of As^{5+} in samples taken from location L6. These predicted forms are commonly found in natural surface and groundwater environments [22]. The concentrations of DO across all analyzed water samples indicated slightly to highly oxic conditions, which potentially led to the predominant of As^{5+}.

Since sorption plays an important role in As speciation, a sorption model was also considered in this analysis. Due to the strong sorption capacities of As on amorphous iron oxides, the Hfo-As sorption model was selected for predicting whether the presence of iron oxides (As, Fe, Mn, and SO_4^{2-}) influenced As mobility in the analyzed samples [21]. Groundwater conditions associated with the presence of As species were illustrated by the Hfo-As sorption model (Fig. S2). As with the resulting speciation model, $H_2AsO_4^-$ and $HAsO_4^{2-}$ were dominant in samples L1-GW to L5-GW and L6-SW, respectively (Table 2). However, hfo_w$OHAsO_4^{3-}$ was found to be the major form detected in L6-GW, L1-SW, and L2-SW.

The As forms under oxic conditions are related to electrostatic repulsion from the negative charge of the oxide surface, resulting in Hfo instability and dissolution [21]. The strong sorption tendency of iron oxides around a neutral pH range in oxic conditions is one of the major causes of low As concentrations in most natural groundwater [21]. However, Fe concentrations in this study were relatively similar; thus, Hfo sorption might not be the major

159 factor causing difference in As speciation among the samples. On the other hand, dissolved As
160 associated with S-species occurs in highly reducing conditions where the As-S mineral limits
161 the dissolved As concentration in high sulfide concentrations [21].

162 **Occurrence of arsenite-oxidizing bacteria in groundwater, surface water, and soil**
163 **samples across low arsenic concentrations**

164 By targeting the functional gene arsenite oxidase (*aioA*), the presence and abundance
165 of arsenite-oxidizing bacteria were examined using PCR-cloning-sequencing and ddPCR,
166 respectively. All 12 samples showed positive amplifications of the *aioA* gene. After 257 clones
167 from 12 libraries were sequenced, representative sequences of each library based on 97% OTU
168 clustering were included in the phylogenetic analysis. Phylogenetic analysis showed that all
169 arsenite-oxidizing bacteria that were detected in the groundwater, surface water, and soil
170 samples across low arsenic concentrations were closely related to *Proteobacteria* (Fig. 1). All
171 *aioA* sequences retrieved from L4-GW and L6-GW were phylogenetically related to only the
172 Betaproteobacterial-like cluster; however, those recovered from L1-GW, L2-GW, L3-GW, and
173 L5-GW fell into both the Betaproteobacterial-like and Alphaproteobacterial-like clusters. The
174 majority of *aioA* sequences recovered from the surface water were affiliated with the
175 Alphaproteobacterial-like cluster, although those closely related to the Betaproteobacterial-like
176 cluster were only found in L6-SW (Fig. 1). Among these three sample types, soil harbored the
177 most diverse arsenite-oxidizing bacteria. Most of the *aioA* sequences retrieved from the soil
178 were associated with the Alphaproteobacterial-like cluster, followed by the
179 Betaproteobacterial-like and Gammaproteobacterial-like clusters, respectively. The
180 Gammaproteobacterial-like cluster was only detected in L2-S and L6-S (Fig. 1).

The abundances of the *aioA* and 16S rRNA genes were quantified by ddPCR. The numbers of the *aioA* and 16S rRNA genes were in the range of $8.0x10^{-1}\pm7.3x10^{-2}$ to $5.0x10^{1}\pm1.0x10^{0}$ and $1.3x10^{2}\pm1.5x10^{0}$ to $6.3x10^{3}\pm1.7x10^{2}$ copies per ng of genomic DNA, respectively (Table S1). The proportion of *aioA* to 16S rRNA genes was in the range of 0.07-4.55%. This indicated that arsenite-oxidizing bacteria were a minor assemblage found in all samples. However, the proportion of *aioA* to 16S rRNA genes was relatively high in the groundwater from the mining area (L2-GW, L3-GW, and L4-GW). The resulting ddPCR was in accordance with the heatmap analysis, showing that arsenite-oxidizing bacteria, such as *Hydrogenophaga* sp., were detected at low abundance (Fig. 3).

Microbial community structures in groundwater, surface water, and soil across low arsenic concentrations

The alpha diversity of the groundwater, surface water, and soil across low arsenic concentrations was revealed by the observed OTUs, Shannon, and Chao1 (Table S2). Rarefaction curves showed that microbial richness in soils was relatively higher than that in groundwater and surface water (Fig. S3). Microbial richness was the highest in L1-S, followed by L2-S, and it was the lowest in L6-GW. Among the water samples, L2-GW and L3-GW showed higher microbial richness, which were comparable to L6-S (Fig. S3).

Microbial compositions in groundwater, surface water, and soil from the gold mining area with low arsenic concentrations were determined by the MiSeq Illumina sequencing of the 16S rRNA gene. The resulting microbial abundance at the phylum level indicated that the phylum *Proteobacteria* was common in all samples, accounting for 13-99% of the total microbial abundance (Fig. S4). The microbial community structure in groundwater, L1-GW to L6-GW, was dominated by the phylum *Proteobacteria*, accounting for 65-95% of the total

204 microbial abundance. The four major phyla that were found in the surface water, L1-SW, L2-
205 SW, and L6-SW, were *Proteobacteria* (13-66%), *Cyanobacteria* (13-36%), *Actinobacteria* (7-
206 28%), and *Bacteroidetes* (6-19%). Microbial compositions in soil samples, L1-S, L2-S, and
207 L6-S, were mainly composed of the phya *Proteobacteria* (25-59%), *Actinobacteria* (9-35%),
208 *Acidobacteria* (5-14%), and *Chloroflexi* (8-9%). The results indicated that microbial
209 community structure was closely associated within each sample type, whereas it changed from
210 one sample type to another (Fig. S4).

