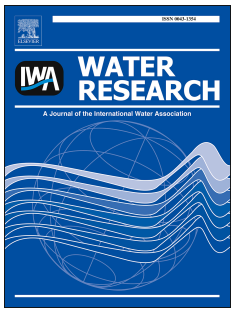
Accepted Manuscript



Analysis of the functional gene structure and metabolic potential of microbial community in high arsenic groundwater

Ping Li, Zhou Jiang, Yanhong Wang, Ye Deng, Joy D. Van Nostrand, Tong Yuan, Han Liu, Dazhun Wei, Jizhong Zhou

PII: S0043-1354(17)30530-4

DOI: 10.1016/j.watres.2017.06.053

Reference: WR 13008

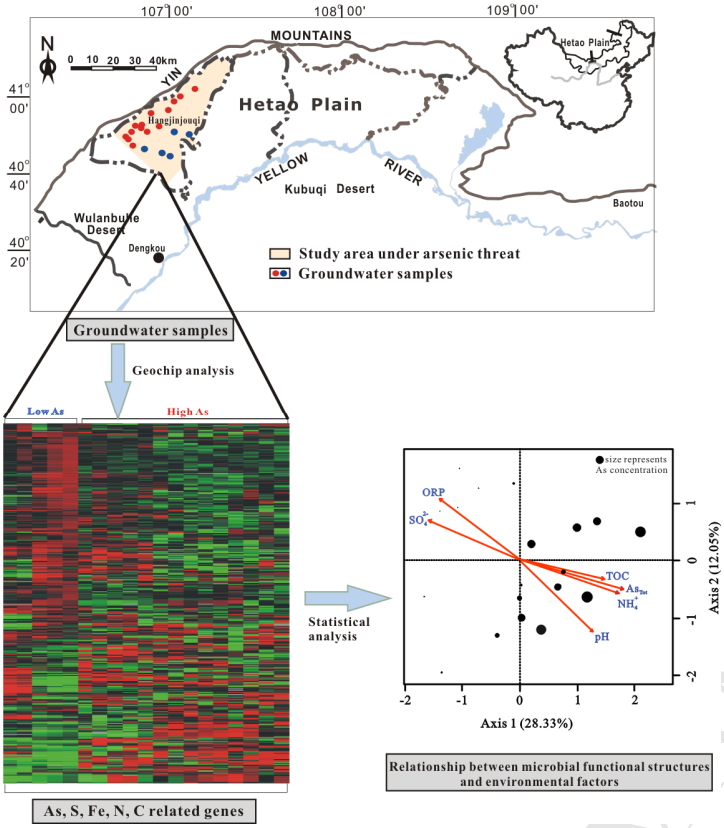
To appear in: *Water Research*

Received Date: 5 February 2017

Revised Date: 18 June 2017

Accepted Date: 19 June 2017

Please cite this article as: Li, P., Jiang, Z., Wang, Y., Deng, Y., Van Nostrand, J.D., Yuan, T., Liu, H., Wei, D., Zhou, J., Analysis of the functional gene structure and metabolic potential of microbial community in high arsenic groundwater, *Water Research* (2017), doi: 10.1016/j.watres.2017.06.053.



Analysis of the Functional Gene Structure and Metabolic Potential of Microbial Community in High Arsenic Groundwater

Ping Li[1*], Zhou Jiang[1, 2], YanhongWang[1], Ye Deng[3], Joy D. Van Nostrand[4], Tong Yuan[4], Han Liu[1], Dazhun Wei[1], Jizhong Zhou[4]

1. State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan 430074, PRC
2. School of Environmental Studies, China University of Geosciences, Wuhan 430074, PRC
3. CAS Key Laboratory of Environmental Biotechnology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences (CAS), Beijing, China
4. Institute for Environmental Genomics and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, 73019, USA.

*Corresponding Author: Ping Li
State Key Laboratory of Biogeology and Environmental Geology,
China University of Geosciences,
Wuhan 430074,
P. R. China
Tel: +86-27-67883606
Fax: +86-27-67883456
E-mail: pli@cug.edu.cn (P. Li)

Abstract

Microbial functional potential in high arsenic (As) groundwater ecosystems remains largely unknown. In this study, the microbial community functional composition of nineteen groundwater samples was investigated using a functional gene array (GeoChip 5.0). Samples were divided into low and high arsenic groups based on geochemical parameters and microbial functional structure in hierarchical clustering analysis. The results showed that arsenic related genes (*arsC*, *arrA*), sulfate related genes (*dsrA* and *dsrB*), nitrogen cycling related genes (*ureC*, *amoA*, and *hzo*) and methanogen genes (*mcrA*, *hdrB*) were highly correlated with arsenic, SO_4^{2-}, NH_4^+ or CH_4 concentrations in groundwater, respectively. Canonical correspondence analysis (CCA) results indicated that some geochemical parameters including As, total organic content, SO_4^{2-}, NH_4^+, oxidation-reduction potential (ORP) and pH were important factors shaping the functional microbial community structure. Alkaline and reducing conditions with relatively low SO_4^{2-}, ORP, and high NH_4^+, as well as SO_4^{2-} and Fe reduction and ammonification involved in microbially-mediated geochemical processes could be associated with arsenic enrichment in groundwater. This study provides an overall picture of functional microbial communities in high arsenic groundwater aquifers, and also provides insights into the critical role of microorganisms in As biogeochemical cycling.

Keywords: Arsenic, Groundwater, Functional genes, GeoChip5.0, Inner Mongolia

1. Introduction

Arsenic (As) in groundwater is a serious environmental issue due to its widespread distribution and high toxicity, which threatens the health of millions of people in many countries such as Bangladesh, West Bengal, China, Cambodia, Japan, Argentina, Chile and USA (Pi et al. 2017; Schaefer et al., 2016; Michael, 2013; Rodríguez-Lado et al., 2013; Nordstrom 2002, Fendorf et al., 2010; Chakraborti et al., 2010). Long term ingestion of arsenic groundwater can result in arseniasis, which causes many chronic diseases including cardiovascular, respiratory diseases, and skin, lung, liver and kidney cancers (Chen, 2014; Chung et al., 2013). Over the last few decades, numerous studies focusing on hydrology, mineralogy, and geochemistry have been undertaken to detect As mobilization and determine transformation mechanisms in high arsenic groundwater aquifers. The explanation of As generation in groundwater aquifers is not trivial due to the complex sets of hydrogeological conditions and biogeochemical processes within the aquifers. The absorbed and sequestered arsenic on Fe

oxides/hydroxides is one of the most common As reservoirs in some sedimentary basins. Reductive dissolution of the Fe oxide minerals in sediments, as well as reduction of As(V) to highly mobile As(III) leads to the release of bound As into the groundwater (Guo et al., 2011; Kocar et al., 2010; Deng et al., 2009). Generation of HS^- by sulfate reduction was proposed to promote As mobilization through HS^- abiotic reductive dissolution of the Fe oxide minerals and formation of As-sulfur compounds in high HS^- concentration water (Wang et al., 2014; Stauder et al., 2005). In addition, anthropogenic activities such as agricultural irrigation which releases chemicals including organic matter, nitrogen, and other chemicals into groundwater systems, were found with important roles on arsenic release in groundwater (Weng et al., 2017; Gao et al., 2014; Neidhardt et al., 2012; Itai et al., 2008; Bose et al., 2002). Previous studies showed that As mobilization and transformation could be ascribed to the complex interactions between microbes and geochemical processes and that microbes are likely to play key roles in driving the biogeochemical cycle in high arsenic groundwater aquifers (Chen et al., 2017; Zhang et al., 2015; Ghosh et al., 2013; Gorra et al., 2012; Mumford et al., 2012; Kocar et al., 2010; Barringer et al., 2010; Sutton et al., 2009).

Understanding microbial community structures and their associations with geochemical processes is one of the central topics in microbial ecology. Previous studies investigated microbially-mediated arsenic transform and release mechanisms using pure microbial cultures under laboratorial conditions (Flynn et al., 2014; Sutton et al., 2009; Saalfield and Bostick 2009). Recently, with the methods of sequencing technologies and qPCR, some studies explored the diversity and structure of in situ microbial communities and functional genes of arsenic reducing and oxidizing bacteria, sulfate reducing bacteria and methanogens in high As groundwater aquifers including Bangladesh, Cache Valley Basin Utah, Jianghan and Hetao Plain of China (Ye et al., 2017; Chen et al., 2017; Mirza et al., 2017; Wang et al., 2016; Wang et al., 2015; Hassan et al., 2015; Li et al., 2015; Li et al., 2014; Sultana et al., 2011). These previous studies provided useful but limited information on the mechanisms of microbially-mediated arsenic transformation and mobilization in groundwater aquifers. It is still not fully understood which functional microbial populations mediate arsenic geochemical processes *in situ*. However, microbial communities in groundwater aquifers contaminated with high levels of arsenic are complex and largely unknown. Therefore, there is an urgent need to more fully investigate the functional potential of microbial communities in these As contaminated aquifers. In this study, we used GeoChip 5.0, a powerful and high-throughput technology which targets more than 1,400 functional gene families, and involves in geochemical cycling (N, C, S, and P), metal homeostasis, and organic

contaminant degradation genes (Van Nostrand et al., 2016) to investigate the functional potential microbial communities in high As groundwater aquifers.

