Diversity and arsenic-tolerance potential of bacterial communities from soil and sediments along a gold tailing contamination gradient

Journal:	<i>Canadian Journal of Microbiology</i>
Manuscript ID	cjm-2017-0214.R2
Manuscript Type:	Article
Date Submitted by the Author:	19-Jun-2017
Complete List of Authors:	Guan, Xiangyu; China University of Geosciences, Beijing Yan, Xu; School of Ocean Sciences, China University of Geosciences, Beijing Li, Youxun; National Oceanographic Center, Qingdao Jiang, Bo; School of Ocean Sciences, China University of Geosciences, Beijing Luo, Ximing; School of Ocean Sciences, China University of Geosciences, Beijing Chi, Xiaoyuan; Shandong Peanut Research Institute
Keyword:	Arsenic resistance, bacterial communities, ArsB, ACR3, environmental factors

SCHOLARONE™
Manuscripts

Diversity and arsenic-tolerance potential of bacterial communities from soil and sediments along a gold tailing contamination gradient

Xiangyu Guan ^{a,*}, Xu Yan ^{a,b}, Youxun Li ^{b,c}, Bo Jiang ^a, Ximing Luo ^a, Xiaoyuan Chi ^{d,*}

^a*School of Ocean Sciences, China University of Geosciences, Beijing 100083, P.R. China*

^b*Beijing Key Laboratory of Water Resources and Environmental Engineering, China University of Geosciences, Beijing 100083, P.R. China*

^c*National Oceanographic Center, Qingdao 266071, P.R. China*

^d*Shandong Peanut Research Institute, Qingdao, 266100, P R China*

* Corresponding author

Tel: +86-010-82320920

Address: School of Ocean Science, China University of Geosciences, Beijing, No.29 Xueyuan Road, Haidian District, Beijing, 100083, PRC.

E-mail: Xiangyu Guan, guanxy@cugb.edu.cn; Xiaoyuan Chi, chi000@126.com

Abstract

Gold tailings often release arsenic (As) contaminants into the surrounding environment. Microorganisms play an important role in the As cycle, whereas the effects of As on bacterial communities remain unclear. To reveal the effects of As on the diversity of bacterial communities and their As-tolerance potential, farmland soil and river sediment samples were collected at varying distances from tailings in the

Dandong area of northeastern China. The bacterial communities were analyzed using high-throughput sequencing of 16S rRNA genes. The membrane transport proteins ArsB or ACR3 pump As(III) out of the cell to resist As toxicity. We studied the abundance and phylogeny of ArsB and ACR3 using PCR-based clone libraries and quantitative PCR. The bacterial community was divided into 10 phyla and 59 genera. The transformation from As (V) to As(III) was predominant, which was coupled with denitrification. Both ArsB and ACR3 likely evolved from different orders of Proteobacteria. The *arsB* gene seems to be more stable in bestowing the capability of bacteria to respond to the As concentration. Moreover, As with iron, manganese and total organic carbon also influenced the clustering relationships of samples and bacterial distribution.

Keywords: Arsenic resistance, bacterial communities, ArsB, ACR3, environmental factors

Introduction

Arsenic (As) is a ubiquitous metalloid element in the Earth's crust and almost all As compounds are strongly toxic. With increasing developments in the modern economy, anthropogenic activities such as mining and industrial sewage are releasing more and more As compounds into the surrounding environments. Smelter emissions, atmospheric fallout, and oxidized and leached solutions of tailings lead to the migration and proliferation of As (Carrizales et al. 2006; Matschullat 2000). Residents living around gold mines are likely to be exposed to As-rich water and foods, and are likely to experience various types of cancer (Niemeyer et al. 2012).

Microbial communities are complex and sensitive to subtle environmental changes,

and thus microorganisms are often regarded as important indicators of the environment (Niemeyer et al. 2012; Valverde et al. 2011). Previous studies have demonstrated that As affects microbial community structure and function in soils, aquifers and groundwater (Lami et al. 2013; Legg et al. 2012; Liao et al. 2011). The effect of As interacting with other environmental factors (carbon, metals and sulfate) has been analyzed using 16S rRNA gene sequences (Freikowski et al. 2013; Legg et al. 2012; Sheik et al. 2012). The results showed that the structure, diversity and abundance of microbial communities were highly influenced by the presence and concentration of As (Sheik et al. 2012). In another study, shifts in functional bacterial communities in acid mine drainage with a high As content were examined. The seasonal fluctuations of sulfate-reducing bacteria community structure showed a clear correlation with As, iron (Fe) and pH levels (Giloteaux et al. 2013a). However, there is a lack of comprehensive knowledge on the structure of bacterial communities involved in As-contaminated farm soils and rivers around gold-mining sites.

Recent studies have also provided evidence that microorganisms can lower the levels of As compounds by mineralization and oxidation (Wang and Mulligan 2009). Firstly, there are bacteria resistant to or capable of transforming As, and these bacteria predominantly belong to the phyla Proteobacteria, Firmicutes and Actinobacteria (Sheik et al. 2012). Secondly, microorganisms have evolved various metabolic responses to As on the basis of redox reactions and/or efflux systems, which could directly or indirectly affect the transformation and migration of As. Arsenite [As(III)] and arsenate [As(V)] are the two most common inorganic As oxidation states. Both

67 As(III) oxidation and As(V) reduction have been identified in various bacterial genera:
68 *Pseudomonas*, *Alcaligenes*, *Thiomonas*, *Herminiimonas*, *Agrobacterium* and *Thermus*
69 for As(III) oxidation (Cai et al. 2009), and *Sulfurospirillum*, *Alkaliphilus*,
70 *Desulfosporosinus* and *Geobacter* for As(V) reduction (Giloteaux et al. 2013).

71 The *ars* genes are the key encoding genes, and are related to the most common
72 pathway for As detoxification (Rosen 2002). A single *ars* transcriptional unit contains
73 three (*arsRBC*) or five (*arsRDABC*) genes (Achour et al. 2007). This system basically
74 involves the enzymatic reduction of As(V) by ArsC to As(III), which is pumped out
75 of the cell by a membrane transport protein from the ArsB or ACR3 families (Achour
76 et al. 2007). Both ArsB and ACR3 showed a similar function of conferring As
77 resistance, whereas these proteins have no significant sequence similarity. ACR3p can
78 be divided into two subfamilies based on their phylogenetic dissimilarities: ACR3(1)
79 and ACR3(2). The As(III) transporter genes *arsB*, *ACR3(1)* and *ACR3(2)*, have been
80 identified in various soil bacteria (Cai et al. 2009; Poirel et al. 2013). However, the
81 relationship between the abundance of *arsB* or *Acr3* genes with As and other
82 environmental factors, and the different responses of the diversity of ArsB or ACR3
83 to As-contaminated environments require further study.