211 To better understand the dynamics of microbial structure in groundwater, surface water,
212 and soil, the 16S rRNA gene sequences associated with these six phyla were examined at the
213 class level (Fig. 2). The phylum *Proteobacteria* was noticeably dominant across the
214 groundwater samples; it was also highly represented in the surface water and soil samples. The
215 five detected classes belonging to the phylum *Proteobacteria* were *Betaproteobacteria* (5-
216 98%), *Alphaproteobacteria* (1-46%), *Gammaproteobacteria* (0-14%), *Deltaproteobacteria* (0-
217 8%), and *Zetaproteobacteria* (0-3%) (Fig. 2a). The phylum *Actinobacteria* was highly
218 represented in the surface water and soil samples but not in groundwater. The detected phylum
219 *Actinobacteria* was dominated by three classes: *Actinobacteria*, *Thermoleophilia*, and
220 *Acidimicrobiia* (Fig. 2b). The phylum *Cyanobacteria* was affiliated with the class
221 *Oxyphotobacteria* and was dominant only in surface waters (Fig. 2c). The phylum
222 *Acidobacteria* was highly detected in the soil samples and was mainly composed of four
223 classes: *Subgroup_6, Blastocatellia, Aminicenantia*, and *Holophagae* (Fig. 2d). Although the
224 phylum *Bacteroidetes* was highly represented in the surface water samples, it was slightly
225 detected in the soil and groundwater samples. The three major classes belonging to the phylum
226 *Bacteroidetes* were *Bacteroidia, Ignavibacteria*, and *Rhodothermia* (Fig. 2e). The phylum

Chloroflexi was highly represented in the soil samples and was composed of the classes *Anaerolineae*, *Chloroflexia*, *KD4-96*, *Ktedonobacteria*, *TK10*, and *Dehalococcoidia* (Fig. 2f).

To elucidate the dominant microbial assemblages detected in each sample, a heatmap was calculated based on the presence of OTUs that were more than 2% abundant (Fig. 3). The resulting dendrogram revealed that the microbial structures in each sample type or habitat were closely related. The microbial assemblages that were highly associated with surface water were Cyanobium_PCC-6307, hgcI_clade, and MWH-UniP1_aquatic_group, belonging to the phyla *Cyanobacteria*, *Actinobacteria*, and *Proteobacteria*, respectively. Microbial community structures in L1-S and L2-S were relatively similar but were different from those in L6-S. *Microvirga* sp., uncultured *Xanthobacteraceae*, *Pseudarthrobacter* sp., and *Streptomyces* sp. were highly detected in L1-S and L2-S, while uncultured *Hydrogenophilaceae*, *Thiobacillus* sp., *Dyella* sp., and *Acidothermus* sp. were highly represented in L6-S (Fig. 3). *Novosphingobium* sp. was commonly found in the groundwater samples, except for L6-GW. Instead, the dominant microbial assemblage detected in L6-GW was related to *Ferriphaselus* sp. Those associated with *Limnohabitans* sp., *Rhodocyclaceae*, and *Hydrogenophaga* sp. were found at high proportions in L1-GW, L3-GW, and L5-GW, respectively (Fig. 3). Among the microbial assemblages detected in the groundwater, surface water, and soil analyzed in this study, those related to *Hydrogenophaga* sp. are involved in As transformation. PCR cloning-sequencing of the *aioA* gene also showed that the arsenite-oxidizing bacteria that were found in these analyzed samples were related to *Hydrogenophaga* sp. (Fig. 1).

Factors that impacted the microbial distribution patterns of the groundwater, surface water, and soil

Ordination analysis based on a Bray-Curtis distance matrix showed that the microbial communities in each habitat grouped together but were distinct from one another (Fig. 4). Jaccard and Theta YC ordination analyses were in accordance with the Bray-Curtis ordination analysis (Fig. S5). The microbial communities in the six groundwater samples grouped together but differed from those in the surface waters and soil samples. As with the groundwater, the microbial communities in the surface water were unique and closely related within their habitat (Fig. 4), although the arrangement of microbial communities in the soil samples were relatively similar. The microbial communities in soil samples L1-S and L2-S were closely related, but they were apart from L6-S.

A canonical correlation analysis (CCA) was conducted to investigate the relationships between geochemical factors and microbial distribution patterns. The results demonstrated that total As influenced the microbial community structures in groundwater samples L1-GW to L5-GW, with the exception of L6-GW (Fig. 5). However, TOC, ORP, and DO showed a positive correlation with the microbial structures in the surface water samples (L1-SW, L2-SW, and L6-SW).

Discussion

Groundwater, surface water, and soil samples were collected in the wet season from the gold mining area, including the upstream and downstream of the gold mine; however, As concentrations were low across all sampling locations. The speciation and sorption models predicted that As^{5+} was dominant in all analyzed samples. Previous surveys reported that As^{5+} was a dominant species found in water affected by mines, including surface water and groundwater collected from mining areas in Thailand [23].