Our study area is located in Hetao Plain, an arid region with annual precipitation 140-180 mm and evaporation from 2,000 to 2,500 mm. About half of the soils are saline, due to strong evapotranspiration (Deng et al., 2009). It has been one of the largest irrigation districts using delivered Yellow River water. The groundwater for drinking is taken from the late Pleistocene and Holocene layer in a complex organic-rich reductive environment which was detected with many gases such as H_2S and CH_4 (Li et al., 2014; Tang et al., 2017). The water chemical types were mostly Na(–Mg)–Cl(–HCO_3) and Na–Mg–HCO_3. Most high-As groundwater in the region occurs at the depth 20-30 m. Groundwater from this area contains As concentrations up to 1.74 mg/L (Deng et al., 2009), which greatly exceeds the upper limit (10 μg/L) recommended by world health organization guidelines (WHO, 2011).

The primary objectives of this study were to: (1) investigate the groundwater chemistry with respect to As and other components, characterize the high arsenic groundwater geochemistry comparing to the low arsenic samples. (2) examine functional diversity and structure of the *in situ* microbial communities; (3) assess the metabolic potential of microbial communities, and (4) determine the relationships between the functional microbial community and environmental factors in arsenic-rich aquifers. To achieve these objectives, a coordinated geochemical and molecular survey was conducted for 19 groundwater samples in Hangjinhouqi County, Inner Mongolia, China. The results of this study fill a much needed knowledge gap regarding the relationship between the functional microbial community and geochemical processes in natural arsenic aquifers and expand the current understanding of microbial ecology in these aquifers.

2. Materials and Methods

2.1 Site description

The study site is located in the western part of the Hetao Basin (the Great Bend of Yellow River) Inner Mongolia, China (Figure. 1) in an area that has been seriously affected by arsenic poisoning (Wei et al., 2016; Wade et al., 2009). The Basin was formed at the end of the Jurassic Period and contains fine clastic sediments. Frequent channel changes deposited sediments and generated oxbow lakes replete with accumulated humus and organic mud. The targets of this study are shallow aquifers which are composed of late Pleistocene and Holocene alluvial and lacustrine deposits. Groundwater is

recharged by lateral flowing groundwater from bedrocks, vertical infiltrating meteoric water, and/or by irrigation return flow and leakage from the Yellow River from the south. Discharge occurs mainly via evapotranspiration and pumping. These aquifers have been widely used for drinking water by local residents. Our case study was performed in Hangjinhouqi County (HC) in the western part of the Hetao Plain. Local residents have been drinking the arsenic contaminated groundwater for over 30 years. Some of the residents have been affected with serious skin cancer which was caused by arseniasis (Wei et al., 2016; Wade et al., 2009).

2.2 Sample collection and field measurements

Nineteen groundwater samples were collected from tube wells with a depth range of 20-30 m in June 2013 from nine towns in Hangjinhouqi County (Figure 1) and named S1-19 based on arsenic concentration. The groundwater samples consisted of one from Manhui (sample S1), three from Shahai (sample S2, 5 and 19), one from Sizhi (sample S3), five from Sandaoqiao (sample S4, S6, S15, S17 and S18), one from Shanba (sample S7), four from Taiyangmiao (sample S8, 9, 10 and 12), one from Shuangmiao (sample S11), two from Erdaoqiao (sample S13 and S14), and one from Menghai (sample S16) (Figure 1). Water samples were pumped out and then filtered or acidified. The tubing was flushed with copious amounts of the sampling groundwater before each use. Water pH, electronic conductivity (EC), oxidation-reduction potential (ORP), total dissolved solids (TDS) and dissolved oxygen (DO) were measured in the field at the site of water collection using a hand-held meter (Horiba W-23X D, Japan). Concentrations of ammonium, sulfide, ferrous iron (FeII) and total iron were also determined in the field with a Hach spectrophotometer (DR850, Hach Corp., USA) according to the manufacturer's instructions. Samples used for laboratory measurements, e.g., cations, anions and dissolved organic carbon (DOC) were filtered through 0.22 μm mixed cellulose ester membranes and then the filtrates were collected into 50 mL acid-washed polypropylene bottles or brown glass bottles. Water samples for cations and DOC were acidified with 1% v/v HNO_3. AsIII (arsenite) and AsV (arsenate) species were separated with an anion exchange cartridges (Supelco, USA) (Li et al., 2013). Microbial samples were collected by on-line filtering of 5-10 L water through 0.22-μm filters (Millipore). Filtered biomass-containing membranes were placed in 50 mL sterile tubes and immediately stored in dry ice. All microbial samples (biomass-containing membranes) were immediately frozen and stored on dry ice in the field and during transportation, and then kept at -80 ^{o}C in laboratory until further analyses.

2.3 Geochemistry measurements

Anion and As measurements were performed according to Li et al. (2013). Briefly, pre-separated AsIII and AsV were determined by ion chromatography-hydride generation atomic fluorescence spectrometry (IC-HG-AFS-9700, Haiguang, China). Anions, including SO_4^{2-}, NO_3^-, NO_2^- and Cl^- were determined by ion chromatography (DX-120, Dionex, USA). Dissolved organic carton (DOC), total carbon (TC) and total nitrogen (TN) of the samples were determined using a TOC analyzer (Vario MICRO cube, Elemental, Germany), respectively. Samples for CH_4 determination were collected in 100 mL brown serum bottles without filtration, which were fully filled with no headspace and sealed with gas impermeable butyl rubber septum stoppers, secured with a 20-mm aluminum seal. Two drops of saturated mercury chloride were added to the samples to retard microbial activity. Samples were kept on blue ice during transportation and then kept in 4 ^oC in laboratory before analysis. The head space was created by purging with pure helium (99.999%) (Johnson et al., 1990). Soluble CH_4 concentrations were analyzed from the head space by gas chromatography (Thermo Trace Ultra) at the Third Institute of Oceanography, State Oceanic Administration. All samples were run in triplicate and then averaged.

2.4 DNA extraction and quantitation

In the laboratory, DNA extractions for all groundwater were conducted using FastDNA spin kits for soil (MP Bio, USA) according to the manufacturer's manual. DNA yield was quantified based on spectral absorbance at 260 nm, ratios of 260/280 nm and 260/230 nm, respectively (Thermo Scientific NanoDrop 2000), and then stored at -80 ^oC until required for PCR amplification. The final DNA concentrations were quantified by PicoGreen, using a FLUO star Optima instrument (BMG Labtech, Jena, Germany).

2.5 GeoChip analysis

The new generation of functional gene array (GeoChip 5.0) was used to analyze the functional potential of microbial samples. The purified DNA (500 ng) was labelled with Cy 3 as described previously (Sun et al., 2014). Briefly, the labeled DNA was re-suspended in hybridization solution, and then hybridized in an Agilent hybridization oven at 67 °C for 24 h. After hybridization, the slides were washed with buffers to remove unbound DNA. The arrays were scanned with a NimbleGen MS200 Microarray Scanner (Roche NimbleGen, Inc., Madison, WI, USA). The images were extracted by the Agilent Feature Extraction program. Poor quality spots with a signal-to-noise ratio of less than 2.0 were removed before statistical analysis. The positive signals were normalized within each sample and across all samples.

2.6 Statistical analysis

Pre-processed GeoChip data were further analyzed by the Vegan package in R 3.1.1 (http://www.r-project.org/) unless otherwise stated. Hierarchical clustering was performed with CLUSTER 3.0 using uncentered correlations and the complete average linkage for both genes and environmental variables, and trees were visualized in TREEVIEW (de Hoon et al., 2004). Functional gene diversity was calculated using Simpson's 1/D, Shannon-Weiner's H' and evenness. The effects of arsenic on functional microbial communities were analyzed by response ratio (RR) using the formula described by Zhang et al (2013). Based on the standard error, the 90% confident interval for each response variable was obtained and the statistical difference between the high arsenic and low arsenic groups was estimated. The total abundance of each gene category or family was simply the sum of the normalized intensity for the gene category or family. Three non-parametric multivariate analyses (analysis of similarities (ANOSIM), non-parametric multivariate analysis of variance (ADONIS) using distance matrices, and multi-response permutation procedure (MRPP)) were used to examine whether arsenic has significant effects on microbial communities. Canonical correspondence analysis (CCA) and partial CCA for co-variation analysis (variation partitioning analysis, VPA) were performed to link microbial communities to environmental variables (Low et al., 2016; Zhou et al., 2008). Selection for CCA modeling was conducted by an iterative procedure of eliminating redundant environmental variable based on variance inflation factor (VIF).