84 To provide comprehensive knowledge on the bacterial community structure and
85 functions involved in As-contaminated environments, we collected samples from
86 farmland soil and river sediments at various distances from a gold-mine tailing site.
87 We focused on (1) the differences in diversity, composition and phylogeny of bacterial
88 communities and ArsB and ACR3 between river sediment and farm soil samples; (2)

whether the bacterial communities were influenced by As and determined the major microbial process of As cycling at sampling sites; and (3) the response of the diversity and distribution of ArsB and ACR3 to As contamination, and which genes were likely to indicate the As concentration.

Materials and Methods

Sample collection and chemical analysis

Sampling sites were around gold mine tailings in the Dandong area of China. The main minerals of this tailing are gypsum and the main components are Fe, Ca and S, and the As is mainly in the form of FeAsS or As₂O₅. The Luo river flowed by the edge of tailing, and corn farmland was distributed by the river. The Luo river flows through a gold mine tailing, and the closest distance between the river and the tailing is 50 meters. Cornfields are located on either side of the river. We sampled surface sediment samples (SE) along the river about 2.0, 4.0, 8.0 and 10.0 km from SE1 (nearest from tailings), and soil samples (SO) were collected from corn farmland in October 2014 (Fig. 1). Sediments under about 30-cm depth of water were taken with stainless steel spoons at the vertex of a triangle (about 0.5 m triangle side), which was placed ~2 m from the edge of the river. At each corn farmland sampling site, we collected 3–5 cm layers of the soil about 1–2 m from the edge of the river. Triplicate samples were taken at the vertex of a triangle (about 1 m triangle side). Soil samples were collected using stainless steel sample tubes (diameter 40 × 40 mm, length 22 cm), and these tubes were immediately put into an incubator with ice packages.

110 Samples from the 5–10-cm depth were taken and mixed under a sterile condition as
111 soon as possible, and then divided into two parts and stored at 4°C for chemical tests
112 and –20°C for DNA extraction within 12 h, respectively.

113 Soil and sediment pH values were measured in a soil/water (1:5 w/v) suspension
114 using a pH meter (Sartorius PB-10, Goettingen, Germany). Sediment total organic
115 carbon (TOC) were measured using the K₂Cr₂O₇ oxidation method (Islam and Weil
116 1998). Total nitrogen (TN) was analyzed using a Eurovector elemental analyzer
117 (Isoprime-EuroEA 3000, Milan, Italy). Total sulfur (TS) was measured using the
118 infrared absorption method after high frequency combustion (High-speed Analyzer
119 HWF-900A, Wuxi, China). Other trace metals were analyzed by ICP-MS (Thermo
120 Fisher X-series, Franklin, MA, USA) and ICP-AES (TJA IRIS-Advantage, Franklin,
121 MA, USA).

122 Soil and sediment samples (soil/water 1:5 w/v) were soaked for 24 h, and then
123 filtered using 0.45 µm-pore-size Millipore GSWP filters. Speciation of As was
124 analyzed using a SA-10 (IC-HG-AFS) atomic fluorescence spectrometry speciation
125 analyzer (Beijing Titan Instruments, Beijing, China), equipped with a LC-10ATVP
126 high pressure liquid pump (Shimadzu Corporation, Kyoto, Japan) and a 7725i six-way
127 injection valve with a 50-µL sample loop (Rheodyne, Cotati, CA, USA) (Keller et al.
128 2014; Luo et al. 2008). The separation column (Hamilton Corporation, Reno, NV,
129 USA) was a PRP (styrene-divinylbenzene copolymer) - X100 (250 mm × 4.1 mm i.d.,
130 10 µm) anion exchange column, with a guard column of the same material (25 mm ×
131 2.3 mm i.d., 12-20 µm). The detection device was a hydride generation-atomic

fluorescence spectrometer (HG-AFS) and the excitation light source was a high-performance As hollow cathode lamp (General Research Institute for Nonferrous Metals) (Wei and Liu 2007). Chemicals: As(III): Standard solution, 1 g/L (National Research Center of Geoanalysis); As(V): Disodium hydrogen arsenate ($\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$); Parameters of atomic fluorescence spectrometry speciation analyzer was as follows: an ion chromatography (IC) mobile phase of 15 mmol/L $(\text{NH}_4)_2\text{HPO}_4$ (pH 6.0) and a flow rate of 1 mL/min. Regents of hydride generation (HG) were 1.5% KBH_4 and 0.5% KOH with a 4.5 mL/min flow rate.

DNA extraction and high-throughput sequencing

Total DNA of each sample (0.5g) was extracted using the PowerSoil DNA Extraction Kit (MoBio Laboratories, Carlsbad, CA, USA). The extracted DNA was stored at -20°C for further use and at -80°C for permanent preservation. The quantity and quality of isolated DNA were evaluated using a Nano Drop spectrophotometer (Thermo Fisher Scientific ND-1000, Waltham, MA, USA) and agarose gel electrophoresis (Bio-Rad, Hercules, CA, USA), respectively.

16S rRNA gene was amplified using primers 515F-806R for the 16S rRNA V4 region. Adapters, indexes (barcodes of 6 nucleotides for mixing and separating different samples) and sequencing primers were integrated into each PCR primer (Bartram et al. 2011), thus the construction of the sequencing libraries was achieved after PCR amplification. All PCR reactions were carried out in 30 μL reactions with 15 μL of Phusion® High-Fidelity PCR Master Mix (New England Biolabs, Beverly,

MA, USA); 0.2 μ M of forward and reverse primers, and about 10 ng templates DNA. Thermal cycling consisted of initial denaturation at 98°C for 1 min, followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 50°C for 30 s, elongation at 72°C for 60 s, and finally 72°C for 5 min. PCR products were mixed in equidensity ratios. Then, mixture PCR products were purified using a GeneJET Gel Extraction Kit (Thermo Scientific). Library quality was assessed on the Qubit@ 2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. Libraries with different indexes were diluted and pooled into one sample according to their estimated concentrations. The pooled libraries were sequenced on an Illumina MiSeq platform and 250 bp/300bp paired-end reads were generated. The WGS sequencing data were deposited in the Sequence Read Archive (SRA) database with accession number SRP058903.

Processing of high-throughput sequencing data

Paired-end reads from the original DNA fragments are merged using FLASH, a very fast and accurate analysis tool that is designed to merge paired-end reads when there are overlaps between reads1 and reads2 (Magoc and Salzberg 2011). Paired-end reads were assigned to each sample according to the unique barcodes. Sequences were analyzed using the QIIME software package (Quantitative Insights Into Microbial Ecology), and in-house Perl scripts were used to analyze alpha diversity (Caporaso et al. 2010). First, reads were filtered using QIIME quality filters. Sequences with $\geq 97\%$ similarity were assigned to the same OTUs. We chose representative sequences for

each OTU and used the RDP classifier to annotate taxonomic information for each representative sequence (Cole et al. 2014).