Although the total As concentrations were relatively low across all sampling stations, arsenite-oxidizing bacteria were detected. All retrieved arsenite-oxidizing bacteria were closely related to *Proteobacteria* (Fig. 1). Arsenite-oxidizing bacteria belonging to *Proteobacteria* have been exclusively detected in mesophilic environments, including groundwater, surface water, and soil [17,24,25,26]. A mesophilic environment may enhance the presence of arsenite-oxidizing bacteria belonging to *Proteobacteria,* and it may be considered a common niche for this group of bacteria. As with the alpha diversity analysis, the soil samples harbored the most diverse arsenite-oxidizing bacteria that were affiliated with *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria*. Although arsenite-oxidizing bacteria associated with *Alphaproteobacteria* were commonly found in all analyzed samples, they were dominant in the soil and surface water samples. The *aioA* sequences related to *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria* were previously discovered in As-contaminated soil and surface water [17,26]. The most dominant arsenite-oxidizing bacteria found in As-contaminated soil belonged to the *Alphaproteobacteria* [13]. However, the arsenite-oxidizing bacteria found in the groundwaters analyzed in this study were dominated by *Alphaproteobacteria* and *Betaproteobacteria.* The cooccurrence of arsenite-oxidizing bacteria related to *Alphaproteobacteria* and *Betaproteobacteria* was observed in other groundwaters in Bangladesh [24] and Taiwan [19].

The quantitative data showed a low abundance of arsenite-oxidizing bacteria across all samples (Table S1). The numbers of *aioA* genes ranged from 8.0×10^{-1} to 5.0×10^{1} copies per ng of genomic DNA. Although the information on the abundance of *aioA* gene copies detected in natural environments is limited, the resulting qPCR survey showed that the abundances of the *aioA* gene in surface water and groundwater upstream and downstream of disused mines in

France were in the range of $8.1x10^3$ to $1.6x10^5$ copies per ng of genomic DNA [26]. The abundance of the *aioA* gene in samples from geothermal areas in China ranged from $1.63x10^1$ to $7.08x10^3$ copies per ng of genomic DNA [20]. To account for biases caused by different biomass, background bacterial abundance, and DNA extraction, the abundance of *aioA* gene copies was normalized to that of the bacterial 16S rRNA gene copies. The ratios of *aioA* to total bacterial 16S rRNA gene copies across samples analyzed in this study ranged from 0.07-4.55 (Table S1). Previous studies reported the ratios of *aioA* to total bacterial 16S rRNA gene copies in surface water and groundwater ranging from 0.01 to 0.14 [26]. Another study showed that the ratios of *aioA* to total bacterial 16S rRNA gene copies across samples from geothermal areas ranged from less than 0.10-15.10 [20]. Although the numbers of *aioA* genes detected across the samples analyzed in this study were relatively low, the ratios of *aioA* to total bacterial 16S rRNA gene copies were comparable to those of previously reported studies. The ratios of *aioA* to total 16S rRNA gene copies were relatively high in groundwaters from the mining area and the PHREEQC model predicted the dominance of As^{5+} in the analyzed samples. These results indicated that the gold mining activities may enhance the presence and activity of arsenite-oxidizing bacteria in the analyzed groundwater.

Metagenomic analysis revealed that *Proteobacteria* (13-99%) were commonly detected at high abundance in the groundwater, surface water, and soil samples analyzed in this study. Previous studies also reported that *Proteobacteria* were dominant in the groundwater of Assam, India (62.6%); groundwater of Rayong, Thailand (37-93%); groundwater of the Nakdong River Bank, South Korea (64-98%); soils from Bangladesh, China, and the United Kingdom (29-38%); heavy metal-contaminated soils from a Pb-Zn mine in Iran (38-49%); and water and sediments from the Apies River in South Africa (69.8%) [1,3,7,27,28,29].

Among these three sample types, the microbial diversity in the soil was the highest, while the diversities in the groundwater and surface water were comparable (Fig. S3). Alpha diversity calculations of samples collected from Alaska, USA, showed that bacterial diversities in soil waters and headwater streams were the highest, whereas those in lake water were the lowest [30]. Metagenomic analysis of samples from a cold lake in China revealed that microbial diversity in sediments was much higher than that in water [31]. The analysis of microbial communities of samples from groundwater aquifers in China also showed that the microbial diversity in high As-contaminated sediments was higher than that in high As-contaminated groundwaters [11]. The higher microbial diversity in the soils/sediments, compared to the diversity in the surface water/groundwater, was possibly due to the higher biomass and higher sample complexity.

The soil microbiomes harbored a variety of taxa, such as *Proteobacteria*, *Actinobacteria*, *Acidobacteria*, *Chloroflexi*, and *Bacteroidetes* (Fig. 2). The relative abundances of these detected taxa were comparable to those of soils from various environments such as forests, vineyards, and grasslands [10,32].

The microbiomes from the surface water samples mainly contained *Proteobacteria*, *Cyanobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Chloroflexi* (Fig. 2). These detected taxa were also found in other surface water environments [30,31]. *Oxyphotobacteria*, members of the phylum *Cyanobacteria*, were highly detected, particularly in surface water (Fig. 2c). A high relative abundance of *Oxyphotobacteria* was found in water from a water purification facility [33]. Overall, except for the high abundance of *Cyanobacteria*, the major microbial compositions of the surface waters were almost similar to those of the soils. It is possible that surface water microbiomes may be built upon initial inocula from soil microbiomes [30].

341 The groundwater microbiomes were dominated solely by *Proteobacteria* (65-95%),
342 particularly the classes *Betaproteobacteria* and *Alphaproteobacteria* (Fig. 2a). Using
343 metagenomic analysis of the 16S rRNA gene, the classes *Betaproteobacteria*,
344 *Alphaproteobacteria*, and *Gammaproteobacteria* were highly represented in alluvial
345 groundwater in South Korea [27] and As-contaminated groundwater in Thailand and India
346 [3,7].