3. Results

3.1 Sample characteristics and chemical composition

The groundwater geochemical parameters were measured. Total arsenic (As_{Tot}) concentrations ranged from 38.93 to 863.42 μg/L (Table S1), and were positively correlated with NH_4^+, TOC, CH_4 and the ratio of Fe(II/III) ($r=0.917$, $p<0.001$; $r=0.785$, $p<0.001$; $r=0.807$, $p<0.001$; $r=0.582$, $p<0.01$, respectively) (Figure 2, Table S2). Negative correlations were observed between As_{Tot} and sulfate and nitrate concentrations, and ORP ($r=-0.699$, $p<0.001$; $r=-0.669$, $p<0.01$; $r=-0.734$, $p<0.01$, respectively) (Figure 2, Table S2). Hierarchical clustering analyses separated the geochemistry of the 19 samples into two groups (Figure 3). The first group contained five samples (S1-5), which were characterized by relatively low arsenic (38.93-92.17 ug/L), relatively high oxidation reduction potential (ORP), and low TOC, FeII, and NH_4^+. The remaining 14 samples were grouped together, and were characterized by relatively high arsenic concentrations (157.71-863.42 μg/L). Most samples in this group were either

neutral or slightly alkaline in pH. Some samples in the group had a strong hydrogen sulfide smell, consistent with their high sulfide concentrations and low ORP (Table 1). Samples with high arsenic concentrations generally had low concentrations of sulfate, negative ORP, and relatively high concentrations of AsIII, FeII, CH_4, NH_4^+ and TOC contents. These geochemical characteristics indicated that strong reducing conditions prevailed in the high arsenic groundwater. The results of pairwise comparisons using ANOSIM (bray-cutis) also suggested that there was a significant difference in geochemistry between these two groups of samples ($r=0.6526$, $p<0.001$) (Table 1).

3.2 Functional microbial community

The functional potential of the microbial communities was measured using a high-throughput functional gene array, GeoChip5.0. Significant differences were observed in functional microbial structure between low and high arsenic groups ($r=0.4776$, $p=0.007$) (Table 1). Consistent with geochemical ordination patterns, CCA showed a differentiation in functional microbial structure between samples with different levels of arsenic (Figure 4). Samples were mainly divided into two groups which were characterized by relatively low and high arsenic concentrations, respectively. The microbial functional structure of samples with low arsenic clustered more tightly than those of the higher arsenic samples (Figure 3). Similarly, hierarchical clustering analysis showed similar clustering of microbial community functional structure (Figure S1).

The numbers of functional gene probes detected in these 19 groundwater samples were all in the magnitude of 10^5 (Table S3). The functional gene distribution among the samples was variable (Figure S1). For example, some genes involved in nitrogen cycling, sulfate reduction and methanogenesis from groups 9, 11 and 12 had a distinctly higher abundance in high arsenic samples than samples with low arsenic. Arsenic related genes were mainly distributed in group 8, 11 and 12. Three complimentary non-parametric multivariate statistical tests including MRPP (multi response permutation procedure), ANOISM (analysis of similarity), and adonis further confirmed the significant differences of geochemical patterns and microbial communities between different sample groups (high and low arsenic groups) (Table 1).

3.3 Overall functional microbial community structure in relation to geochemistry

CCA ordination results showed that six variables, As_{Tot}, SO_4^{2-}, NH_4^+, pH, ORP, and TOC were the most significant environmental variables shaping the microbial community structure ($p=0.021$, Figure 4). The length and direction of vectors represent the influence of a given geochemical vector to samples. For microbial community samples with high arsenic, As_{Tot}, TOC and NH_4^+ were the most

significantly correlated environmental factors, while ORP and SO_4^{2-} were correlated with low arsenic samples. To better understand the influence of the relative contributions from environmental variables to the microbial community structure, a CCA-based variation partitioning analysis was performed (Figure S2)(Zhou et al., 2008). Six environmental variables were divided into three group, including group 1 (SO_4^{2-}, ORP), group 2 (As_{Tot}, TOC and NH_4^+) and group three (pH). Both groups 1 and 2 showed significant correlation with the functional gene structure of the community. Group 1 (SO_4^{2-}, ORP) explained 22.63% (p=0.010), and group 2 (As_{Tot}, TOC and NH_4^+) explained 24.98% (p=0.010), the pH environmental variable (group 3) independently explained 17.64% (p=0.005) of the observed variation (Figure S2). About 58.80% of the community functional variation remained unexplained by the above selected variables.

3.4 Functional genes for arsenic cycling

Arsenic related genes including arsenic-resistance (*arsC*/*arsB*), dissimilatory arsenate-reducing (*arrA*), arsenic-oxidation (*aoxB*) and arsenic methylation genes (*arsM*) were detected in these samples. Compared with the other arsenic related genes, arsenic-resistance *arsC*/*arsB* genes were distinctly more diverse. A total of 855 arsenic-resistance *arsC*/*arsB* gene sequences derived from 418 bacterial genera were detected. Of these, 438 genes were shared among all the samples. *ars*C genes with percentages greater than 2% of the total intensity of *ars*C genes were from *Rhodanobacter*, *Rhodococcus*, *Stenotrophomonas*, *Aurantimonas*, *Nitrobacter*, Xantho*bacter*, *Agrobacterium*, *Rhodococcus*, *Acinetobacter*, *Thioalkalivibrio*, and *Mycobacterium* (Table S4). *arsB* genes with percentages higher than 2% of the total intensity of *arsB* genes were from *Methylobacterium*, *Micromonospora*, *Ethanoligenens*, *Acidovorax*, *Methylobacterium*, and *Oceanimonas*. Thirty-nine dissimilatory arsenate-reducing *arrA* genes were detected mainly from *Azoarcus*, *Aromatoleum*, *Chlorobium*, *Thauera*, *Geobacter*, *Alkaliphilus*, *Desulfonatronospira*, *Desulfitobacterium* and some uncultured bacteria (Table S4). Of these, 15 genes were shared among all the samples. Thirty-nine arsenic methylation genes *arsM* were detected and 15 genes were shared among all the samples. A majority (81%) of the *arsM* probes were derived from *Conexibacter*, *Desulfohalobium*, *Thiocapsa*, *Nitrosomonas*, *Cupriavidus*, *Leptonema*, *Halomonas* and *Pelotomaculum*. The 96 arsenic-oxidization genes, *aoxB*, detected were mainly from *Vibrio*, *Acidiphilium*, *Halomonas*, *Variovorax* and *Burkholderia*. Of these, 50 were shared among all the samples. Multiple arsenic-related genes from the same dissimilatory arsenate-reducing microorganisms were detected. For example, *arrA* and *arsCBM* derived from the genus *Geobacter* were detected in the same sample. Similarly, multiple genes derivedfrom *Aromatoleum (arrA* and

arsC), Azoarcus (arrA and *arsCB), Chlorobium* and *Thauera (arrA* and *arsB), Alkaliphilus (arrA* and *arsM),* and *Desulfitobacterium (arrA* and *arsB)* were also detected. The predominated arsenic metabolic related populations were significantly different with those in high As groundwater of Cache Valley Basin Utah and Bangladesh (Mirza et al., 2017; Hassan et al., 2015).

Most of the arsenic related genes including the arsenic detoxification genes *arsC* and *arsB*, and the respiratory arsenate (AsV) reductase gene *arrA* had higher relative abundances in the high arsenic group than low arsenic group (Figure 5). Positive correlations were observed between arsenic concentration and the number of detected arsenic related genes *arsC* ($r=0.462$, $p<0.01$) and *arrA* ($r=0.486$, $p<0.01$). There was no correlation between *arsB*, AsIII S-adenosylmethionine methyltransferase gene *arsM*, arsenite (AsIII) oxidase gene *aoxB* and arsenic concentrations. More *arsC/arsB* genes were detected than the other genes (Figure 5), such as *arrA* (around 5-8 folds), indicating that the most common detoxification mechanism was reduction of AsV to AsIII rather than respiratory arsenate reduction in the high arsenic groundwater aquifers, which was consistent with results based on phylogenic information detected with 454 (Li et al., 2015). Similar results were observed for paddy soils in a previous study (Zhao et al., 2013).

3.5 Functional genes for sulfur cycling

Of the 1211 sulfur related genes detected in this study, most were for sulfur reduction (the average gene intensity for the *dsrAB* probes was 53.48%, and was 25.33% for *soxY*). Microbial populations including *Desulfovibrio*, *Desulfobulbus*, *Desulfohalobium*, *Desulfotomaculum*, *Chlorobium*, *Clostridium*, *Geobacter*, *Magnetospirillum*, *Thiobacillus*, *Thioalkalivibrio*, *Syntrophobacter*, *Pyrobaculum*, and *Acetobacterium* were frequently detected in samples, while most of the detected *dsrAB* were from some uncultured sulfate-reducing bacteria. Both arsenic related genes and sulfate reductions genes were detected from *Desulfohalobium*, *Geobacter*, *Thioalkalivibrio*, and *Chlorobium*. The diversity of the sulfur reduction population in this study was much higher than that of high arsenic groundwater in Bangladesh (Gorra et al., 2012). There were positive correlations between arsenic and sulfur reduction genes (*dsrA* and As: $r=0.455$, $p<0.01$; *dsrB* and As: $r=0.327$, $p<0.05$). These results were consistent with our previous studies using qPCR (Li et al., 2014).