PCR amplification of *ACR3* and *arsB* gene fragments, clone library construction and phylogenetic analysis

Genes involved in arsenite detoxification (*arsB* and *ACR3*) were amplified by PCR from the DNA isolated from soil (SO1 and 4) and sediment (SE1 and 4) samples of this study. The primers set *acr5F* (5'-TGATCTGGGTCATGATCTTCCC VATGMTGVT-3') and *acr4R* (5'-CGGCCACGGCCAGYTCRAARAARTT-3') were used to amplify the *ACR3* genes, and primers *darsb1F* (5'-GGTGTGGAACATCGTCTGGAAYGCNAC-3') and *darsb1R* (5'-CAGGCCGT ACACCACCAGRTACATNCC-3') were used on soil DNA to amplify the *arsB* genes (Poirel et al. 2013; Sheik et al. 2012). All 25 µL reactions contained 1 ng purified DNA template, 5 µL 1X GoTaq Flexi buffer, 2.5 µL MgCl₂ solution, 0.5 U Taq DNA polymerase, and 1 µM of each primer (Promega, Madison, WI, USA). PCR was performed using the Veriti PCR apparatus (Applied Biosystems, Carlsbad, CA, USA) under the following conditions: 5 min denaturation at 94°C followed by 35 cycles of 94°C for 45 seconds, annealing at 57–52°C with a 0.5°C decrement per cycle during the first 10 cycles (45 seconds) and then 72°C for 30 seconds with a final extension of 72°C for 7 minutes. Successful PCR amplicons were purified using a PCR Purification Mini Kit (Bio Basic Inc., Markham, Canada) and cloned using the pMD18-T cloning kit (TaKaRa, Dalian, China). Transformants were screened by

195 blue-white selection on agar plates containing 5-bromo-4-chloro-3-indolyl
196 -D-galactopyranoside (X-Gal, 20 µg/mL) (Sangon, Shanghai, China),
197 isopropyl-D-thiogalactopyranoside (IPTG, 24 µg/mL) (Sangon) and 100 µg/mL of
198 ampicillin (Sangon). White colonies were randomly picked and tested by PCR, and
199 then the positive clones were sequenced by BGI (Beijing, China).

200 The sequences of *arsB* and *ACR3* genes were first checked for chimeras using the
201 BLAST program against the NCBI database (<http://www.ncbi.nlm.nih.gov/>), and then
202 only highly-qualified members were translated to amino acid sequences using
203 BIOEDIT Version 7.1.9 (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). The closest
204 relatives of our sequences were retrieved from the GenBank database as the reference
205 sequences in further phylogenetic analysis. Sequence alignments were constructed
206 using ClustalX Version 1.8 (Thompson et al. 1997) and grouped with a 2% distance
207 cutoff by DOTUR version 1.53. Phylogenetic trees were constructed for *ArsB* and
208 *ACR3* amino acid sequences using MEGA6 with Maximum Parsimony (MP) method
209 with branch-and-bound algorithm (1,000 replications) (Tamura et al. 2013).

210 The *arsB* and *ACR3* gene sequences have been deposited in the NCBI database
211 under accession numbers KT191613 to KT191690 and KT191489 to KT191612,
212 respectively.

213 **Quantification of *ACR3*, *arsB* and 16S rRNA genes**

214 QPCR for 16S rRNA, *arsB* and *ACR3* was performed on an ABI Prism 7500 (Applied
215 Biosystems, Foster City, CA, USA) with a reaction mixture containing 10 µL of

SYBR FAST Universal 2 × qPCR Master Mix from the KAPA SYBR FAST qPCR kit (Kapa Biosystems, Woburn, MA, USA), 1 μM of each primer, 2 μL of diluted DNA template corresponding to ~10 ng of total DNA, and the appropriate amount of RNase-free water to obtain a 20 μL volume. Standard curves were obtained with serial dilutions of a known amount of plasmid DNA containing a fragment of 16S rRNA, *ACR3* and *arsB* (Hallin et al. 2009; Ritalahti et al. 2006). Standard curves for the *ACR3* and *arsB* genes were based on serial dilutions of target genes (between 10⁸ and 10² copies), and the 16S rRNA gene (between 10¹² and 10⁶ copies). These numbers were calculated from the concentration of extracted plasmids measured using a Nano Drop ND-2000 spectrophotometer. Thermal cycling conditions used for each reaction of 16S rRNA were as follows: 95°C for 5 min; 40 cycles of 95°C for 30 s, 60°C for 30 s, 72°C for 30 s; and 72°C for 10 min. Primers for 16S rRNA were 341F (5'-CCTACGGGAGGCAGCAG-3') and 534R (5'-ATTACCGCGGCTGCTGGCA-3'). Cycling conditions for the *arsB* and *ACR3* genes consisted of 95°C for 5 min; 40 cycles of 95°C for 30 s, 53°C for 30 s, 72°C for 30 s; and 72°C for 10 min. Primers darsB1F/darsB1R (5'-GGTGTGGAACATCGTCTGGAAYGCNAC-3'/5'-CAGGCCGTACACCACCA GRTACATNCC-3') and dacr1F/dacr1R (5'-GCCATCGGCCTGATCGTNATGATGTAYCC-3'/5'-CGGCGATGGCCAGCTC YAAAYTTYTT-3') were used for *arsB* and *ACR3*, respectively (Achour et al. 2007; Poirel et al. 2013).

Statistical analyses and visualization

In order to compute alpha diversity, we rarified the OTU and calculated the Chao1 (estimates the species abundance) and Shannon index of ArsB and ACR3 amino acid sequences. We performed unweighted unifracs principal coordinate analysis (PCoA) to evaluate the similarity among bacterial communities of all the samples based on the data of relative abundance of OTUs of 16S rRNA by QIIME (Caporaso et al. 2010). We then used canonical correlation analysis (CCA) using Windows 4.5 to measure the relationship between environmental parameters and bacterial community structure in Canoco (Li et al. 2008). We examined the relationships between the main environmental parameters (As, Fe, Mn, TOC, TN, TS and pH levels) and bacterial communities, and between the main environmental parameters and several dominant genera in this research, such as *Clostridium*, *Sulfuricurvum*, *Thiobacillus*, *Flavobacterium* and so on.