347 The heatmap of the most abundant OTUs showed that the most dominant taxa of each
348 sample were unique (Fig. 3). These findings imply that different geochemical parameters and
349 environmental conditions may shape the indigenous microbial structures in each habitat.
350 Heatmap and PCoA analyses revealed that sample types or habitats primarily governed the
351 microbial structures in the analyzed samples. The microbial structures of the same habitat were
352 closely clustered but were distant from one another (Fig. 4). Using metagenomics and statistical
353 analyses, previous studies demonstrated that microbial community structures in groundwaters
354 and sediments were significantly different [11]. Distinct microbial clusters were also observed
355 when the microbial structures of waters and sediments were compared [1,31]. Clustering
356 analysis revealed that the microbial community compositions in soils clustered together, distant
357 from lake waters [34].

358 Total As concentrations influenced all the groundwater microbiomes (L1-GW to L5-
359 GW) collected from the monitoring wells. However, one groundwater microbiome (L6-GW)
360 that was collected from the pumping well showed a unique pattern that was distinct from those
361 retrieved from the monitoring stations (Fig. 5). The groundwater residence time and flowrate
362 of the pumping well were shorter and higher, respectively, compared to the monitoring well.
363 These characteristics possibly accounted for the unique pattern of the groundwater microbiome

in L6-GW. Hydrological characteristics, including residence time, flow system, and pumping, influenced the microbial community structure in the groundwater [35]. The analysis of the 16S rRNA gene sequences revealed that intensive groundwater pumping was responsible for the absence of taxa in the analyzed sample [35]. Flowrate and water residence time also affected the microbial community structures in river water [8,9]. In addition to the effect of hydrodynamics on the groundwater microbial community, As concentration appeared to be a major factor that differentiated the groundwater microbial community in the low-As and high-As assemblages [36]. The As concentration was one of the geochemical parameters that influenced the microbial community in the groundwater of Inner Mongolia, China [11].

While As, residence time, and groundwater flowrate potentially influenced the microbial structure in the groundwater, TOC, ORP, and DO showed positive correlations with the microbial structure in the surface waters (Fig. 5). TOC was found to be one of the geochemical factors affecting the microbial communities in groundwaters [11, 36]. By denaturing gradient gel electrophoresis (DGGE) and 16S rRNA gene sequencing, the microbial community in surface water of the Jiulong River, China, was correlated with water temperature, conductivity, PO_4-P, and TN/TP instead of with TOC [37]. By next generation sequencing analysis of the 16S rRNA gene, it was found that ORP and DO impacted the microbial community in the subsurface groundwater of the Nakdong River Bank in South Korea [27]. In addition, ORP was one of the geochemical parameters influencing the microbial community in the groundwater of the Hetao Plain in Inner Mongolia [36].

Conclusions

386 The abundance of arsenite-oxidizing bacteria and the microbial community structures
387 in the groundwater, surface water, and soil from the gold mining area, including the upstream
388 and downstream areas of the mine, were explored. Arsenite-oxidizing bacteria in the
389 groundwater, surface water, and soil samples were closely related to *Proteobacteria*, indicating
390 common arsenite-oxidizing bacterial taxa found in a mesophilic environment. Together with
391 the results of the ddPCR analysis, the results suggested that arsenite-oxidizing bacteria were
392 widespread at low abundance and corresponded with the low As concentrations. Nonetheless,
393 the PHREEQC model predicted that As^{5+} was dominant in all analyzed samples, potentially
394 indicating the oxidation of As^{3+} to As^{5+} by the activity of arsenite-oxidizing bacteria. The
395 phylum *Proteobacteria* was highly represented in all analyzed samples. The soil and surface
396 water microbiomes were mainly composed of the phyla *Proteobacteria*, *Actinobacteria*,
397 *Bacteroidetes*, and *Chloroflexi*. *Oxyphotobacteria*, members of the phylum *Cyanobacteria*,
398 were dominant in surface waters but not in soil or groundwater. The groundwater microbiomes
399 were dominated exclusively by *Proteobacteria*, especially the classes *Betaproteobacteria* and
400 *Alphaproteobacteria*. The heatmap of the most abundant OTUs revealed unique bacterial taxa
401 in each sample. These dominant bacterial taxa likely play roles in the carbon, nitrogen, and
402 sulfur cycles. Cluster analyses showed that habitat primarily governed the microbial structures
403 in the analyzed samples. The environmental factors that potentially influenced the microbial
404 structure in the groundwater were As, residence time, and groundwater flowrate, while those
405 affecting the microbial structure in the surface water were TOC, ORP, and DO. However,
406 arsenite-oxidizing bacteria and microbial community structure were monitored only in a single
407 season, and they should be further analyzed at different time points to better understand the
408 effects of spatial and temporal variations. Overall, this study provides insights into the

409 groundwater, surface water, and soil microbiomes within and surrounding a gold mine and
410 expands the current knowledge of the diversity and abundance of arsenite-oxidizing bacteria,
411 which could play a role in As transformation in impacted environments.
412
413 **Materials and methods**
414 **Sampling site description and sample collection**
415 The study area was a gold mine located in Phetchabun (L1 to L2) and Pichit (L3 to L6)
416 Provinces, Thailand (Fig. 6, Table S3). The direction of groundwater flow was from L1 to L6.
417 Groundwater, surface water, and soil samples were collected from a gold mining area (L2-L5),
418 including upstream (L1) and downstream (L6) of the gold mining area. In a total of 12 analyzed
419 samples, six groundwater (GW) samples were collected from L1 to L6, together with three
420 surface water (SW) and three soil (S) samples from L1, L2, and L6 (Fig. 1 and Fig. S1). The
421 details of sample collection are shown in the supplementary information.
422 **Geochemical analyses of groundwater and surface water**
423 The geochemical parameters measured on site were temperature, dissolved oxygen
424 (DO), pH and oxidation-reduction potential (ORP). Temperature and DO were measured using
425 a portable meter (Hach, USA). ORP and pH were measured using a dissolved oxygen meter
426 (WTW, USA). Arsenic (As) concentrations were analyzed using atomic absorption hydride
427 (continuous) ZEEnit 700P (Analytik Jena AG, Germany) according to an atomic absorption
428 spectrometric method [38]. TOC was analyzed using a total organic carbon analyzer (TOC-
429 VCPH; Shimadzu, Japan) following a previously published protocol [39]. Total phosphorus
430 (TP) and total nitrogen (TKN) were analyzed by the ascorbic acid method and macro-Kjeldahl

method, respectively [39]. The geochemical parameters of six groundwater and three surface water samples are shown in Table 1.