3.6 Functional genes for the nitrogen cycling

Nitrogen (N) cycling, an important process in high arsenic groundwater, is generally of concern for its roles of microbial activity (Kurosawa et al., 2013; Natarajan et al., 2009). Previous studies found that microbial nitrate-dependent oxidation of FeII enhances the immobilization of As in the anoxic

environments, ammonium-N could enhance microbial activity, and then improve As release in a reducing condition (Weng et al., 2017; Kurosawa et al., 2008). There are many genes involved in N cycling, including those for denitrification (*narG*, *nirS*, *nirK*, *norB*, and *nosZ*), nitrification (*amoA* and *hao*), dissimilatory N reduction (*napA* and *nrfA*, *nasA* and *nir*), ammonification (*ureC*, *gdh*), anammox (*hzo*) and N fixation (*nifH*). The dissimilatory N reduction gene (*nasA*) was positively correlated with arsenic ($r=0.357$, $p<0.01$). The *ureC* gene that transforms organic nitrogen to NH_4^+ was positively correlated with arsenic and NH_4^+ ($r=0.571$, $p<0.01$; $r=0.323$, $p<0.01$, respectively). The NH_4^+ oxidization gene, *amoA*, was marginal negatively correlated with arsenic ($r=-0.311$, $p<0.05$), while the anammox gene, *hzo*, was positively correlated with arsenic concentration ($r=0.452$, $p<0.05$), which implies that anaerobic bacterial oxidation might be a survival strategy oxidizing ammonia in high arsenic groundwaters. The signal intensities were significantly different ($p<0.05$) between the high and low arsenic groups for the genes *ureC*, *gdh*, *amoA*, *hzo*, *narG* and *norB* (Figure 6). The signal intensities of the other nitrogen related genes exhibited no significant correlation between arsenic and nitrate or NH_4^+ and the intensities in high and low arsenic groups were not significantly different (Figure 6).

3.7 Methanogenesis genes

The methanogenesis genes detected (248) were mainly from *Methanosaeta*, *Methanoculleus*, *Methanocella*, *Desulfatibacillum*, and *Moorella*. Among these genes, 59-65% were *mcrA*, encoding the α subunit of methyl coenzyme M reductase, and *hdrB*, encoding cytoplasmic heterodisulfide reductase. Most of the methanogenesis genes were more abundant in the high arsenic group samples than in the low arsenic group. The average abundance of methanogenesis genes was 1.37% in the low arsenic group and 3.06% in the high arsenic group. Positive correlations were observed between methanogenesis genes and arsenic (As_{Tot}) and methane concentrations (*mcrA*: $r=0.549$, $p<0.05$; $r=0.314$, $p<0.05$, and *hdrB*: $r=0.387$, $p<0.05$; $r=0.389$, $p<0.05$, respectively). These results indicated that methanogenesis might be an important metabolic process and accelerate As release and accumulation in high arsenic aquifers, consistent with our previous study (Wang et al., 2015). The composition of methanogens in this study is totally different with high arsenic groundwater in Bangladesh (Gorra et al., 2012). The Fifty two methane oxidation genes were detected, mainly from *Methylosinus*, *Methylocapsa* and some uncultured bacteria (*pmoA*), as well as *Methylomonas*, *Citreicella*, *Mesorhizobium* and *Methylocella* (*mmox*). There was no correlation between methane oxidation gene abundance and arsenic concentration.

330 3.8 Genes related to phosphorus, carbon degradation, organic remediation and mental resistance

331 There was no correlation between arsenic concentration and the quantity of carbon degradation
332 genes, although arsenic was positive correlated with TOC contents (Figure 2). A potential explanation
333 is that dissolved organic matter, such as fulvic or humic acids could form stable complexes with
334 mineral surfaces, effectively blocking arsenic from adsorption onto minerals, which would improve
335 arsenic keeping in the aqueous phrase (Natarajan et al., 2009). Similarly, no obvious changes in relative
336 abundance across the nineteen samples were detected between arsenic and phosphorus, organic
337 remediation genes, or metal resistance genes (data not shown), indicating that organic remediation and
338 metal resistance mechanisms contributed little to the change observed in the functional microbial
339 communities.

340

341 **4. Discussion**

342 In this study, a large number of genes with a tremendous diversity of sequences were detected in
343 our samples. Such functional gene information is useful for assessing the impacts of microbially-
344 mediated As geochemistry. Our study highlights that the geochemistry is a strong driver of a microbial
345 community's functional structure, with arsenic being one of the key environmental factors contributing
346 to the differences in the geochemistry and microbial community structure in groundwater. Groundwater
347 samples could be divided into well-defined high and low arsenic groups based on both geochemistry
348 and the functional composition of the microbial communities. Similar results were found in our
349 previous phylogeny based studies on microbial communities in this sites using high throughout and
350 traditional sequencing technologies (Wang et al., 2016; Li et al., 2015; Wang et al., 2014; Li et al.,
351 2013). This could be explained by arsenic release and mobilization. Our study provides evidence for
352 possible links between functional potential of *in situ* microbial communities and As, S, N and C
353 chemicals reduction which promoted As release in groundwater. The results are in agreement with
354 previous studies that reported arsenic enrichment was related to nitrogen, carbon, sulfur and iron
355 geochemistry, all of which are usually considerable drivers of specific metabolic pathways in microbes,
356 and the composition of associated microbial communities (ThomasArrigo et al., 2016; Wang et al.,
357 2016; Li et al., 2015; Ghosh et al., 2015; Wang et al., 2014; Li et al., 2013; Nordstrom, 2002; Kocar et
358 al., 2010; Xiong et al., 2010). The current investigation supported the finding that alkaline and
359 reducing conditions with relatively low ORP could be linked to natural arsenic enrichment (Rodríguez-
360 Lado et al., 2013). Our study indicated that alkaline pH was one of the most important environmental

variables shaping the functional microbial community structure in high arsenic samples (Figure 4 and S2). The elevated pH might be associated with the rise of NH_4^+ and TOC in high arsenic groundwater. The pH value of most of the high arsenic samples in this study were slightly alkaline (7.77-8.48). Water chemical types were mostly Na(–Mg)–Cl(–HCO_3) and Na–Mg–HCO_3. Most groundwater samples in our study area are unsaturated with gypsum and close to saturation with respect to calcite and dolomite (Deng et al., 2009). The rise of NH_4^+ increased microbial activities (Kiyoshi et al., 2008), which might in turn promote the biotic dissolution of calcite and dolomite minerals, and then increases pH value in groundwater. Reducing conditions associated with arsenic, sulfate and iron reduction, have long been considered the prevailing geochemical processes in arsenic enrichment mechanisms in many high arsenic groundwater aquifers and some other anoxic natural systems (Wang et al., 2014; Ehlert et al., 2014; Kocar et al., 2010; Razzak et al., 2009; Deng et al., 2009; Stauder et al., 2005). Reductive dissolution of Fe oxide minerals and the oxidation of pyrite and mineral phase transition in sediments were demonstrated to promote the release of bound As into the groundwater in our study area (Zhang et al., 2017; Guo et al., 2016; Guo et al., 2011). Additionally, As(III) reducing, sulfate reducing and methanogenesis were proposed to promote As mobilization in the study sites (Dai, et al., 2016; Wang et al., 2015; Li et al., 2014). Our current study connected the reducing geochemical conditions with functional gene profiles including As, S, C and N related functional genes in the high arsenic groundwater aquifers.

In this study our data implied that arsenic, sulfate and iron reduction might occur in high arsenic samples at the same sites (Figure 2). In our previous study demonstrated that indigenous arsenic/iron reducing bacteria from the high arsenic aquifer could release arsenic from sediments to aqueous phase (Dai et al., 2016). Moreover, the positive correlation between *dsrAB* gene abundance and As concentration was found in our study. Some arsenic and sulfate reduction genes from the same organism were also found, which was in agreement with previous findings that some microorganisms are capable of reducing both arsenic and sulfate (Macy et al., 2000; Newman et al., 1997). In high arsenic groundwater aquifers, arsenic, sulfur and iron often coexist and can be important for arsenic transformation and mobilization (Kocar et al., 2010; Saalfield and Bostick 2009). The homeostasis of dissolved arsenic may be controlled by biotic reduction of Fe-(hydr)oxides and FeIII minerals and the solubility of sulfide phases (Wang et al., 2014; Xiong et al., 2010). Biotic sulfate reduction could drive reductive dissolution of Fe-(hydr)oxides as well as arsenic reduction, resulting in increased As release and mobilization when As is not incorporated into iron sulfides (Li et al., 2014).