Results

Geochemical characteristics of samples

The geochemical characteristics of ten samples in this study are shown in Table 1. The As content reached 1,023 mg/kg at site SO1, which was higher than in other soil samples. However, the As contents were higher in the five sediment samples than in the soil samples. Even at site SE5, which was tens of kilometers from the tailings, the As content was 182 mg/kg.

Of all the samples, TOC was also highest in SO1. The highest concentrations of manganese (Mn) and Fe (and the lowest pH) occurred at SO5 and SE5. Mn and Fe

concentrations in all samples were positively correlated (Pearson rank correlation test, $P < 0.01$). Meanwhile, As concentration was negatively correlated with S concentration ($P < 0.05$).

Distribution and diversity of bacterial communities

The sequencing of 10 samples resulted in 283,280 sequences with an average length of 250 bp after quality screening from 334,727 raw paired-reads, and an average of 1,864 OTUs was obtained based on a 97% cutoff. For each sample, the number of OTUs was in the range 528–1154.

There was a large variability in OTU richness among the samples, whereas the variability characteristics of the soil and sediment samples differed. Alpha diversities were relatively lower for SO than SE samples (Fig. 2). Whereas, for soil samples, SO1 with the highest As concentration had the lowest alpha diversity. However, alpha diversity was more variable among SO than SE samples (Fig. 2).

Taxonomic analyses placed all sequences obtained in this study into 10 phyla, which were divided into 59 genera. Proteobacteria, Bacteroidetes, Gemmatimonadetes, Acidobacteria, Actinobacteria and Firmicutes were the most abundant phyla, accounting for 88% of bacterial community composition (Fig. 3a). Proteobacteria, accounting for 40.4%–54.1% of the total, was the most dominant phylum, similar to findings in an As-contaminated soil in Lahore, Pakistan, and a high-As aquifer in Araihaizer, Bangladesh (Legg et al. 2012; Sheik et al. 2012). At SO1, Chloroflexi was high, accounting for 6.0% of the total, compared to 0.5%–3.0% in other samples,

280 whereas Chloroflexi was approximately 11% of the community in a high-As aquifer
281 (Legg et al. 2012). Moreover, Acidobacteria (< 1%) were rare in As- and
282 chromium-contaminated soil (Sheik et al. 2012), which was different to the present
283 study (average of 6.7%). At the genus level, *Sulfuricurvum*, *Clostridium*, *Thiobacillus*,
284 *Luteolibacter* and *Flavobacterium* were the five most abundant genera (Fig. 3b).
285 *Clostridium* was the dominant genus at SO1, concurrent with the highest
286 concentration of As. SE1 and SE5 shared similar abundant genera of *Sulfuricurvum*,
287 *Clostridium*, *Gillisia* and *Sulfurimonas*. SO1 and SO5 also shared similar abundant
288 genera of *Sulfuricurvum*, *Clostridium*, *Flavobacterium*, *Gemmatimonas*, *Sulfurimonas*
289 and *Geobacter*.

290 Several genera with the ability to reduce As(V) to As(III) were identified in this
291 area by sequencing and taxonomic analyses (Fig. 3b). For example, *Clostridium* with
292 Arr operon was the dominant genus, capable of reducing arsenate to As(III) using
293 nitrate as a preferred electron (Stolz et al. 2006). *Sulfuricurvum* and *Sulfurimonas* can
294 utilize sulfur compounds and nitrate as an electron acceptor for their growth in
295 arsenic-rich neutral mine drainage tailings and As-contaminated lakes (Kodama and
296 Watanabe 2004; Majzlan et al. 2011). *Flavobacterium* was also isolated as an
297 As-resistant genus from As-contaminated soil and groundwater, and shown to rapidly
298 reduce As(V) under aerobic conditions (Wang and Mulligan 2006). *Thiobacillus*
299 *ferrooxidans* with *arsB*, *arsC*, *arsH*, and a putative *arsR* gene, were identified to resist
300 As contamination by nitrate-dependent oxidation of Fe(II) and U(IV) (Beller et al.
301 2006; Butcher et al. 2000).

There are also genera capable of oxidizing As(III) to As(V), but their abundance and diversity are lower than for As(V)-reducing bacteria. For example, *Gemmatimonas* in a chlorate-reducing microbial consortium oxidized As(III) in an anaerobic sludge bed (Sun et al. 2011). *Hydrogenophaga*, a sulfur-oxidizing bacterium, directly or indirectly transformed As(III) and thioarsenates to As(V) during growth (Fisher et al. 2008). *Arthrobacter* oxidizes As(III) to As(V) using a membrane-bound As(III) oxidase (Prasad et al. 2009).

Diversity, phylogenetic and spatial analyses of functional genes

ArsB

After quality filtering and removal of chimeras, 386 *arsB* gene sequences were translated into amino acid sequences, and were identified as 78 phylogenetic types (Fig. 2). Sequences of samples of SO1, SO4, SE1 and SE4 represented 16, 9, 28 and 25 phylogenetic types, respectively. These results implied that ArsB phylogenetic types were more abundant in the river sediment samples than in farmland soil samples.

Phylogenetic analysis based on translated amino acid sequences revealed that all phylogenetic types of ArsB were clustered into six clades (Fig. 4). Clade I consisted of 25 phylogenetic types: 21 sequences from river sediment and only four from soil samples. Their similar sequences in the NCBI database were arsenical pump membrane proteins of Betaproteobacteria such as *Massilia alkalitolerans*, *Janthinobacterium* and *Herminiimonas*. Clade II comprised 20 phylogenetic types, with river sediment and farm soil samples contributing 10 sequences each. The closest relatives of this clade were arsenical pump membrane proteins of *Lysobacter*

325 *arseniciresistens*, arylsulfatase of Proteobacteria and As(III) trans-membrane
326 pump-like proteins of *Acidithiobacillus caldus* (Tuffin et al. 2005). Clade III
327 contained 13 phylogenetic types: 10 from sediment and three from farm samples.
328 Their closest relatives in the NCBI database were As(III) efflux pump ArsB of
329 bacterium enrichment culture clone NMarsB_2 from wasteland soil and *Pseudomonas*.
330 Clade IV consisted of seven phylogenetic types, all from river sediment samples, and
331 all had a close relationship with the arsenical pump membrane proteins of *Serratia*
332 *liquefaciens*, *Enterobacter cloacae* and *Klebsiella* of the Gammaproteobacteria. Clade
333 V had only three phylogenetic types, all from sample SE4, which have a close
334 relationship with the arsenical pump membrane proteins of *Sphingomonas*
335 *parapaucimobilis*. Clade VI contained six phylogenetic types, all from sample site
336 SO4, which were close to *Halomonas zhanjiangensis*, as determined by BLAST
337 analysis.