Prediction of As speciation in groundwater using the geochemical model PHREEQC

The As speciation in groundwater was calculated by using the geochemical model PHREEQC [40,41]. The geochemical parameters included in the model were total As, pH, ORP, temperature, and DO (Table 1). Other than these analyzed geochemical parameters, the concentrations of Fe^{2+}, Mn^{2+}, and SO_4^{2-}, which are the secondary data from the previous study [42], were also included (Table S4). All these parameters were loaded into the PHREEQC geochemical model using the WATEQ4F.dat database. The thermodynamic data, including dissolution values, enthalpy, and dissociation equation, were provided by the program [43]. To determine the effect of iron oxides on As sorption in natural environments, the Hfo-As sorption model was calculated by Phreeplot extension through the geochemical model PHREEQC [40,41].

DNA extraction

Genomic DNA of water and soil samples was extracted using a FastDNA™ SPIN Kit for Soil (MP Biomedicals, USA), following the manufacturer's protocol. For groundwater and surface water samples, the filters (Sigma-Aldrich, USA) were cut into very tiny pieces before placing into a bead-beating tube provided by the extraction kit. The quantity and quality of extracted DNA were examined by a NanoDrop spectrophotometer ND-100 (Thermo Fisher Scientific, USA) and agarose gel electrophoresis, respectively. The extracted DNA of each sample was subsequently diluted to 5 ng μl^{-1} to use as a PCR template.

***aioA* gene clone library construction**

Twelve samples (6 groundwater, 3 surface water, and 3 soil samples) were screened for the presence of the *aioA* gene using specific primers aoxBM1-2F-ND and aoxBM2-1R-ND [26]. Positive *aioA* amplicons of 550 bp were generated using a T100™ Thermal Cycler (Biorad, USA). The details of *aioA* gene amplification are shown in the supplementary information. PCR products were subsequently verified using agarose gel electrophoresis. Before ligation, PCR products were purified to remove impurities using a NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Germany), according to the manufacturer's protocols. The purified PCR products were ligated and cloned into pGEM®-T Easy Vector Systems (Promega, USA) and XL1-Blue supercompetent cells (Agilent, USA), respectively, according to the manufacturer's protocols. Approximately 20-25 white colonies were randomly selected for sequencing at Macrogen Inc., South Korea (ABI 3730XL sequencer). Nucleotide sequences reported in this study were deposited in the GenBank database under accession numbers MK751216-MK751301.

Phylogenetic analysis

All analyzed *aioA* sequences were searched against the GenBank database with BLAST to obtain closely related reference sequences. Then, all sequences from each library were clustered at 97% operational taxonomic units (OTUs) using a CD-HIT program [44]. Each representative OTU of each library, reference *aioA* sequences, and an outgroup (*Synechocystis* sp. NR076327) were included in the phylogenetic tree construction using the MEGA package, version 7.0.21 [45]. All analyzed sequences included in the phylogenetic analysis were aligned using MUSCLE [46]. Neighbor-joining tree with 1,000 bootstrap replicates was constructed using the maximum composite likelihood model.

Droplet digital PCR (ddPCR)

The abundance of arsenite-oxidizing bacteria and total bacteria was estimated through the quantification of the *aioA* and 16S rRNA gene copies, respectively, using the ddPCR approach. Two primer sets, aoxBM1-2F-ND/aoxBM2-1R-ND [26] and 341f/518r [47], were used to estimate the numbers of *aioA* and 16S rRNA genes, respectively. The details of ddPCR mixture and ddPCR condition are shown in the supplementary information. All amplifications were conducted in triplicate in a T100™ Thermal Cycler (Biorad, USA). Clone with the *aioA* gene fragment and purified bacterial genomic DNA were included as positive controls for the *aioA* and 16S rRNA genes, respectively. After the amplification process, the numbers of detected genes across all samples were analyzed by QuantaSoft analysis software (Bio-Rad, USA).

Illumina MiSeq library preparation

Extracted genomic DNA of 12 samples was amplified in triplicate for each sample using a T100™ Thermal Cycler (Biorad, USA). The V3 and V4 regions of the 16S rRNA gene were amplified with the previously published primers [48]. The detail of PCR condition is shown in the supplementary information. For each sample, approximately 400 ng μl^{-1} of purified PCR product was used for the Illumina library preparation using the MiSeq Reagent Kit V3, 600 cycles (2x300 bases; Illumina, USA), following the manufacturer's protocol. Raw sequence data were deposited in the Sequence Read Archive (SRA) under the BioProject accession PRJNA528471.