Ammonification is an important biological process providing N in high arsenic groundwater by transforming organic nitrogen to inorganic nitrogen. In high arsenic groundwater aquifers, NH_4^+-N was found in high concentration. The gene *ureC* was in relatively high abundance in high arsenic groundwater samples in this study (Figure 6). Although there is no direct evidence of NH_4^+-N in groundwater from agriculture in our study sites, previous studies indicated that the NH_4^+-N (mainly urea) in some other high arsenic groundwater aquifers is from fertilizer in agriculture practices (Weng et al., 2017; Uddin and Kurosawa 2014; Mayorga et al., 2013; Itai et al., 2008). Our current evidence of NH_4^+ accumulation in high arsenic groundwater might be due to the higher abundance of genes such as *ureC* that transform organic nitrogen to NH_4^+ and the lower abundance of NH_4^+ utilization genes (*hzo* and *gdh*) (Figure 6). The rise in ammonium-N concentration, as well as elevated TOC, might enhance microbial activities, which in turn would lower ORP, and then enhance arsenic reduction, microbial reduction of Fe(III) oxides/hydroxides, and sulfate reduction which improved As mobilization through dissolution of Fe(III) oxide minerals or formation of As-sulfur compounds in high HS^- concentration water (Kurosawa et al., 2008; Deng et al; 2009; Wang et al., 2014). This might explain why NH_4^+ was significantly correlated with arsenic concentrations in groundwater aquifers. In addition, organic matter from agricultural irrigation was reported to release into groundwater in this study area (Gao et al., 2014). Previous studies proposed high concentrations of TOC might be in favor of microbial mediated As enrichment (Uddin and Kurosawa 2014; Mayorga et al., 2013; Barringer et al. 2010; Rezza et al. 2010). These results suggested again that human agriculture activities might be one of the most important factors of improving microbial activities, causing arsenic contamination in groundwater (Gao et al., 2014; Neidhardt et al., 2012).

It is worth noting that genes specific for anammox were detected, indicating that ammonium and nitrite can be converted directly into nitrogen gas via anammox pathways in high arsenic groundwater. Anammox can happen without requiring substantial amounts of organic-C, reduced iron, or mineral phase electron donors. Groundwater hydrology could create expanded zones suitable for anammox communities and widespread spatial distribution of anammox activity (Smith et al., 2015). Arsenic concentrations were positively correlated with anammox gene numbers, suggesting anammox was a survival strategy oxidizing ammonia in high arsenic groundwaters. However, the importance of anammox in groundwater aquifers had not been considered previously. Further tracer studies of the anammox pathway in the groundwater might help elucidate nitrogen cycling in this ecosystem.

5. Conclusion

Our results reveal that environmental variables account for the majority of variation in microbial functional potential. This study provides evidence for possible links between As, S, N and C related microbial functional gene abundances and As geochemistry. Alkaline and reducing conditions as well as SO_4^{2-} and Fe reduction and ammonification associated with microbially-mediated geochemical processes could be linked to arsenic enrichment in groundwater in this study area. These results suggest that the indigenous microbial communities could have a significant role in arsenic release and transformation in high arsenic groundwater aquifers of the Hetao Basin, Inner Mongolia. This study expands our current understanding of microbial ecology in high arsenic aquifers.

Acknowledgements

This work was financially supported by National Natural Science Foundation of China (grant numbers 41372348, 41521001), and Fundamental Research Funds for the Central Universities, China University of Geosciences (CUG170642).

Supporting Materials

Additional information: Figure S1-2, Table S1-4.

441 **Reference**

442 Barringer, J. L., Mumford, A., Young, L. Y., Reilly, P. A., Bonin, J. L., & Rosman, R. 2010. Pathways
443 for arsenic from sediments to groundwater to streams: biogeochemical processes in the Inner
444 Coastal Plain, New Jersey, USA. Water Res. 44(19), 5532-5544.

445 Bose, P., Sharma, A. (2002). Role of iron in controlling speciation and mobilization of arsenic in
446 subsurface environment. Water Res. 36(19), 4916-4926.

447 Chakraborti, D., Rahman, M.M., Das, B., Murrill, M., Dey, S., Mukherjee, S.C., Dhar, R.K., Biswas,
448 B.K., Chowdhury, U.K., Roy, S., Sorif, S., Selim, M., Rahman, M. Sorif, S. 2010. Status of
449 groundwater arsenic contamination in Bangladesh: a 14-year study report. Water Res. 44(19),
450 5789-5802.

451 Chen, C.-J. 2014. Health hazards and mitigation of chronic poisoning from arsenic in drinking water:
452 Taiwan experiences. Rev. Environ. Health 29(1-2), 13-19.

453 Chen, X., Zeng, X. C., Wang, J., Deng, Y., Ma, T., Guoji, E., Mu, Y., Yang, Y., Li, Hao., Wang, Y.
454 2017. Microbial communities involved in arsenic mobilization and release from the deep
455 sediments into groundwater in Jianghan plain, Central China. Sci. Total Environ. 579, 989-999.

456 Chung, C. J., Huang, Y. L., Huang, Y. K., Wu, M. M., Chen, S. Y., Hsueh, Y. M., & Chen, C. J. 2013.
457 Urinary arsenic profiles and the risks of cancer mortality: a population-based 20-year follow-up
458 study in arseniasis-endemic areas in Taiwan. Environ Res. 122(2), 25-30.

459 Dai, X., Li, P., Tua, J., Zhang, R., Wei, D., Li, B., Wang, Y. and Jiang, Z. 2016. Evidence of arsenic
460 mobilization mediated by an indigenous iron reducing bacterium from high arsenic groundwater
461 aquifer in Hetao Basin of Inner Mongolia, China. Int. Biodeterior. Biodegrad.
462 http://dx.doi.org/10.1016/j.ibiod.2016.05.012.

463 de Hoon, M.J., Imoto, S., Nolan, J. and Miyano, S. 2004. Open source clustering software.
464 Bioinformatics 20(9), 1453-1454.

465 Deng, Y., Wang, Y., Ma, T., Gan, Y. 2009. Speciation and enrichment of arsenic in strongly reducing
466 shallow aquifers at western Hetao Plain, northern China. Environ. Geol. 56(7), 1467-1477.

467 Ehlert, K., Mikutta, C., & Kretzschmar, R. 2014. Impact of birnessite on arsenic and iron speciation
468 during microbial reduction of arsenic-bearing ferrihydrite. Environ. Sci. Technol. 48(19), 11320-
469 11329.

470 Fendorf, S., Michael, H.A. and van Geen, A. 2010. Spatial and temporal variations of groundwater
471 arsenic in South and Southeast Asia. Science 328(5982), 1123-1127.

472 Flynn, T.M., O'Loughlin, E.J., Mishra, B., DiChristina, T.J. and Kemner, K.M. 2014. Sulfur-mediated
473 electron shuttling during bacterial iron reduction. Science 344(6187), 1039-1042.

474 Gao, C., Liu, W., Feng, C., Chen, Y., Zhang, Guo., Song, J. 2014. Research on the formation
475 mechanism of high arsenic groundwater in arid and semi-arid regions: A case study of Hetao Plain
476 in Inner Mongolia. Earth Sci. Front. 21(4), 013-029.

Ghosh, S., Sar, P. 2013. Identification and characterization of metabolic properties of bacterial populations recovered from arsenic contaminated ground water of North East India (Assam). Water Res. 47(19), 6992-7005.

Ghosh, D., Routh, J. and Bhadury, P. 2015. Characterization and microbial utilization of dissolved lipid organic fraction in arsenic impacted aquifers (India). J. Hydro. 527, 221-233.

Gorra, R., Webster, G., Martin, M., Celi, L., Mapelli, F. and Weightman, A.J. 2012. Dynamic microbial community associated with iron–arsenic co-precipitation products from a groundwater storage system in bangladesh. Microb Ecol. 64(1), 171-186.

Guo, H., Zhang, B., Li, Y., Berner, Z., Tang, X., Norra, S., Stüben, D. 2011. Hydrogeological and biogeochemical constrains of arsenic mobilization in shallow aquifers from the Hetao basin, Inner Mongolia. Environ. Pollut. 159(4), 876-883.

Guo, H., Zhang, Y., Jia, Y., Zhao, K., Li, Y., & Tang, X. 2013. Dynamic behaviors of water levels and arsenic concentration in shallow groundwater from the Hetao Basin, Inner Mongolia. J. Geoch. Explor. 135, 130-140.

Guo, H., Jia, Y., Wanty, R. B., Jiang, Y., Zhao, W., Xiu, W., Shen, J., Li, Y., Cao, Y., Wu, Y., Zhang, D., Wei, C., Zhang, Y., Cao, W., Foster, A. 2016. Contrasting distributions of groundwater arsenic and uranium in the western Hetao basin, Inner Mongolia: Implication for origins and fate controls. Sci. Total Environ. 541, 1172-1190.

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 Microbiol. Ecol. 91(4), fiv026.

Itai, T., Masuda, H., Seddique, A. A., Mitamura, M., Maruoka, T., Li, X., Kusakabe, M., Dipak, B. K., Farooqi, A., Nakaya, S. 2008. Hydrological and geochemical constraints on the mechanism of formation of arsenic contaminated groundwater in Sonargaon, Bangladesh. Appl. Geoch. 23(11), 3155-3176.