338 **Acr3**

339 After quality filtering and removal of chimeras, 333 sequences of *acr3* genes were
340 obtained in this study, which were translated into amino acid sequences, and resulted
341 in 124 phylogenetic types (Fig. 5). Sequences of samples of SO1, SO4, SE1 and SE4
342 represented 55, 25, 29 and 15 phylogenetic types, respectively. The phylogenetic tree
343 only included type ACR3(2), which were more abundant in farm soil than river
344 sediment samples, contradictory to the distribution of the ArsB phylogenetic tree in
345 two types samples.

346 Phylogenetic analysis of translated amino acid sequences grouped all ACR types
347 into five clades (Fig. 3). Clade I consisted of 38 ACR phylogenetic types, which were
348 closely related to sequences in the NCBI database of As transporters of *Leptothrix*

349 *cholodnii* SP-6, *Azohydromonas australica*, *Pseudomonas aeruginosa*, *Rubrivivax*
350 *gelatinosus* and *Thiobacillus denitrificans*. Clade II contained 26 phylogenetic types
351 of ACR: 17 from farm soil and the other nine from river sediment samples. These
352 sequences shared identities with As transporters of *Burkholderia oklahomensis*,
353 *Methyloglobulus morosus* and *Nevskia soli*. Clade III comprised 33 phylogenetic
354 types: 29 sequences (85.7%) from farm soil and only four (14.3%) from river
355 sediment samples. The closest relatives of sequences in clade III were As transporters
356 of *Variovorax paradoxus*, *Rhodanobacter fulvus* and *Sphingomonas melonis*, and
357 As(III) efflux pump ACR3 of *Cupriavidus* sp. HMR-1. Clade IV only had two
358 sequences; these were not closely related and were probably new sequences. Clade V
359 contained 25 phylogenetic types, which were closely related to As transporters of
360 *Acidovorax* sp. NO-1, *Sphingobium* sp. *ba1* and other uncultured bacteria. This clade
361 was closely related to sequences from phyla Bacteroidetes, Firmicutes and
362 Verrucomicrobia – displaying a different phylogenetic relationship to the other clades,
363 which had close relationships with As transporter proteins belonging to phylum
364 Proteobacteria.

365 **Abundance of functional genes *arsB* and *ACR3*, and 16S rRNA genes**

366 The abundances of 16S rRNA genes in river sediment samples were 2.70×10^9 to
367 4.64×10^9 copies per μg DNA, and were slightly higher than in the soil samples. The
368 abundances of *arsB* and *ACR3* genes in the river sediment samples were 10^6 copies
369 per μg DNA, with no obvious differences among all samples; and their abundance in
370 soil samples fluctuated within the ranges of 10^3 – 10^6 and 10^4 – 10^6 copies per μg DNA,
371 respectively. The abundances of *arsB* and *ACR3* genes in SO1 (with the highest As

concentration) were nearly 4- and 20-fold those for other soil samples, respectively. The abundances of 16S rRNA, *arsB* and *ACR3* genes were positively correlated with As concentration ($P = 0.01$, $P < 0.01$ and $P = 0.01$, respectively) (Fig. 6). Interestingly, S concentration was negatively correlated with the abundance of *arsB* ($P < 0.05$), but showed no obvious correlation with abundances of 16S rRNA and *ACR3* genes. Among the three genes, the abundance of 16S rRNA gene was positively correlated with the abundance of *ACR3* (Pearson rank correlation test: $r = 0.45-0.59$, $P < 0.01$) (Fig. 6).

PCoA and CCA

PCoA was performed, based on normalized taxonomic composition, to elucidate similarities in the bacterial communities between the samples. PC1 and PC2 explained 58.7% and 21.4% of the total variance in bacterial community composition, respectively (Fig. 7). SO1, SO4, SE1 and SE4 were separated from other samples along PC1, and the other samples except SO5 were respectively clustered.

CCA was performed to explore the relationship between environmental factors and bacterial community composition (Fig. 8a). CCA1 distinguished the respective bacterial communities of sample SO1 and samples SE1, SE4 and SO4 from other samples, whereas CCA2 distinguished bacterial communities of samples SE1, SE4 and SO4 from the other seven samples, and demonstrated that TOC was correlated with the communities of samples SE1, SE4 and SO4. SO1 was separated from the cluster of soil samples, mainly correlated with high As concentration, whereas the

393 communities of SO₅ were correlated with Fe and Mn concentrations.

394 The CCA of some particular bacterial taxa and the geochemical parameters
395 obtained in this study are presented in Fig. 8b. The predominant genus, *Clostridium*,
396 was positively correlated with As, whereas *Sulfuricurvum*, *Sulfurimonas* and
397 *Gemmatimonas* were closely associated with TOC. *Geobacter*, *Halomonas* and
398 *Flavobacterium* were positively correlated with Fe and Mn concentrations, and
399 negatively correlated with TOC and pH value. Meanwhile, *Thiobacillus*,
400 *Desulfobulbus* and *Arthrobacter* were positively correlated with S concentration.

401 Discussion

402 It is well known that tailings of gold mining contaminate surrounding environments.
403 As concentration was related to the distance of sites from the tailings, indicating that
404 the river played an important role in transporting As in this area. Based on
405 high-throughput sequencing and PCR-based clone library analyses, this study
406 revealed the bacterial diversity and composition, and the diversity and phylogeny of
407 *ArsB* and *ACR3* genes, of two representative river sediment and farmland soil samples
408 around As-contaminated gold tailings at Dandong, China. The alpha diversity indices
409 and the most dominant genera were obviously affected by As concentration, and the
410 bacterial distribution may have been affected by environmental factors, including the
411 sample type.

412 Resistance to As plays an important role in the biogeochemical cycling of As in
413 nature (Oremland and Stolz 2003). In our study, various As-resistant genera were
414 found, possibly in response to long-term As contamination. In this study, *Clostridium*,

415 *Sulfuricurvum*, *Sulfurimonas* and *Thiobacillus* were the abundant genera, which are
416 capable of reducing As(V) to As(III) using nitrate as a preferred electron donor and
417 acetate, sulfide and Fe(II), among others, as electron donors (Stolz et al. 2006). We
418 suggested that the process from As(V) to As(III) was predominant, coupled with
419 denitrification (Beller et al. 2006; Butcher et al. 2000; Inagaki et al. 2003; Kodama
420 and Watanabe 2004; Majzlan et al. 2011). Nitrate functions widely as an electron
421 donor, and electron donors were various, which benefited As reduction from As(V) to
422 As(III). There are also genera capable of oxidizing As(III) to As(V) such as
423 *Gemmatimonas*, *Hydrogenophaga* and *Arthrobacter*, but their abundance and
424 diversity are lower than for As(V)-reducing bacteria (Fisher et al. 2008; Prasad et al.
425 2009; Sun et al. 2011). When the major biogeochemical process is from As(V) to
426 As(III), H_3AsO_3 is the main form in the reducing environment ($\text{pH} < 9.2$). Thus,
427 As(III) was only detected in the river water. As(III) is active and migrates easily,
428 which could cause serious health problems for nearby residents.