Bioinformatic, statistical and multivariate analyses

After retrieving the raw 16S rRNA (V3-V4) metagenomic reads, adapters and primer sequences were removed using Trimmomatic version 0.38 [49]. The amplicon-based metagenomic analysis was performed using mothur version 1.41.1 [50]. Paired-end reads were

first merged into contiguous contigs. Contaminated sequences, including ambiguous sequences, nontargeted amplicon size, homopolymers, chimeric sequences, and off-target sequences were removed. The qualified contigs were *de novo* clustered into operational taxonomic units (OTUs) based on 97% sequence similarity. Singleton and doubleton OTUs were considered sequencing errors and were removed. The taxonomy of the derived OTUs was assigned based on the SILVA database version 132 [51]. Taxonomic profiles were drawn using relative abundance at different taxa levels. To compare the metagenomic profiles between samples, OTU abundances were normalized using the scaling method. Rarefaction curves were calculated based on normalized abundance to observe species richness in the analyzed samples. The Shannon index was used to measure community diversity within a sample. A heatmap was generated using the seaborn Python package representing the quantification of diversity between samples. Principal coordinates analysis (PCoA) was performed based on the Bray-Curtis dissimilarity metric for estimating microbial compositions between samples. Canonical correspondence analysis (CCA) was analyzed using the vegan R package [52] to determine the dominant environmental parameters that were correlated with a linear relationship to the microbiome. In-house Python and R scripts were used for visualization.

References

1. Abia, A. L. K., Alisoltani, A., Keshri, J., & Ubomba-Jaswa, E. Metagenomic analysis of the bacterial communities and their functional profiles in water and sediments of the Apies River, South Africa, as a function of land use. *Sci. Total Environ.* **616**, 326-334 (2018).

2. Epelde, L., Lanzen, A., Blanco, F., Urich, T., & Garbisu, C. Adaptation of soil microbial community structure and function to chronic metal contamination at an abandoned Pb-Zn mine. *FEMS Microbiol. Ecol.* **91**, 1; 10.1093/femsec/fiu007 (2015).

525 3. Sonthiphand, P. *et al*. Metagenomic insights into microbial diversity in a groundwater
526 basin impacted by a variety of anthropogenic activities. *Environ. Sci. Pollut. Res*. **26**, 26765–
527 26781 (2019).
528
529 4. Boonsrang, A., Chotpantarat, S., & Sutthirat, C. Factors controlling the release of metals
530 and a metalloid from the tailings of a gold mine in Thailand. *Geochem-Explor. Environ. Anal.*
531 **18**, 109-119 (2017).
532
533 5. Fashola, M., Ngole-Jeme, V., & Babalola, O. Heavy metal pollution from gold mines:
534 environmental effects and bacterial strategies for resistance. *Int. J. Environ. Res. Public*
535 *Health* **13**, 1047; 10.3390/ijerph13111047 (2016).
536
537 6. Jackson, T. A., Vlaar, S., Nguyen, N., Leppard, G. G., & Finan, T. M. Effects of
538 bioavailable heavy metal species, arsenic, and acid drainage from mine tailings on a
539 microbial community sampled along a pollution gradient in a freshwater ecosystem.
540 *Geomicrobiol. J.* **32**, 724-750 (2015).
541
542 7. Das, S., Bora, S. S., Yadav, R. N. S., & Barooah, M. A metagenomic approach to decipher
543 the indigenous microbial communities of arsenic contaminated groundwater of Assam.
544 *Genomics Data* **12**, 89-96 (2017).
545
546 8. Read, D. S. *et al*. Catchment-scale biogeography of riverine bacterioplankton. *ISME J.* **9**,
547 516-526 (2015).
548
549 9. Wang, P., Chen, B., Yuan, R., Li, C., & Li, Y. Characteristics of aquatic bacterial
550 community and the influencing factors in an urban river. *Sci. Total Environ.* **569**, 382-389
551 (2016).
552
553 10. Castañed,a L. E., & Barbosa, O. Metagenomic analysis exploring taxonomic and
554 functional diversity of soil microbial communities in Chilean vineyards and surrounding
555 native forests. *PeerJ* **5**, e3098; 10.7717/peerj.3098 (2017).
556
557 11. Li, P. *et al*. Microbial community in high arsenic shallow groundwater aquifers in Hetao
558 Basin of Inner Mongolia, China. *PLOS ONE* **10**, e0125844; 10.1371/journal.pone.0125844
559 (2015).
560
561 12. Li, P. *et al*. Analysis of the functional gene structure and metabolic potential of microbial
562 community in high arsenic groundwater. *Water Res.* **123**, 268-276 (2017).
563
564 13. Sun, W. *et al*. Response of soil microbial communities to elevated antimony and arsenic
565 contamination indicates the relationship between the innate microbiota and contaminant
566 fractions. *Environ. Sci. Technol.* **51**, 9165-9175 (2017).
567
568 14. Bhattacharya, P. *et al*. Arsenic in environment: biology and chemistry. *Sci. Total Environ.*
569 **379**, 109-120 (2007).
570

15. Inskeep, W. P., McDernlott, T. R., & Fendorf, S. In *Environmental chemistry of arsenic - Arsenic (V)/(III) cycling in soils and natural waters: chemical and microbiological processes*. (ed. Frankenberger, W.) 203-236 (CRC Press, 2001).

16. Jackson, C. R., Jackson, E. F., Dugas, S. L., Gamble, K., & Williams, S. E. Microbial transformations of arsenite and arsenate in natural environments. *Recent Res. Dev. Microbiol.* **7**, 103-118 (2003).

17. Quéméneur, M. *et al*. Diversity surveys and evolutionary relationships of *aoxB* genes in aerobic arsenite-oxidizing bacteria. *Appl. Environ. Microbiol.* **74**, 4567-4573 (2008).

18. Santini, J. M., Sly, L. I., Schnagl, R. D., & Macy, J. M. A new chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine: phylogenetic, physiological, and preliminary biochemical studies. *Appl. Environ. Microbiol.* **66**, 92-97 (2000).