Jia, Y., Guo, H., Jiang, Y., Wu, Y., Zhou, Y. 2014. Hydrogeochemical zonation and its implication for arsenic mobilization in deep groundwaters near alluvial fans in the Hetao Basin, Inner Mongolia. J. Hydrol. 518, 410-420.

Johnson, K. M., Hughes, J. E., Donaghay, P. L., Sieburth, J. M. (1990). Bottle-calibration static head space method for the determination of methane dissolved in seawater. Anal. Chem. 62(21), 2408-2412.

Kocar, B.D., Borch, T. and Fendorf, S. 2010. Arsenic repartitioning during biogenic sulfidization and transformation of ferrihydrite. Geochim. Cosmochim. Ac. 74(3), 980-994.

Kurosawa, K., Egashira, K. and Tani, M. 2013. Relationship of arsenic concentration with ammonium–nitrogen concentration, oxidation reduction potential and pH of groundwater in arsenic-contaminated areas in Asia. Phys. Chem. Earth, Parts A/B/C 58-60(4), 85-88.

513 Kurosawa, K., Egashira, K., Tani, M., Jahiruddin, M., Moslehuddin, A.Z.M. and Zulfikar Rahman, M.
514 2008. Variation in arsenic concentration relative to ammonium nitrogen and oxidation reduction
515 potential in surface and groundwater. Commun. Soil Sci. Plant Anal. 39(9-10), 1467-1475.
516 Li, P., Li, B., Webster, G., Wang, Y., Jiang, D., Dai, X., Jiang, Z., Dong, H. and Wang, Y. 2014.
517 Abundance and diversity of sulfate-reducing bacteria in high arsenic shallow aquifers.
518 Geomicrobiol. J. 31(9), 802-812.
519 Li, P., Wang, Y., Dai, X., Zhang, R., Jiang, Z., Jiang, D., Wang, S., Jiang, H., Wang, Y. and Dong, H.
520 2015. Microbial Community in High Arsenic Shallow Groundwater Aquifers in Hetao Basin of
521 Inner Mongolia, China. PLoS One 10(5), e0125844.
522 Li, P., Wang, Y., Jiang, Z., Jiang, H., Li, B., Dong, H. and Wang, Y. 2013. Microbial Diversity in High
523 Arsenic Groundwater in Hetao Basin of Inner Mongolia, China. Geomicrobiol. J. 30(10), 897-909.
524 Low, A., Ng, C., He, J. 2016. Identification of antibiotic resistant bacteria community and a GeoChip
525 based study of resistome in urban watersheds. Water Res. 106(9), 330-338.
526 Macy, J.M., Santini, J.M., Pauling, B.V., O'Neill, A.H. and Sly, L.I. 2000. Two new arsenate/sulfate-
527 reducing bacteria: mechanisms of arsenate reduction. Arch. Microbiol. 173(1), 49-57.
528 Mayorga, P., Moyano, A., Anawar, H.M. and García-Sánchez, A. 2013. Temporal variation of arsenic
529 and nitrate content in groundwater of the Duero River Basin (Spain). Phys. Chem. Earth, Parts
530 A/B/C 58-60, 22-27.
531 Michael, H.A. 2013. An arsenic forecast for China. Science 341(6148), 852-853.
532 Mirza, B. S., Sorensen, D. L., Dupont, R. R., & McLean, J. E. 2017. New arsenate reductase gene
533 (arrA) PCR primers for diversity assessment and quantification in environmental samples. Appl.
534 Environ. Microbiol. 83(4), e02725-16.
535 Mumford, A. C., Barringer, J. L., Benzel, W. M., Reilly, P. A., Young, L. Y. 2012. Microbial
536 transformations of arsenic: mobilization from glauconitic sediments to water. Water Res. 46(9),
537 2859-2868.
538 Natarajan, S., Stamps, R.H., Saha, U.K. and Ma, L.Q. 2009. Effects of Nitrogen and Phosphorus Levels,
539 and Frond-Harvesting on Absorption, Translocation and accumulation of arsenic by Chinese
540 Brake Fern (Pteris Vittatal.). Inter. J. Phytoremediation 11(4), 313-328.
541 Neidhardt, H., Norra, S., Tang, X., Guo, H. and Stüben, D. 2012. Impact of irrigation with high arsenic
542 burdened groundwater on the soil–plant system: Results from a case study in the Inner Mongolia,
543 China. Environ. Pollut.163, 8-13.
544 Newman, D.K., Kennedy, E.K., Coates, J.D., Ahmann, D., Ellis, D.J., Lovley, D.R. and Morel, F.M.M.
545 1997. Dissimilatory arsenate and sulfate reduction in Desulfotomaculum auripigmentum sp. nov.
546 Arch. Microbiol. 168(5), 380-388.
547 Nordstrom, D.K. 2002. Worldwide occurrences of arsenic in ground water. Science 296(5576), 2143-
548 2145.

549 Pi, K., Wang, Y., Xie, X., Ma, T., Liu, Y., Su, C., Zhu, Y. Wang, Z. 2017. Remediation of arsenic-
550 contaminated groundwater by in-situ stimulating biogenic precipitation of iron sulfides. Water
551 Res. 109(1), 337-346.
552 Razzak, A., Jinno, K., Hiroshiro, Y., Halim, M.A. and Oda, K. 2009. Mathematical modeling of
553 biologically mediated redox processes of iron and arsenic release in groundwater. Environ.Geol.
554 58(3), 459-469.
555 Reza, A. S., Jean, J. S., Lee, M. K., Liu, C. C., Bundschuh, J., Yang, H. J., Lee, J. F., Lee, Y. C. 2010.
556 Implications of organic matter on arsenic mobilization into groundwater: evidence from
557 northwestern (Chapai-Nawabganj), central (Manikganj) and southeastern (Chandpur)
558 Bangladesh. Water Res. 44(19), 5556-5574.
559 Rodríguez-Lado, L., Sun, G., Berg, M., Zhang, Q., Xue, H., Zheng, Q. and Johnson, C.A. 2013.
560 Groundwater arsenic contamination throughout China. Science 341(6148), 866-868.
561 Saalfield, S.L. and Bostick, B.C. 2009. Changes in iron, sulfur, and arsenic speciation associated with
562 bacterial sulfate reduction in ferrihydrite-rich systems. Environ. Sci. Technol. 43(23), 8787-8793.
563 Schaefer, M.V., Ying, S.C., Benner, S.G., Duan, Y., Wang, Y. and Fendorf, S. 2016. Aquifer arsenic
564 cycling induced by seasonal hydrologic changes within the Yangtze River Basin. Environ. Sci.
565 Technol.
566 Smith, R.L., Bohlke, J.K., Song, B. and Tobias, C.R. 2015. Role of anaerobic ammonium oxidation
567 (anammox) in nitrogen removal from a freshwater aquifer. Environ. Sci. Technol. 49(20), 12169-
568 12177.
569 Stauder, S., Raue, B., Sacher, F., 2005. Thioarsenates in sulfidic waters. Environ. Sci. Technol. 39,
570 5933 – 5939.
571 Sultana, M., Härtig, C., Planer-Friedrich, B., Seifert, J., Schlömann, M. 2011. Bacterial communities in
572 Bangladesh aquifers differing in aqueous arsenic concentration. Geomicrobiol. J. 28(3), 198-211.
573 Sun, Y., Shen, Y. X., Liang, P., Zhou, J., Yang, Y., Huang, X. 2014. Linkages between microbial
574 functional potential and wastewater constituents in large-scale membrane bioreactors for
575 municipal wastewater treatment. Water Res. 56(3), 162-171.
576 Sutton, N.B., van der Kraan, G.M., van Loosdrecht, M., Muyzer, G., Bruining, J. and Schotting, R.J.
577 2009. Characterization of geochemical constituents and bacterial populations associated with As
578 mobilization in deep and shallow tube wells in Bangladesh. Water Res. 43(6), 1720-1730.
579 Tang, J., Bian, J., Li, Z., Li, Y., Yang, W., Liang, S. 2017. Comparative study on the hydrogeochemical
580 environment at the major drinking water based arsenism areas. Applied Geochemistry, 77, 62-67.
581 ThomasArrigo, L.K., Mikutta, C., Lohmayer, R., Planer-Friedrich, B. and Kretzschmar, R. 2016.
582 Sulfidization of organic freshwater flocs from a minerotrophic peatland: speciation changes of
583 iron, sulfur, and arsenic. Environ. Sci. Technol. 50(7), 3607-3616.
584 Uddin, M.S. and Kurosawa, K. 2014. Effects of peat and water quality parameters on groundwater
585 arsenic contamination in Bangladesh. Water Environ. J. 28(2), 165-172.