429 Our functional data revealed that As-resistant genes likely evolved mainly in
430 different orders of Proteobacteria. In this study area, this phylum had an important
431 role among microbes adapted to an As-contaminated environment. We suggest that
432 horizon gene transfer was more likely within the same phylum. A high diversity of
433 functional genes for As resistance was observed in different groups, e.g. bacteria in
434 the same order such as Betaproteobacteria could encode different functional types. In
435 addition, different phylogenetic types of functional genes were clustered into different
436 clades, did not show correlation with the phylogenetic relationships drawn by 16S
437 rRNA analysis, the distributions of clusters in this study were more closely related to
438 As-contaminated environments such as soils, aquifers and groundwater. The
439 abundance of OTUs of ArsB and ACR3 from farm soil and river sediment samples

clearly differed. All the above phenomena inspire curiosity concerning the core elements of the As-resistant mechanism despite the taxonomic differences.

Primers targeting *arsB* and *ACR3* were previously shown to be highly specific and enabled amplification of As(III) transporter gene sequences from a taxonomically wide range of bacteria (Achour et al., 2007). From the abundance of *arsB* and *ACR3* genes, our data demonstrate that the capability of As resistance increased with As concentration. This indicates that high fluxes of As lead to bacterial communities enriched in a membrane transport protein that pumps As(III) out of the cell. In the present study, the abundance of the *arsB* gene was more sensitively influenced by As concentration than the *ACR3* gene. It was reported that ACR3(1) and ACR3(2) appeared to be mostly associated with high As resistance since they were only identified from the high and intermediate As-contaminated sites. However, the diversity of the *acr3* gene declined with high As concentrations (Cai et al. 2009). In this study, *arsB* seems to be more stable in bestowing the capability of a microbial community to respond to As concentration. The *arsB* gene seemed more prevalent among prokaryotes than *ACR3*, and bacteria with a pathway of As(V) oxidation were dominant in this study area. However, ACR3 may have a higher affinity and velocity for extruding As(III) than ArsB and thus seems more effective (Cai et al. 2009). To better reflect the abundance of functional genes involved in microbial As resistance with concentrations of As, some suitable and specific primers for diverse *ACR3(1)* and *ACR3(2)* genes are necessary.

Among environmental parameters, As with Fe, Mn and TOC were the dominant

factors influencing clustering of samples and bacterial distribution. *Clostridium* was clearly the dominant member of the bacterial community in soil with high As. The high Fe and Mn levels in the present study may have favored bacteria that can detoxify As, since *Halomonas* and *Flavobacterium* (both capable of reducing As(V)) were positively correlated with Fe and Mn concentrations. The potential for Deltaproteobacteria to mediate As detoxification, Fe-reduction, Mn-reduction and humic substance-reduction pathways could explain their higher abundance (Legg et al. 2012). Furthermore, some bacteria were respectively correlated with TOC concentrations. TOC is a main nutrient source, and may influence observable community shifts (Fierer et al. 2007). The types of TOC involved in different metabolisms of bacteria may be important for the electron shuttling that promotes As mobilization (Fendorf et al. 2010). Normally, nitrogen can affect the bacterial community, as most As-resistant bacteria are involved in denitrification, but this was not apparent in our results. We suggest that nitrogen was not a limiting factor for bacterial growth in our study site. Bacterial community structure was significantly related to location, sediment grain size and sediment carbon in an As-contaminated aquifer (Legg et al. 2012). Differing properties of dominant genera and clustering relationships of samples suggested that As may be the factor with the most effect on shaping the dominant genus; nevertheless, other environmental components are likely to be co-metabolized with As, such as Fe, Mn and S, which also play important roles in shaping the bacterial communities. Reduced As may be released to solution by the reductive dissolution of host Fe and the solubility of sulfide phases, and then the ratio

of reactive Fe to S controlled the transformation of As, which is influenced by microbial processes (O'Day et al. 2004). The concentration and phase of Fe, Mn and S could have important impacts on the As phase and valence state through the bacteria associated with Fe, Mn and S cycles. Thus, the phase and ratio of Fe, Mn to S should be considered in As remediation of this area.

Acknowledgements

This research was supported by The Fundamental Research Funds for the Central Universities (2652015166, granted to XYG) and National Natural Science Foundation of China (41172226, granted to XML).

References

- Achour, A.R., Bauda, P., and Billard, P. 2007. Diversity of arsenite transporter genes from arsenic-resistant soil bacteria. *Res Microbiol* 158(2): 128-137.
- Bartram, A.K., Lynch, M.D., Stearns, J.C., Moreno-Hagelsieb, G., and Neufeld, J.D. 2011. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Applied and environmental microbiology* 77(11): 3846-3852.
- Beller, H.R., Chain, P.S., Letain, T.E., Chakicherla, A., Larimer, F.W., Richardson, P.M., Coleman, M.A., Wood, A.P., and Kelly, D.P. 2006. The genome sequence of the obligately chemolithoautotrophic, facultatively anaerobic bacterium *Thiobacillus denitrificans*. *J Bacteriol* 188(4): 1473-1488.

-
- 504 Butcher, B.G., Deane, S.M., and Rawlings, D.E. 2000. The chromosomal arsenic
505 resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer
506 increased arsenic and antimony resistance to *Escherichia coli*. *Applied and*
507 *Environmental Microbiology* 66(5): 1826-1833.
- 508 Cai, L., Liu, G.H., Rensing, C., and Wang, G.J. 2009. Genes involved in arsenic
509 transformation and resistance associated with different levels of arsenic-contaminated
510 soils. *Bmc Microbiol* 9.
- 511 Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello,
512 E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T.,
513 Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D.,
514 Pirrung, M., Reeder, J., Sevinsky, J.R., Tumbaugh, P.J., Walters, W.A., Widmann, J.,
515 Yatsunenko, T., Zaneveld, J., and Knight, R. 2010. QIIME allows analysis of
516 high-throughput community sequencing data. *Nat Methods* 7(5): 335-336.
- 517 Carrizales, L., Razo, I., Tellez-Hernandez, J.I., Torres-Nerio, R., Torres, A., Batres,
518 L.E., Cubillas, A.C., and Diaz-Barriga, F. 2006. Exposure to arsenic and lead of
519 children living near a copper-smelter in San Luis Potosi, Mexico: Importance of soil
520 contamination for exposure of children. *Environmental research* 101(1): 1-10.
- 521 Cole, J.R., Wang, Q., Fish, J.A., Chai, B.L., McGarrell, D.M., Sun, Y.N., Brown, C.T.,
522 Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. 2014. Ribosomal Database Project:
523 data and tools for high throughput rRNA analysis. *Nucleic Acids Research* 42(1):
524 633-642.
- 525 Fendorf, S., Michael, H.A., and van Geen, A. 2010. Spatial and Temporal Variations