19. Das, S. *et al*. Depth-resolved abundance and diversity of arsenite-oxidizing bacteria in the groundwater of Beimen, a blackfoot disease endemic area of southwestern Taiwan. *Water Res.* **47**, 6983-6991 (2013).

20. Jiang, Z. *et al*. Diversity and abundance of the arsenite oxidase gene *aioA* in geothermal areas of Tengchong, Yunnan, China. *Extremophiles* **18**, 161-170 (2014).

21. Smedley, P. L. In *Sources and distribution of arsenic in groundwater and aquifers*. - *Arsenic in groundwater: a world problem*. (ed. Appelo, T.) 4-32 (International Association of Hydrogeologists Publication, 2008).

22. Pal, P. *Groundwater arsenic remediation: treatment technology and scale up*. (Elsevier, 2015)

23. Williams, M. Arsenic in mine waters: an international study. *Environ. Geol.* **40**, 267-278 (2001).

24. Hassan, Z., Sultana, M., van Breukelen, B. M., Khan, S. I., & Röling, W. F. Diverse arsenic-and iron-cycling microbial communities in arsenic-contaminated aquifers used for drinking water in Bangladesh. *FEMS Microbiol. Ecol.* **91**, fiv026; 10.1093/femsec/fiv026 (2015).

25. Inskeep, W. P. *et al*. Detection, diversity and expression of aerobic bacterial arsenite oxidase genes. *Environ. Microbiol.* **9**, 934-943 (2007).

26. Quéméneur, M. et al. (2010). Population structure and abundance of arsenite-oxidizing bacteria along an arsenic pollution gradient in waters of the Upper Isle River Basin, France. *Appl. Environ. Microbiol.* **76**, 4566-4570.

27. Lee, J. H., Unno, T., & Ha, K. Vertical and horizontal distribution of bacterial communities in alluvial groundwater of the Nakdong River bank. *Geomicrobiol. J.* **35**, 74-80 (2018).

28. Gu, Y. *et al*. Bacterial community and arsenic functional genes diversity in arsenic contaminated soils from different geographic locations. *PLOS ONE* **12**, e0176696; 10.1371/journal.pone.0176696 (2017).

29. Hemmat-Jou, M. H., Safari-Sinegani, A. A., Mirzaie-Asl, A., & Tahmourespour, A. Analysis of microbial communities in heavy metals-contaminated soils using the metagenomic approach. *Ecotoxicology* **27**, 1281-1291 (2018).

30. Crump, B. C., Amaral-Zettler, L. A., & Kling, G. W. Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *ISME J.* **6**, 1629; 10.1038/ismej.2012.9 (2012).

31. Fang, L. *et al*. Planktonic and sedimentary bacterial diversity of Lake Sayram in summer. *MicrobiologyOpen* **4**, 814-825 (2015).

32. Janssen, P. H. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.* **72**, 1719-1728 (2006).

33. Stamps, B. W. *et al*. Characterization of the Microbiome at the World's Largest Potable Water Reuse Facility. *Front. Microbiol.* **9**, 2435; 10.3389/fmicb.2018.02435 (2018).

34. Monard, C., Gantner, S., Bertilsson, S., Hallin, S., & Stenlid, J. Habitat generalists and specialists in microbial communities across a terrestrial-freshwater gradient. *Sci. Rep.* **6**, 37719; 10.1038/srep37719 (2016).

35. Ben Maamar, S. *et al*. Groundwater isolation governs chemistry and microbial community structure along hydrologic flowpaths. *Front. Microbiol.* **6**, 1457; 10.3389/fmicb.2015.01457 (2015).

36. Wang, Y. *et al*. Microbial community of high arsenic groundwater in agricultural irrigation area of Hetao Plain, Inner Mongolia. *Front. Microbiol.* **7**, 1917; 10.3389/fmicb.2016.01917 (2016).

37. Liu, L., Yang, J., Yu, X., Chen, G., & Yu, Z. Patterns in the composition of microbial communities from a subtropical river: effects of environmental, spatial and temporal factors. *PLOS ON*E **8**, e81232; 10.1371/journal.pone.0081232 (2013).

38. United States Department of Agriculture (USDA). In *Food Safety and Inspection Service - Determination of arsenic by atomic absorption spectroscopy*. (Office of Public Health Science, 2016)

660 39. American Public Health Association (APHA) *Standard Methods for the Examination of*
661 *Water and Wastewater* 22nd ed. (American Water Works Association, 2012)
662
663 40. Parkhurst, D. L., & Appelo, C. A. J. (1999). User's guide to PHREEQC (Version 2): A
664 computer program for speciation, batch-reaction, one-dimensional transport, and inverse
665 geochemical calculations. *Water-Resources Investigations Report* **99**, 4259 (1999).
666
667 41. Parkhurst, D. L., & Appelo, C. A. J. *Description of input and examples for PHREEQC*
668 *version 3: a computer program for speciation, batch-reaction, one-dimensional transport,*
669 *and inverse geochemical calculations*. https://doi.org/10.3133/tm6A43 (2013).
670
671 42. Department of Primary Industries and Mines (DPIM) *The analysis of environmental*
672 *impacts within a gold mining area in Pichit, Phetchabun, and Phitsanulok Provinces*.
673 (Chulalongkorn University, 2016)
674
675 43. Nordstrom, D. K., & Archer, D. G. In *Arsenic in ground water - Arsenic thermodynamic*
676 *data and environmental geochemistry*. (eds. Welch, A. H., & Stollenwerk, K. G.) 1-25
677 (Springer, 2003)
678
679 44. Li, W., & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of
680 protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659 (2006).
681
682 45. Kumar, S., Stecher, G., & Tamura, K. MEGA7: molecular evolutionary genetics analysis
683 version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870-1874 (2016).
684
685 46. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
686 throughput. *Nucleic Acids Res.* **32**, 1792-1797 (2004).
687
688 47. Muyzer, G., De Waal, E. C., & Uitterlinden, A. G. Profiling of complex microbial
689 populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-
690 amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **59**, 695-700 (1993).
691
692 48. Klindworth, A. *et al*. Evaluation of general 16S ribosomal RNA gene PCR primers for
693 classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1;
694 10.1093/nar/gks808 (2013).
695
696 49. Bolger, A. M., Lohse, M., & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
697 sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
698
699 50. Schloss, P. D. *et al*. Introducing mothur: open-source, platform-independent, community-
700 supported software for describing and comparing microbial communities. *Appl. Environ.*
701 *Microbiol.* **75**, 7537-7541 (2009).
702
703 51. Quast, C. *et al*. The SILVA ribosomal RNA gene database project: improved data
704 processing and web-based tools. *Nucleic Acids Res.* **41**, D590-D596; 10.1093/nar/gks1219
705 (2012).