586 Van Nostrand J D, Yin H, Wu L, et al. 2016. Hybridization of environmental microbial community
587 nucleic acids by geochip. Microb. Environ. Genomics, 1399(1), 183-196.
588 Weng, T. N., Liu, C. W., Kao, Y. H., & Hsiao, S. S. Y. 2017. Isotopic evidence of nitrogen sources and
589 nitrogen transformation in arsenic-contaminated groundwater. Sci. Total Environ. 578, 167-185.
590 Wade, T.J., Xia, Y., Wu, K., Li, Y., Ning, Z., Le, X.C., Lu, X., Feng, Y., He, X. and Mumford, J.L.
591 2009. Increased mortality associated with well-water arsenic exposure in Inner Mongolia, China.
592 Int. J. Env. Res. Public Health 6(3), 1107-1123.
593 Wang, Y., Li, P., Jiang, Z., Sinkkonen, A., Wang, S., Tu, J., Wei, D., Dong, H. & Wang, Y. 2016.
594 Microbial community of high arsenic groundwater in agricultural irrigation area of Hetao Plain,
595 Inner Mongolia. Front Microbiol, 7: 1917.
596 Wang, Y., Li, P., Dai, X., Zhang, R., Jiang, Z., Jiang, D. and Wang, Y. 2015. Abundance and diversity
597 of methanogens: potential role in high arsenic groundwater in Hetao Plain of Inner Mongolia,
598 China. Sci. Total Environ. 515, 153-161.
599 Wang, Y., Li, P., Li, B., Webster, G., Weightman, A. J., Jiang, Z., Jiang, D. Deng, Y., Wang, Y. 2014.
600 Bacterial diversity and community structure in high arsenic aquifers in Hetao Plain of Inner
601 Mongolia. Geomicrobiol. J. 31(4), 338-349.
602 Wang, Y., Xie, X., Johnson, T.M., Lundstrom, C.C., Ellis, A., Wang, X., Duan, M. and Li, J. 2014.
603 Coupled iron, sulfur and carbon isotope evidences for arsenic enrichment in groundwater. J.
604 Hydrol. 519, 414-422.
605 Wei, B., Yu, J., Li, H., Yang, L., Xia, Y., Wu, K., Gao, J., Guo, Z. and Cui, N. 2016. Arsenic
606 metabolites and methylation capacity among individuals living in a rural area with endemic
607 arseniasis in Inner Mongolia, China. Biol. Trace Elem. Res. 170(2), 300-308.
608 Weng, T. N., Liu, C. W., Kao, Y. H., & Hsiao, S. S. Y. 2017. Isotopic evidence of nitrogen sources and
609 nitrogen transformation in arsenic-contaminated groundwater. Sci. Total Environ. 578, 167-185.
610 WHO, 2011. Guideline for drinking-water quality, fourth ed. World Health Organization, Geneva,
611 Switzerland.
612 Xiong, J., Wu, L., Tu, S., Van Nostrand, J.D., He, Z., Zhou, J. and Wang, G. 2010. Microbial
613 communities and functional genes associated with soil arsenic contamination and the rhizosphere
614 of the arsenic-hyperaccumulating plant Pteris vittata L. Appl. Environ. Microbiol. 76(21), 7277-
615 7284.
616 Ye, H., Yang, Z., Wu, X., Wang, J., Du, D., Cai, J., Lv, K., Chen, H., Mei, J., Chen, M., Du, H. 2017.
617 Sediment biomarker, bacterial community characterization of high arsenic aquifers in Jianghan
618 Plain, China. Sci. Rep. 7, 42037.
619 Zhang, D., Guo, H., Xiu, W., Ni, P., Zheng, H., Wei, C. 2017. In-situ mobilization and transformation
620 of iron oxides-adsorbed arsenate in natural groundwater. J. Hazard. Mater. 321, 228-237.
621 Zhang, S.-Y., Zhao, F.-J., Sun, G.-X., Su, J.-Q., Yang, X.-R., Li, H. and Zhu, Y.-G. 2015. Diversity
622 and abundance of arsenic biotransformation genes in paddy soils from southern China. Environ.
623 Sci. Technol. 49(7), 4138-4146.

624 Zhang, Y., Xie, J., Liu, M., Tian, Z., He, Z., van Nostrand, J. D., Ren, L., Zhou, J. Yang, M. 2013.
625 Microbial community functional structure in response to antibiotics in pharmaceutical wastewater
626 treatment systems. Water Res. 47(16), 6298-6308.
627 Zhao, F.J., Harris, E., Yan, J., Ma, J., Wu, L., Liu, W., McGrath, S.P., Zhou, J. and Zhu, Y.G. 2013.
628 Arsenic methylation in soils and its relationship with microbial arsM abundance and diversity,
629 and as speciation in rice. Environ. Sci. Technol. 47(13), 7147-7154.
630 Zhou, J., Kang, S., Schadt, C.W. and Garten, C.T. 2008. Spatial scaling of functional gene diversity
631 across various microbial taxa. Proc. Natl. Acad. Sci. U. S. A. 105(22), 7768-7773.
632

633

634 **Tables**

635 Table 1. Significance test of the overall microbial community structures and geochemical patterns
636 between different sample groups using three statistical analyses. Samples S1-5 with relatively low
637 concentration of arsenic were grouped together and sample S6-19 with higher arsenic were included
638 into another group.

Method	Geochemical parameters [I]		Microbial community	
	Statistic	P-value	Statistic	P-value
MRPP[II]	0.4708	0.002	0.0469	0.018
ANOSIM[III]	0.6526	0.001	0.4776	0.007
Adonis[IV]	0.3852	0.001	0.1264	0.060

639 Abbreviations: ANOSIM, analysis of similarity; MRPP, multi response permutation procedure. All
640 three tests are non-parametric multivariate analyses based on dissimilarities among samples.
641 [I]Geochemical parameters included DO (dissolved oxygen), ORP (oxidation reduction potentials), As_{Tot}
642 (total arsenic), AsIII (arsenite), Fe_{Tot} (total dissolved iron), FeII (ferrous iron), TOC (total organic
643 carbon) and etc (see table S1).
644 [II]Multiple response permutation procedure.
645 [III]Analysis of similarities.
646 [IV]Non-parametric multivariate analysis of variance (MANOVA) with the adonis function.

Figure captions

Fig. 1. Geographic map of Hangjinhouqi County (HC), Inner Mongolia of China. Map showing location of study area (HC) and the sampling sites. Numbers 1-19 refer to groundwater samples S1-19.

Fig. 2. Variation of arsenic concentrations with (a) AsIII, SO_4^{2-}, $NO_3^-$and NH_4^+ (b) pH and ORP, (c)TOC, CH_4 and ratios of FeII to FeIII in the groundwater sample S1-19 from Inner Mongolia.

Fig. 3. Hierarchical cluster analysis of geochemical parameters of sample S1-19 from Inner Mongolia. Results were generated in CLUSTER and visualized using TREEVIEW. Red indicates signal intensities below background, whereas blue indicates signal intensities above background. Brighter blue color indicates higher signal intensities. Sample S1-5 with lower arsenic concentrations clustered together and were well separated from other samples S6-19 with higher arsenic concentrations. EC: electrical conductivity; DO: dissolved oxygen, ORP: oxidation reduction potentials; TDS: total dissolved solids, As_{Tot}: total arsenic; AsIII: arsenite; Fe_{Tot}: total dissolved iron, FeII: ferrous iron; TOC: total organic carbon, C/N: ratio of total carbon to total nitrogen.

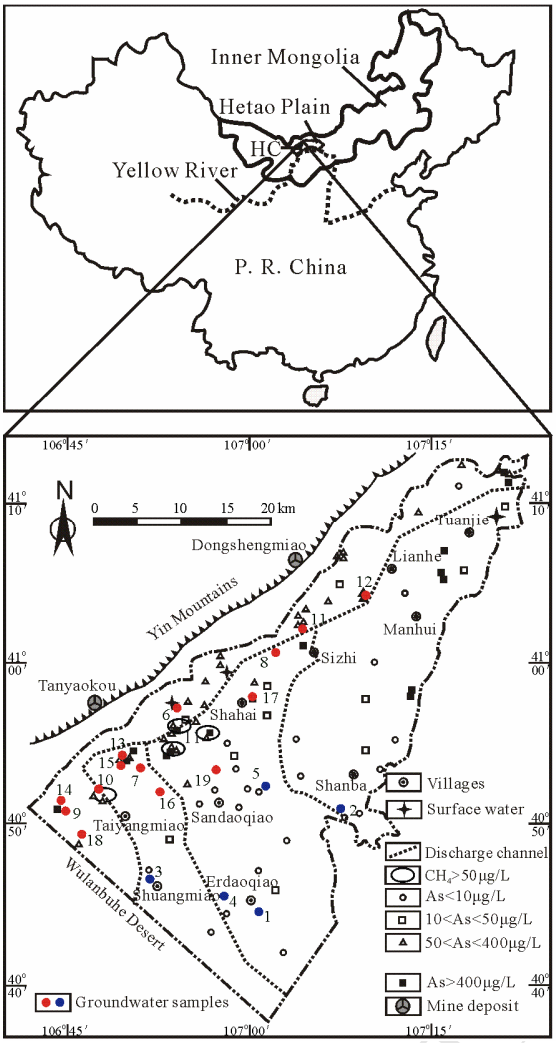
Fig. 4. Canonical correspondence analysis (CCA) analysis of GeoChip data (symbols) and environmental variables (arrows). Environmental variables were chosen based on significance calculated from individual CCA results and variance inflation factors (VIFs) calculated during CCA. The percentage of variation explained by each axis is shown, and the relationship is significant (P=0.021).