- 526 of Groundwater Arsenic in South and Southeast Asia. *Science* 328(5982): 1123-1127.
- 527 Fierer, N., Bradford, M.A., and Jackson, R.B. 2007. Toward an ecological
528 classification of soil bacteria. *Ecology* 88(6): 1354-1364.
- 529 Fisher, J.C., Wallschlager, D., Planer-Friedrich, B., and Hollibaugh, J.T. 2008. A new
530 role for sulfur in arsenic cycling. *Environ Sci Technol* 42(1): 81-85.
- 531 Giloteaux, L., Holmes, D.E., Williams, K.H., Wrighton, K.C., Wilkins, M.J.,
532 Montgomery, A.P., Smith, J.A., Orellana, R., Thompson, C.A., Roper, T.J., Long, P.E.,
533 and Lovley, D.R. 2013. Characterization and transcription of arsenic respiration and
534 resistance genes during in situ uranium bioremediation. *Isme Journal* 7(2): 370-383.
- 535 Hallin, S., Jones, C.M., Schlöter, M., and Philippot, L. 2009. Relationship between
536 N-cycling communities and ecosystem functioning in a 50-year-old fertilization
537 experiment. *Isme Journal* 3(5): 597-605.
- 538 Inagaki, F., Takai, K., Kobayashi, H., Nealson, K.H., and Horikoshi, K. 2003.
539 *Sulfurimonas autotrophica* gen. nov., sp. nov., a novel sulfur-oxidizing
540 epsilon-proteobacterium isolated from hydrothermal sediments in the Mid-Okinawa
541 Trough. *Int J Syst Evol Microbiol* 53(6): 1801-1805.
- 542 Islam, K.R., and Weil, R.R. 1998. A rapid microwave digestion method for
543 colorimetric measurement of soil organic carbon. *Commun. Soil Sci. Plant Anal.* 29:
544 2269-2284.
- 545 Keller, N.S., Stefansson, A., and Sigfusson, B. 2014. Arsenic speciation in natural
546 sulfidic geothermal waters. *Geochim Cosmochim Acta* 142: 15-26.
- 547 Kodama, Y., and Watanabe, K. 2004. *Sulfuricurvum kujiense* gen. nov., sp. nov., a

- 548 facultatively anaerobic, chemolithoautotrophic, sulfur-oxidizing bacterium isolated
549 from an underground crude-oil storage cavity. *Int J Syst Evol Microbiol* 54(Pt 6):
550 2297-2300.
- 551 Legg, T.M., Zheng, Y., Simone, B., Radloff, K.A., Mladenov, N., Gonzalez, A.,
552 Knights, D., Siu, H.C., Rahman, M.M., Ahmed, K.M., McKnight, D.M., and
553 Nemergut, D.R. 2012. Carbon, metals, and grain size correlate with bacterial
554 community structure in sediments of a high arsenic aquifer. *Front Microbiol* 3: 82.
- 555 Luo, X.M., Zhao, L.Q., and Xu, B.Y. 2008. Application of IC-HG-AFS Determination
556 Method for Arsenic in Gold Mine Tailings. *Spectroscopy and Spectral Analysis*
557 28(1): 206-208.
- 558 Magoc, T., and Salzberg, S.L. 2011. FLASH: fast length adjustment of short reads to
559 improve genome assemblies. *Bioinformatics* 27(21): 2957-2963.
- 560 Majzlan, J., Lalinska, B., Chovan, M., Blass, U., Brecht, B., Gottlicher, J., Steininger,
561 R., Hug, K., Ziegler, S., and Gescher, J. 2011. A mineralogical, geochemical, and
562 microbiological assessment of the antimony- and arsenic-rich neutral mine drainage
563 tailings near Pezinok, Slovakia. *Am Mineral* 96(1): 1-13.
- 564 Matschullat, J. 2000. Arsenic in the geosphere--a review. *The Science of the total*
565 *environment* 249(1-3): 297-312.
- 566 Niemeyer, J.C., Lolata, G.B., de Carvalho, G.M., Da Silva, E.M., Sousa, J.P., and
567 Nogueira, M.A. 2012. Microbial indicators of soil health as tools for ecological risk
568 assessment of a metal contaminated site in Brazil. *Appl Soil Ecol* 59: 96-105.
- 569 O'Day, P.A., Vlassopoulos, D., Root, R., and Rivera, N. 2004. The influence of sulfur

570 and iron on dissolved arsenic concentrations in the shallow subsurface under changing
571 redox conditions. *Proc Natl Acad Sci U S A* 101(38): 13703-13708.

572 Oremland, R.S., and Stolz, J.F. 2003. The ecology of arsenic. *Science* 300(5621):
573 939-944.

574 Poirer, J., Joulain, C., Leyval, C., and Billard, P. 2013. Arsenite-induced changes in
575 abundance and expression of arsenite transporter and arsenite oxidase genes of a soil
576 microbial community. *Res Microbiol* 164(5): 457-465.

577 Prasad, K.S., Subramanian, V., and Paul, J. 2009. Purification and characterization of
578 arsenite oxidase from *Arthrobacter* sp. *Biometals : an international journal on the role*
579 *of metal ions in biology, biochemistry, and medicine* 22(5): 711-721.

580 Ritalahti, K.M., Amos, B.K., Sung, Y., Wu, Q., Koenigsberg, S.S., and Löffler, F.E.
581 2006. Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes
582 simultaneously monitors multiple Dehalococcoides strains. *Appl Environ Microbiol*
583 72(4): 2765-2774.

584 Rosen, B.P. 2002. Biochemistry of arsenic detoxification. *Febs Lett* 529(1): 86-92.

585 Sheik, C.S., Mitchell, T.W., Rizvi, F.Z., Rehman, Y., Faisal, M., Hasnain, S.,
586 McInerney, M.J., and Krumholz, L.R. 2012. Exposure of soil microbial communities
587 to chromium and arsenic alters their diversity and structure. *PLoS One* 7(6): e40059.

588 Stolz, J.F., Basu, P., Santini, J.M., and Oremland, R.S. 2006. Arsenic and selenium in
589 microbial metabolism. *Annu Rev Microbiol* 60: 107-130.