52. Oksanen, J. *et al*. Vegan: community ecology package. R package version, 2-3 (2015)

Acknowledgements

This work was funded principally by the Thailand Research Fund (TRF) Grant for New Scholar (MRG6180127). This research project was also supported by the Thailand Toray Science Foundation (TTSF) through the Science & Technology Research Grant and by the Faculty of Science, Mahidol University. We would also like to thank King Mongkut's University of Technology Thonburi through the KMUTT 55th Anniversary Commemorative Fund for supporting high performance computing.

Author contributions

PS designed the experiments, interpreted the results, and wrote the main manuscript. SK and SC conducted the geochemical PHREEQC model. SK and EL involved in the analysis and discussion of arsenite-oxidizing bacteria and microbial community structures. SP, KK, PU, and CR conducted bioinformatic, statistical and multivariate analyses. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Figure Legends

Fig. 1. Neighbor-joining phylogenetic tree of *aioA* gene sequences retrieved from groundwater (GW), surface water (SW), and soil (S) across 6 sampling locations (L1-L6). The tree was constructed with 1,000 bootstrap replicates, and only bootstrap values >50% are shown. Bold text and color indicate *aioA* sequences recovered from each habitat: groundwater (blue), surface water (green) and soil (brown).

Fig. 2. Abundance of microbial compositions at the class level associated with six major phyla (a-f) detected in groundwater (GW), surface water (SW), and soil (S) from 6 sampling locations (L1-L6).

Fig. 3. Clustering of the samples using a heatmap based on the presence of OTUs that were more than 2% abundant. The color intensity shows the proportion of a genus found in each sample.

Fig. 4. Principal coordinate analysis (PCoA) of microbial compositions in groundwater (GW), surface water (SW), and soil (S) across 6 sampling locations (L1-L6) based on the Bray-Curtis dissimilarity matrix.

Fig. 5. Canonical correspondence analysis (CCA) of microbial structures and geochemical parameters. Arrows indicate the correlation and magnitude of geochemical parameters associated with bacterial community structure.

Fig. 6. Map showing the sampling locations (L1-L6). GW, SW, and S refer to the groundwater, surface water, and soil samples, respectively.

Tables

Table 1. Geochemical parameters of 6 groundwater (GW) and 3 surface water (SW) samples

sample ID	pH	Temp. (°C)	DO (mg l^{-1})	ORP (mV)	TOC (mg l^{-1})	TKN (mg $l^{-1)}$	TP (mg l^{-1})	Total As (µg l^{-1})
L1-GW	6.50	29.2	3.87	212.2	1.24	0.3	0.02	3.97
L2-GW	6.87	32.3	2.84	59.4	1.34	0.4	0.02	3.52
L3-GW	6.39	30.5	3.29	115.4	3.24	0.1	0.05	4.45
L4-GW	6.70	31.8	2.95	88.9	1.40	0.3	0.06	3.50
L5-GW	4.90	31.0	2.92	296.9	1.22	0.1	0.01	3.12
L6-GW	7.21	32.3	4.79	149.5	1.43	0.5	0.04	2.05
L1-SW	6.34	31.0	7.63	390.3	4.19	0.1	0.05	2.29
L2-SW	6.98	33.4	7.14	308.1	2.92	0.5	0.02	2.31
L6-SW	8.22	34.0	7.19	269.1	16.62	0.3	0.03	2.66

Table 2. As speciation calculated by using the geochemical model PHREEQC

Sample ID	As^{3+} (µg l^{-1})	As^{5+} (µg l^{-1})	Dominant form (speciation model)	Dominant form (Hfo-As sorption model)
L1-GW	0.00	3.97	$H_2AsO_4^-$	$H_2AsO_4^-$
L2-GW	0.08	3.45	$H_2AsO_4^-$	$H_2AsO_4^-$
L3-GW	0.06	4.40	$H_2AsO_4^-$	$H_2AsO_4^-$
L4-GW	0.03	3.47	$H_2AsO_4^-$	$H_2AsO_4^-$
L5-GW	0.00	3.12	$H_2AsO_4^-$	$H_2AsO_4^-$
L6-GW	0.00	2.05	$HAsO_4^{2-}$	$hfo\_wOHAsO_4^{3-}$
L1-SW	0.00	2.29	$H_2AsO_4^-$	$hfo\_wOHAsO_4^{3-}$
L2-SW	0.00	2.31	$H_2AsO_4^-$	$hfo\_wOHAsO_4^{3-}$
L6-SW	0.00	2.66	$HAsO_4^{2-}$	$HAsO_4^{2-}$