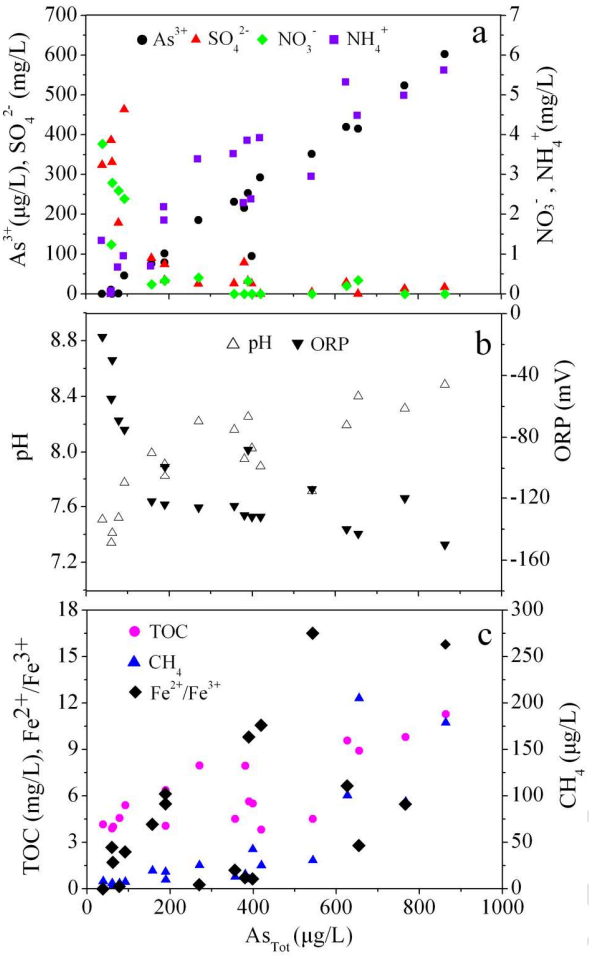
Fig. 5. Relative abundance of the genes involved in arsenic metabolisms in the groundwater samples. The signal intensity for each gene was the average of the total signal intensity in each group. L: low arsenic group samples; H: high arsenic group samples.

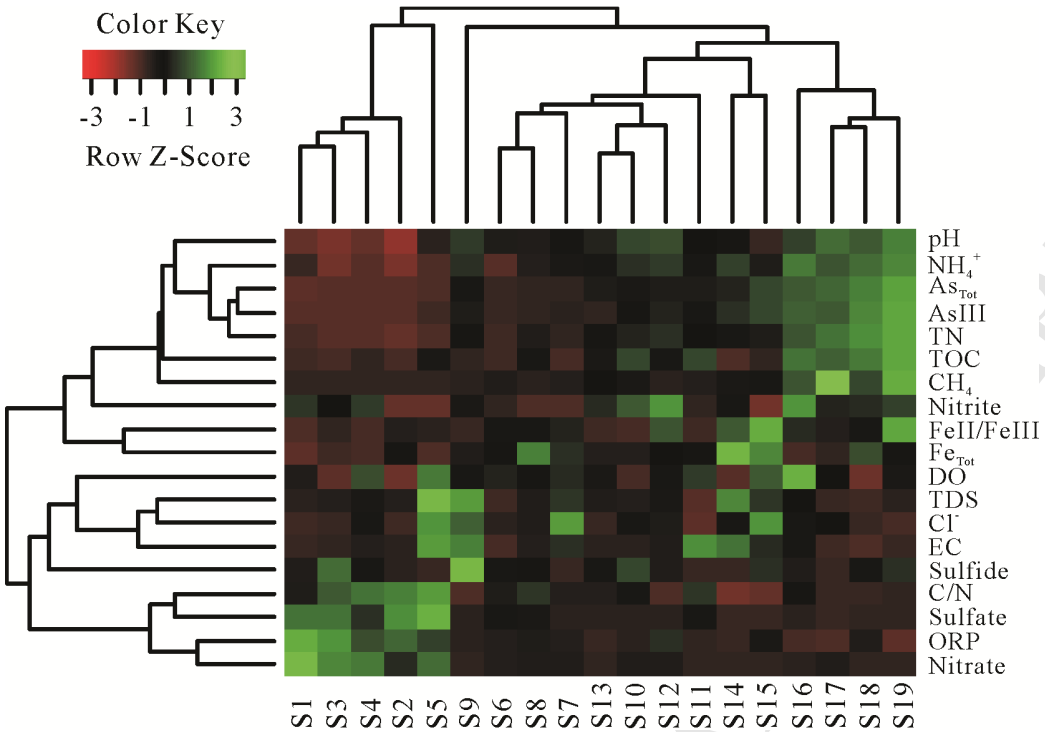
Fig. 6. The relative change of detected nitrogen cycling genes from the high arsenic group samples. The signal intensity for each gene was normalized by the mean signal of all detected gene sequences in the high arsenic samples. Percentages indicate the normalized total intensity of the functional genes. Red color indicates the difference in gene abundance between high and low arsenic groups was

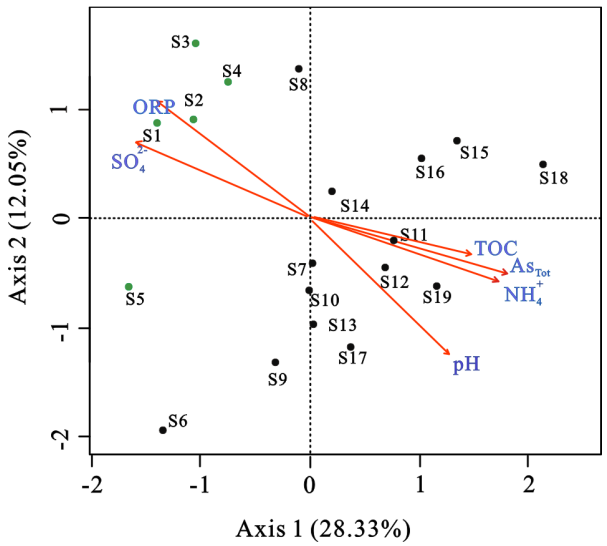
672 significant (**p<0.01, *p<0.05). Orange arrows indicate ammonium source processes, green arrows

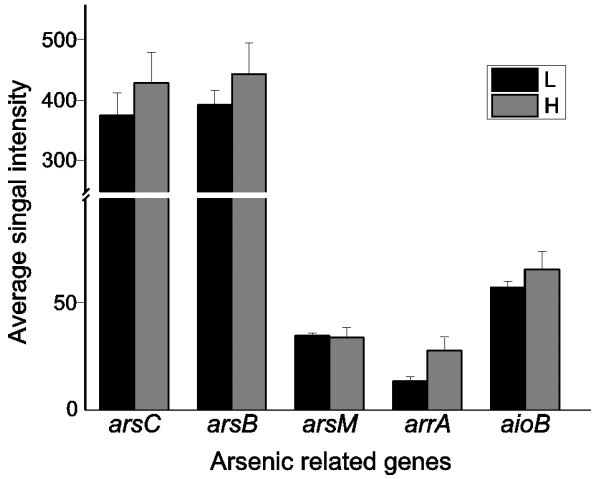
673 indicate ammonium utilization processes.

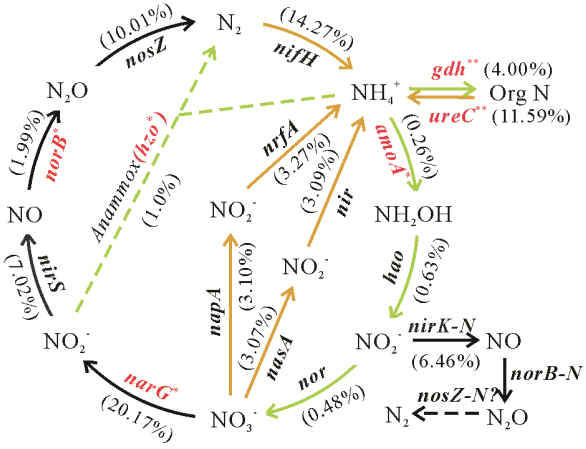












Highlights

1. Samples were divided into low and high arsenic groups based on geochemical parameters and microbial functional structures

2. Genes *arsC*, *arrA*, *dsrA*/*B*, *ureC*, *amoA*, *hzo*, *mcrA*, *hdrB* were highly correlated with arsenic, SO_4^{2-}, NH_4^+ or CH_4, respectively.

3. As, TOC, SO_4^{2-}, NH_4^+, ORP and pH were important factors shaping the functional microbial community structure.

4. Alkaline and reducing conditions could be associated with arsenic enrichment in groundwater.

5. An overall picture of functional microbial communities in high arsenic aquifers is provided.