590 Sun, W.J., Sierra-Alvarez, R., and Field, J.A. 2011. Long term performance of an
591 arsenite-oxidizing-chlorate-reducing microbial consortium in an upflow anaerobic

- 592 sludge bed (UASB) bioreactor. *Bioresource Technology* 102(8): 5010-5016.
- 593 Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6:
594 molecular evolutionary genetics analysis version 6.0. *Molecular biology and*
595 *evolution* 30(12): 2725-2729.
- 596 Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997.
597 The CLUSTAL_X Windows Interface: Flexible Strategies for Multiple Sequence
598 Alignment Aided by Quality Analysis Tools. *Nucleic Acids Research* 25(24):
599 4876-4882.
- 600 Tuffin, I.M., de Groot, P., Deane, S.M., and Rawlings, D.E. 2005. An unusual
601 Tn21-like transposon containing an ars operon is present in highly arsenic-resistant
602 strains of the biomining bacterium *Acidithiobacillus caldus*. *Microbiology* 151(9):
603 3027-3039.
- 604 Valverde, A., Gonzalez-Tirante, M., Medina-Sierra, M., Santa-Regina, I.,
605 Garcia-Sanchez, A., and Igual, J.M. 2011. Diversity and community structure of
606 culturable arsenic-resistant bacteria across a soil arsenic gradient at an abandoned
607 tungsten-tin mining area. *Chemosphere* 85(1): 129-134.
- 608 Wang, S., and Mulligan, C.N. 2006. Natural attenuation processes for remediation of
609 arsenic contaminated soils and groundwater. *J Hazard Mater* 138(3): 459-470.
- 610 Wei, C.J., and Liu, J.X. 2007. A new hydride generation system applied in
611 determination of arsenic species with ion chromatography-hydride generation-atomic
612 fluorescence spectrometry (IC-HG-AFS). *Talanta* 73(3): 540-545.
- 613
- 614

Figure Legends

Figure 1 Sampling sites.

SE1-5 samples (green dots) are river sediment samples according to the distance from the tailing;
SO1-5 samples (yellow dots) are corn farmland soil samples, located next to the river sampling
sites.

Figure 2 Alpha diversity of bacterial communities of 10 samples and functional genes of SE1&4 and SO 1&4.

Figure 3 Relative abundance of the identified bacterial phyla (a) and genera (b) in the 10 samples.

SE1-5, surface sediment samples and SO1-5, soil samples

Figure 4 Phylogenetic tree of the ArsB sequences derived from SE1&4 and SO 1&4 in this study as well as from cultured bacterial and other typical environmental clones.

Bootstrap values of >50% (from 1000 replicates), scale bar representing 5% sequence divergence,
and Clades (I-VI) are indicated. SE1 colored blue; SE4 colored green; SO1 colored brown; SO4
colored red. The genus followed by α , β , γ -Proteobacteria meant it belong to these Classes,
respectively.

Figure 5 Phylogenetic tree of the Acr3 sequences derived from SE1&4 and SO 1&4 in this study as well as from cultured bacterial and other typical environmental clones.

Bootstrap values of >50% (from 1000 replicates), scale bar representing 5% sequence divergence,
and Clades (I-V) are indicated. SE1 colored blue; SE4 colored green; SO1 colored brown; SO4

colored red. The genus followed by α , β , γ -Proteobacteria meant it belong to these Classes, and the genus followed by Bacteroidetes, Firmicutes, and Verrucomicrobia meant it belong to this Phyla, respectively.

Figure 6 Principal coordinate analysis (PCoA) of various soil samples based on the composition and abundance of their bacterial communities.

The first two PC axes (PC1 and PC2) explained 58.7% and 21.4% of the variance in the bacterial communities 'composition and abundance based on 16S rRNA gene sequence analysis.

Figure 7 Canonical correspondence analysis (CCA) based on normalized environmental factors and bacterial communities (a) and dominant genera (b).

The abundance of bacterial taxa was based on the taxonomic abundance, as determined using 16S rRNA gene sequences analysis. The first two CCA axes explained 69.3% of the total variance in the bacterial composition-environmental factors relationship (CCA1 48.9% and CCA2 20.4%) and 71.3% of the cumulative variance of the dominant genera-environmental factors relationship (CCA1 51.8.9% and CCA2 19.5%).

Figure 8 Correlation between concentration of As and the abundance of genes 16S rRNA, *arsB* and *acr3* in 10 samples.

Pearson index was used to evaluate the level of correlation (** means $P < 0.001$ and * means $P < 0.05$).

123°39' E 40°44' N

123°40' E 40°44' N

123°39' E 40°42' N

123°40' E 40°42' N

PCoA - PC1 vs PC2

Table 1 Geochemical characteristics of 10 samples in the study.

Samples	TN (%)	TS (%)	P (%)	TOC (%)	As (mg/kg)	Fe (mg/kg)	Mn (mg/kg)	Cu (mg/kg)	Zn (mg/kg)	Cd (mg/kg)	Cr (mg/kg)	Pb (mg/kg)	Hg (mg/kg)	pH	Mean grain size (µm)
SE1	10.54	15.00	0.09	2.71	787.00	344.43	527.00	94.50	160.00	0.60	106.00	104	0.092	8.12	59.38
SE2	6.56	1.10	0.05	0.75	949.00	222.30	379.00	16.90	64.20	0.20	42.80	99.8	0.011	7.96	306.66
SE3	13.36	2.00	0.07	1.50	623.00	290.16	492.00	21.70	87.10	0.27	64.30	124	0.025	8.13	92.36
SE4	18.08	4.00	0.08	2.10	272.00	346.86	540.00	33.00	95.60	0.30	81.70	76.8	0.036	8.40	47.38
SE5	14.46	2.00	0.06	1.40	182.00	365.31	709.00	28.80	79.60	0.23	74.00	70.9	0.083	7.75	47.00
SO1	9.97	3.60	0.07	1.40	1023.00	344.43	587.00	159.00	169.00	0.71	92.90	97	0.084	8.10	124.08
SO2	9.29	20.00	0.06	1.02	23.50	345.33	582.00	54.90	229.00	0.87	85.50	148	0.168	8.05	44.41
SO3	14.76	30.00	0.07	1.82	33.00	344.43	450.00	47.10	198.00	0.74	85.10	126	0.119	8.08	46.30
SO4	11.71	11.00	0.05	1.17	140.00	354.33	469.00	40.60	136.00	0.46	87.70	88.4	0.075	8.20	73.51
SO5	11.82	6.30	0.08	1.20	35.40	577.71	1460.00	51.00	116.00	0.44	78.60	76.2	0.114	8.00	37.93

SE means river sediment sample, SO means corn farmland sample.